3D Image segmentation for modelling the patient-specific anatomy of congenital heart disease

Byrne, Nick

Awarding institution:
King's College London

School of Biomedical Engineering and Imaging Sciences
4th Floor Lambeth Wing
St. Thomas' Hospital
Westminster Bridge Road
London, SE1 7EH

3D Image segmentation for modelling the patient-specific anatomy of congenital heart disease

---

Nick Byrne

Programme:
Imaging Sciences and Biomedical Engineering MPhil/PhD

Supervisors:
Dr Andrew King, Prof Giovanni Montana, Dr Israel Valverde

PhD Thesis
2018 – 2022

To Kar Lok and Jos,

Abstract

The marriage of cutting edge technologies for 3D visualisation (including 3D printing and virtual reality) with volumetric medical imaging admits the construction and representation of high-fidelity models of patient-specific anatomy. These capture the structural insights of 3D scan data - those critical to the care of patients with congenital heart disease (CHD) - in a form that is accessible not only to the imaging or radiological specialist, but also to the remainder of the multidisciplinary team. In relatively small-numbered studies, this type of enhanced communication has fostered improved consensus decision-making and personalised treatment planning, amongst a host of clinically related applications. Despite their promise, we argue that the wider application of patient-specific models has been limited by the technical burden of manual image segmentation, an unavoidable step in their determination from medical images. In response, this thesis investigates methods from the burgeoning field of deep learning, in pursuit of automated solutions to the segmentation of CHD anatomy from 3D cardiac magnetic resonance (CMR) data. More specifically, we make a clinically focused appraisal of state of the art convolutional neural networks (CNNs), a family of non-linear models of high statistical capacity.

Dependent on an underlying set of parameterised functions, CNNs can be tuned to the task of discriminative classification through data-driven optimisation. Observing the paucity of training examples appropriate to our task, we curate the Evelina London Children's Hospital (ELCH) dataset, including: isotropic CMR volumes and 4D contrast enhanced scans of 150 patients with CHD; each labelled according to a clinically meaningful manual segmentation protocol expressing the haemodynamic continuity of up to eighteen cardiovascular structures (including the congenital defects therein) by pixel adjacency. In a comprehensive clinical

characterisation and comparative analysis, we confirm the ELCH dataset as a quantitatively and qualitatively unique resource for both CNN training, and, more generally, for advancing our collective 3D understanding of the heart.

Leveraging this dataset within an assessment of CNN-based segmentation, we investigate different modes for combining 3D and 4D scan data within the U-Net architecture, observing inclusion of the latter to be associated with marginal gains in spatial overlap performance. More significantly, we extend our analyses beyond those encountered in the bulk of the technical literature. Presenting novel, clinically focused metrics sensitive to the presence of defects, we highlight limitations in conventional CNN optimisation: that the application of pixelwise loss functions, ignorant of extended spatial context, can result in predictions that lack coherence, and which fail to describe image data in a clinically meaningful fashion.

Interpreting these metrics through the lens of topology, we extend existing persistent homology (PH)-based loss functions for binary segmentation to the multi-class setting. Within a combinatorial framework sensitive to the topology of both individual and combined multi-class labels, these expose the differences between a predicted segmentation and a prior specification of topology according to abstract Betti numbers. We demonstrate the capacity of such losses to reliably make statistically significant improvements in multi-class segmentation topology across a range of 2D and (thanks to our highly efficient implementation based on cubical complexes and parallel execution) 3D cardiac image segmentation tasks, for the first time. Critically we show that our compact, multi-class description of topology informs patient-specific CHD diagnosis. Accordingly, by optimising our PH-based loss functions, CNNs learn a clinically meaningful representation of cardiac defects, overcoming the shortcomings of conventional pixelwise losses.

Though we cannot claim our work heralds an automated solution to the segmentation of patient-specific CHD anatomy from volumetric CMR, we believe that we have made valuable contributions in pursuit of this goal. Whether through our unique training dataset, keen clinical assessments or highly generalisable topological loss functions, we anticipate many applications and extensions of our work. I am incredibly grateful to have had the opportunity to make these contributions, and hope that they drive innovation in the personalised care of all members of the CHD population in the future.

Contents

List of figures

List of tables

List of acronyms

Chapter(s)

AAM	active appearance model	3
ACDC	Automatic Cardiac Diagnosis Challenge (Bernard et al., 2018)	5, 7
Ao	aorta	2, 3, 5
AS	congenital aortic stenosis	2
ASC	2018 Atrial Segmentation Challenge (Xiong et al., 2021)	4
ASD	atrial septal defect	2, 4, 6, 7
ASM	active shape model	3, 4
AV	aortic valve	2, 3
AVSD	atrioventricular septal defect	2
BE	Betti error	7
BP	blood pool	2, 3
BT	Blalock-Taussig	5
CAD	computer-aided design	2, 3
CCA	connected component analysis	6, 7
CE	cross-entropy	4, 7
CHD	congenital heart disease	1–8
CMR	cardiac magnetic resonance	1–8
CNN	convolutional neural network	1, 3–8
CoA	coarctation of the aorta	2
COVID-19	coronavirus disease (2019)	7
CPU	central processing unit	7
CRF	conditional random field	3
CT	computed tomography	1–8

CTA	computed tomography angiography . . . 3
DE	discontinuity error . . . 6
DICOM	Digital Imaging and Communications in Medicine . . . 5
DILV	double inlet left ventricle . . . 2, 7
DKS	Damus-Kaye-Stansel . . . 5, 6
DORV	double outlet right ventricle . . . 2, 5, 6
DSC	Dice similarity coefficient (Dice, 1945) . . . 3, 4, 6, 7
E-M	expectation-maximisation . . . 3
ECG	electrocardiogram . . . 5
ELCH	Evelina London Children's Hospital . . . 1, 4–8
FCN	fully convolutional network . . . 4
fMRI	functional MRI . . . 6
GDSC	generalised Dice similarity coefficient (Crum et al., 2006) . . . 6, 7
GMM	Gaussian mixture model . . . 3
GPU	graphics processing unit . . . 4, 6, 7
GRU	gated recurrent unit . . . 6
HDD	Hausdorff distance (Huttenlocher et al., 1993) . . . 6, 7
HLHS	hypoplastic left heart syndrome . . . 2, 5, 6
HRA	Health Research Authority . . . 5
HRHS	hypoplastic right heart syndrome . . . 5
HVSMR	Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease (Pace et al., 2015) . . . 3–6, 8
IRAS	Integrated Research Application System . . . 5
IVC	inferior vena cava . . . 2, 5, 6
KNN	K-nearest neighbours . . . 3

LA	left atrium 2, 5–7
LAA	left atrial appendage 3
LASC	2013 Left Atrium Segmentation Challenge (Tobon-Gomez et al., 2015) 4
LH	left heart 6
LPA	left pulmonary artery 2, 5, 6
LPS	left-posterior-superior 5
LPV	left pulmonary vein 2, 5, 6
LSTM	long short-term memory 4, 6
LSVC	left superior vena cava 5, 6
LV	left ventricle 2, 5–7
M&Ms	2020 Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (Campello et al., 2021) 5
MA	mitral valve atresia 7
MAPCA	major aortopulmonary collateral artery 2
MAS	multi-atlas segmentation 3, 4
MM-WHS	2017 Multi-Modality Whole Heart Segmentation (Zhuang et al., 2019) 4–7
MPA	main pulmonary artery 2, 3, 5, 6
MRF	Markov random field 3
MRI	magnetic resonance imaging 2–6
MS	mitral valve stenosis 2
MSE	mean squared error 7
MUPPS	multiple path propagation and segmentation 3
MV	mitral valve 2
MY	myocardium 2, 3, 5, 7
PA	pulmonary atresia 2
PACS	picture archive and communications system 1, 5
PC-MRI	phase contrast MRI 3
PCA	principal component analysis 6
PDA	patent ductus arteriosus 2, 5

PDM	point distribution model . . . 3
PET	positron emission tomography . . . 6
PH	persistent homology . . . 7, 8
PLF	patch-based label fusion . . . 4
PPVI	percutaneous pulmonary valve implantation . . . 2
PS	pulmonary stenosis . . . 2
PV	pulmonary valve . . . 2
R-CNN	region-based convolutional neural network . . . 4
RA	right atrium . . . 2, 5–7
REC	Research Ethics Committee . . . 5
ReLU	rectified linear unit . . . 4
REV	réparation à l'étage ventriculaire . . . 5
RF	random forest . . . 3
RGB	red-green-blue . . . 4
RH	right heart . . . 6
RNN	recurrent neural network . . . 4, 6
ROI	region of interest . . . 4
RPA	right pulmonary artery . . . 2, 5, 6
RPV	right pulmonary vein . . . 2, 5, 6
RSVC	right superior vena cava . . . 2, 5, 6
RV	right ventricle . . . 2, 3, 5–7
RVOT	right ventricular outflow tract . . . 2, 3
RVSC	2012 Right Ventricle Segmentation Challenge (Petitjean et al., 2015) . . . 5
SCOT-HEART	Scottish Computed Tomography of the Heart (The SCOT-Heart Investigators, 2015) . . . 4
SE	shunt error . . . 6
SGD	stochastic gradient descent . . . 4
SSFP	steady state free precession . . . 3, 5, 6
STAPLE	simultaneous truth and performance level estimation . . . 3
STEPS	similarity and truth estimation for propagated segmentations . . . 3

SVM	support vector machine	3
TA	tricuspid valve atresia	2
TAPVD	total anomalous pulmonary venous drainage	2
TAVI	transcatheter aortic valve implantation	2
TCPC	total cavopulmonary connection	5, 6
TeTrIS	template transformer networks for image segmentation	4
TGA	transposition of the great arteries	2, 5, 7
TOF	tetralogy of Fallot	2, 3
TP	topological post-processing	7
TR-MRA	time-resolved magnetic resonance angiography	1, 3, 5, 6, 8
TS	topological success rate	7
TV	tricuspid valve	2
VBI	VSD boundary intersection	6
VGG	Visual Geometry Group	4
VSD	ventricular septal defect	2, 3, 5–7
WD	weight decay	4
WH	whole heart	6

Acknowledgments

This report is independent research supported by the National Institute for Health and Care Research (NIHR Doctoral Research Fellowship, Mr Nicholas Byrne, DRF-2017-10-085). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

I am greatly indebted to my supervisor, Andy King.
He is a much better supervisor, than I am a student.

Chapter 1

Introduction

1.1 Motivation

Congenital heart disease (CHD) describes the collection of structural heart defects arising during gestation and which present from birth. The function of affected anatomy (including any of the heart's chambers, valves or associated vasculature) can be severely compromised, necessitating expert clinical management throughout life, including the possibility of surgical or cardiac catheter intervention (Moons et al., 2009).

This already complex clinical challenge is exacerbated by the highly variable presentation of defects. These might: characterise the incomplete growth or development of anatomy, compromising cardiac output; give rise to anomalous connections between normally isolated components of the heart and the mixing of blood; pathologically interrupt the circulation; or in the most complex cases, result in anatomy of indeterminate morphology or which diverges from the normal asymmetry of the thoracic structures. Though bearing on the circulation variably, all diagnoses affect the passage of blood to, through or from the heart.

Coupled with this broad range, select malformations can reasonably affect different parts of the heart. For example (and as explained in Section 2.1), septal defects may permit the communication of the left and right atria, ventricles, or all four cardiac chambers. Moreover, in addition to their localisation to different anatomical sub-structures, defects take a wide variety of shapes, sizes and posi-

tions. In total, these modes of variation combine to realise a patient population that demonstrates significant structural heterogeneity. Hence, in delivering the specialised and personalised care on which these patients rely, a faithful appreciation of patient-specific anatomy and disease morphology is critical (Kim et al., 2008a).

Diagnostic and pre-procedural 3D medical imaging is increasingly used to gain these insights. Assimilating a 3D understanding of anatomy from a 2D, tomographic reconstruction of cardiac magnetic resonance (CMR) or X-ray computed tomography (CT), however, demands significant interpretive expertise. Whilst the make up of multi-disciplinary teams accounts for this challenge through the inclusion of imaging and radiology experts (who develop such skills through specialist experience and training), the geometric complexity of both the healthy and defective heart challenge this paradigm. Even where the specialist might be able to appreciate the 3D anatomy captured by volumetric acquisition, the organic forms characterising cardiac morphology often elude verbal description. In these cases, it is difficult for the expert to convey their understanding to the remaining members of the team, possibly including the surgeons or interventional cardiologists relied upon to deliver treatments.

In response, 3D models of patient-specific anatomy have been proposed as a solution, seeking to capture the imaging expert's understanding of a CMR, CT or other 3D scan in a form that is accessible to all (Byrne et al., 2016). Though a model might ultimately be realised in virtual reality or physically fabricated using a 3D printer, irrespective of its mode of downstream presentation, all such examples are derived from medical image data by image segmentation. Typically involving a laborious and time-consuming computer-based exercise, such pixel-wise labelling demands specialist software and training. Neither the expertise, nor the time required by the segmentation of volumetric data are consistent with the working practices of busy clinical staff. As we shall argue, this gap has limited the growth of patient-specific 3D modelling outside of the largest teaching hospitals and medical research centres. Were its associated burden reduced, expedited segmentation would not only extend the use of 3D models to the care of more patients, but also generate research findings associated with their clinical effectiveness, and further our academic understanding of cardiac anatomy.

The last decade has seen deep learning technologies garner great acclaim. In particular, convolutional neural networks (CNNs), a class of high capacity, non-linear statistical model, have received an enormous amount of research attention. In the medical image processing field, their successes have led some to consider such networks as the state of the art solution to many analysis tasks, including segmentation (Litjens et al., 2017). Their efficacy owes a huge amount to their dependence on data-driven training: rather than on a set of rules or heuristics based on domain knowledge, CNNs are optimised against a set of inputs and outputs which exemplify the task at hand. Once trained, the parameterised operations on which they are based allow for predicted outputs to be determined for unseen cases, including those that are yet to be encountered in the clinic.

However, whilst reports of their strong technical performance are ubiquitous, their clinical adoption, though growing, remains relatively limited. In this work, we firstly seek to understand whether CNN-based segmentation is a credible solution to extract patient-specific CHD anatomy from 3D medical imaging data. Secondly, by paying close attention to the requirements of this task, we present novel methodologies to promote clinically meaningful CNN-based segmentation.

1.2 Contributions

In so doing, we make the following contributions:

The ELCH Dataset

As per any CNN-based image processing solution, supervised optimisation demands learning examples. Having identified a paucity of training data describing the segmentation of congenital cardiac anatomy from isotropically high spatial resolution 3D CMR, we curate the Evelina London Children's Hospital (ELCH) dataset. Searching the picture archive and communications system (PACS) at our local institution, we make a judicious selection of 150 clinically representative patients, straddling five broad diagnostic categories. For each CMR volume, we manually segment a rich, multi-class description of sub-structural, cardiac anatomy, including up to eighteen different semantic labels. Every case is additionally ac-

companied by 4D time-resolved magnetic resonance angiography (TR-MRA). In all of these respects, the ELCH dataset is both qualitatively and quantitatively unique, containing an order of magnitude more patients than its closest CMR counterpart, and being larger than a related CHD dataset concerned with 3D CT.

Clinically sensitive assessments

In the largest study of its kind, we leverage the ELCH dataset to assess the application of CNN-based methodologies, to the segmentation of patient-specific and highly detailed representations of CHD anatomy, from 3D CMR data. Moreover, we investigate novel strategies for learning from combined structural (isotropic 3D CMR) and dynamic (4D TR-MRA) medical images, performing controlled experiments to compare different approaches. In all cases, we assess performance via a range of novel, clinically focused metrics, sensitive to the presence and faithful representation of congenital defects. These reveal the limitations of state of the art methods for CNN-based segmentation, including its propensity to make predictions which lack spatial coherence and fail to represent the most clinically salient features of anatomy, including congenital heart defects. In so doing, we expose topology as a lens through which multi-class labelling (and the patient-specific CHD anatomy described) can be considered, graded and optimised.

Topological loss functions

Carrying forth this observation, we extend existing, persistent homology-based, topological loss functions to the task of multi-class image segmentation. These admit CNN optimisation against an abstract, prior specification of multi-class topology, granting sensitivity to higher-order topological relationships, including class adjacency, hierarchy and containment. We present and leverage an efficient and highly generalisable implementation based on cubical complexes and parallel execution across a range of 2D and, for the first time, 3D segmentation tasks. Critically, this includes assessment against cases drawn from the ELCH dataset, for which we construct a patient-specific topological prior, descriptive of congenital diagnosis and interventional history. In all cases, our topological loss functions make statistically significant improvements in multi-class topology.

1.3 Outline

This work is presented in the following chapters:

Chapter 2 Patient-specific anatomical models of congenital heart disease

We expand on the clinical and technical background presented in Section 1.1, as required to appreciate how patient-specific models support the personalised care of patients with CHD. We review the literature concerning clinical and research applications, and present the workflow for reverse engineering patient-specific anatomy from image data, discussing the practical challenges posed by its implementation, and segmentation in particular.

Chapter 3 Cardiac image segmentation

We conduct a staged literature review of conventional cardiac image segmentation methodologies (those developed prior to the advent of deep learning). In the first stage, we focus on methods adopted by *clinical* exponents of patient-specific heart models, and review their associated operator burden. Secondly, we broaden our scope to the wider literature on *technical* image processing, including a discussion of methodological limitations when applied to patient-specific modelling of CHD.

Chapter 4 Deep learning for image segmentation

Promising to overcome these limitations, we present a theoretical foundation to CNN-based segmentation. We review CNN architectures and their application to the segmentation of 3D cardiac anatomy, observing: (1) a paucity of training data concerning CHD; and (2) deficiencies in CNN optimisation and performance assessment, associated schemes being largely ignorant of the clinically salient features of segmented anatomy. We set out our experimental ambitions in response.

Chapter 5 The ELCH dataset

Addressing the first of these, we describe our curation of the ELCH dataset (as outlined in Section 1.2). Reported characteristics include: its associated manual segmentation protocol; and its clinically relevant demographic, diagnostic, patient

history, and imaging characteristics. In each of these respects, we perform a comparative analysis to establish the unique quantities and qualities that distinguish our data from existing resources.

Chapter 6 CNN segmentation of congenital heart defects

We leverage the ELCH dataset to assess the application of state of the art CNN-based methods to the segmentation of CHD anatomy from CMR images. As previewed in Section 1.2, we ensure the clinical relevance of our findings by: developing bespoke metrics sensitive to the representation of CHD; and explore strategies for incorporating routinely acquired 3D and 4D data. We find that predicted segmentations often lack spatial coherence, hindering clinically meaningful modelling.

Chapter 7 Topological loss functions

In response, we motivate the development of clinically meaningful models by consideration of anatomical topology. This chapter presents our formulation of multi-class, topological loss functions (trailed in Section 1.2), and reports our experiments on 2D and 3D CMR segmentation. Critically, we show that multi-class topology provides a description of cardiac structure that, for cases from the ELCH dataset, we can optimise against our prior knowledge of patient-specific CHD.

Chapter 8 Conclusions

We summarise the work completed and its impact, and identify directions for future investigation. Finally we draw conclusions.

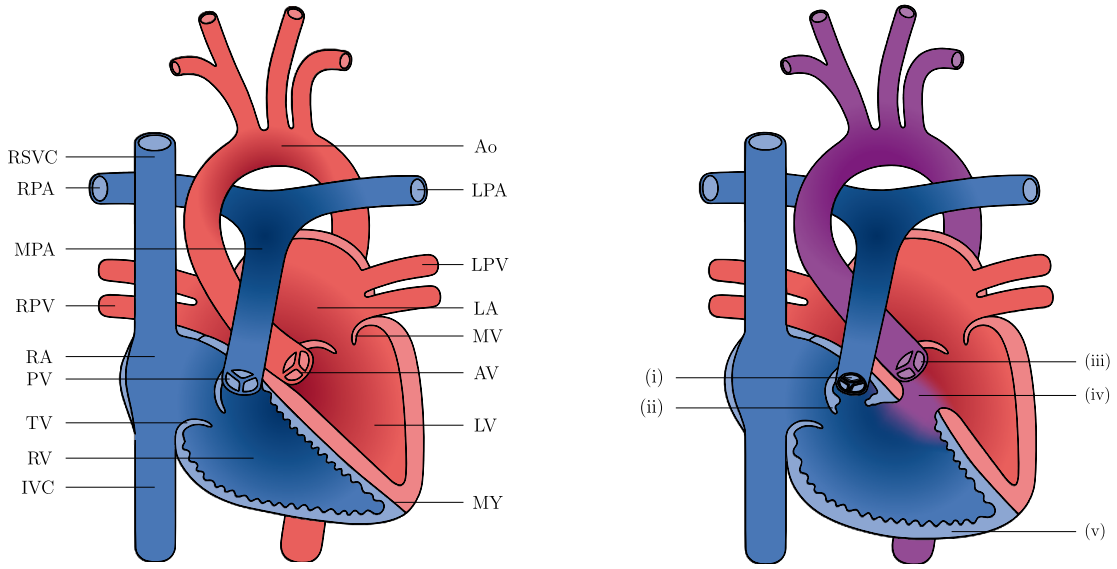
Chapter 2

Patient-specific anatomical models of congenital heart disease

2.1 Congenital heart disease

Estimates suggest that approximately one percent (Triedman and Newburger, 2016) of infants are born with a cardiac abnormality that will have a moderate to severe effect on their health and require expert care (Hoffman and Kaplan, 2002). This group of defects is collectively known as congenital heart disease (CHD). Accounting for nearly one third of all major abnormalities (Van Der Linde et al., 2011), they are the most common type of congenital disorder in new-borns (Van Der Bom et al., 2011). Though many have been proposed, a highly cited definition of this condition is given by: "a gross abnormality of the heart or intra-thoracic great vessels that is actually or potentially of functional significance" (Mitchell et al., 1971). Whilst surgical and medical management of CHD has improved over time, mortality remains high (Oster et al., 2013).

Cardiogenesis describes the normal prenatal development of the heart and culminates in the arrangement of chambers, valves and associated vasculature schematically shown in Figure 2.1a. The healthy heart is consistent with two networks of haemodynamic circulation: (i) deoxygenated blood drains to the right atrium via the systemic veins, crosses the tricuspid valve, and is pumped by the right ventricle, to the lungs, via the main pulmonary artery; (ii) oxygenated blood



(a) Normal heart

(b) Tetralogy of Fallot

Figure 2.1: (a) The normal heart includes many anatomical structures which compose the small (blue) and great (red) circulations. *The small circulation*: the right-sided systemic, caval veins (right superior vena cava (RSVC), and inferior vena cava (IVC)) drain to the right atrium (RA). Guarded by the tricuspid valve (TV), the right-sided atrioventricular junction permits diastolic filling of the right ventricle (RV). During systole, contraction of the myocardium (MY) ejects deoxygenated blood through the pulmonary valve (PV) and, via the main pulmonary artery (MPA) and subsequently branched pulmonary arteries (left pulmonary artery (LPA) and right pulmonary artery (RPA)), to the pulmonary vasculature. *The great circulation*: the left and right-sided pulmonary veins (left pulmonary veins (LPVs) and right pulmonary veins (RPVs), respectively) collect oxygen rich blood from the pulmonary bed, draining to the left atrium (LA). Guarded by the mitral valve (MV), the left-sided atrioventricular junction allows the passage of blood to the left ventricle (LV). Subsequently, systolic contraction of the myocardium forces the ejection of blood through the systemic ventriculoarterial junction (guarded by the aortic valve (AV)), the ascending, arched and descending portions of the aorta (Ao) delivering oxygenated blood to the systemic organs. (b) The cardiac anatomy of a patient with tetralogy of Fallot exhibits a number of defects, disrupting these circulations. Here, the right ventricular outflow tract is obstructed, possibly by (i) infundibular or (ii) pulmonary valve stenosis. Moreover, (iii) the aorta overrides a (iv) ventricular septal defect, allowing blood to mix between circulations and reducing the oxygen saturation of systemic flow (purple). To overcome these defects the wall of the right ventricle thickens. Illustrations taken from `http://www.chd-diagrams.com` (New Media Centre - University of Basel, 2021).

Anomalous connections	Drainage, confluence or connection between the extracardiac vasculature or great vessels and the heart, impinging upon the small and great circulations.
Discordant connections	Paired, anomalous association of small and great circulations at either atrioventricular or ventriculoarterial junctions.
Double connections	Double connections at the ventriculoarterial (atrioventricular) junctions allow blood to pass from a single (both) cardiac chamber(s) to the segments of both (either) small and (or) great circulations.
Hypoplasia	A description of any anatomical segment that has not grown to attain normal size or morphology, remaining small or diminutive (frequently used to describe incomplete ventricular development).
Isomerism	Isomerism of the atrial appendages describes the deviation from the normal arrangement of thoracic and abdominal organs in which both atria share a common left or right morphology.
Obstructive defects	Any defect which obstructs the flow of blood by the narrowing or stenosis of intracardiac tracts, valvular orifices or extracardiac blood vessels.
Septal defects	Holes within the intracardiac septa that normally isolate the small and great circulations, allowing the shunting (and therefore mixing) of blood between the left and right heart.
Situs inversus	Also known as mirror image defects, situs inversus describes the inverse lateralisation of the morphologically left and right atria to opposing sides of the body.
Valvular atresia	An atretic valve causes the absence or closure of a natural atrioventricular or ventriculoarterial junction, interrupting circulation.

Figure 2.2: Types of congenital heart defect.

returns to the left atrium via the pulmonary veins, crosses the mitral valve, and is pumped by the left ventricle, to the systemic organs, through the aorta. The respective small and great circulations act in sequence, without communication or mixing of blood (Thiene and Frescura, 2010).

The presence of a congenital heart defect disrupts these circulations in some way, affecting the haemodynamic continuity of circulation. Though of great structural variety, the *types* of defect listed in Figure 2.2 frequently present in associated combinations to allow widely recognised diagnoses to be made. Tetralogy of Fallot (TOF), for example, combines: obstruction of the right ventricular outflow tract (RVOT); the aorta over-riding a ventricular septal defect (VSD); and right ventricular hypertrophy (see Figure 2.1b). Consequently, pulmonary blood flow is often reduced and, depending on the degree to which the RVOT is narrowed, blood is shunted from the right to the left ventricle (Apitz et al., 2009). Oxygenated and deoxygenated blood mix at the site of the VSD, resulting in cyanotic symptoms: a bluish discolouration of the skin and mucous membranes, caused by elevated levels of deoxyhaemoglobin in the blood.

Grouping congenital heart defects by these features (whether cyanotic or acyanotic) gives rise to the pathophysiological classification of disease shown in Table 2.1. This demonstrates a significant range of defects and diagnoses that is further complicated by the substantial anatomical variation presented by almost every case of CHD (Moore et al., 2018). This results in a structurally heterogeneous patient population, in which no two patients and no two defects are alike. Contending with such anatomical diversity, clinicians have sought increasingly personalised management of the spectrum of disease encountered in the clinic.

2.1.1 Developments in congenital heart disease care

Modern cardiothoracic surgery and the development of minimally invasive cardiac catheter intervention are the products of fifty years' progress in the fields of descriptive anatomy, physiology, surgical technique (Schaffer, 2013), medical imaging and device innovation. Moreover, the development of cardiac surgery was predicated on the emergence of cardiopulmonary bypass (Lillehei, 1957). These advances have fostered the rapid evolution of care for those with CHD, including

	Classification	Examples	Abbreviation
Acyanotic	Left-to-right shunts	Atrial septal defect	ASD
		Ventricular septal defect	VSD
		Atrioventricular septal defect	AVSD
		Aortopulmonary window	—
		Patent ductus arteriosus	PDA
	Left-sided obstructive lesions	Coarctation of the aorta	CoA
		Congenital aortic stenosis	AS
		Interrupted aortic arch	—
		Mitral (valve) stenosis	MS
Cyanotic	Decreased pulmonary blood flow (right-to-left shunts)	Tetralogy of Fallot	TOF
		Pulmonary (valve) stenosis	PS
		Pulmonary (valve) atresia	PA
		Tricuspid (valve) atresia	TA
		Ebstein's anomaly	—
	Increased pulmonary blood flow (complete mixing)	Transposition of the great arteries	TGA
		Double outlet right ventricle	DORV
		Total anomalous pulmonary venous drainage	TAPVD
		Truncus arteriosus	—
	Single-ventricle physiology	Hypoplastic left heart syndrome	HLHS
		Double inlet left ventricle	DILV

Table 2.1: Functional classification of congenital heart lesions (St. Louis, 2008). Note that the acronym PA is often overloaded to stand for both pulmonary atresia (as in this table) and pulmonary artery, its meaning needing to be inferred from context.

the increasing use of novel therapeutic options (Moons et al., 2009). Owing to these advances, the population of adults with CHD is growing, forming more than 60% of the total CHD population (Marelli et al., 2014; Hunter and Swan, 2016). The startling improvement in outcomes for babies born with CHD is one of the success stories of modern medicine (Apitz et al., 2009).

To contend with the spectrum of structural CHD, these innovations have relied on a detailed knowledge of patient-specific disease morphology. The ability of non-invasive imaging modalities to define structural abnormalities has become paramount (Kim et al., 2008a). For this purpose, advances in 3D image acquisition and reconstruction have made tools such as 3D cardiac magnetic resonance (CMR) (Razavi et al., 2003), routinely available. In fact, surgeons in many centres now

prefer CMR or X-ray computed tomography (CT) to a conventional angiographic approach, when planning the most complex operations (Heathfield et al., 2013).

Despite this progress, garnering a 3D, structural appreciation of the congenitally malformed heart *in vivo*, remains hampered by its tomographic representation in imaging data. Visualising volumetric data as a series of effectively 2D slices demands a high level of interpretive expertise if the 3D nature of the anatomy is to be understood (Kim et al., 2008b). Furthermore, even for those imaging (or radiology) specialists able to inspect a tomographic presentation of data, the organic forms and surfaces which define cardiovascular anatomy and CHD morphology frequently elude verbal description. This hampers any attempt to convey expert understanding to the remainder of the multi-disciplinary team. Although these data can be represented in 3D using tools such as volume rendering (Sorensen et al., 2003; Ehret et al., 2018; Schneider et al., 2019), such methods are limited by their dependence on computing resources, their lack of haptic feedback, their incompatibility with surgical or catheter-based simulation (Schmauss et al., 2015) and their unrefined representation of image data of limited quality.

These observations prompt the investigation of advanced forms of 3D representation including: 3D printing (Valverde et al., 2017a), virtual reality (Ong et al., 2018; Deng et al., 2021) and holography (Brun et al., 2019). Given our previous work (Byrne et al., 2016; Valverde et al., 2017a), we primarily rely on 3D printing to motivate this thesis, but our principles are equally applicable to any form of patient-specific anatomical modelling for which 3D representation is derived from medical images by surface rendering. Whilst our focus is on applications within CHD, in the following section we review other use cases from the wider field of cardiology.

2.2 Patient-specific anatomical modelling

2.2.1 Applications of patient-specific 3D printing in CHD

The marriage of 3D printing with computer-aided design (CAD) and medical imaging allows for complex anatomical structures to be reproduced (Mankovich et al., 1990) with an array of mechanical and visual properties (see Figure 2.3). We refer

Figure 2.3: Example patient-specific 3D printed models. Each was derived from medical images and printed to aid clinical decision-making and interventional or surgical planning. All models were printed by Nicholas Byrne or Israel Valverde at St Thomas' Hospital.

to patient-specific 3D printing, an application that seeks to demonstrate anatomy with a structural fidelity consistent with the patient's actual disease processes (Garcia et al., 2018), at least as far as they are represented in image data or can be reasonably inferred by the radiological expert.

Since the first translation of this technology to clinical cardiovascular disease in 2006 (Ngan et al., 2006; Noecker et al., 2006), patient-specific 3D printed models have been employed in a multitude of applications. Published research spans the hierarchy of evidence, from individual case reports to controlled trials. For a complete account, we refer the reader to any of the reviews listed in Table 2.2. However, for the self-contained consistency of this thesis, we present our own review of the literature in the following areas:

Surgical planning

Some of the earliest reports of patient-specific 3D printing being used in cardiology concerned surgical applications. These primarily focused on individual case reports, or small numbered studies, making initial assessments of feasibility, accuracy and the qualitative benefits of this technique. One of the first surgical reports was presented by Ngan et al. (2006), detailing the use of patient-specific 3D printed models of pulmonary atresia with VSD and major aortopulmonary collateral arteries (MAPCAs). Based on their findings from six patients, post-operative questionnaires suggested that models could accurately represent the vast majority of MAPCAs, easing the identification of vascular anatomy. Another early work completed by Vranicar et al. (2008) also examined a case series, printing models

Table 2.2: Reviews of patient-specific 3D printing for clinical cardiovascular care.

Year	Authors	Title
2016	Giannopoulos et al.	Applications of 3D printing in cardiovascular diseases
2017	Foley et al.	3D-printing: applications in cardiovascular imaging
	Grant and Olivieri	The role of 3-D heart models in planning and executing interventional procedures
	Kuk et al.	3D printing from cardiac computed tomography for procedural planning
	Meier et al.	Structural and congenital heart disease interventions: the role of three-dimensional printing
	Otton et al.	3D printing from cardiovascular CT: a practical guide and review
	Valverde	Three-dimensional printed cardiac models: applications in the field of medical education, cardiovascular surgery, and structural heart interventions
	Vukicevic et al.	Cardiac 3D printing and its future directions
2018	Anwar et al.	3D printing provides a precise approach in the treatment of tetralogy of Fallot, pulmonary atresia with major aortopulmonary collateral arteries
	El Sabbagh et al.	The various applications of 3D printing in cardiovascular diseases
	Farooqi and Mahmood	Innovations in preoperative planning: insights into another dimension using 3D printing for cardiac disease
	Hangge et al.	Three-dimensional (3D) printing and its applications for aortic diseases
	Kiraly	Three-dimensional modelling and three-dimensional printing in pediatric and congenital cardiac surgery
	Moore et al.	Three-dimensional printing and virtual surgery for congenital heart procedural planning
	Anwar et al.	3D printing is a transformative technology in congenital heart disease
	Shin and Truong	Manufacturing better outcomes in cardiovascular intervention: 3D printing in clinical practice today
	Uccheddu et al.	3D printing of cardiac structures from medical images: an overview of methods and interactive tools
	Wang et al.	Innovations in cardiac surgery: techniques and applications of 3D printing
	Yoo and Van Arsdell	3D printing in surgical management of double outlet right ventricle
2019	Batteux et al.	3D-printed models for surgical planning in complex congenital heart diseases: a systematic review
	Fan et al.	Three-dimensional printing in structural heart disease and intervention
	Forte et al.	Living the heart in three dimensions: applications of 3D printing in CHD
	Harb et al.	Three-dimensional printing applications in percutaneous structural heart interventions
	Sun et al.	Personalized three-dimensional printed models in congenital heart disease
	Tuncay and van Ooijen	3D printing for heart valve disease: a systematic review
2020	Bateman et al.	Cardiac patient–specific three-dimensional models as surgical planning tools
	Byl et al.	Moving beyond two-dimensional screens to interactive three-dimensional visualization in congenital heart disease
	Ferrari et al.	Three-dimensional printing in adult cardiovascular medicine for surgical and transcatheter procedural planning, teaching and technological innovation
	Gardin et al.	Recent applications of three dimensional printing in cardiovascular medicine
	Garg and Zahn	Utility of three-dimensional (3D) modeling for planning structural heart interventions (with an emphasis on valvular heart disease)
	Hermsen et al.	Three-dimensional printing in congenital heart disease
	Levin et al.	3D printing applications for transcatheter aortic valve replacement
	Sun	Clinical applications of patient-specific 3D printed models in cardiovascular disease: current status and future directions
	Wang et al.	3D printing in adult cardiovascular surgery and interventions: a systematic review
	Wang et al.	Three-dimensional printing for cardiovascular diseases: From anatomical modeling to dynamic functionality
2021	Gharleghi et al.	3D printing for cardiovascular applications: from end-to-end processes to emerging developments
	Ma et al.	Three-dimensional printing for heart diseases: clinical application review
	Wang et al.	3D printing, computational modeling, and artificial intelligence for structural heart disease

for twelve patients exhibiting anomalies of the aortic arch (including coarctation of the aorta (CoA) and vascular ring), their inspection enhancing surgical planning. In more limited reports, Sodian et al. (2007), Jacobs et al. (2008), Sodian et al. (2008a) and Sodian et al. (2008b), all derived 3D representations from the CT data of at most two patients, printing models to aid in the planning of surgery to correct right aortic arch with abnormal left subclavian artery, resection of left ventricular aneurysm, aortic valve replacement and heart transplantation after failed Fontan palliation, respectively.

In the final of these, the anatomical model for one of two examined cases was derived from CMR data. Soon after, the same modality provided the basis of two larger case series interrogating the use of 3D models to plan surgical intervention addressing complex CHD. Whilst Sørensen et al. (2008) relied on virtual models of 42 patients, Riesenkampff et al. (2009) printed heart models for eleven patients, achieving biventricular repair in five cases for whom consensus decision remained unequivocal after conventional radiological review.

Since this time and to this day, case reports on the surgical application of patient-specific 3D printing continue to be published. For example, Valverde et al. (2015b) found this approach to be extremely helpful when planning a Nikaidoh repair for a patient with transposition of the great arteries (TGA) and VSD. Similar findings were reported in the context of: cardiac tumour resection (Al Jabbari et al., 2016), VSD closure (Bhatla et al., 2017a), transplant (Smith et al., 2017; Yoo et al., 2020), septal myectomy (Hermsen et al., 2017; Andrushchuk et al., 2018), TOF (Olejnik et al., 2018) with pulmonary atresia (Averkin et al., 2022), arteriovenous malformation (Carberry et al., 2019), pulmonary venous baffle obstruction post atrial switch (Schneider et al., 2019), coronary artery fistula (Zhang et al., 2019a; Aroney et al., 2019), double-chambered right ventricle (Mokkarala et al., 2020), and left ventricular outflow tract obstruction post mitral valve replacement (Sodian et al., 2021).

Thankfully, however, larger case series have also become more prevalent, now making up the bulk of high impact, published work. For example, Ma et al. (2015) demonstrated the utility of patient-specific printed models derived from CT data, successfully planning surgical repair for 35 patients with TOF, and establishing dimensional agreement with intra-operative measurement. In respective studies of

five patients with double outlet right ventricle (DORV), Dydynski et al. (2016) and Garekar et al. (2016) found that 3D models revealed the relationship and separation of VSD from the ventricular outflows. In the same sized cohort, Kappanayil et al. (2017) made similar findings across a range of complex CHD, including DORV, criss cross heart and congenitally corrected TGA. Others have focused on particular methodological improvements, such as the use of super-flexible printed models for twenty patients with CHD (Hoashi et al., 2018), or device sizing for twenty cases requiring surgical aortic valve replacement (Faletti et al., 2018).

The final group of publications detail concerted efforts to assess the impact of patient-specific 3D printing on the outcomes of cardiovascular surgery. For example, reflecting their three-year experience, Ryan et al. (2018) performed a retrospective comparative analysis, contrasting the care of 33 CHD patients, each of which benefited from the preparation of a printed anatomical model, with 113 contemporaneous cases which did not. Whilst their results suggested an association between the use of 3D printing and reduced surgical time, this was not statistically significant. Perhaps surprisingly, in a similar study design, Zhao et al. (2018) found patient-specific 3D printing to be associated with a statistically significant reduction in mechanical ventilation time and duration of stay in the intensive care unit. Despite the promise of these findings, they are compromised by a lack of prospective or randomised sampling and substantial differences between the surgically relevant covariables expressed by the experimental and controlling arms of the study. Being eight and seventeen, respectively, the considered sample sizes appear inadequate to represent the underlying population of eligible patients with DORV. Alternate study designs have also been considered: Han et al. (2019) sought age and diagnostically matched controls for eighteen consecutive patients with CHD. Deriving printed anatomical models from each of echocardiography, CT and CMR, they observed a trend toward shorter surgical times, but without statistical significance.

To this point, the preceding citations have sought to expose differences in quantitative metrics of surgical execution (operating or cardiac bypass time, for example). Whilst attractive for their evidence-based association with patient outcome (Al-Sarraf et al., 2011), each is compromised by the statistics of their underlying distribution within normal care. Due to the substantial heterogeneity of the CHD

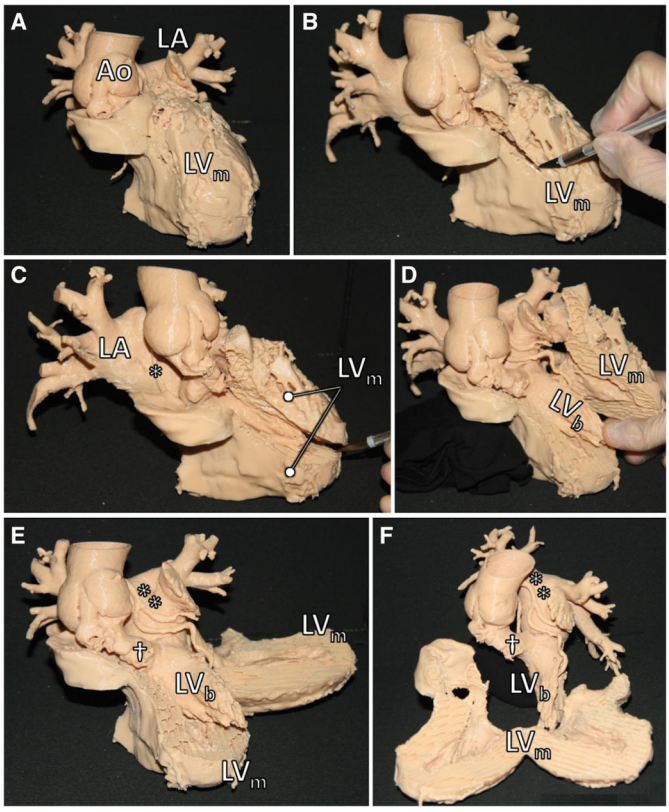


Figure 2.4: Incision and tactile manipulation of patient-specific 3D printed models have been found to enhance decisions made by multi-disciplinary consensus and surgical planning. See Valverde et al. (2017a) (from where this figure is reproduced) for full details of the steps (A to F) of the anatomical dissection.

population, such metrics frequently demonstrate high standard deviation (or other measures of spread), so as to limit the applicability of conventional quantitative study designs to prohibitively large samples. In response, several authors have instead opted to assess the impact of patient-specific 3D printing on pre-surgical planning, measuring the rate at which anatomical models prompt a change in operational approach.

From a group of six cases, Bhatla et al. (2017b) found that patient-specific 3D printing altered clinical management of complex DORV. Inspired by Riesenkampff et al. (2009), Valverde et al. (2017a) found that in an international, multi-centre sample of forty patients with complex CHD, inspection of a 3D printed model modified the surgical decision in nineteen cases. Critically, in four patients where conventional radiological review had indicated univentricular palliation, enhanced 3D planning allowed successful biventricular repair, a decision with far reaching consequences for the individuals concerned (see Figure 2.4). Similar findings have been made recently: Kiraly et al. (2021) found that 3D printing improved the sur-

gical approach or made modification to the planned biventricular repair in thirteen of fifteen paediatric CHD cases; Tiwari et al. (2021) observed a change in planned surgical management in eight out of ten patients; in a group of eighteen patients with complex CHD, Yıldız et al. (2021) found 3D models altered the planned approach in one third of cases; and finally, albeit in a cohort of fourteen adults with structural heart disease, Borracci et al. (2021) found that 3D printing refined surgical planning in six, and verified device delivery in three patients.

Planning of cardiac catheter intervention

The incorporation of patient-specific 3D printing within the planning of catheter-based intervention approximately coincided with the surgical application of this technique. However, unlike surgical investigation - where a sense of clinical impact has been investigated relatively recently - early reports from the interventional cardiology arena sought this understanding from the outset. Publications from around 2007 considered the highly variable morphology of the RVOT, as relevant to percutaneous pulmonary valve implantation (PPVI). Schievano et al. (2007) described the impact of 3D printed models on clinical-decision making. In a retrospective evaluation of twelve cases, they found that compared with a conventional presentation of magnetic resonance imaging (MRI) data, patient-specific models improved the ability of two observers to assess PPVI suitability. Critically, with the benefit of a printed model, both were able to predict the failure of PPVI in two patients in which this procedure was attempted.

Even if their innovative study design would not be replicated until years later, their approach to *in vitro* testing (Armillotta et al., 2007) proved a highly attractive means of enhancing interventional planning. Such procedures depend on the judicious selection and sizing of balloon catheters, stents, occlusion, valvular or other devices. Accordingly, 3D printed models afford an opportunity to practically deploy and test such devices, ascertaining their suitability to patient-specific anatomy prior to intervention. Sulaiman et al. (2008) deployed and tested an endovascular prosthesis within a patient-specific model of aortic aneurysm, their simulation benefiting from perfusion using an extracorporeal pump, and fluoroscopic guidance. To this day, further case reports of pre-operative simulation

include consideration of mitral valve annuloplasty (Dankowski et al., 2014); stenting to relieve hypoplasia (Valverde et al., 2015a), CoA (Ghisiawan et al., 2016) or aneurysm (Meyer-Szary et al., 2019) of the aortic arch; atrial septal defect (ASD) closure (Chaowu et al., 2016) in the setting of complex CHD (Imai et al., 2018); transcatheter aortic valve implantation (TAVI) (Fujita et al., 2017); planning to confirm (Little et al., 2016; Bagur et al., 2018) or reject percutaneous mitral valve implantation (Lavie-Badie et al., 2021); percutaneous aneurysmal closure and valve implantation within the RVOT (Jivanji et al., 2019); and closure of left ventricular pseudoaneurysm (Quimby Jr et al., 2022). Cases presented by So et al. (2017) and Oliveira-Santos et al. (2018) are notable for their derivation of anatomical models from lesser used (transoesophageal) echocardiography and rotational angiography. The former leveraged improved visualisation of the atrial septum to plan ASD closure; the latter used a model of the coronary arteries to guide complex intervention. Lastly, Pracon et al. (2018) published an incredible case report detailing the selection of an Amplatzer septal occluder to close a VSD sustained after a stab wound.

Thankfully, and as per advances in the surgical field, reports of larger case series have also emerged. Forte et al. (2017) presented four patients (both adult and paediatric) with coronary artery fistula. Multi-disciplinary discussion was enhanced by inspection of flexible, patient-specific printed models, revealing the tortuous course of dilated fistulous anatomy. *In vitro* simulation identified optimal points for occluder deployment, allowing percutaneous solutions to be planned in three patients and discouraged in favour of surgery in the fourth. In a study of the same size, Aroney et al. (2019) made identical findings. Vukicevic et al. (2017b) demonstrated the use of both CT and echocardiography to model the mitral valve, using benchtop simulation to plan mitral valve catheter intervention in three patients. In a retrospective study of eighteen patients, Qian et al. (2017) used printed models to simulate TAVI, conceiving a novel annular bulge index for the prediction of paravalvular leak with statistical significance. Similar work comparing *in vitro* and *in vivo* measurement was undertaken by Sommer et al. (2020): using 3D models of 52 patients, they established the consistency of CT and bench top assessments of coronary fractional flow reserve. Li et al. (2020) used patient-specific 3D printing to investigate and then plan the closure of multiple

ASDs using a single device, relying on transthoracic echocardiography rather than fluoroscopic guidance at the point of intervention. Comparing thirty procedures planned in this way with 32 undergoing normal care, they found that 3D planning yielded statistically significant reductions in the presence of residual shunts and the rate of re-occlusion.

To close our review of catheter-based intervention, we wish to highlight two clinical applications, each demonstrating different benefits conferred by the application of patient-specific 3D printing. Whilst in the context of acquired rather than CHD, the first uses 3D models to plan occlusion of the left atrial appendage, an established strategy to manage complications of atrial fibrillation (Iriart et al., 2018). Case reports nicely summarise the advantages of patient-specific printing, including the selection and sizing of devices in the context of anatomy with highly variable structural presentation (Fan et al., 2016; Khalili et al., 2017). In eight patients, Liu et al. (2016) established the feasibility of replicating the left atrial appendage from transoesophageal echocardiography. Several larger studies have also been able to demonstrate the clinical impact of this approach observing statistically significant reductions in radiation exposure (Li et al., 2017a), procedure time and anaesthetic time, and residual peri-device leak (Obasare et al., 2018); and improved device sizing (Hachulla et al., 2019). However, these conclusions have not always been reproduced so emphatically. Conti et al. (2019) and Hudec et al. (2021) found that devices selected after *in vitro* simulation matched those clinically deployed in only 35% and 37% of cases, respectively.

Finally, we review the literature relevant to partial anomalous pulmonary venous drainage and sinus venosus ASD. In this setting, Velasco Forte et al. (2018) demonstrated the ability of 3D models to not only enhance interventional planning, but also to conceptualise and validate a novel catheter-based solution to a lesion typically treated surgically (see Figure 2.5). Case reports published by Thakkar et al. (2018) and Huang et al. (2019) also endorse this application of *in vitro* simulation. This innovative body of work has culminated in a credible alternative to surgery, successfully implemented in 25 patients (Riahi et al., 2018; Hansen et al., 2020); patient-specific 3D printing playing a critical role in its development.

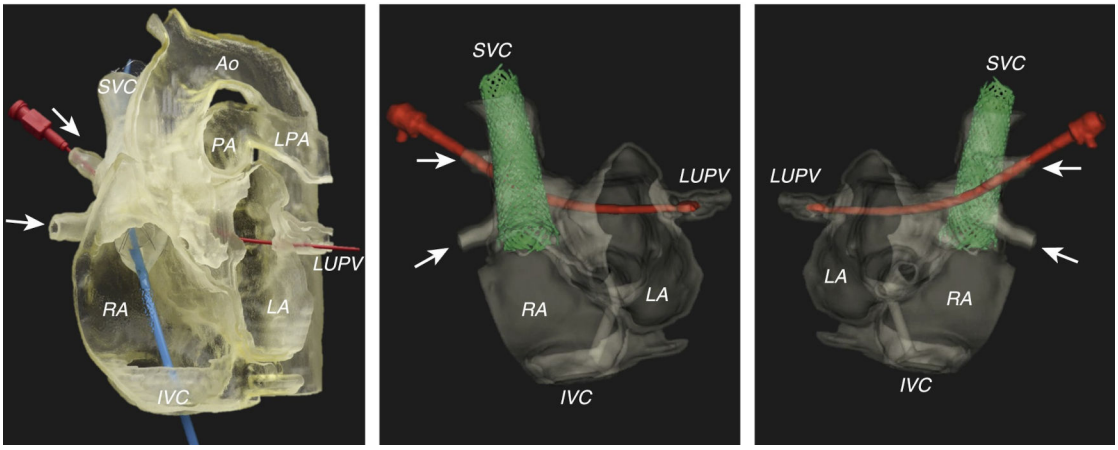


Figure 2.5: Patient-specific 3D printed models permit *in vitro* simulation, including device deployment and planning of catheter-based intervention. They have also fostered and justified bespoke interventions such as in work by Velasco Forte et al. (2018) (from where this figure is reproduced), who developed a novel catheter-based solution to partial anomalous pulmonary venous drainage and sinus venosus atrial septal defect.

Medical education and training

Given the capacity of printed models to demonstrate the detailed and patient-specific anatomy of the structurally variable CHD population, they naturally find application in medical education and training. Classically, our understanding of CHD anatomy is developed and taught by cardiac morphologists, who in turn develop their knowledge through inspection of cadaveric cardiac specimens. However, requiring constant maintenance, such specimens represent an an ultimately perishable resource. Consequently, researchers have examined the feasibility and accuracy of using CT (Greil et al., 2007) and MRI (Kiraly et al., 2019) to "digitally preserve" the *ex vivo* heart, optionally providing a basis for 3D printing. Deakyne et al. (2019) recently presented their experiences of using high resolution MRI to prepare a library of over a hundred virtual heart models, also reflecting on the possibility of device implantation *in silico*.

Avoiding the technical challenges of developing a virtual simulation environment, others have investigated the use of patient-specific models in simulation-based training (distinct from the pre-procedural simulation for pre-operative planning reviewed previously). Early work demonstrated the feasibility of this ap-

proach to simulate catheter-based intervention (Abdel-Sayed et al., 2009) and surgical simulation, including the use of printed materials that allow for incision and suture (Shiraishi et al., 2010). More recently, researchers have sought to assess the *impact* that such models can have on training. In surgical applications: Costello et al. (2014, 2015) quantified the incremental value of simulating closure of five common VSD sub-types after conventional teaching; Yoo et al. (2017) developed "Hands On Surgical Training" for a range of operations relevant to CHD, receiving positive feedback from fifty surgeons; and Chen et al. (2018b) and Hussein et al. (2021) both simulated reconstruction of the aortic arch as relevant to the Norwood procedure, concluding that such an approach was feasible, and could improve surgical skill through practice. From these studies, important feedback highlighted the limited capacity of 3D printing materials to replicate the mechanical properties of cardiac tissue, and the sub-optimal representation of valvular anatomy. Perhaps for these reasons, in catheter-based applications, Jang et al. (2020) recently reported a mixed reality procedural simulator, combining virtual and printed models. For more comprehensive overviews of the use of patient-specific 3D printing in simulation-based, CHD training, we refer the reader to Subat et al. (2018) and Hussein et al. (2020).

Finally, a slew of recent publications have used patient-specific printed models for anatomical demonstration, considering their incorporation within classrooms, workshops, lectures, seminars or similar teaching environments (see Figure 2.6). Biglino et al. (2017b) enhanced nurse education through nine 3D printed examples, administering a survey to elicit participant feedback. Simpler studies of impact have relied on a before and after, or crossover design: both Smerling et al. (2019) and Lee and Lee (2020) captured a range of CHD with several printed models, finding statistically significant improvement in the students' self-graded, subjective knowledge and understanding of anatomy; Valverde et al. (2022) focused on the particularly complex criss cross heart, also recording a statistically significant improvement in learners' objective test scores after inspection of 3D models. More complex designs have employed randomisation, comparing conventional classroom teaching (control group) with that enhanced by 3D printed hearts. Some have reported that patient-specific models significantly improve self-reported knowledge and skill acquisition, and student satisfaction. These include educational courses

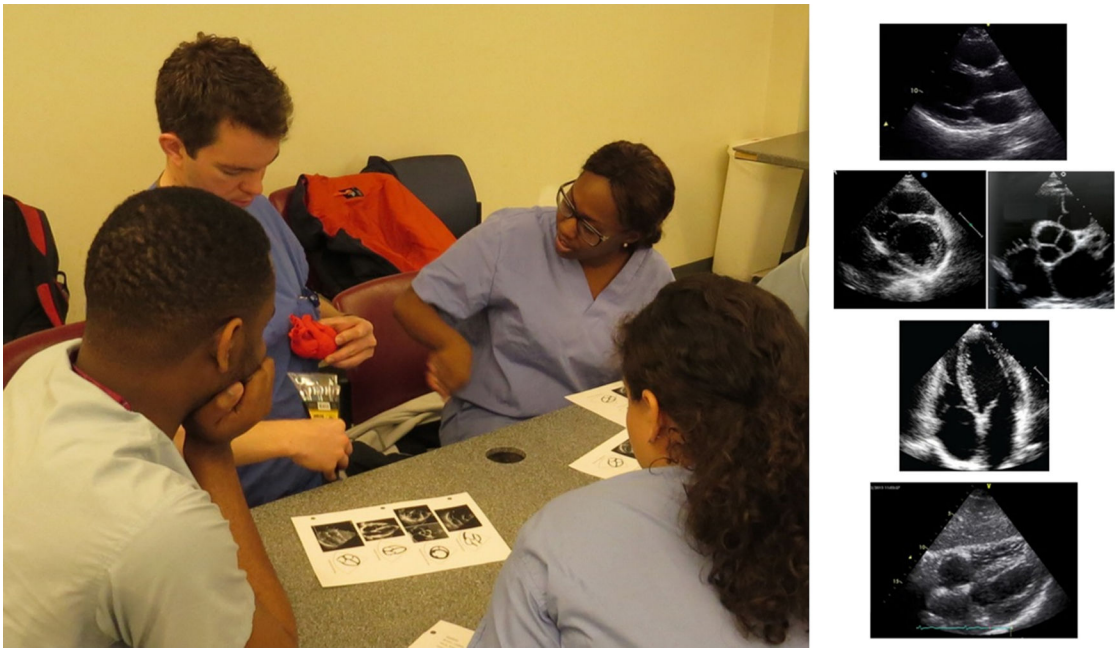


Figure 2.6: Patient-specific 3D printed models have been used to demonstrate cardiac anatomy, augmenting conventional medical training and education. In this example (reproduced from the work of Ochoa et al. (2019)), models were used to contextualise echocardiographic examination.

on VSD (Su et al., 2018), ASD, CoA and TOF (Karsenty et al., 2021). In other studies, however, the efficacy of this approach was found to be mixed (Loke et al., 2017); or make no difference to student learning (Wang et al., 2017b; Ochoa et al., 2019). As observed by Smerling et al. (2019), it is possible that the benefit conferred is related to the complexity of disease.

Patient communication

Clinicians and clinical trainees are not the only ones who stand to benefit from a detailed and 3D knowledge of patient-specific anatomy. Though drawn from a smaller section of the research literature, Biglino et al. (2015) were the first to publish their experience of using printed heart models to aid consultant-patient communication. Randomising 103 parents of children with CHD to either conventional clinical consultation or that enhanced by a model specific to their child, they made objective assessments of parental knowledge and sought subjective feedback

Figure 2.7: Patient-specific 3D printed models have been used to communicate cardiac anatomy to patients and members of the public. In this example (reproduced from the work of Biglino et al. (2019)), models were presented as the centrepiece of an artistic installation concerning CHD.

prior to and following attendance in clinic. Whilst they found no statistically significant difference in knowledge acquisition between groups, questionnaire feedback prompted the conclusion that patient-specific models can enhance parental engagement and communication with cardiologists. Biglino et al. (2017c) extended this approach to a pilot study of adolescent patients with CHD. On the other hand, Zablah et al. (2021) observed statistically significant improvements in parental understanding of proposed cardiac catheter intervention, associated with inspection of a heart model specific to their child. However, these findings are limited by the use of non-randomised design, subjects serving as their own control via an entirely subjective questionnaire completed before and after 3D model review.

Finally, Biglino et al. (2019) examined the wider capacity of printed models to communicate and convey the experiences of patients and families affected

by CHD. Collaborating with an artist, they canvassed the response of viewers attending an exhibition which included patient-specific hearts, observing empathetic engagement with the topic (see Figure 2.7).

Research applications

The capacity of 3D models (printed or otherwise) to capture and demonstrate patient-specific anatomy extends their application beyond purely clinical use, also enabling scientific research. Whilst not our main motivation, these models have found utility within both *in silico* and *in vitro* modelling of either cardiac physiology and its disease processes, or device development and procedural simulation. Bodies of work in their own right, for comprehensive reviews of these topics, we direct the reader to Niederer et al. (2019) (concerning applications in cardiology more generally) and to Biglino et al. (2017a) (concerning CHD in particular). Here we provide only a brief and incomplete review of citations thought most relevant to CHD, cardiac care or their association with 3D printing.

Most immediately, printed anatomical models provide a basis for *in vitro* simulation and haemodynamic measurement. For example, in the context of different strategies for single ventricle palliation (including Norwood (Biglino et al., 2012b) and Sano (Biglino et al., 2013) procedures), researchers have measured pressure within printed models of the aorta, also considering the impact of coincident CoA. Rather than by direct, instrumented measurement, others have assessed *in vitro* haemodynamics through advanced MRI acquisition. Using 4D flow MRI (Medero et al., 2017), Ha et al. (2016) examined the association between aortic valve angle and haemodynamic performance. Meanwhile, Falk et al. (2018) sought to understand flow patterns within printed models of prenatal cardiac abnormalities. Given the visual quality of the models relied upon, incredibly, each was derived from prenatal echocardiography (see Figure 2.8). Also using 4D flow, Markl et al. (2012) quantified velocity, flow distribution, vorticity and kinetic energy within total cavopulmonary connection, their primary motivation being to establish methods for validating *in silico* models of fluid dynamics. This ambition was tackled more directly in works investigating aortic flow (Canstein et al., 2008; Wen et al., 2010), including in the presence of CoA (Biglino et al., 2014).

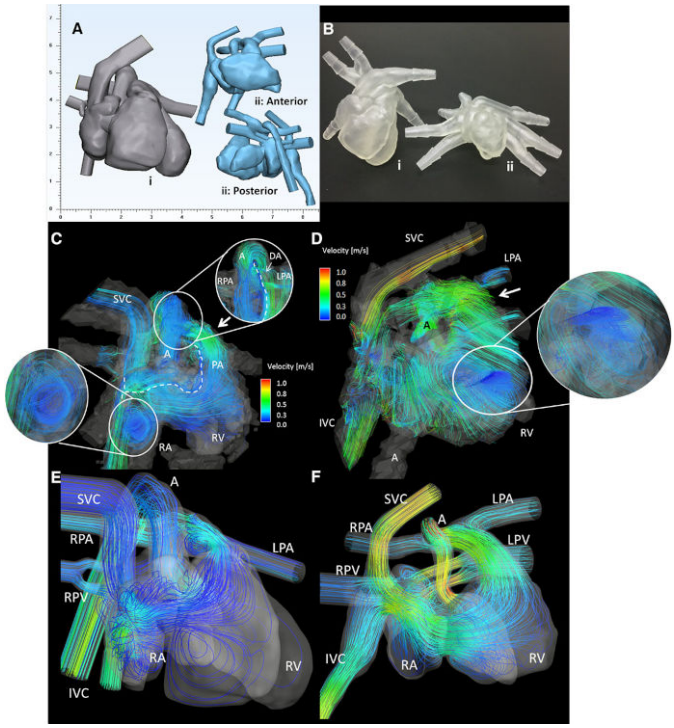


Figure 2.8: Patient-specific 3D printed models have been used to undertake scientific research concerning the anatomy and physiology of the developing heart. In this example (reproduced from the work of Falk et al. (2018)), printed models derived from foetal echocardiography were incorporated within a circulatory simulation. 4D flow MRI subsequently revealed the distorted haemodynamics associated with CHD.

Finally, patient-specific models provide suitable geometries for testing and developing medical devices that must conform to or modify anatomy: Saeed et al. (2008) virtually implanted a developmental left ventricular assist device, assessing the degree of tissue interference associated with different placement positions; Sodian et al. (2009) used a printed model to understand aneurysmal disease of the aorta, manufacturing a custom-made occluder in response to their findings; and Capelli et al. (2012) used finite element analysis to understand the mechanical stresses associated with TAVI, demonstrating wider feasibility in borderline cases. A number of works have relied on patient-specific representation of the RVOT to analyse and develop novel devices for PPVI. These include *in silico* (Capelli et al., 2010) or *in vitro* modelling (Schievano et al., 2010) or both (Biglino et al., 2012a).

Summary

The common threads amongst the preceding body of work establish the mechanisms by which patient-specific 3D printing and anatomical modelling might enhance the care of patients with CHD. In clinical applications (our focus), patient-

specific 3D printed models afford a tactile interaction with image data. Through haptic feedback, direct manipulation and simulation, these models foster an enhanced structural understanding of cardiovascular anatomy and disease morphology. Unlike a tomographic presentation of data, this understanding is achieved rapidly by inspection, independently of the viewer's (be they medical, technical, patient or lay observer) expertise in medical image interpretation. This level of accessibility admits a common understanding of disease morphology, one that is shared amongst the entire multi-disciplinary team. We suggest that such enhanced communication provides the conditions necessary to facilitate optimal multi-disciplinary consensus and subsequent treatment planning, providing opportunities for highly personalised care.

In view of these qualities, it is tempting to promote patient-specific anatomical modelling as a "democratisation" of medical imaging: one which makes the expert interpretation of data available to all (Hermsen et al., 2020). However, and as we explore in the following sections, this is to overlook the technical and clinical demands of the patient-specific modelling workflow, and the associated expertise it currently relies upon.

2.2.2 Workflow

The ambition of the patient-specific 3D printing workflow is to extract a 3D representation of the heart from an image volume, in a form that is suitable for downstream fabrication as a 3D printed model. Hence it can be described as anatomical reverse engineering. Clinical applications of this technique can largely be summarised by the five major steps shown in Figure 2.9:

Image acquisition

Any 3D printed model is fundamentally limited by the appearance of anatomy within the imaging volume on which it is based. At minimum this must expose the cardiac and vascular blood pool separately from the myocardium and remaining anatomy (Yoo et al., 2016). Minimally acceptable image quality must be considered in the context of the clinical indications that motivated printing. Consequently, trends relating imaging modalities to anatomical targets have emerged.

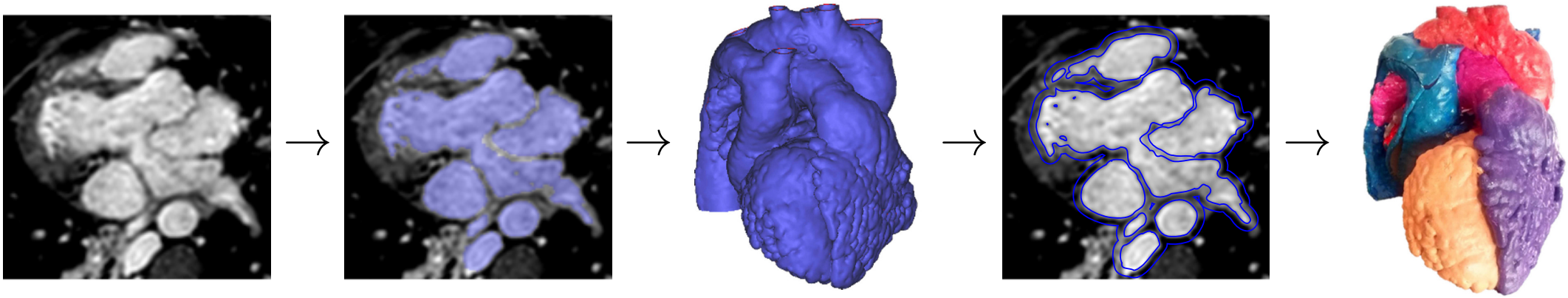


Figure 2.9: The patient-specific 3D printing workflow. Note that for those steps illustrated by medical images (annotated or otherwise), the associated operations are completed throughout the imaging volume. In clinical applications, the validation completed in the fourth step ensures that the 3D model reflects the anatomical features salient to downstream care.

CT (Kuk et al., 2017; Otton et al., 2017) and CMR lend themselves to the representation of the intracardiac blood pool and the myocardium. Extracardiac vasculature is well-represented in contrast-enhanced CT and time-resolved magnetic resonance angiography (Giannopoulos et al., 2016). Finally, whilst echocardiography has more often been used to replicate valvular anatomy (Witschey et al., 2014; Vukicevic et al., 2017a; Muraru et al., 2017; Daemen et al., 2019), its range of applications continues to expand (Birbara et al., 2019; Mowers et al., 2021).

Detailed acquisition desiderata can be found in Giannopoulos et al. (2016); Yoo et al. (2016). However, from our experiences within paediatric CHD, we find that independent of imaging modality, an isotropic pixel spacing of at most 1 mm is required. Combined with both electrocardiogram gating and respiratory navigation, this constraint admits sufficient spatial resolution for the representation of the majority of fine structure (such as the atrial septum). Beyond this requirement, we advocate the use of routinely acquired image data as far as its quality remains consistent with the motivations for 3D modelling. To account for the presence of artefacts, we find that suitability is best assessed on a case by case basis.

Finally, we stress the importance of using multi-modal imaging to inform 3D representation. Adjuvant to a tomographic volume to be segmented, complementary imaging sources can enhance 3D models. Without requiring 3D acquisition or spatial registration, these might only implicitly inform downstream segmentation of CAD adaptation. For example, of those modalities presented, conventional 2D echocardiography best visualises the thin tissue interface comprising the atrial septum, and reveals defects therein. Therefore, where tomographic acquisition might be compromised, insights gained from the inspection of echo data can shape the segmentation of CHD anatomy. This approach draws on the conventional role of the radiologist or cardiac imaging specialist, leveraging their capacity to mentally assimilate anatomical structure from a diverse set of imaging data.

The natural extension of this principle, so-called “hybrid models”, rely on spatial registration to combine the findings of disparate image data explicitly (Kurup et al., 2015). Most often, these have leveraged the high fidelity visualisation of valvular anatomy (including the sub-valvular apparatus and chordae tendinae) by 3D echocardiography (Gosnell et al., 2016; Gomez et al., 2020), making up for their sub-optimal appearance within both CT and CMR data (Anwar et al., 2018b).

Image segmentation

For the purpose of patient-specific 3D printing, image segmentation describes the task of partitioning the volume into a set of clinically meaningful foreground objects against a background, assigning class labels at the level of the individual voxel. This description is consistent with a semantic classification task, each element of the image necessarily belonging to a single class. Depending on their motivations for printing, exponents of patient-specific anatomical models have taken different approaches to the formulation of this task. These might differ in the foreground segments targeted; or the granularity with which hierarchically or laterally organised anatomical features are discriminated.

Guided by those reports concerning printed anatomy, for the most part users have shown interest in two varieties of model. The first bounds the anatomical surfaces of the blood pool using a uniform wall thickness; the second also demonstrates the mural thickness of the myocardium (see Figure 2.10). Though the "RepliCast" and "RepliCardio" terms respectively coined to describe these varieties of model (Mottl-Link et al., 2008) have not survived, interest in both, and their comparison, remains (Farooqi et al., 2016; Liang et al., 2021). Each variant relies on a different formulation of the segmentation task. Models of the blood pool require only a binary specification that separates the blood pool from the background (see Figure 2.10b). Whereas, myocardial models additionally demand delineation of the epicardial surface, most often achieved via segmentation of the myocardium (see Figure 2.10c).

These formulations provide a basis for the majority of models of or derived from whole heart anatomy: including the chambers, intracardiac tissues, muscle and associated vasculature. However, clinical applications frequently only demonstrate a subset of these structures (for example, consider models of the RVOT, as relevant to the planning of PPVI (Schievano et al., 2007)). In these cases an anatomically-limited task specification might be applicable, focusing on individual sub-components of the cardiovascular blood pool. Taken together, these foreground objects describe the multi-class blood pool (see Figure 2.10a). Distinct from the whole heart blood pool, a multi-class specification differentiates each chamber or vessel as a separate class of the segmentation. In addition to providing

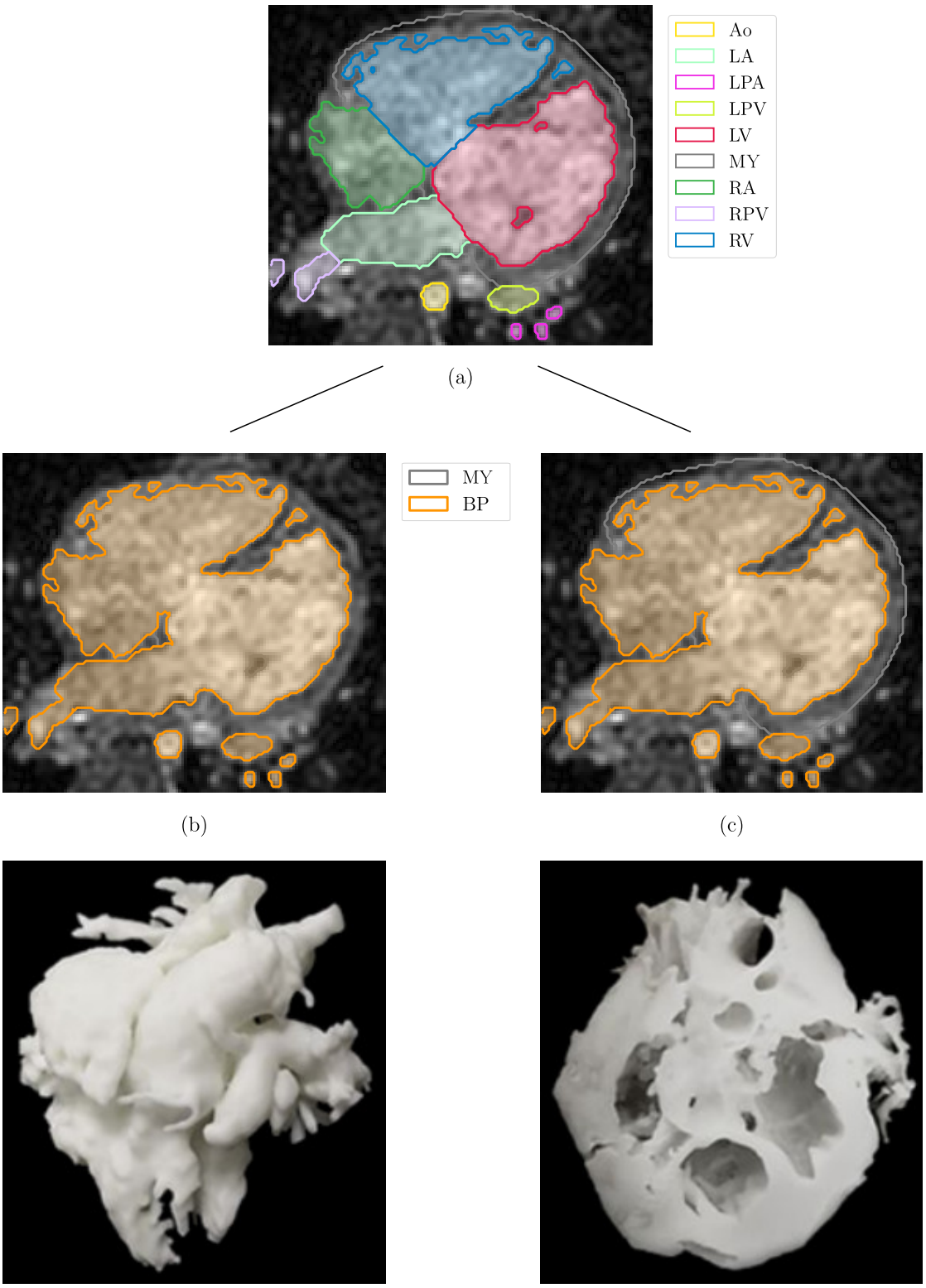


Figure 2.10: The formulation of a relevant segmentation task specification should match the motivations for printing and the eventual model to be fabricated. (a) Foundationally, a range of task specifications can be established by the selection and union of multi-class, granular cardiovascular anatomy, including but not limited to: the myocardium (MY); aorta (Ao); left atrium (LA), left pulmonary artery (LPA), left pulmonary veins (LPVs), and left ventricle (LV); and right pulmonary veins (RPVs) and right ventricle (RV). (b) One formulation reflects printing of the endocardial and internal surfaces of the cardiovascular blood pool (BP). (c) Another formulation allows for models to be printed which also demonstrate the mural thickness of the MY. Photos of printed models reproduced from (Liang et al., 2021).

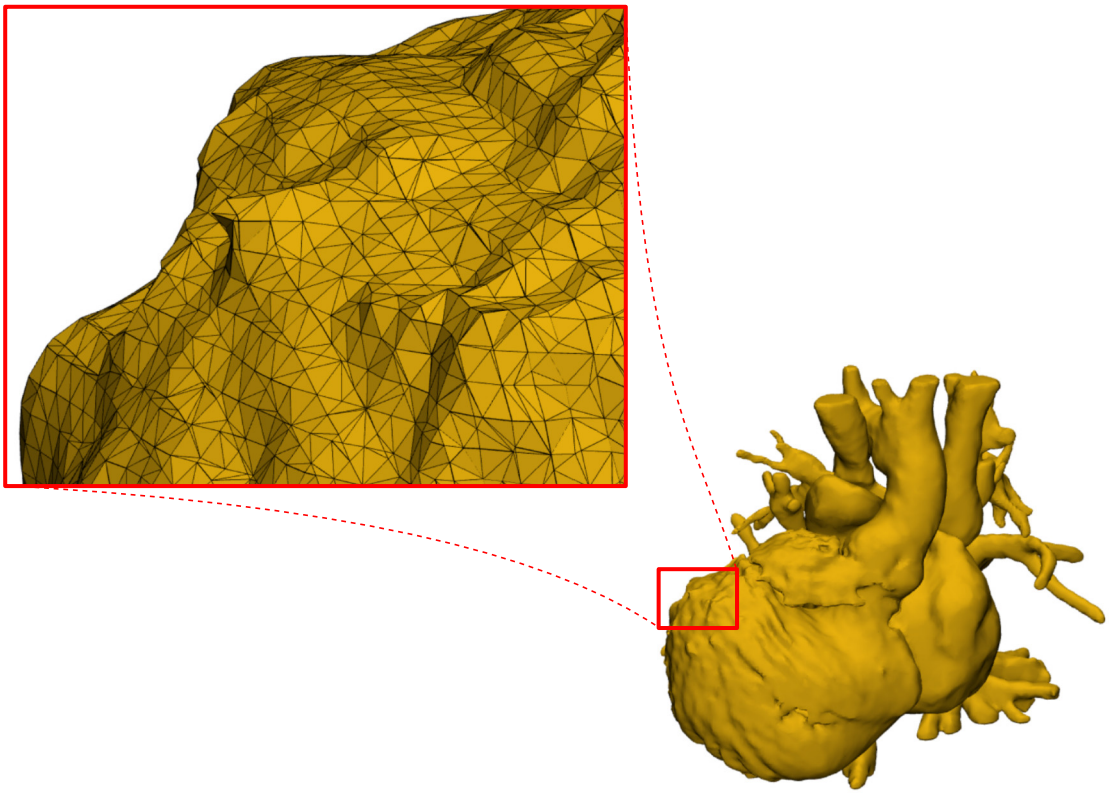


Figure 2.11: Surface rendering describes an anatomical segmentation by a 2D manifold of tessellating polygons, often triangles. This case demonstrates a patient with dextrocardia, with the cardiac apex pointing to the patient's right hand side.

a basis for specific anatomical targets, multi-class anatomy presents a greater degree of flexibility to any post-processing or design adaptation, in aid of a broader array of downstream applications.

Irrespective of task formulation, volumetric segmentations are used to render anatomical surfaces, providing the basis of downstream application. Per class, algorithms for extracting the isosurface of a binary segmentation include classical marching cubes (Lorensen and Cline, 1987) and its extensions, and those based on dual contouring methods (Ju et al., 2002). Each represents segmented anatomical surfaces by a 2D manifold of tesselating polygons (most frequently triangles) embedded within 3D space (see Figure 2.11).

Computer-aided design

A surface-rendered representation of the segmented image might be appropriate for 3D printing. More frequently, however, some adaptation by CAD is required (Giannopoulos et al., 2016). This might be motivated by technical considerations: to comply with the requirements of downstream printer hardware or overcome mesh errors, for example. Alternatively, CAD affords the opportunity to add designed features to the segmented surface, enhancing the functionality of the eventual patient-specific model. The operator may wish to introduce a wall thickness to surround a hollow void; proximally terminate extracardiac vessels which might otherwise obscure anatomy (Pace et al., 2015) or introduce sections to ease visualisation of intracardiac defects (see Figure 2.10c).

These are relatively simple options amongst the array of operations familiar to exponents of CAD. Moreover, the reliance of the manufacturing industry on software-based design processes has ensured the availability, continuous development and support of cutting edge CAD packages. Commercially available options include SOLIDWORKS (Dassault Systèmes) and various software sold under the Autodesk family (Autodesk, inc.).

It must be remembered, however, that these resources have developed alongside associated manufacturing technologies. Most often, reductive manufacturing processes have been limited to the realisation of parts of relatively simple morphology, their complexity being built by the combination of primitive geometrical polyhedra. Whereas, the advent of additive manufacturing processes (including 3D printing) affords the capability to fabricate parts comprising organic, freeform surfaces. This has prompted the development of tools allowing the application of CAD operations to such models, both within the aforementioned software, and within dedicated packages including Geomagic Freeform (3D Systems).

These requirements are shared by the inherently complex organic surfaces which define patient-specific, CHD anatomy. Aside from the generic CAD resources described, there also exist solutions dedicated to medical applications. The Mimics Innovation Suite (Materialise NV) provides tools for the reverse engineering of anatomy from image data and subsequent CAD adaptation.

Clinical validation

A crucial step in the patient-specific 3D printing workflow, clinical validation seeks to ensure that the salient features of image data are faithfully represented by their associated model. Typically, this assessment is made by an expert in CHD imaging. Depending on local arrangements, this might be a consultant cardiologist or radiologist. Whilst equipped with unrivalled expertise in clinical imaging, it is important to recognise that comprehensive experience of 3D printing remains rare amongst these staff. Therefore, it is not reasonable to expect that their understanding of image data naturally extends to a practical appreciation of model validity. Familiarity with software interfaces must be developed through practice, strengthening an association between overlaid model contours and the specialist's inherent understanding of image data. Where these are aligned, a model is clinically valid.

As illustrated by Figure 2.9 (note the inner and outer walls of the model are superimposed), and as performed at our own centre, we suggest that models are validated against the surfaces of the final model (that intended for printing), not the voxelised segmentation. Ultimately, the latter remains an intermediate result of the workflow. Whilst surface rendering and CAD manipulation provide an opportunity to enhance 3D representation, they also risk the incorporation of anomalous features that might misrepresent the underlying image data. This is not to say that clinical input should be delayed until the end of the reverse engineering workflow. On the contrary, we advocate close working between all involved in the derivation of patient-specific 3D printed models and their application. Development of a strong, multi-disciplinary team is critical (Farooqi and Mahmood, 2018) if clinically useful models are to be reliably achieved.

3D Printing

3D printing is a variety of additive manufacture, an umbrella term for a group of fabrication methods that construct objects through a layer-by-layer deposition technique (Negi et al., 2014). This approach avoids the technical challenges associated with the geometrical complexity of 3D parts, simplifying structures through an incremental treatment of effectively 2D slices. Its primary advantage over re-

ductive fabrication, 3D printers are able to fabricate almost any complex shape or geometric feature (Rengier et al., 2010). These include the organic forms which make up the anatomical constituents of the body such as the heart.

As per CAD, the field of 3D printing (and additive manufacturing more generally) increasingly represents an area of applied specialism in and of itself. There now exists a wide variety of commercially available 3D printing technologies. The choice of an appropriate printer depends on the clinical motivations of the model or modelling service. In our experience, we find that the mechanical properties of those materials compatible with a particular printer strongly influence selection. For example, in aid of surgical simulation, the availability of soft, rubber-like materials which allow for direct incision, is desirable. On the other hand, patient-specific models used in education might need to be more durable, in which case a rigid and more robust polymer would be preferable.

In this sense, the variety of applications reviewed in Section 2.2.1 presents a challenge for adopters of this technology: how to acquire a printer with the versatility to match their range of ambitions. Given the substantial investment required by high end technology, centres with limited resource have effectively investigated the possibility of lower cost options, most often demonstrating the success of such approaches (Abdelkarim et al., 2018; Perens et al., 2020). Alternatively and to a varying extent, others have outsourced aspects of the 3D printing workflow to external providers. Whilst some rely on resources outside of their academic or clinical centre to deliver the entire workflow, we endorse the use of external 3D printing (as opposed to modelling) companies. Where possible, developing and maintaining the multi-disciplinary expertise necessary to perform image segmentation, CAD and validation within the hospital enhances and makes best use of available clinical skills. Ultimately, this commitment is most likely to deliver clinically useful models, those with the best chance of improving care.

Where financial resource is available, and depending on their needs, operators can now choose from the following classes of additive manufacture: vat polymerisation, material jetting (or polyjet printing), binder jetting, material extrusion, powder bed fusion and the exciting new technology of digital light processing. In the context of patient-specific cardiac modelling, a review of these printing technologies (including excellent graphics) is provided by Otton et al. (2017).

2.3 Practical challenges of image segmentation

The previous sections have presented an account of why and how patient-specific 3D printing should be incorporated within the care of those with CHD. Section 2.2.1 presented a wealth of applications supported by significant research publication and associated activity. Building on these successes, the field shares collective ambitions for the future. Chief amongst these is an aspiration to expand the application of 3D printing outside of the largest centres (those fortunate to benefit from more significant investment or specialist expertise captured within attached research facilities), to become part of routine, albeit specialised, care. Widely considered the most significant obstacle to this objective, many authors have highlighted the need to establish rigorous evidence of clinical effectiveness (Giannopoulos et al., 2016; Foley et al., 2017; Biglino and Milano, 2018; Fan et al., 2019; Levin et al., 2020; Ma et al., 2021). Without these findings, gaining financial and ethical support from the commissioners of healthcare services has been, and will continue to be, challenging. Related hurdles include significant start up and ongoing costs (Bramlet et al., 2017); and a lack of procedural standardisation that might establish a route to regulatory compliance or accreditation (Shin and Truong, 2018). Whilst overcoming these challenges may unlock the desired expansion of patient-specific 3D printing, each is simultaneously best solved by an increase in the number of active centres and the number of models printed. Every patient considered boosts the statistical power of quantitative findings; the experiences of each new centre enriches procedural standardisation. We sense that this bind has limited progress.

We assert that conceivably, the literature reflects this stagnation. Though the review contained in Section 2.2.1 suggests a vibrant and active group of researchers, we suggest that there is significant redundancy within this body of work. This is perhaps best illustrated by the quantity of review articles published since 2017 and listed in Table 2.2. Whilst for the most part each of these present a sound analysis, we are not convinced that all make an incremental and unique contribution. Many are enhanced by their focus on particular applications of patient-specific 3D printing, such as surgical (Bateman et al., 2020) or catheter-based intervention (Harb et al., 2019); their consideration of particular imaging modalities (Otton

et al., 2017); or their attention to relevant topics advising on the development of hospital-based 3D printing services (Kiraly, 2018) or even bioprinting (Gardin et al., 2020). However, others present a formulaic recipe that includes: a reproduction of the patient-specific 3D printing workflow; a narrative (and incomplete) review of the state of the art; a limited number of case studies; and an outlook for the future[1]. We suspect that progress in the field does not warrant the number of such reviews that have been published since the end of 2016 (being at least 38) and anticipate a significant overlap in the references collectively cited.

We stress that we are not questioning the veracity or value of these reports in isolation, and recognise the insightful comment presented by each. As ably identified by many, we highlight the technical and clinical demands of the patient-specific 3D printing workflow (Batteux et al., 2019). Going a step further, we assert that these challenges not only restrict the growth of this technology, but are the main obstacle to its expansion. As illustrated by Figure 2.9, there are several technical steps to this pipeline, each of which might be optimised in pursuit of greater efficiency. Amongst these, and in response to the calls from Valverde (2017) and Forte et al. (2019), we choose to focus on limitations in current image segmentation methodologies. Our rationale reflects the observation that this step is strongly influential in determining the quality (and therefore clinical utility) of resulting models (Anwar et al., 2018a; Yoo and Van Arsdell, 2018). After all, it is through segmentation that we expose the clinically salient features of image data. It is in this step that clinical insight can be incorporated, realising models that not only reflect anatomy, but which capture the understanding of the imaging expert in a communicable and accessible form. Whilst it is true that similar gains can be made through judicious CAD adaptation, for the most part these can only be performed blindly, without direct reference to the source image data.

Authors' complaints centre around the significant operator burden of existing approaches (Harb et al., 2019). In addition to an expert knowledge of CHD anatomy (Grant and Olivieri, 2017), these tend to rely on significant expertise in 3D image interpretation and segmentation methodologies (Farooqi and Mahmood, 2018; Wang et al., 2020a). Lastly, due to a lack of automation, each of these must

[1]We are aware of the irony of this critique in the context of this chapter! However, we consider our content necessary to ensure the self-contained consistency of this thesis.

be expressed through the operation of software, demanding familiarity with highly specialised tools and interfaces. Problematically, these skills and experience are hard earned (Bramlet et al., 2017) and are rarely possessed by an individual: the clinician being best placed to interpret the presentation of CHD anatomy within image data; the medical physicist, clinical engineer or imaging scientist being more acquainted with image processing methodologies and their application. Without automated solutions (Bhatla et al., 2017b), the result is a lengthy and laborious task (El Sabbagh et al., 2018; Ferrari et al., 2020), one prone to significant subjectivity and variability (Meier et al., 2017), and that comes at significant expense. These requirements are not consistent with the working practices, clinical commitments or training of staff in the hospital (Byrne et al., 2016) and limit the use of an already strained human resource to only the most complex patients.

In response, some have called for the development and application of machine learning or artificial intelligence to this task (Byl et al., 2020; Gharleghi et al., 2021; Sun et al., 2019; Wang et al., 2021a). In the remainder of this thesis we take up this challenge, seeking to understand whether such methods might reduce the operator burden of image segmentation and deliver the desired expansion in patient-specific 3D printing.

Chapter 3

Cardiac image segmentation

3.1 Introduction

Across a range of reviewed applications, Chapter 2 presented 3D printing as an innovative, accessible and exciting means of understanding cardiac imaging. Central to our clinical motivation, this approach reveals the patient-specific anatomy and disease morphology of patients with congenital heart disease (CHD), informing multi-disciplinary decision-making and enhancing treatment planning. However, our review also highlighted the limitations and technical challenges associated with the patient-specific 3D printing workflow. We argued that in particular, image segmentation presents an obstacle to the continued uptake of this technology, and its consolidation within routine care.

In this chapter, we explore such image processing methods in detail, underscoring some of the assertions made previously. Firstly, we attempt to understand techniques that have been applied to the task at hand (segmentation of cardiac anatomy from volumetric image data for the downstream purpose of anatomical modelling and visualisation), and review their performance in respect of operator burden and duration. Thereafter, we broaden our investigation, taking in methodologies from the wider literature on 3D cardiac image segmentation and consider their suitability and possible limitations for patient-specific 3D printing.

3.2 Methods for patient-specific 3D printing

Between 2015 and 2016, we conducted a systematic review of the image segmentation methodologies applied to this task (see Byrne et al. (2016) for a full account of the search criteria). Rather than examining the latest developments from the image processing literature, this explored the methods deployed in practice. Considering clinical and technical reports of patient-specific 3D printing, we extracted data relevant to: the modalities considered for[1]; the methodologies (and their implementation in software) applied to; and the operator burden of, image segmentation. We summarise our historical findings relevant to each of these aspects in the following sections.

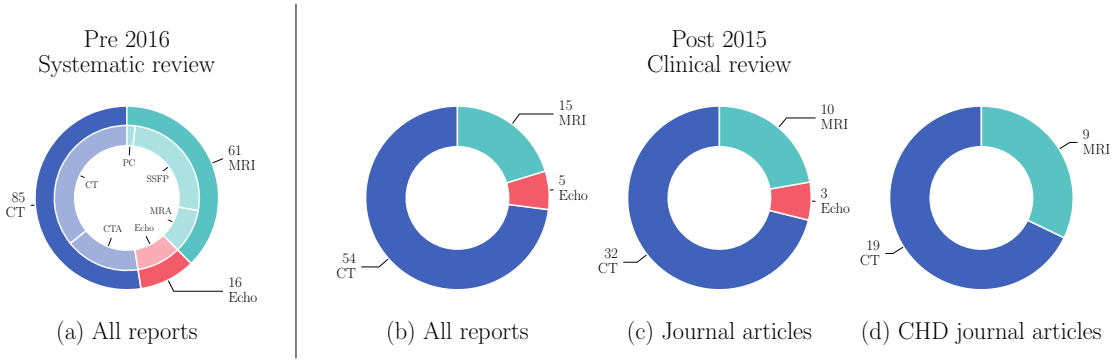
To update this account we gather the same information from modern publications: those published during and after 2016. To draw attention to the clinical application of this technology, we limit our survey to those reports referenced within the "Surgical planning" and "Planning of cardiac catheter intervention" sections of Section 2.2.1. Whilst we cannot claim that these have been sought *systematically* (as might be required to comply with the PRISMA 2020 reporting standard (Page et al., 2021)), they have been compiled *methodically*. Between the start of 2016 and December 2021, relevant reports returned by a Google Scholar alert for any work satisfying the following search were included.

("additive manufacturing" OR "rapid prototyping" OR "rapid prototype" OR "3d printing" OR "3d printer" OR "3d printed" OR "stereolithography" OR "fdm")

AND

(heart OR cardiac OR cardiovascular OR cardiothoracic OR surgery OR radiology OR medicine)

[1]Due to inconsistent reporting, we make no attempt to understand the number of cases segmented from each modality per publication. Rather, a publication can rely on computed tomography (CT), cardiac magnetic resonance (CMR), echocardiography, or a combination of all three, in a mutually binary fashion.



(a) All reports (b) All reports (c) Journal articles (d) CHD journal articles

Figure 3.1: The imaging modalities used to derive patient-specific heart models within the published literature: computed tomography (CT), computed tomography angiography (CTA); magnetic resonance imaging (MRI), magnetic resonance angiography (MRA), phase contrast MRI (PC-MRI) and steady state free precession (SSFP). Compared with previous systematic review (a, reproduced from Byrne et al. (2016)), our survey of recent reports (b) suggests a clinical preference for CT segmentation. However, after excluding case reports of single patients (c), and in particular after focusing on congenital heart disease (CHD) (d), the share of publications which leverage MRI data increases.

3.2.1 Imaging modalities

Spanning journal articles, case and technical reports prior to 2016, over 90% of the works gathered by Byrne et al. (2016), derived patient-specific anatomy from either CT or CMR. Figure 3.1a suggests a historical preference for the former, CT data being segmented in 52% of publications. Only 38% relied on magnetic resonance imaging (MRI). Of those models segmented from CMR data, electrocardiogram-gated balanced steady state free precession (SSFP) was employed three times more frequently than non-gated time-resolved magnetic resonance angiography (TR-MRA). The remaining 10% of studies segmented anatomical geometries from 3D echocardiography, most often in the preparation of valvular models.

Since 2015, our review of clinical applications requires careful consideration. Across all clinically centred citations (including individual case reports), CT appears the predominant subject of 3D cardiac segmentation, being leveraged in over 70% of publications. In contrast, CMR data are segmented in just one fifth of studies (see Figure 3.1b). A more fine-grained analysis, however, suggests an association between the choice of imaging modality and the relevant patient co-

hort. Within journal articles concerning larger studies (more than a single case report, see Figure 3.1c), and in particular within those also concerning CHD, the imbalance between CT and CMR is reduced. In this group of publications, patient-specific models of CHD are retrieved from CMR data in almost one third of articles (see Figure 3.1d). Compared with the literature concerning acquired heart disease (in which only 6% of reports segment CMR data), exact binomial test suggests this difference in proportion is statistically significant ($p < 10^{-8}$).

A comparison of the two imaging modalities helps explain this result. Whilst both CT and CMR provide a 3D representation of anatomy, the richer physiological insights returned by CMR greatly inform the management of patients with CHD (Ntsinjana et al., 2011). Moreover, and despite its evolving demographics, the CHD population maintains a largely paediatric membership. Given their youth and increased radiosensitivity, avoiding the radiation dose associated with repeated CT scanning further enhances the use of CMR within this cohort.

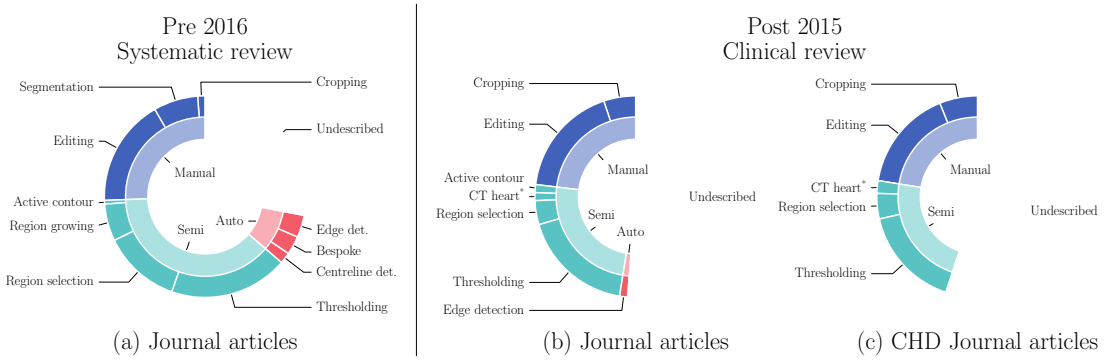
3.2.2 Image segmentation methods

Given the formatting limitations imposed by publishers of conference proceedings, or case reports, it is perhaps unrealistic to expect reports from the grey literature to provide a comprehensive account of image segmentation methodology. Hence, in this section we limit our review to full journal articles concerning patient-specific 3D printing.

Our previous systematic review found a reliance on manual and semi-automated methods, with only few accounts involving fully automated techniques (see Table 3.1 and Figure 3.2a). Even where methods such as edge or centreline detection have been employed, they have most often been hindered by compromised image quality. Limited contrast resolution, contrast- and signal-to-noise ratios, and the presence of artefacts inevitably ensure that the intensity distribution of the foreground and background classes overlap (Byrne et al., 2016). This means that frequently, automated approaches have had to be combined with subsequent manual editing before an acceptable segmentation of patient-specific anatomy is achieved (for example in Jacobs et al. (2008)). They also struggle to contend with the variety of CHD anatomy encountered in the clinic. Lastly and whilst infrequently

Table 3.1: Legend for Figure 3.2, a glossary of image segmentation methods including manual, semi-automated and automated tools.

Methods	Description	Citation
Manual		
Segmentation	Contours are drawn round structures to be segmented on each slice of the 3D volume. Pixels within each contour are included within the segmentation.	
Editing	Pixels are manually added or removed from the segmentation in order to correct for any errors introduced by use of more involved methods.	
Cropping	A partially complete segmentation is spatially cropped, usually by reducing the size of a cuboidal region of interest. Only segmented regions that are contained within the envelope of the region of interest are maintained.	
Semi-automated		
Thresholding	Pixels are selected according to their brightness compared to a user-defined window. Pixels that have a brightness which falls inside of the window are segmented.	Sezgin and Sankur (2004)
Region selection	The connected components comprising an intermediate segmentation are determined by pixel adjacency. A seed point is manually placed within the desired component; only regions connected to the seed point are maintained.	He et al. (2017b)
Region growing	A segmentation is iteratively grown from a manually placed seed point. Growth is controlled by manually tuned constraints that are imposed upon the statistical distribution of pixel intensities within the expanding region.	Adams and Bischof (1994)
Active contour	Often initialised with a manually defined segmentation, the bounding contour of this estimate is iteratively updated within an energy minimisation scheme. Evolution is sensitive to local changes in pixel intensity.	Kass et al. (1988), Caselles et al. (1997), Chan and Vese (2001)
CT heart	A proprietary algorithm developed by Materialise NV. Provided a set of user-defined seeds and thresholds, multi-class anatomical segments of the normal heart are automatically segmented from contrast-enhanced CT data.	Farotto and Maes (2019)
Automated		
Edge detection	Regions of the image are segmented according to their separation either side of significant gradients in intensity. Common implementations rely on convolutional filters, so-called edge detectors.	Canny (1986)
Centreline detection	Primarily for the extraction of vascular structures, software can be used to automatically extract a trajectory that tracks along and through the centre of blood vessels.	Aylward and Bullitt (2002)



(a) Journal articles (b) Journal articles (c) CHD Journal articles

Figure 3.2: The image segmentation methods (see Table 3.1) used to derive patient-specific heart models within the published literature. Compared with previous systematic review (a, reproduced from Byrne et al. (2016)), our survey of recent reports (b), including those focused on congenital heart disease (CHD) (c), suggests that methodological reporting remains poor, with many publications lacking any description whatsoever. Those which do report their approach rely primarily on a combination of manual and semi-automated methods with very few examples of fully automated techniques. *CT heart is a proprietary segmentation tool provided by Materialise (Materialise NV, Leuven, Belgium).

observed, authors have combined conventional image processing operations within automated pipelines. Riesenkampff et al. (2009) adopted a strategy combining thresholding with simplex meshes.

Perhaps a more significant finding of our systematic review concerns the standard of methodological reporting in the surveyed body of work. Figure 3.2 indicates that over one quarter of publications fail to describe their means of image segmentation. This observation prompted us to suggest reporting guidelines relevant to patient-specific 3D printing. At minimum, these required a description of: the imaging modality segmented; any software used; the segmentation tools and methods deployed; a brief description of how these were executed; and an indication of how long segmentation took to complete (Byrne et al., 2016).

Despite our suggestion, and notwithstanding others' calls for improved image processing methods (Valverde, 2017; Forte et al., 2019), reporting quality does not appear to have improved since 2016. At least within clinical reports of patient-specific 3D printing, more than half of publications fail to describe their segmentation methods (see Figure 3.2b); including eleven of the eighteen clinical articles

concerning CHD (see Figure 3.2c). Amongst these, three describe their approach as semi-automatic, without further elaboration (Dydynski et al., 2016; Qian et al., 2017; Hachulla et al., 2019). Whilst it is likely that the difference in reporting quality is explained by the mix of publications sampled (those surveyed in 2016 including reports focused on technical developments as well as the clinical applications reviewed since this time), such superficial reporting of image segmentation methodologies, hampers the algorithmic optimisation and development sought by others (Valverde, 2017; Forte et al., 2019; Batteux et al., 2019), see Section 2.3.

Of those reports providing a faithful description, and as per our previous findings, segmentation has relied on manual and semi-automated methods. In particular, intensity thresholding, manual editing and region selection predominate. An excellent account of their judicious combination is provided by Schievano et al. (2007). Briefly, pixels are differentiated firstly by their intensity, returning a crude representation of target anatomy; spurious, false positive results are subsequently isolated from the true anatomical object; and can then be rejected by analysis of component connectivity, selecting only those pixels connected to a user-provided seed. Given its popularity, it is unsurprising that these tools constitute the segmentation protocol advocated and implemented by the developers of the Mimics software (Materialise NV, Leuven, Belgium). Whilst our survey of clinical reports encountered other platforms (including open source, freeware and proprietary options), various versions of the Materialise software were used in thirteen out of eighteen CHD reports (see Table 3.2).

Amongst the methods used, we draw attention to the "CT heart" tool, provided as a paid-for extension to the Mimics software. This functions in two modes of operation: (1) fully automatic segmentation of contrast-enhanced CT data into the blood pool chambers of the normal heart and its associated vasculature; (2) semi-automatic segmentation of the same targets, as constrained by a series of user-defined anatomical seeds and thresholds in voxel intensity. Being a proprietary algorithm, additional details of its technical formulation are not made public by its associated white paper (Farotto and Maes, 2019).

Whilst its impartiality is difficult to ascertain, this report does make a quantitative assessment of performance, considering seventy adult cardiac CT scans. This suggests strong performance, successfully segmenting anatomy in 65 cases in

Table 3.2: List of image segmentation software used within clinical reports of patient-specific 3D printing. Software details are extracted from the publications referenced in Section 2.2.1 and which concern planning of surgical or catheter intervention.

Software	**Developer**	**Frequency**
Mimics Innovation Suite	Materialise NV	30
ITK-SNAP	Yushkevich et al. (2006)	4
3D Slicer	Kikinis et al. (2014)	3
AYRA	Fernandez-Alvarez et al. (2014)	
Osirix	Rosset et al. (2004)	
Amira	Thermo Fisher Scientific	2
CT Auto Valve	Siemens AG	1
MARACAS	Hernández-Hoyos et al. (2002)	
MeVisLab	MeVis Medical Solutions AG; Fraunhofer MEVIS	
MITK	Wolf et al. (2005)	
Extended Brilliance Workstation	Philips NV	
TeraRecon	TeraRecon, Inc.	
Vitrea	Vital Images, Inc.	
Ziostation	Ziosoft, Inc.	

a median time of 113 s. Within these cases, however, the authors observe reduced accuracy in the segmentation of the right heart structures when compared with their counterparts of the great circulation. In part, Farotto and Maes (2019) attribute this degradation, to anatomical variation. This may be problematic in the context of CHD, in which we observe substantial heterogeneity in cardiac structure and morphology. Moreover, the test patients considered did not include congenital anomalies, the authors only citing occasional anatomical variation associated with aneurysm, tumour, or calcification: hallmarks of acquired cardiac disease. Despite this challenge, the CT heart tool was employed by Kappanayil et al. (2017) (see Figure 3.2), in their application of patient-specific 3D printing to five patients with complex CHD. Outside of its intended domain of operation (considering both CHD and CMR data), it is perhaps unsurprising that the results of automated segmentation required further manual refinement.

3.2.3 Operator burden of image segmentation

Many authors have highlighted the limitations of the patient-specific 3D printing workflow, including the time consumed in its completion (Dydynski et al., 2016; Farooqi and Mahmood, 2018; Yıldız et al., 2021). Within the steps of this pipeline, our experience matches the assertion of Bateman et al. (2020): that segmentation is the most protracted. Moreover, within this image processing task itself, manual editing has been suggested as the most time consuming component (Yoo and Van Arsdell, 2018). This is perhaps unsurprising since it is within this step that errors within semi-automated segmentation are refined, and expert interpretation brought to bear, realising a clinically veracious representation of image data.

The vast majority agree that segmentation is a labourious, time-consuming and tedious activity (Bhatla et al., 2017b; Yoo et al., 2017; Bertolini et al., 2021). The operator time spent editing a patient-specific segmentation of data, however, is unclear. Reflective of a more general lack of any standardised approach to modelling, attempts to establish this burden from the literature must contend with a highly heterogeneous collection of reports. These vary in their clinical application; the type of model printed and the anatomy targeted; the experience and subjective performance of the operator; and the standard of data collection (being mostly anecdotal) and reporting. Even where operator times are reported, these most often reflect the duration of the entire patient-specific 3D printing workflow, without reflecting the intermediate steps comprised, including image segmentation (Tuncay and van Ooijen, 2019).

As demonstrated by Table 3.3, this heterogeneity generates estimates of operator burden that are measured in both minutes and days. Whilst quantitative estimates are scant, our interpretation is that for the most part, segmentation of the anatomical components needed to construct a model of the whole heart (all chambers and associated vasculature) requires at least an hour of expert time and often as many as three. Where more granular anatomy provides the sole focus, segmentation of individual structures such as the aortic root can be much faster. Whilst there are outliers at both ends of this distribution, further details (including for the incredibly rapid, whole heart segmentation performed by Thakkar et al. (2018)), are often not provided.

Table 3.3: Estimates of the operator burden of 3D image segmentation. Each is sourced from a search of the citations concerning clinical applications of patient-specific 3D printing (planning of surgical and cardiac catheter interventions) and the contents of Table 2.2. Reporting formats: x = anecdotal point estimate; $\leq x$ = upper bound; (x_1, x_2) = (lower bound, upper bound); $x(x_1, x_2)$ = mean(lower bound, upper bound); $x \pm \sigma$ = mean $\pm$ standard deviation. Abbreviations: aortic valve (AV), right ventricular outflow tract (RVOT), left atrial appendage (LAA).

Target	**Citation**	**Estimate**	**Unit**
Whole heart	Kim et al. (2008b)	$(2, 5)$	days
	Yıldız et al. (2021)	0.5	
	Vukicevic et al. (2017a)	≤ 12	hours
	Grant and Olivieri (2017)	≤ 10	
	Jacobs et al. (2008)	$(3, 8)$	
	Kiraly et al. (2021)	$(2, 4)$	
	Valverde et al. (2015a)	3	
	Forte et al. (2017)	3	
	Velasco Forte et al. (2018)	$(2, 3)$	
	Valverde et al. (2015b)	2	
	Riesenkampff et al. (2009)	$(0.6, 3)$	
	Valverde et al. (2017a)	75 ± 32	minutes
	Sørensen et al. (2008)	$59(42, 85)$	
	Meyer-Szary et al. (2019)	45	
	Thakkar et al. (2018)	≤ 5	
Myocardium	Andrushchuk et al. (2018)	2	hours
Aorta	Sulaiman et al. (2008)	40	minutes
RVOT	Schievano et al. (2007)	$(2, 3)$	hours
LAA	Song et al. (2017)	$(15, 25)$	minutes
	Hachulla et al. (2019)	≤ 15	
AV annulus	Qian et al. (2017)	$(5, 10)$	minutes
	Faletti et al. (2018)	≤ 5	

It should be noted that in some cases where authors report a segmentation time short of sixty minutes, this has been followed by lengthy computer-aided design (CAD). For example, although Meyer-Szary et al. (2019) required only 45 minutes for image processing, this was followed by almost three hours, using CAD to correct "mesh errors". Perhaps the only rigorous estimate, made by Valverde et al. (2017a), also allows for a further ninety minutes of CAD adaptation. In our local practice, we suspect that the majority of these adaptations would have been made with respect to the image data, via prolonged segmentation. The variable division of labour between steps of the patient-specific 3D printing workflow further demonstrates the lack of standardisation, in both practice and reporting.

Rather than detailing segmentation time, authors appear to favour reporting the time consumed during model printing. This duration, however, is associated with the operation of an autonomous manufacturing technology. Apart from for those interested in non-elective or emergency care, we suggest that this is less relevant to the development of a hospital-based, patient-specific 3D printing service. Whereas, an understanding of operator time can inform plans for staffing. In addition to the temporal burden of image segmentation, many have flagged the reliance of image segmentation on specialist experience, and the challenge of gaining and developing associated expertise (Bramlet et al., 2017; Forte et al., 2019; Sun et al., 2019). This supports our belief that it is the clinical understanding of data, and its expression within segmented images, that is key to the value of patient-specific modelling; not the performance of a particular printing technology.

3.3 Methods from the wider literature

The previous section found a clinical reliance on relatively basic, manual and semi-automated segmentation methods. In this section we expand our scope, reviewing the cardiac image segmentation literature to understand the advanced methods that have been developed in the context of, or that have been applied to, 3D cardiac anatomy. Moreover, we hope to glean why, for the most part, these have not found use within clinical applications of patient-specific 3D printing, modelling, or associated software packages. We also highlight the imprecision of referring to cardiac image segmentation generically, realising that this field of research covers a

Table 3.4: Reviews of conventional cardiac image segmentation methodologies.

Year	Authors	Title
2011	Petitjean and Dacher	A review of segmentation methods in short axis cardiac MR images
2012	Kang et al.	Heart chambers and whole heart segmentation techniques
2013	Tavakoli and Amini	A survey of shaped-based registration and segmentation techniques for cardiac images
2013	Zhuang	Challenges and methodologies of fully automatic whole heart segmentation: a review
2016	Peng et al.	A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging
2020	Habijan et al.	Overview of the whole heart and heart chamber segmentation methods

wide variety of tasks and imaging modalities or acquisition protocols. Whilst some of these applications share characteristics with the segmentation of patient-specific CHD anatomy, others are no more relevant than those concerning any other organ of the body. Consequently, in summarising the scientific literature, the following review seeks to reveal the interaction between different methodologies, and the image data and segmentation tasks to which they have been applied. We begin by describing the technical basis of such approaches in Section 3.3.1 and Section 3.3.2. Section 3.3.3 reviews the applications in which these methods have been employed.

Prior to the advent of deep learning methodologies (to be discussed in Chapter 4), a variety of approaches have been applied to cardiac image segmentation. Though focusing on different applications, the review articles listed in Table 3.4 recount developmental progress, each attempting to divide methods into families sharing common characteristics. As acknowledged by Tavakoli and Amini (2013), such categories are not well defined. Despite this challenge, and though authors may have adopted different terminology, techniques have most often been separated by their inclusion and treatment of prior information. In the following, and as per Peng et al. (2016), we consider methods to be largely: *image-driven*, primarily leveraging low level features of data and exploiting limited priors that can be straightforwardly derived from pixel intensity, or that reference basic geometrical structure; or *model-driven*, relying on the spatial adaption, deformation or registration of a prior model of cardiac anatomy to the image at hand.

Within this scheme, it is also useful to compare techniques that are *unsupervised* with those that are *supervised*. Unsupervised methods rely solely on abstract prior knowledge or statistical arguments to discriminate between segmentation classes, most often through clustering algorithms such as K-means, fuzzy C-means and other expectation-maximisation (E-M) classifiers. Whereas, supervised methods rely on training data. In the context of image segmentation, training data serve two main functions: (1) to optimise the parameters of an image-driven, parametric classifier such as K-nearest neighbours, random forests (RFs) and support vector machines (SVMs) (see Section 3.3.1); or (2) to form a prior representation of anatomy, perhaps captured within a statisical shape model or cardiac atlas (see Section 3.3.2). At test time, the performance of either approach depends strongly on the quality and quantity of the training data.

3.3.1 Image-driven approaches

Image-driven methods rely on relatively low level features to discriminate between segmentation classes. Most fundamentally, raw pixel (or voxel) intensity itself often reveals image content, particularly in cases where the acquisition protocol is tuned to deliver strong contrast between the anatomical objects of interest. This might be through the selection of imaging modality (consider the contrast developed by the differential X-ray attenuation properties of bone and soft tissues, for example) or through the administration of an exogenous contrast agent (consider gadolinium-enhanced TR-MRA). Limitations in image quality, and in scanner performance more generally, however, often preclude intensity thresholding as a ubiquitous segmentation solution.

To counter such cases, researchers have handcrafted bespoke features, engineered to expose higher level semantic information. Often exposed through the application of digital filters, the spatial extent of associated convolutional kernels confers a pixel-wise sensitivity to extended image gradients, including geometric structures such as edges (Canny, 1986) or ridges (Frangi et al., 1999). Alternatively, examples such as the median or Gaussian filters serve to reduce noise, and reveal semantically homogeneous portions of the image. Lastly, to expose image textures (or alternatively, the presence of spatial frequencies), bespoke convolu-

tional operators such as Haar-like (Viola and Jones, 2001) or Gabor (Jain and Farrokhnia, 1991) filters, or transforms such as by wavelets, have been designed. Each of these can be tailored to a range of orientations, frequencies and spatial scales. Choosing a suitable feature extractor and tuning its parameters according to the anticipated properties of the image and segmentation task, is sometimes described as feature engineering.

Rarely, however, is a single feature sufficient to discriminate between the anatomical targets of a complex segmentation task. Instead, pixels (or perhaps their collection within amorphous superpixels or rectangular regions) are more often characterised by a feature vector. For example, a pixel might be described by: its intensity, its value after wavelet transformation and its Gabor filter response. This combination provides an example of a multi-dimensional feature space, to which downstream classifiers are subsequently applied. Whilst this vector space need not be constructed explicitly, image-driven methods typically depend on a range of features similar to those described.

In general, and being free from prior constraint, image-driven techniques exhibit the flexibility to adapt to heterogeneous anatomy. Simultaneously however, and without the influence of a strong prior model, they are susceptible to limitations in image quality, including the presence of artefacts. This weakness can yield segmentations of limited spatial and anatomical coherence. For this reason, image-driven methods have more frequently been combined as components of a broader image segmentation pipeline (perhaps for region of interest localisation, initialisation or refinement), rather than providing a comprehensive end-to-end solution in and of themselves.

The following sections describe the various image-driven methods encountered in the cardiac image segmentation literature. For those described previously (in Table 3.1), we provide further details necessary to appreciate their application in Section 3.3.3.

Elementary operations

Thresholding (Sezgin and Sankur, 2004)
As per Table 3.1, thresholding describes the selection of pixels that fall within an intensity interval defined with reference to the statistics of the image histogram. In the image processing literature, this principle has been extended to consider: procedures for automatically defining the window width and position (or adaptive thresholding); both local and global image statistics; the sequential application of multiple thresholding to separate or accumulate candidate regions; and the distribution of pixels in a derived feature (rather than intensity) space. Thresholding has found particular success in CT imaging where pixel intensity quantitatively reflects the physical properties of tissue.

Edge detection (Canny, 1986)
Edge detection describes the delineation of boundaries within the image, according to local intensity gradients. It is critical to the localisation of the statistical shape models developed within the cardiac image processing literature (and as described in Section 3.3.2). Within this family of methods, boundaries are most often localised according to a spatially varying (over the model surface) and multi-dimensional (sensitive to a feature vector of directed 1D gradients) parametric model, learned from example data. Forces associated with edge detection are also frequently included within active contour segmentation schemes.

Region growing and diffusion models

Region growing (Adams and Bischof, 1994)
As per Table 3.1, region growing describes a dynamic, iterative process in which a segmentation expands from an initial seed point or region. At each iteration, expansion is constrained by a parameterised comparison between the distribution of pixel intensities contained by the growing foreground and its neighbours. Within the technical literature, this generic scheme has been enhanced by its formulation as a variational problem, the incorporation of shape priors and the development of adaptive parameter thresholds.

Random walkers (Grady, 2006)
Random walker segmentation models the diffusion of a particle, biased to avoid crossing sharp intensity gradients. Class probability can subsequently be inferred from the frequency with which random walkers, starting (or terminating) at a manually labelled pixel or seed point, visit an unlabelled location. These diffusion methods have been enhanced through the incorporation of shape constraints, and have been employed within segmentation post-processing to smooth predicted regions.

Active contours

Snakes (Kass et al., 1988)
As per Table 3.1, the term *active contours* describes an iterative process of segmentation by dynamic curve evolution. Various energy minimisation schemes have been designed to attract the snake (represented by a spline-interpolated series of spatial coordinates) towards object boundaries within the image. External forces sensitise evolution to local changes (respecting both boundaries and regions) in pixel intensity; internal forces regularise updates by governing smoothness and topological consistency. Pivotal to snake evolution, researchers have focused on the development of novel force terms, seeking to impose geometrical constraints, and incorporating modes of statistical shape variation.

Level set (Caselles et al., 1997; Chan and Vese, 2001)
The level set formulation of active contour segmentation implicitly captures the evolving curve within the zero level set of a scalar function, most frequently a signed distance map. More robust to larger deformations in shape than snakes, the level set framework can also cope with changes in topology and competition between co-evolving contours. Both snake and level set formulations rely on an initialisation that approximates the target anatomy. Hence, they have often been employed as a means of refinement, adapting a prior segmentation to local changes in image intensity.

Graphical methods

Graph cut (Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006)
Graphical methods consider the pixels of an image as nodes of a graph, locally connected to their immediate neighbours. The weight of each connection is determined by the similarity or *affinity* of associated nodes according to relevant features such as pixel intensity or location. In this setting, segmentation amounts to cutting edges of the graph to achieve disjoint subsets, the nodes of which constitute the resulting image segments. Accordingly, graph cut segmentation seeks to find the cut for which the sum of removed edge weights is minimised.

Markov random field (MRF) (Li, 1994; Krähenbühl and Koltun, 2011)
Representation of the image as a graph of locally connected pixels induces the Markov criterion: informally, that the probability of a pixel belonging to a particular class of the segmentation depends only on its neighbours. The resulting MRF describes the factorisation of the joint probability distribution of both segmentation and image, $p(\mathbf{Y}, \mathbf{X})$, into the individual and pair-wise potentials of the random variables at each pixel location. Judicious design (such as the Potts model) sensitises these potentials to spatial variation in image features. Provided these ingredients, optimal segmentation is achieved through dedicated algorithms for probabilistic optimisation. A popular variant of the MRF model, the conditional random field (CRF) is more naturally suited to semantic labelling tasks such as image segmentation. The CRF models the conditional probability of the underlying segmentation given the input data, $p(\mathbf{Y} \mid \mathbf{X})$, directly, making no attempt to understand the probability of the observed image. It does, however, demand paired image-label examples for supervised parameter optimisation prior to inference. CRF models are often used to refine the estimates made by a probabilistic classifier, enforcing spatial coherence through post-processing.

Unsupervised clustering

K-means (Hartigan and Wong, 1979)
K-means clustering partitions the members of a dataset into K exclusive groups according to their separation from the mean or centroid of each group, in feature

space. Each item is associated with the cluster in closest proximity. In this context, K-means seeks to find a set of K cluster means which minimise this distance when summed across the entire dataset (or equivalently the intra-class variances). Within image segmentation, the dataset is composed of single pixels, each taking a location in a feature space, handcrafted to expose differences between the semantic classes of the image. Combined with an informed setting of K, optimisation proceeds through alternating, iterative refinement: (1) pixels are assigned according to the present locations of the respective centroids; (2) the centroid of each cluster is updated to reflect the most recent assignment.

Fuzzy C-means (Dunn, 1973; Bezdek et al., 1984)
Fuzzy C-means extends the K-means clustering algorithm by considering partial or probabilistic class membership. Rather than require each data point belong to a single segmentation class, this approach weighs the degree of association with all possible labels. Per pixel, membership is determined by comparing the distances to each centroid in feature space. This is clearly advantageous in multi-label segmentation, but may also be preferable where clusters demonstrate substantial feature overlap. In such cases, fuzzy membership permits the rich representation of transition zones between classes. As per the K-means algorithm, optimisation proceeds by iterative refinement and does not depend on training data.

Expectation-maximisation (E-M) (Dempster et al., 1977)
Both K-means and fuzzy C-means clustering are special cases of the more general E-M framework. In our context, E-M provides a means of estimating the parameters of a statistical model that explains the distribution of observed pixels (frequently described by a feature vector) by their association to different segmentation classes via hidden variables. Using this terminology, E-M optimisation proceeds iteratively, alternating between: Expectation, estimating the hidden variables given the current model; and Maximisation, updating the model to optimise the likelihood of current parameter estimates. A popular approach models the conditional probability of observed feature vectors, given their association to a particular segmentation class, using a mixture of multivariate Gaussian distributions. In the so-called Gaussian mixture model (GMM) which results, clusters

are parameterised by a mean vector and covariance matrix, with the contribution to each balanced by a vector of mixing factors. In these terms, both K-means and fuzzy C-means correspond to E-M optimisation of a GMM for which all covariances are identically zero, a constraint that assumes feature clusters to be spherical. In contrast, the more expressive, multivariate models of a GMM better describe clusters of anisotropic geometry.

Supervised classifiers

K-nearest neighbours (KNN) (Cover and Hart, 1967; Fix, 1985)
In contrast with previous methods, and in common with the remaining techniques detailed in this section, KNN classification depends on the availability of a training dataset of image-label pairs. Collectively, the exemplar images contained define a feature space of pixels, superpixels or patches, each of which is associated with a known ground truth class or segmentation. At test time, provided an image region to segment, searching the feature space for the KNN returns a collection of K image-label pairs from the training set. Strategies for fusing the K exemplar labels involve pixel-wise majority voting or, more commonly, schemes for weighting their relative contribution according to the proximity of their associated image region, to the test region, in feature space.

Random forest (RF) (Ho, 1998; Breiman, 2001)
Conventional decision trees describe a cascade of binary queries, interrogating the features of data to map an observation to a class. In the context of image segmentation, tracking the passage of a pixel through the nodes of the tree predicts its label. Determining the cascade of queries, including the features against which each splits the tree, is achieved through training and relies upon paired image-label examples. At each node, the feature that best divides the data is determined by metrics such as the Geni impurity or information gain. Though attractive for their interpretability and straightforward training procedure, decision trees have a tendency to overfit the training data, predicting class labels with high variance. RF classifiers extend the notion of decision trees, improving their generalisation to unseen data. This is achieved, by training a "forest" of individual trees, introducing

stochasticity through bagging (the combination of bootstrapping and aggregation) and random feature selection. Bootstrapping synthesises numerous copies of the original training dataset, by repeated random sampling with replacement. Each *bootstrapped* sample is used to train a distinct decision tree, their predictions being *aggregated* to return a final result. The training procedure is also modified by limiting the features considered at each node to a random subset: typically $\sqrt{D}$ features, where D is the dimensionality of the space.

Support vector machine (SVM) (Cortes and Vapnik, 1995)
The SVM is a supervised model for binary classification, with both linear and non-linear modes of operation. In the linear case, solving the SVM objective function amounts to finding the hyperplane that divides feature space into regions of positive and negative class prediction, with the greatest possible margin between the binary classes expressed by the training set. Relying on the dot product as a measure of feature similarity, and being a maximal margin classifier, the hyperplane is defined entirely by the feature vectors (of both positive and negative training examples) in closest proximity: the so-called support vectors. However, in the event that the features expressed by positive and negative training examples are not linearly separable, the SVM objective is instead solved in a higher dimensional feature space, achieved by non-linear transformation of the original features. Judicious selection of the transformation admits the application of the kernel trick, allowing for dot product similarity within the transformed space to be computed efficiently from the original features. In the context of image segmentation, SVM training relies upon a set of image-label pairs, and considers each labelled pixel as a point within feature space. Limited to binary classification, SVM predictors must be applied in combination to perform multi-class segmentation.

3.3.2 Model-driven approaches

As presented by Zhuang (2013), model-driven segmentation methods rely on two main components, each of which has been the subject of research: (1) construction of a prior model of cardiac anatomy or physiology; and (2) a fitting procedure to align the model with the test case at hand. Whilst different types of prior

model have been considered, each is motivated by a desire to enforce the spatial coherence of segmentation, most often by constraining predictions to anatomically or pathologically plausible label configurations. Strong priors make segmentation methods robust against artefacts and limitations in image quality, normally at the expense of flexibility: performance may suffer when confronted with a test case that is poorly represented by the prior model (and the underlying training data used in its construction).

The two most popular approaches rely on atlas-based segmentation and statistical shape models. Variants of each approach are described in the following sections.

Atlas-based segmentation

Single atlas segmentation (Collins and Evans, 1997; Fischl et al., 2002)
In its most basic description, an atlas couples an intensity image with a set of segmentation labels, delineating the targets relevant to a downstream task specification. More advanced models provide a probabilistic atlas (in which class membership is continuously valued), constructed from a training set of spatially aligned labels maps, and accompanied by a mean intensity image. Irrespective of its form, atlas construction is motivated by a desire to provide a faithful and generalisable representation of anatomy; one that captures the characteristics (spatial configuration, complex morphology or topology) of plausible segmentation, and that can be transmitted to the unseen test case by spatial transformation. Label propagation, therefore, depends on the availability of image registration algorithms (including the selection of appropriate similarity metrics) to align the test image with its atlas counterpart. By virtue of their dependence on a constrained transformation of the atlas labels, predictions are robust to limitations in image quality (which might otherwise hinder image-driven methods) and faithfully reproduce anatomically plausible characteristics. Both the atlas model and the registration algorithm have been widely researched throughout the image processing literature.

Multi-atlas segmentation (MAS) (Rohlfing et al., 2004; Heckemann et al., 2006)
For all the advantages of atlas-based segmentation, performance declines in the presence of anatomical targets that exhibit substantial structural differences compared with the atlas model. To at least partially overcome this challenge, researchers have developed the MAS framework. Rather than rely on label propagation from a single atlas, MAS increases the diversity of compatible anatomical appearances by registering a collection of image-label pairs to every test case. Once co-registered to a common space, the MAS framework adopts a strategy for pixel-wise label fusion, resolving conflicts between the segmentations propagated from each atlas. This might be as straightforward as majority voting. However, schemes for ranking and then weighting the contribution of each atlas, perhaps according to the pixel-wise similarity metric, have demonstrated significant performance gains. Such strategies allow for predicted segmentations to be biased on the most similar members of the atlas set (for which anatomical configuration, morphology and hence alignment, are presumed most informative), reducing the influence of irrelevant examples of divergent anatomy.

Statistical shape modelling

Active shape model (ASM) (Cootes et al., 1995)
Active shape modelling seeks a generative model of anatomical geometry, one that reflects both the mean shape expressed by a labelled training set of images, and its distribution according to the most common modes of variation. These are established by training the so-called point distribution model (PDM) on paired image-label examples, a procedure that involves: (1) affine transformation, aligning manual training labels to a common coordinate space; (2) determining correspondences between each training example by establishing a common set of meaningful landmarks; (3) constructing the mean shape and covariance matrix of landmark coordinates; and (4) deploying dimensionality reduction to reveal the dominant modes of shape variation. By a linear, weighted combination of the principle eigen vectors, the PDM generates new anatomies from the modelled distribution of training shapes. In this context, image segmentation amounts to finding the parameters that best fit the PDM to an unseen test image. The associated transfor-

mation can be broken down into two parts: the first affecting the affine pose of the model, the second reflecting non-rigid deformation of the mean anatomical shape. Unlike naive deformable models, the latter is parameterised according to the principle modes of variation expressed by the PDM, with deformations restricted by constraining the associated weights. Hence, predicted segmentations are globally sensitive to the statistical distribution of anatomically plausible shapes. Active shape model (ASM) fitting proceeds iteratively. Evolution is typically governed by the presence of edge-like features (or other textures that can also be learned from the training set), searching the collection of 1D profiles, normal to the PDM surface and local to each landmark.

Active appearance model (AAM) (Cootes et al., 2001; Mitchell et al., 2002)
Active appearance modelling extends the generative capacity of ASMs to also include the distribution of image textures. Whilst local appearance knowledge can be included within an ASM, typically this is of restricted spatial extent, limited to the vicinity of each shape landmark. In contrast, an AAM is trained to synthesise both label shape and a complete image, according to their distribution within training data. A mean appearance model and its principle modes of variation is established analagously to the ASM, considering the synthesis of image vectors formed from the eigen decomposition of the training set. For segmentation purposes, this approach differs from ASM in that parameter fitting is informed by the difference between the entire synthetic and test images, rather than just a subset of image features associated with the landmarks of current shape estimate. Where edge-like textures are reliably and locally associated with object boundaries, ASM alone may provide an accurate anatomical delineation. However, where the appearance of edge profiles is degraded by compromised image quality or the presence of noise, model fitting may benefit from the global constraint enforced by AAM, even if a decline in local boundary accuracy is incurred.

3.3.3 Applications

The image segmentation methodologies described have been deployed in a range of clinical applications. Given our focus, this section primarily reviews their de-

velopment and suitability within 3D image segmentation tasks, those extracting a high fidelity representation of cardiac anatomy. As per their use within patient-specific 3D printing, the resulting geometries have found qualitative application within clinical decision-making and operative planning.

However, whilst technically far removed from our 3D image processing task of interest, we first touch briefly on the segmentation of short axis cine CMR data. We consider a basic appreciation of this application important for its clinical ubiquity, but also to understand its influence on the development of related cardiac image processing tasks, including our own. By comparing the motivations and imaging characteristics of each, we hope to understand the performance of differing segmentation methods.

Ventricular volumetry

Composed of a series of (around ten) 2D slices arranged at regular intervals along the long axis of the heart, short axis CMR cine data are frequently segmented to label the left ventricular blood pool and myocardium. Given the 2D pixel dimension and spacing between slices (typically around 10 mm), delineation of the endocardial surface reveals the left ventricular volume. Coupled with a temporally resolved image series, this approach returns a host of quantitative indices, serving as biomarkers of cardiac health (Peng et al., 2016). Adding the epicardial surface admits assessment of the left ventricular myocardium, including its mass. Less frequently, the same approach has been applied to the right ventricular cavity. Conventionally, manual segmentation has provided the basis of these clinical indices (see Figure 3.3), including within the investigation of CHD: Lorenz et al. (1995) studied the late adaptation of the left and right ventricles to repair of transposition of the great arteries by atrial switch. Thanks to its position as the gold standard technique for ventricular volumetry, however, short axis segmentation has been the subject of significant research investment. More than any other cardiac segmentation task, and motivated by a desire to reduce both operator burden and inter-observer variability, innovators have sought to leverage a variety of automated techniques. Prior to the advent of deep learning, a mixture of image- and model-driven methods have been investigated.

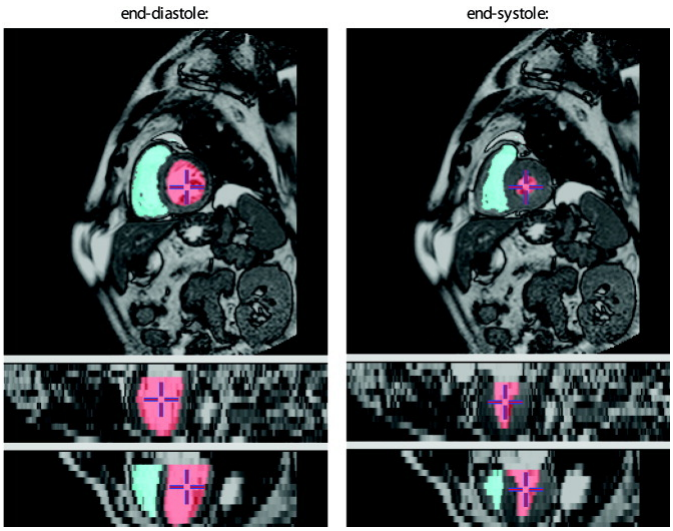


Figure 3.3: The segmentation of 2D cine, short axis CMR data admits volumetry of the left (pink) and right (cyan) ventricles. Note that the in plane spatial resolution far exceeds the spacing between slices. Reproduced from Cocosco et al. (2008).

Image-driven methods

Most often, image-driven methods have been combined within multi-faceted, sequential processing pipelines. For example, Cocosco et al. (2008) described an entirely unsupervised, hand-engineered procedure to isolate the left and right ventricular blood pools within a 2D short axis slice. This involved: temporal smoothing, maximum intensity projection, Gaussian blurring, Otsu thresholding (Otsu, 1979), morphological operation, region growing and connected component analysis. Grosgeorge et al. (2011) deployed 2D active contours without edges, allowing a single initial contour to undergo topological division into left and right ventricular components, prior to morphological smoothing. In their clinical validation on 59 patients, spatial overlap performance degraded towards the apex of the heart and at end systole. Addressing this challenge, Ringenberg et al. (2014) introduced spatial and temporal consistency constraints on right ventricular appearance in neighbouring short axis images. These were expressed in a pipeline that combined advanced thresholding techniques with difference of Gaussians feature extraction and morphological operations.

Perhaps revealed by the complexity of purely image-driven pipelines, their underlying methodologies are susceptible to the variation in patient presentation and image quality encountered in the clinic. Motivated by this challenge, researchers have sought to augment these techniques by injecting prior information concerning anticipated anatomical shape and arrangement. For example, Wu et al. (2013)

imposed a circulatory constraint on the evolution of an active contour model. Mahapatra (2014) used RFs to establish a feature selection and weighting strategy, designing a graph cut cost function that was sensitive to both right ventricular texture and shape. Finally, Queirós et al. (2014) employed 2D elliptical template matching and circular cues to identify the concentric epicardial and endocardial boundaries.

Model-driven methods

Whilst these steps boost the capacity of image-driven pipelines to resist variable anatomical appearances and image artefacts, further improvement demands the incorporation of model-driven methods. Mitchell et al. (2001) were the first to deploy a slice-wise, 2D biventricular statistical shape model, combining both ASM and AAM within a hybrid approach. Subsequently, they extended their work to 3D, developing an AAM to segment left ventricular structures from short axis CMR and echocardiographic data (Mitchell et al., 2002). Critical to the expansion of such techniques, Frangi et al. (2002) alleviated the challenge of manually assigning PDM landmarks, demonstrating that meaningful correspondences could be automatically retrieved by non-rigid deformation of a short axis atlas mesh. This development fostered the growth of statistical shape models of the heart, including: AAM segmentation of isotropic cardiac CT (Lapp et al., 2004); the use of AAM to model the medial skeleton of the left and right ventricles (Sun et al., 2010); and the combination of an ASM of label distance maps with graph cut segmentation (Grosgeorge et al., 2013).

The utility of segmentations achieved by statistical shape model extends beyond volumetry, several authors demonstrating the sensitivity of PDMs to pathological diagnoses. For example, Suinesiaputra et al. (2009) identified wall motion abnormalities in 2D cine MRI as outliers from their modelled distribution of left ventricular shape. Related work has applied similar principles to CHD (see Figure 3.4). Zhang et al. (2009) used AAM for comprehensive 4D segmentation and volumetry of fused short and long axis cine, further leveraging novel shape features to identify patients with repaired tetralogy of Fallot. Within the same disease population, Mansi et al. (2011) developed a generative statistical model of right ventricular shape to describe cardiac remodelling.

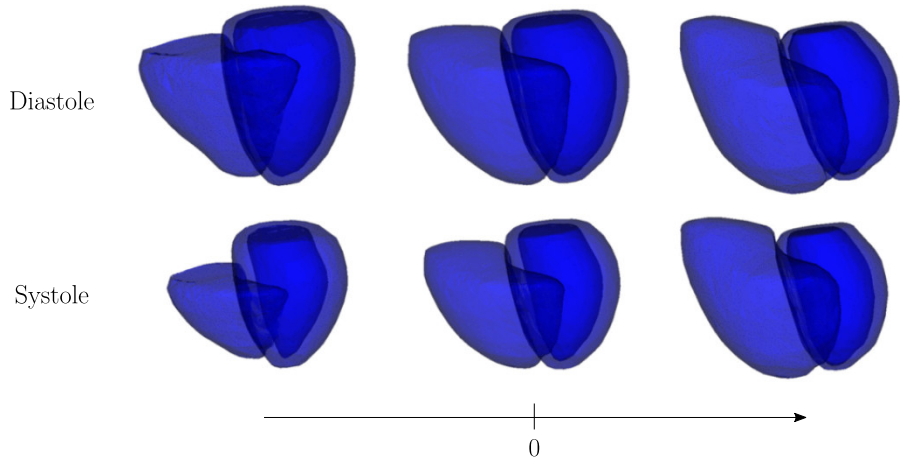


Figure 3.4: Statistical shape models deploy dimensionality reduction to establish the global modes of anatomical variation. These surface-rendered representations of the left and right ventricles are generated by taking samples along the principle axis of variation. Both the mean shape (centred at the axis origin, 0) and the modes of variation are established from a set of co-registered image-label training examples. In this graphic (adapted from Zhang et al. (2009)), increases in the first modal index describe ventricular remodelling after surgical repair of tetralogy of Fallot. Right ventricular dilation is particularly apparent.

Atlas-based methods have also proved successful routes to automate aspects of short axis segmentation. Impressive early work presented by Lorenzo-Valdés et al. (2004) used the predictions of a transformed 4D probabilistic atlas (constructed from the CMR data of fourteen healthy volunteers and varying in both space and time) to initialise the parameters of a GMM. Predicted segmentations were subsequently refined by E-M, updating parameters of both GMM and 4D MRF models to determine spatially and temporally consistent class membership. The importance of a 4D spatio-temporal approach to cardiac registration was reinforced by Peyrat et al. (2010). Decoupling their task into respective 4D mappings of physiological states and physical point trajectories, they presented a multi-channel formulation of the Diffeomorphic Demons algorithm (Vercauteren et al., 2009) to propagate epicardial and endocardial contours to all phases of 4D CT acquisition.

MAS formulations to this task followed. Addressing the 2012 Right Ventricle Segmentation Challenge (Petitjean et al., 2015): Bai et al. (2012) and Zuluaga et al. (2013) both presented MAS solutions, constructing respective multi-atlases

from the training set of sixteen labelled short axis stacks. Later, Bai et al. (2015b) demonstrated improved left ventricular segmentation, using leave-one-out cross-validation to realise a multi-atlas of 82 subjects. They proposed an augmented label fusion scheme that used a SVM to interrogate a rich feature vector of contextual image cues. Lastly, and although eventually combined into a single probabilistic atlas of biventricular short axis anatomy, Bai et al. (2015a) leveraged high resolution, 3D cine data of over a thousand healthy cases, demonstrating the value conferred by atlas quantity and quality. The extensive atlas permitted statistical parametric mapping to study the distribution of cardiac shape and physiology in the spatio-temporal domain.

Whole heart and multi-class segmentation

As outlined in Chapter 2, in the setting of structural and CHD, a holistic appreciation of cardiac anatomy and disease morphology is critical to patient management. So whilst left ventricular volume, dynamic function and geometry remain important considerations, the structure of atrial and vascular components can play an equally sized role in shaping care (Zhuang, 2013). Hence, in this section we review the methods applied to the comprehensive 3D segmentation of multi-class and whole heart anatomy.

Though applied inconsistently, the "multi-class" and "whole heart" descriptors imply a set of anatomical targets that extend beyond the left and right ventricles and ventricular myocardium. At minimum, both will further delineate the left and right atria, and might also include the great arteries, and systemic and pulmonary venous drainage. Variably, these components are either isolated as distinct labels of what we describe as multi-class segmentation, or are collected within a single, whole heart blood pool (see Figure 2.10). In either case (and where of interest), the myocardium is typically labelled as a distinct object, variably capturing the muscular component of the left or right ventricles, or the unified ventricular mass. Unlike ventricular volumetry, for which short axis cine CMR provides an established reference standard, 3D multi-class and whole heart segmentation tasks have no commonly accepted, optimal imaging modality. Instead, segmentations have most often been made of spatially isotropic structural CMR or CT acquisition.

Image-driven methods

Amongst the earliest reports, Makowski et al. (2002) adopted both balloon and snake variants of 2D active contours, synthesising 3D, multi-class anatomy from the concatenation of axial planes. Despite strong qualitative performance, the authors acknowledge that their approach required manual supervision, including parameter tuning per case.

Graph cuts have also been presented as a means of isolating the 3D whole heart blood pool from isotropic CT data: Lombaert et al. (2005) adopted an interactive, multi-scale approach, making dramatic efficiency gains and realising qualitatively impressive results; Funka-Lea et al. (2006) presented region growing as a means of initialisation, assessing their approach within seventy patients. Larrey-Ruiz et al. (2014) also depended on a highly engineered workflow, segmenting the whole and left heart blood pools from 3D CT through the manipulation of various binary thresholds. Lastly, Bui et al. (2018) developed a probabilistic segmentation from the diffusion of weighted random walkers originating from a small number of manually labelled slices. Combining this approach with thresholding and morphological operations to automate seed initialisation, they demonstrated segmentation of the left and right heart from 58 CT studies.

Model-driven methods

In spite of the methodological contributions made by the cited body of work, the performance of purely image-driven methods remains susceptible to limitations in image quality, overlapping intensity distributions of different anatomical structures, and the presence of artefacts. Inevitably, these confounding features prompt predicted segmentations that comprise clinically implausible cardiac anatomy. For example, results might present cardiac sub-structures with spurious geometry or morphology, or might fail to capture the structurally complex connectivity of the isolated great and small circulations. In response to these challenges, model-driven techniques have garnered greater success and popularity than their image-driven counterparts.

Authors have previously attempted to construct models of anatomical shape, seeking to leverage their results within whole heart and multi-class cardiac segmentation. Captured within a surface mesh of tessellating triangles, these have

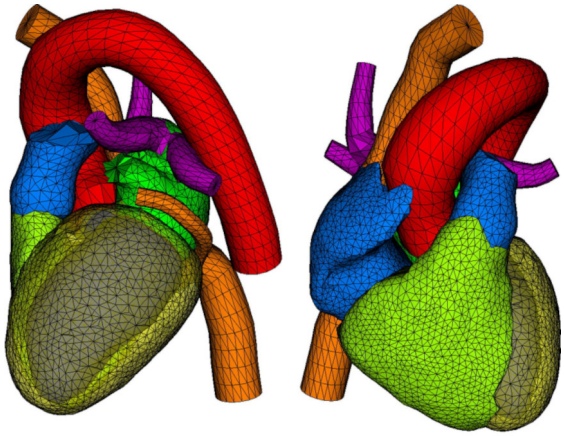


Figure 3.5: A prior model of 3D multi-class cardiac anatomy, encapsulated within a triangulated, surface mesh. Reproduced from Ecabert et al. (2011).

typically been deployed within a, deformable model framework (see Figure 3.5). At test time, the nodes of the mesh are attracted to target points (via an external energy term), identified by searching the image feature profiles normal to the associated triangles, for suitable boundary candidates. Rather than rely on a single edge detector, however, the identification of target points is greatly enhanced by simulated search (Peters et al., 2010), a procedure which selects a locally optimal detection function from a parameterised family. This means that per node of the base model, the identification of target points can be sensitised to the anticipated edge direction (light-to-dark, or dark-to-light) or absolute intensities. Finally, deformation of the initial shape model is constrained by prior knowledge, encapsulated within the internal energy of the evolving geometry. A variety of associated formulations have been explored.

Although inspired by the work of Weese et al. (2001) (who demonstrated a means of incorporating prior knowledge of statistical shape variation), early work presented by Ecabert et al. (2005) relied on a hand engineered heuristic to constrain the deformation of anatomical geometry during model fitting. As per Lorenz and von Berg (2006), their model of multi-class geometry was derived from cardiac CT data, and separately described the four chambers, the left ventricular myocardium, the great arterial trunks and the confluence of pulmonary and systemic veins with the left and right atria, respectively. Ignorant of the distribution of anatomical shape, they instead relied on a purely geometrical constraint, assuming that plausible variations in anatomy could be modelled by a per sub-component, affine transformation. In their active surface segmentation procedure, (and after initial-

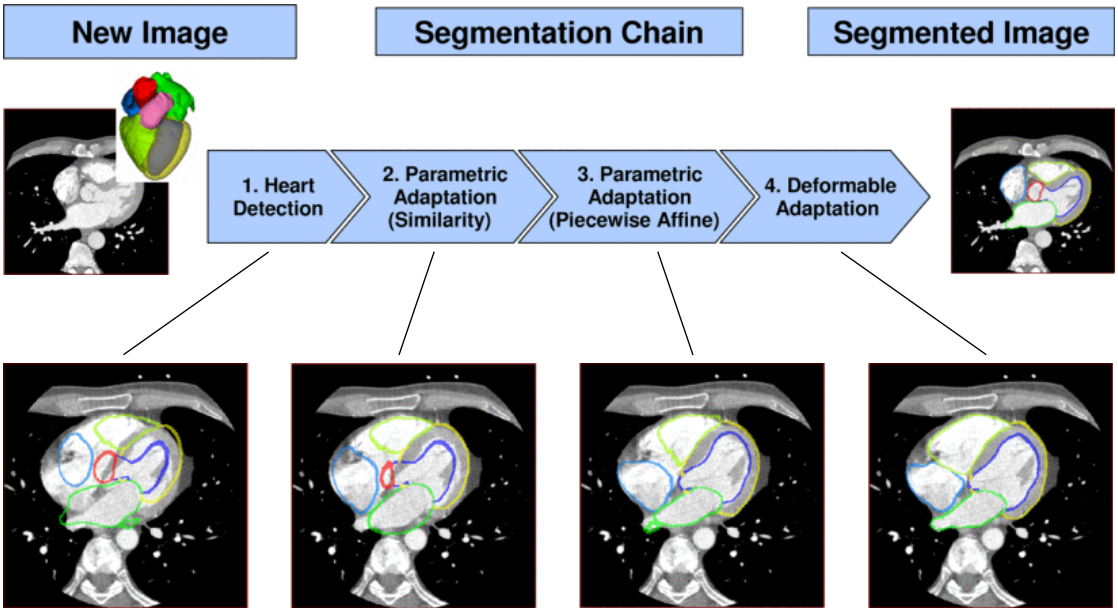


Figure 3.6: Researchers have developed various schemes to fit a prior model of cardiac anatomy to the unseen test image, including: automatic region extraction; rigid, similarity transformation; parametric adaption; and deformable adaptation. In this example, constructed from the work of Ecabert et al. (2008), the authors deploy a triangulated surface model of the four cardiac chambers, the left ventricular myocardium and the ventricular outflows. Note that both the prior model and final segmentation approximate the envelope of each anatomical component only, and are devoid of any high-resolution structural features.

isation by Hough (Duda and Hart, 1972) and global similarity transforms) they constructed an associated internal energy term to penalise deformations that could not be closely described by this model.

Soon after, they extended their approach to also consider the statistical distribution of anatomical training shapes. Incorporating the PDM within a coarse-to-fine cascade, Ecabert et al. (2008) presented a model-driven procedure that sequentially included: global similarity transform; parametric adaptation of shape models expressed by anatomically piece-wise affine transform, a linear combination of principal components of shape variation or both; and final deformable adaptation of mesh vertices according to associated external and internal energies (see Figure 3.6). They observed, however, that the introduction of statistical shape priors failed to improve segmentation accuracy, either when serving as a param-

eterised model of shape variation, or as a reference for establishing an internal energy penalty. The authors speculated that the additional degrees of freedom associated with the piece-wise affine model allowed for more flexible adaptation to the test image at hand; and that perhaps owing to its basis on limited training data (27 CT scans) the statistical model lacked the expressive power to conform to unseen cases. In a leave-one-out experiment, their deformable, piece-wise affine model of anatomy achieved sub-millimetric surface-to-surface error within a test set of 28 images.

Applying the same approach to volumetric CMR, and in a four-fold cross-validation on 42 cases, Peters et al. (2007) demonstrated an equivalent level of performance. Also within this scheme, and considering the same CT and CMR databases, Peters et al. (2010) presented a thorough analysis of the value conferred by simulated search, observing its ability to increase the capture range and ultimate accuracy of multi-class segmentation by model deformation. Lastly, the model fitting and deformable segmentation pipeline was extended by Ecabert et al. (2011), providing for the sequential adaptation of the great arteries.They also introduced low resolution mesh operation and sub-mesh freezing to realise gains in the computational efficiency of their so-called "adaptation engine".

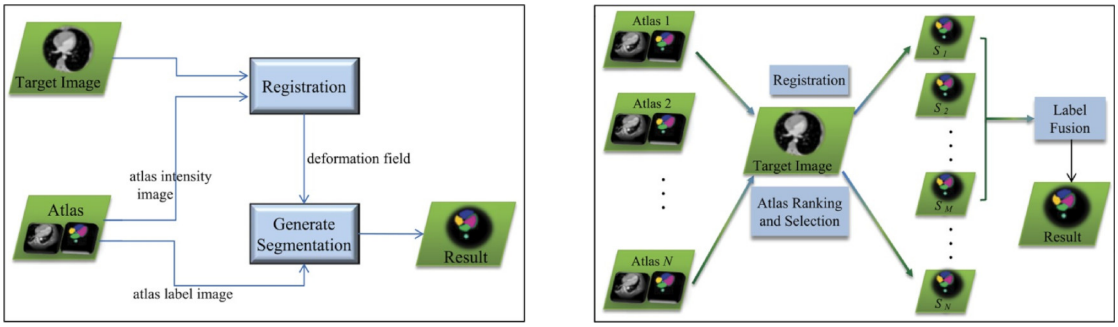
Work published by either Ecabert et al. or Peters et al. was collectively undertaken by various research teams associated with the industrial conglomerate Philips NV. At around the same time, a closely related approach to multi-class segmentation was developed and presented by a group within a research group from Siemens AG. Also within cardiac CT, Zheng et al. (2007, 2008) made several novel contributions, including: a four-chambered, multi-class model of the heart, enhanced by the inclusion of meaningful landmarks defining, for example, the atrioventricular valve annuli; an alternative, geometric scheme for establishing correspondence between the nodes of the PDM, one that avoided the inter-subject registration demanded by the approach of Frangi et al. (2002); marginal space learning; and steerable features. Whilst the overall procedure shared the principles underlying the deformable segmentation approach presented by Weese et al. (2001), the final two of these developments altered the process by which the prior cardiac model was fit to test data.

Zheng et al. (2008) formulated spatial transformation as a classification problem, asking whether there was an appropriately rotated and scaled anatomical object centred at each voxel. Rather than search the 9D parameter space of possible similarity transforms, they used marginal space learning to take a sequential approach: successively determining optimal translation (3D), translation-orientation (6D) and finally translation-orientation-scaling (9D) parameters. At each step, only a subset of candidates from the previous sub-transformation are considered, reducing the size of the overall search space by orders of magnitude. Moreover, within this scheme they introduce steerable features. Essentially determined on a sparse 3D lattice, rotated and scaled with respect to the voxel under consideration as an anatomical reference point, this approach avoids computationally costly image rotation and resampling.

Under marginal space learning, steerable features were input to a probabilistic boosting tree classifier (Tu, 2005) to guide all stages of model fitting, including: cardiac localisation, boundary detection and subsequent non-rigid deformation. The final of these was regularised by projecting the deformed model onto the shape manifold determined by the principle modes of shape variation expressed by a PDM. In a four-fold cross-validation of 323 CT volumes drawn from 137 patients (of unknown health), their approach performed well, only trailing the work of Ecabert et al. (2005) by a small margin[2].

Unlike the conclusion of Ecabert et al. (2008), Zheng et al. (2008) fully incorporated an ASM to constrain deformable model fitting. Similarly, Haak et al. (2013) developed an ASM of the four chambers and aorta, learning two PDMs from the labelled CT data of 151 patients. The first captured 90% of the shape variation of the collective anatomy; the second set of models respectively captured 98% of the shape variation of each sub-structure. This sequential approach was required to contend with the complex geometry of the heart, its segments, and their variation. In the context of 3D transoesophageal echocardiography, and after initial, landmark-based transformation, they fit their deformable models to a probabilistic map of label membership, determined by a gamma mixture model.

[2]Differences in image data and performance assessment (including the suggestion that Ecabert et al. (2005) rejected gross segmentation failures from their analysis) preclude direct comparison.



(a) (b)

Figure 3.7: A comparison of single atlas and multi-atlas segmentation (MAS), as applied to 3D multi-class, cardiac image segmentation. (a) In the single case, an atlas comprising an intensity image is spatially registered to the unseen test data. The predicted segmentation is subsequently established by propagating the associated atlas labels under the resulting spatial transformation. (b) In MAS, each member of a multi-atlas of N independent, intensity images is registered to the test case. Propagating each associated atlas label set in turn establishes multiple segmentation predictions. These are subsequently resolved to a single estimate via the process of label fusion. Reproduced from Zhuang et al. (2015).

Whilst shape modelling has provided a basis for rich investigation, the geometrical complexity of the whole heart (at least compared with the left ventricle) presents a challenge to those seeking an ASM of anatomy. Free from strong shape constraint, atlas-based methods (see Figure 3.7a) provide greater flexibility in the face of unseen cardiac anatomy. This assertion is borne out by the work of Ecabert et al. (2011), whose deformable approach, in the absence of statistical shape priors, bears striking similarity with the atlas-based formulation. Where deformable models most often describe mean geometry by a triangulated mesh, atlas-based methods capture anticipated anatomical shape within ground truth label maps.

Making novel contributions to the formulation of both the rigid and non-rigid components of the typical framework for atlas transformation, Zhuang et al. (2010b) applied atlas-based segmentation to multi-class labelling of 3D CMR. They relied on a simple atlas coupling a mean intensity image (derived from the spatially aligned CMR data of ten healthy volunteers) and discrete multi-class label map. Whilst their methodological contributions reduced segmentation error, it is noteworthy that the worst-case performance observed was associated with a patient

with tetralogy of Fallot. Soon after, the same group applied this approach to 3D echocardiography, investigating a novel similarity metric sensitive to local phase and geometric features (Zhuang et al., 2010c).

Segmentation by single, simple atlas registration and label propagation has endured since this time, including its application to multi-class delineation of 3D CT. Cai et al. (2017a) presented a scheme to dynamically update their atlas (implied to be a single, manually selected and segmented example) to reflect the temporal phases of the cardiac cycle. Subsequently using the same atlas and test set of fourteen volunteers, they investigated the effect of image pre-processing operations, including dynamic range windowing and Gaussian filtering, observing related improvements in the delineation of the epicardial contour (Cai et al., 2017b). Most recently, Galisot et al. (2018) considered a variant of the single atlas-based approach, developing per class, locally bounded mean intensity and probabilistic label templates from twenty 3D CMR and CT scans. Embedding the class-specific atlases within a relational graph capturing relative anatomical position, their procedure is sensitised to global cardiac structure.

The popularity, and differential superiority of *multi*-atlas strategies (see Figure 3.7b) has grown since its early application to 3D cardiac CT. Isgum et al. (2009) developed an atlas from fifteen manually segmented, low dose, high 3D resolution CT scans. At test time, they fused atlas predictions via a spatially varying, weighted decision scheme, informed by the local image similarity achieved after registration. Segmenting the union of whole heart blood pool and myocardial anatomy, and the aorta, they observed statistically significant performance gains compared with a range of single atlas formulations. Performance, however, declined in the presence of abnormalities including aortic calcification and metallic surgical devices. Extending this task specification to a four-chambered, multi-class segmentation, Kirişli et al. (2010) performed an extensive multi-vendor and multi-centre assessment of over 1420 contrast-enhanced CT scans, drawn from almost 900 patients. They registered an atlas comprised of eight manually labelled CT volumes to each test case, resolving predicted segmentation by majority voting. Despite the size of their experiment, their performance assessment was limited by the unavailability of ground truth labels, precluding quantitative analysis for all but a minority subset of test patients. Nonetheless, in conjunction with their

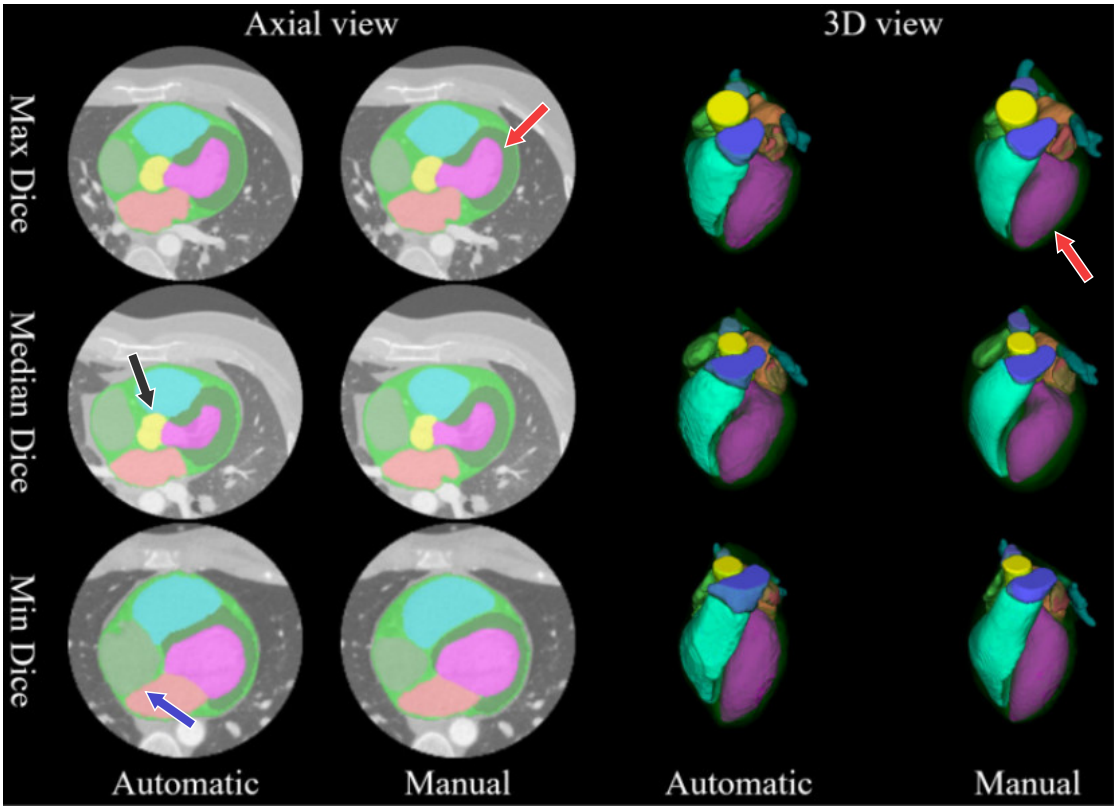


Figure 3.8: An adapted summary of the results achieved by Bui et al. (2020b), using their multi-atlas of 36 CT images, each manually segmented into twelve distinct cardiac classes. Note that whilst predicted segmentations are visually plausible, their quality (as inherited from the associated multi-atlas) fails to describe the fine, structural details that might characterise CHD anatomy and disease morphology. This quality manifests in: the smooth appearance of the endocardium, free from trabeculation (red arrow); and apparent communication of left and right heart structures normally separated by thin interfaces, giving the false impression of congenital defects such as right-sided atrioventricular discordance (black arrow) and atrial septal defect (blue arrow).

qualitative assessment, they demonstrated impressive performance that they determined to be consistent with inter-observer variability.

In closely related task specifications and 3D CT acquisitions, extensions to the MAS approach include: investigation of conditional entropy for atlas ranking and selection (Zhuang et al., 2015); application in radiation oncology planning, including the delineation of the coronary arteries and isolation of the great vessels

within a multi-atlas of twelve cases (Zhou et al., 2017); multi-class segmentation of non-contrast-enhanced CT via an atlas of eight manually segmented contrast-enhanced scans (Shahzad et al., 2017a); region of interest detection prior to MAS (Bui et al., 2020a); a comprehensive atlas of 36 cases, each segmented into twelve and four cardiac and non-cardiac structures, respectively (Bui et al. (2020b), see Figure 3.8); and integration within the Bayesian framework (Ghosh et al., 2021).

MAS also has a long track record of development within its application to 3D CMR data. Observing that individual members of a clinically derived multi-atlas may be corrupted by image artefacts (and that this in turn may compromise atlas registration), Zhuang et al. (2010a) devised the multiple path propagation and segmentation (MUPPS) strategy. This sought to leverage the complementary strengths of single atlas (improved registration with a synthetic mean image, for which noise is reduced) and multi-atlas (statistical gains in accuracy achieved through multiple predictors) segmentation. At test time, MUPPS seeks to register the mean CMR image to the unseen test case, *via* each of the spatial transforms established during atlas construction by co-registration (and which collectively express the manifold of training shapes). Whilst the authors observed that MUPPS was associated with statistically significant improvements over the conventional multi-atlas approach, these were largely reduced by the inclusion of advanced label fusion strategies.

Perhaps for this reason, later applications have favoured conventional MAS. Within a multi-class task specification, Zuluaga et al. (2013) presented a ranking strategy based on the local normalised correlation coefficient, selecting optimal examples from separate 3D CMR and CT multi-atlases. In a leave one out design, they investigated the influence of training set size, attributing superior performance on the CMR data to its larger associated atlas ($n = 22$), when compared with the equivalent CT examples ($n = 8$). Others have also investigated novel strategies for multi-atlas label fusion, including: by consideration of local patch similarity at multiple spatial scales (Zhuang and Shen, 2016); and alternatively, by non-local patch similarity (Heinrich and Oster, 2018). Whilst the latter successfully corrected minor registration errors, it also necessitated segmentation post-processing by smoothing random walk.

Patient-specific models of 3D CHD anatomy

Given the plethora of image-, and in particular, model-driven methods for segmenting multi-class and whole heart anatomy from 3D image data, it is natural to ask how these have been applied to CHD. In so doing, we refer to the Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease (Pace et al., 2015) (HVSMR) Challenge. As per other segmentation challenges, this initiative sought to evaluate the performance of different methods on a particular task, and stimulate associated research activity. The primary clinical motivation of this challenge was patient-specific 3D printing of CHD anatomy and disease morphology, for pre-operative planning and consensus decision-making. To this end, the organisers of HVSMR provided a dataset of balanced SSFP, CMR volumes of high isotropic resolution. The twenty patients included reflected a range of CHD diagnoses and associated structural interventions. For ten of these cases, manually segmented, ground truth labels were provided for algorithmic training, tuning or development. These described the whole heart anatomy of the cardiac blood pool within a single class (including the left and right atria, left and right ventricles, aorta, pulmonary veins, pulmonary arteries, and the superior and inferior vena cavae) and the muscular ventricular myocardium (surrounding both left and right cavities). Pulmonary vascular components were terminated proximally, the organisers reasonably arguing that their dense branching structure can obscure other anatomy and in any case, are not often represented in printed models. Likewise, and since the coronary arteries are less often salient to the direction of care than they might be in adult or acquired cardiac disease, their intra-muscular course was included within the myocardial label. In order to expose the clinically salient features of anatomy, including the presence of defects or interventional modification (see Figure 3.9), ground truth labels were provided at the limit of spatial resolution, realising a highly detailed result.

Scientists and researchers were subsequently invited to submit their solutions to this task, segmenting the remaining ten test cases, for which the ground truth labels were withheld. Owning the manual segmentations of the test data, the organisers are able to grade the results attained by each submission, assessing performance using a number of metrics including the Dice similarity coefficient (Dice,

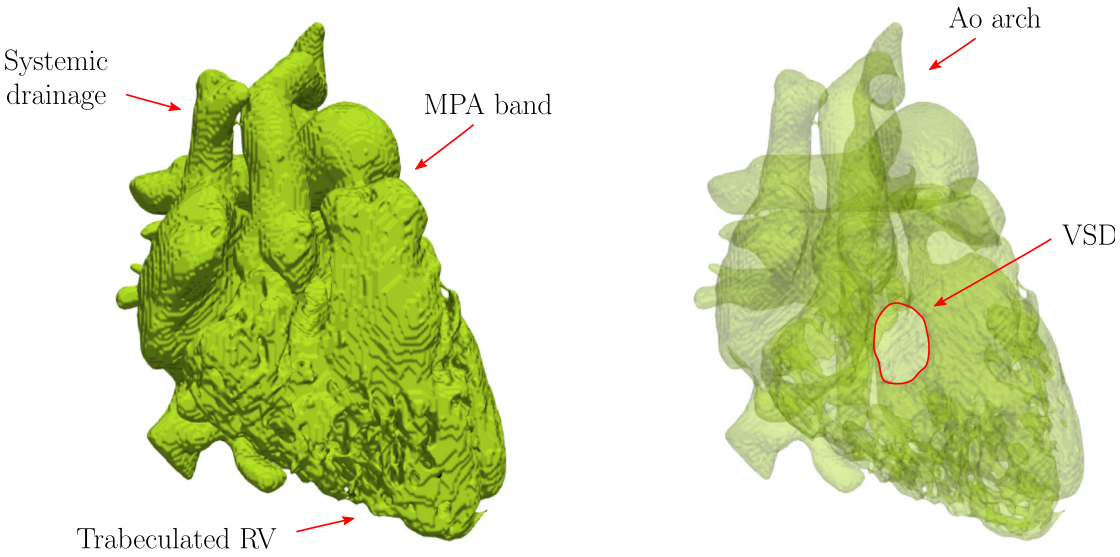


Figure 3.9: The ground truth labels provided by the Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease (Pace et al., 2015) (HVSMR) Challenge expose a highly detailed and clinically relevant representation of 3D CHD anatomy and disease morphology. In this example whole heart blood pool drawn from the challenge training set, such salient features include: proximal extension of the superior and inferior venous drainage; trabeculation of the right ventricular (RV) endocardium; the presence of a surgical band constricting the main pulmonary artery (MPA); irregularities in the aortic (Ao) arch; and a ventricular septal defect (VSD).

1945) (DSC) (which scores the spatial overlap between predicted and ground truth segmentations in the range $[0, 1]$, reflecting zero and perfect agreement, respectively). Table 3.5 collects and ranks the results of different segmentation methodologies assessed against the HVSMR task, including those submitted to the original challenge (prior to October 2016) and those published in the time since.

Image-driven methods

As reflected in Section 3.2, a variety of image-driven methods have been applied to the segmentation of 3D patient-specific CHD morphology. However, in the majority of clinical reports, these have largely been deployed to achieve a crude representation of anatomy, one that subsequently requires significant manual editing. In technical reports drawn from the image processing literature, researchers

Table 3.5: In order of descending spatial overlap, a performance ranking of published segmentation methodologies when applied to the HVSMR test set and its task of isolating CHD anatomy and disease morphology from 3D CMR data. DSC_{BP} and DSC_{MY} record the Dice similarity coefficient (Dice, 1945) (DSC) for the whole heart blood pool and ventricular myocardium, respectively. Note that whilst other reports have been published since the HVSMR Challenge, to ensure fair comparison, we restrict our summary to those that report results on the full test set, rather than, for example, a cross-validation on the training data. *semi-automated methodology; CNN stands for convolutional neural network.

Citation	**Method**	DSC_{BP}	DSC_{MY}
Lösel and Heuveline (2017)	Random walkers*	0.957	0.832
Zheng et al. (2019a)	CNN	0.942	0.837
Zhang et al. (2019b)	CNN	0.941	0.839
Rezaei et al. (2020)	CNN	0.940	0.860
Zheng et al. (2020)	CNN	0.937	0.830
Liang et al. (2019)	CNN	0.936	0.828
Zheng et al. (2019b)	CNN	0.936	0.823
Yu et al. (2017a)	CNN	0.931	0.786
Yu et al. (2017b)	CNN	0.921	0.821
Ran et al. (2018)	CNN	0.929	0.813
Dou et al. (2017)	CNN	0.928	0.739
Yang et al. (2018a)	CNN	0.928	0.739
Min et al. (2020)	CNN	0.926	0.821
Wolterink et al. (2017)	CNN	0.926	0.802
Zuluaga et al. (2017)	Atlas-based	0.900	0.730
Shahzad et al. (2017a)	Atlas-based	0.885	0.747
Li et al. (2017b)	CNN	0.873	0.517
Tziritas (2017)	Markov random field	0.876	0.612
Wang et al. (2017a)	Random forest, active contour	0.856	0.664
Mukhopadhyay (2017)	Random forest	0.794	0.495

have sought to expand the scope and reliability of image-driven techniques through their combination, often within complex processing pipelines.

For example, in their submission to HVSMR, Tziritas (2017) make an initial segmentation prediction by MRF parameter optimisation. This was subsequently refined by probabilistic intensity modelling and anatomical feature tracking to achieve a blood pool overlap score of 0.876. Whilst far from a clinically acceptable approach, their pipeline did out perform two challenge submissions, each of which relied upon supervised classifiers and are described immediately below.

Wang et al. (2017a) employed RFs at different stages of a multi-faceted pipeline. At the outset, a classifier was used to detect anatomical landmarks and inform the alignment of a prior model of whole heart anatomy (whilst the authors describe this as being a statistical shape model, it is unclear whether the variation in anatomical geometry is incorporated within their approach, see discussion in the following section on model-driven methods). The registered mean model was used to initialise active contour refinement which, rather than evolve according to external forces based on pixel intensity, was sensitised to the class membership probabilities inferred by a multi-class RF classifier, trained on second order Haar-like features. Observing that the clinically acquired HVSMR training data were degraded by the presence of noise, Mukhopadhyay (2017) first minimised an energy criterion based on total variation, before RF prediction.

Unfortunately, works based on RFs placed last in our ranking of HVSMR submissions. In the worst case Mukhopadhyay (2017) achieved a score of just 0.794 for the spatial overlap between predicted and ground truth blood pools. Whilst some have previously suggested a Dice score of 0.7 as indicative of strong segmentation agreement (Collins et al., 1998), in the HVSMR (and related 3D CMR whole heart segmentation tasks) we assert that scores closer to, or in excess of, 0.95 are required to approach clinically acceptable levels of performance. Such a high bar is necessitated by the increased volume-to-surface area ratio of the whole heart blood pool, a property which biases the DSC toward the large number of interior voxels, those remote from the endocardial boundaries that ultimately define anatomical geometry.

Spatial overlap performance in this range is achieved by the image-driven approach published by Lösel and Heuveline (2017). They develop a probabilistic

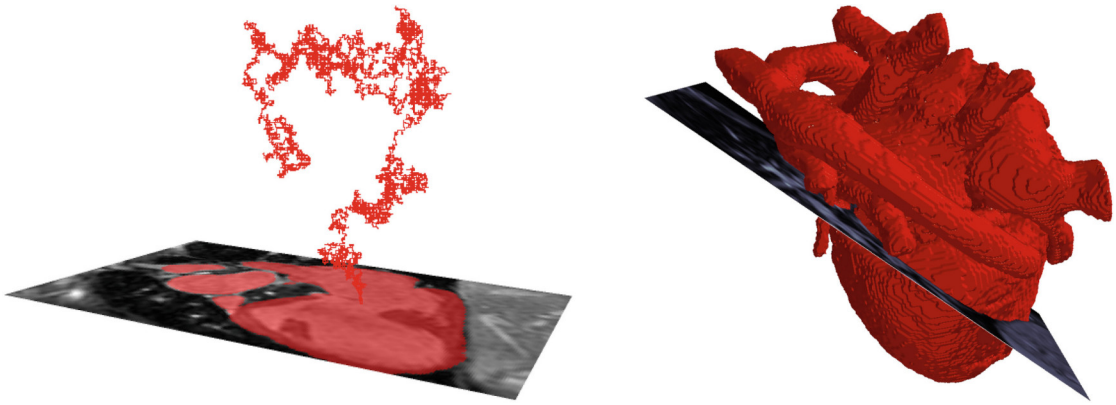


Figure 3.10: Presently leading the ranking of published submissions to HVSMR, the semi-automated procedure of Lösel and Heuveline (2017) develops a probabilistic segmentation according to the diffusion of random walkers, emanating from a subset of manually labelled slices.

segmentation from the diffusion of weighted random walkers originating from a small number of manually labelled slices (see Figure 3.10). Albeit on only a subset of four cases from the HVSMR test set (and for this reason their work is not included in Table 3.5), Pace et al. (2015) achieve similarly strong performance, deploying KNN classification within an interactive segmentation framework. Relying of a subset of manually labelled slices (requiring approximately an hour of operator time), they trained an individualised, patch classifier per patient, achieving state of the art results. Whilst limited by the size of their respective test sets (it seems unlikely that ten or fewer cases could faithfully represent the range of anatomy and CHD morphology encountered in the clinic), both of these image-driven approaches achieve mean overlap scores in excess of 0.95, topping the ranking of published HVSMR analyses. In common, however, they each rely on a subset of manually delineated slices to inform downstream algorithmic execution. Apart from the operator burden associated with labelling, segmentation of a 2D slice in isolation can be perceptually challenging, especially when drawn from a 3D volume containing complex cardiac anatomy.

Model-driven methods

Outside of deep learning solutions, improvements in fully automated segmentation of the HVSMR test set have been achieved by atlas-based, and in particular MAS,

methods. Two submissions to the challenge adopted this framework. Though each of the training volumes was made available with a corresponding manual segmentation of the whole heart blood pool and ventricular myocardium, Shahzad et al. (2017b) elected to construct their atlas from a separate, external source. They employed the eight manually labelled CT volumes provided by the work of Kirişli et al. (2010). Given the mismatch in acquisition (CMR versus CT), task specification (whole heart and ventricular myocardium, versus multi-class including the left ventricular myocardium), and level of anatomical detail, this appears a strange choice. Not only did this necessitate inter-modality registration, but also required bespoke post-processing (including by Hough transform (Duda and Hart, 1972) and active contour) to provide for the aortic and right ventricular components: both of which were demanded by the HVSMR Challenge specification, but absent from their chosen atlas. Perhaps surprisingly, however, and despite these obstacles, their pipeline out-performed those submissions relying on fully automated image-driven methods.

In a like-for-like comparison, however, their results were not as strong as those of Zuluaga et al. (2017), who achieved a Dice score of 0.9. Constructing their atlases from the provided training data, they used the HVSMR Challenge as the basis for a comprehensive investigation of different design choices within the MAS framework, comparing: different atlas constructions (multi-atlas versus single atlas and associated discrete, majority voted label map); label specification (whole heart blood pool and ventricular myocardium versus their union); multi-atlas fusion schemes (majority voting versus similarity and truth estimation for propagated segmentations (STEPS) (Cardoso et al., 2013) versus simultaneous truth and performance level estimation (STAPLE) (Warfield et al., 2004)) and post-processing (none versus E-M of a GMM). They found MAS to be far superior to the single atlas equivalent. Although less substantial, further gains were associated with the use of E-M post-processing. Lastly, the benefit conferred by a given label specification or fusion strategy was unclear, with no significant differences observed between the design choices considered.

More generally, Zuluaga et al. (2017) make keen observations surrounding the applicability of MAS to congenitally malformed anatomy. In common with previous reports, they note a drop in segmentation accuracy associated with significant

anatomical variations. In the context of MAS, they attribute this observation to two sources: (1) the challenge of amassing a multi-atlas that is representative of the highly heterogeneous CHD population; and (2) the limitations imposed by existing image registration methods on the space of possible atlas transformations. Particularly relevant to the presence of congenital defects, they cite the inability of current registration strategies to cope with changes in anatomical topology. For example, the presence of a ventricular septal defect constitutes a discrete change in the topology of the blood pool, one that cannot be expressed by continuously varying spatial transformation. Combined, these factors limit the suitability of atlas-based segmentation as a means of addressing CHD anatomy.

This is not to say that atlas-based analyses are without application in other, closely related, CHD image processing tasks. Zuluaga et al. (2015) consider the application of this approach to disease classification, inferring CHD diagnosis by the relative similarity between a test image and a set of spatially registered atlases, each representative of a different defect or condition. In a class-balanced sample of sixty 3D CMR acquisitions, they sought to separate patients into three groups, those: with normal anatomy; that had undergone arterial switch; and that had undergone atrial switch to correct transposition of the great arteries. This insightful approach achieved an impressive diagnostic classification accuracy of 97.3%.

Lastly, we address the dearth of publications developing statistical shape models of 3D CHD anatomy. We have found only a single, vague description of their application to the anatomy of the whole heart or its multi-class description. Wang et al. (2017a) claim to have developed and deployed a statistical shape model of the whole, congenitally malformed heart. Unfortunately, however, their account does not make clear how this was constructed from the ten training cases made available by HVSMR. In the absence of further detail, we can only speculate that they established a low-resolution, mean whole heart shape, and that this was used to initialise the active contour refinement described in the preceding section. Hence it is difficult to ascertain where and how the distribution of training shapes is considered, if at all. Perhaps more problematically, we also fundamentally disagree with their method's underpinning assertion that "... the overall shape of [*the congenitally malformed heart*] appears relatively similar from patient to patient"[3].

[3]Later, please see Figure 5.4 for a visual refutation.

3.4 Challenges posed by 3D CHD modelling

Despite the wealth of cited research, Section 3.2 found that for the most part, clinical exponents of the patient-specific 3D printing workflow have largely relied on manual methods for image segmentation and editing. It is natural, therefore, to ask why the wealth of methodologies cited in this review of the technical, image processing literature, have not found application in clinical practice. This disparity is perhaps most simply explained by the publication bias that favours reports of novel segmentation methodologies that demonstrate positive results. However, this section also seeks to explain this gap by examining the methodological challenges posed by 3D CHD modelling and whole heart and multi-class segmentation. Importantly, we also differentiate between the hurdles shared by related cardiac image segmentation tasks, in which the majority of developmental work has been conducted, and those unique to patient-specific modelling of CHD. Addressing each in turn, we examine challenges associated with: motivation and corresponding task specification; image acquisition and data quality; and anatomical heterogeneity.

Motivation and task specification

The preceding review traced the development of cardiac image segmentation methods, from their roots in assessing ventricular volume using 2D short axis images, to the geometrical representation of patient-specific 3D anatomy described by volumetric CMR or CT data. Between these extremes, clinical motivations vary significantly. In the former, we seek to effectively regress an imaging dataset to a single scalar index of cardiac health. Therefore, so long as spatial overlap between predicted and ground truth segmentations is sufficient, their detailed geometry is largely unimportant. In the latter, we seek insights as to the size, morphology and configuration of the heart, in order to enhance the selection and planning of structural intervention.

These different motivations are borne out in the divergent specifications of each application's associated segmentation task. In its simplest formulation, volumetry might require delineation of the left ventricular cavity alone. Focusing on a single anatomical target, model-driven methodologies have flourished (Tavakoli and Amini, 2013). Whereas, patient-specific modelling demands the isolation of whole

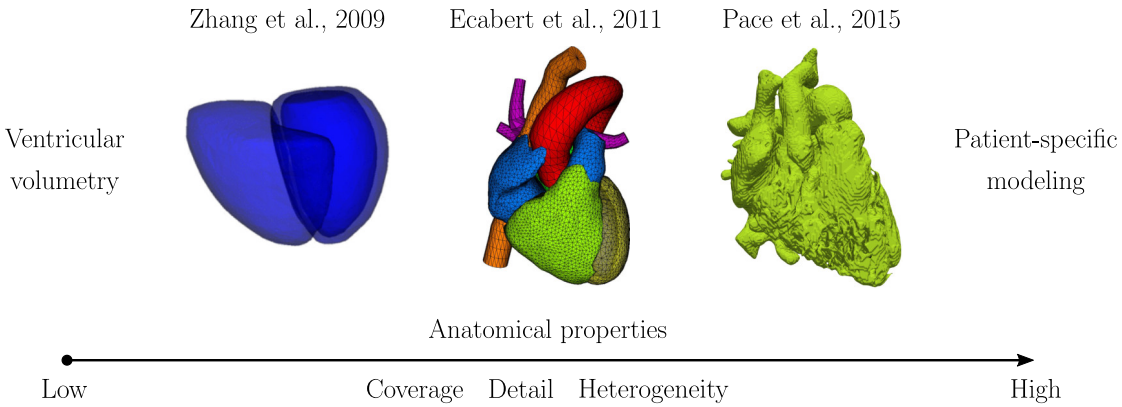


Figure 3.11: The underlying motivations of ventricular volumetry have shaped the development of conventional, model-driven methods for 3D cardiac image segmentation. With respect to anatomical coverage, detail and heterogeneity, these are in conflict with the requirements of those seeking patient-specific models of CHD morphology.

heart cardiac anatomy, a freeform, organic surface composed by sub-structural segements, and whose morphological complexity outstrips the conical appearance of the left ventricle.

Differences in task specification, however, are not limited to the semantic classes composing segmented data, but extend to the level of anatomical detail represented (see Figure 3.11). Given that the presence or absence of congenital defects can be defined by thin tissue interfaces (often at the limit of spatial resolution; consider holes in the atrial septum, for example), accurate and detailed representation of cardiac anatomy is key to patient-specific 3D modelling. Accordingly, where short axis segmentations of the left ventricle need not describe the complex trabeculation of the endocardial surface, patient-specific models of CHD must provide structural representations at the limit of spatial resolution (Kim et al., 2008a).

Many of the cited technical works examining 3D whole heart or multi-class segmentation suggest treatment planning as one of their motivators. However, outside of those addressing the HVSMR Challenge, rarely has such detailed anatomy been examined. Some of the most successful reports (Ecabert et al., 2011; Zheng et al., 2008; Zhuang and Shen, 2016) have instead sought largely featureless representations of the heart, describing the anatomical envelope of each chamber and associated vasculature. Undoubtedly, these models remain highly relevant to the plan-

ning and guidance of catheter ablation, where the spatial uncertainty in tracked position might preclude more detailed representation (Tobon-Gomez et al., 2015). Nevertheless, we speculate that the research community's focus on low resolution anatomical approximation, also stems from its roots in short axis segmentation. Owing to its ubiquitous position as the clinical gold standard for ventricular volumetry, labelling of 2D short axis cine CMR has received the vast majority of research attention and methodological development (Habijan et al., 2020). Hence, when authors came to consider the segmentation of whole heart anatomy from isotropic 3D data, they naturally sought the same level of detail. In some cases, this approach limits the generalisation of their findings to clinical challenges such as CHD, where the fullest possible appreciation of patient-specific anatomy and disease morphology is paramount. As identified by Gao et al. (2011), the visual difference between the two approaches, and as demonstrated by Figure 3.11, is stark.

For our interests, it is unfortunate that so much of the associated methodological development has taken place within a low resolution, coarsely detailed context. Whilst model-driven approaches such as statistical shape modelling and atlas-based segmentation have advanced the state of the art, their application to highly detailed anatomical targets may be hampered by associated technical challenges. In particular, the success of the deformable models developed by Ecabert et al. (2011) and Zheng et al. (2008) may be difficult to replicate at higher spatial resolution. Firstly, this shift would increase the number of model landmarks required to capture detailed anatomy. Secondly, some of these would need to describe the subtle differences and inconsistent appearances of thin tissue interfaces, including the trabeculated endocardium. In combination, the challenge of drawing correspondences between such a dense set of anatomical features likely precludes meaningful shape modelling. These obstacles perhaps also inform the community's focus on low resolution models and might explain the lack of any concerted attempt to address the HVSMR Challenge using either statistical or deformable priors.

Image acquisition and data quality

Irrespective of modality, many of the challenges associated with limitations in image quality are shared between different cardiac image segmentation tasks and patient groups. Cardiac CT data rely primarily on the administration of exogenous agents to generate contrast between the cardiac blood pool and soft tissues. As a result, the timing of acquisition relative to the the passage of contrast strongly influences the derived image quality. Where this is synchronised correctly, the high spatial resolution (potentially sub-millimeter) of CT acquisition faithfully exposes the tissue interfaces that divide the heart and in which defects might occur. However, where this is sub-optimal, the lack of endogenous contrast between the blood pool and myocardium can obfuscate visualisation of intracardiac structures.

Compared with cardiac CT, CMR carries no associated ionising radiation burden, a particularly important consideration within the paediatric CHD population (Ntsinjana et al., 2011). These patients are not only more likely to require repeated diagnostic imaging during infancy, but also, due to their rate of growth and development, are more sensitive to the radio-biological effects of X-ray exposure. Owing to their age, they also have more time to express associated morbidities. CMR benefits from exceptional soft tissue contrast, exposing the blood pool of the cardiovascular system separately from the containing myocardial and vascular tissues. However, since its spatial resolution typically trails CT (typically being only isotropically millimetric), the interfaces between neighbouring anatomical structures can appear indistinct, posing a challenge for those pursuing their automated delineation (Zhuang, 2013).

Both imaging modalities suffer from the presence of image artefacts associated with implanted metallic devices (resulting in X-ray scatter within CT; and signal dephasing within CMR) and patient motion (including that associated with the cardiac and respiratory cycles). In the case of CMR, the passage of blood can also lead to signal dephasing: associated hypo-intense artefacts often arising at the confluence of the pulmonary veins and left atrium or through stenotic vessels, outflows and valves. Coupled with the presence of noise and other limitations in image quality, these artefacts motivate the development of model-driven techniques. Predictions made by this family of methods are bolstered by their dependence on prior

models, reducing the likelihood of spurious, anatomically implausible prediction in the vicinity of confounding deficiencies in image quality.

In common, and as most often examined in the whole heart and multi-class 3D image segmentation literature, typical CT and CMR acquisitions share isotropically high spatial resolution. This marks a shift from the dependence of ventricular volumetry on short axis CMR images, conventionally acquired as a series of 2D slices distributed at intervals along the long axis of the heart. Given that the in plane resolution of cine data far exceeds the slice offset, associated methodologies have been developed and applied in 2D (Mitchell et al., 2001). Thankfully, the majority of reviewed citations from the reviewed literature on whole heart labelling reflect the dimensionality of the target images and anatomy, investigating 3D methodologies (including 3D shape models and atlases).

Independently of the dimensionality or basis of the approach, however, it is difficult to ascertain whether potential differences in image acquisition or quality can explain the lack of methodological integration within clinical reports of patient-specific 3D printing. It is possible that eligibility criteria or screening processes deployed within the technical literature might preferentially select test cases more amenable to automated or semi-automated segmentation; and that perhaps by excluding clinically acquired examples of compromised image quality, algorithmic effectiveness has been inflated. These assertions, however, remain highly speculative. Just as problematically as their quality, the quantity of test cases used in the assessment of novel methods has often been lacking. This is particularly problematic in the setting of the structurally heterogeneous CHD population, to which we now turn.

Anatomical heterogeneity

The CHD population presents a structurally heterogeneous group in which no two patients share the same anatomy and disease morphology. Fundamentally, such diversity results from the presence of heart defects, and the variation in their size, morphology and location. Defects take a number of different forms (see Figure 2.2), but in addition to describing continuously varying changes in anatomical size and morphology (such as stenosis and hypoplasia), also characterise discrete topological

changes in cardiac structure. Septal defects, for example, associate the normally isolated great and small circulations; discordant or doubled connections alter the expected haemodynamic continuity of the two. This variety of lesions can occur within the heart or affect the extracardiac vasculature. Adding to this complex picture, structural interventions (either surgical or catheter-based) also introduce discrete changes in anatomy, such as through the creation of shunts between the left and right heart or perhaps most significantly, through palliative conversion to a univentricular circulation. This diversity poses a challenge to the model-driven, state of the art methods for whole heart and multi-class segmentation.

Deformable models of statistical shape and atlas-based segmentation both build prior representations of the heart from training data, their underlying ambition being to capture the distribution of anatomy encountered within the clinic. Some have suggested that in the setting of acquired disease, this aim is extremely challenging (Koikkalainen et al., 2008; Tavakoli and Amini, 2013), and likely impossible in the presence of congenital abnormality (Zuluaga et al., 2017). Where this ambition fails, several authors (Isgum et al., 2009) note the susceptibility of such priors to out of sample test cases, those that present anatomy far removed from, or outside of the training distribution. In such cases, dependence on strong priors limits the extent to which the atlas or statistical model can adapt to new anatomy - including the presence of congenital defects (Zhuang et al., 2010b; Shahzad et al., 2017b; Zuluaga et al., 2017) - compromising performance. Even in their investigation of patients with acquired heart disease, Ecabert et al. (2011) found constraining model deformation to the principle modes of shape variation to be too restrictive, eventually favouring a piece-wise affine model of anatomical deviation.

Attempts have been made to address this challenge, but have been limited to shape models of short axis data. For example, Albà et al. (2015) sought to extend the applicability of a generic 3D statistical shape model to the cardiac anatomy of patients with acquired hypertrophic cardiomyopathy and pulmonary hypertension. To do so, they employed a landmark-based, non-rigid transformation of short axis data to the canonical and pathology-free space of the prior model. Though free from statistical shape constraint, atlas-based segmentation also depends on this family of transformations to adapt a prior model to patient-specific anatomy.

As such, neither statistical shape modelling nor atlas-based segmentation provide a means of handling the discrete structural changes associated with CHD. Problematically, these changes in anatomical topology are not incidental to the representation of anatomy, but are associated with the pathological defects whose presence provide the primary clinical motivation for patient-specific modelling and segmentation. For this reason, we argue that neither of the conventional model-driven methodologies that might previously have been considered state of the art, are suited to 3D CHD image segmentation. Moreover, they have most often been tested against the normal anatomy of homogeneous cohorts of healthy volunteers (Zhuang, 2013), or patients for which structural changes are limited to continuous variations in morphology. We assert that this disparity best explains the mismatch between the methods developed within the technical image processing literature, and those deployed in the clinical practice of patient-specific anatomical modelling of CHD.

3.5 Conclusion

Our review of conventional cardiac image segmentation methodologies has revealed a disparity between the techniques employed in the clinically focused literature concerning patient-specific modelling of CHD, and those advanced by the technical image processing community. Where clinical applications have almost entirely relied on interactive segmentation workflows, including a time-intensive manual editing component, scientific researchers have presented automated, model-driven approaches, claiming to provide robust and reliable solutions. Examining the applications in which these methods have been developed and applied, we have argued that their basis in ventricular volumetry limit their ability to segment 3D CHD morphology in a clinically useful fashion. As summarised by Figure 3.11, fundamental differences between the two tasks include the anticipated level of anatomical coverage and detail, and the structural heterogeneity of the associated patient population. A historical focus on the coarse, relatively homogeneous ventricular anatomy of those with acquired disease is at odds with our clinical ambition. These differences render the model-driven approaches founded on statistical shape modelling and atlas-based segmentation inadequate for our purpose.

Our conclusion is not based on conjecture alone, but is supported by the findings of the HVSMR Challenge, in which model-driven methods (whilst out performing their conventional image-driven counterparts) rank towards the bottom of considered techniques. Hamstrung by prior models of at best limited representational capacity, and dependent on fitting procedures incapable of describing the discrete changes in anatomy associated with congenital defects, these methods have been surpassed by the introduction of convolutional neural networks (CNNs). The obvious omission from this chapter, and the interloper in Table 3.5, CNNs are a family of primarily image-driven, non-linear classifiers with extreme representational capacity. Drawn from the field of deep learning methodologies, CNN solutions presently lead the HVSMR ranking of automated segmentation methods, and will be our focus in the next chapter and remainder of this work.

Chapter 4

Deep learning for image segmentation

4.1 Introduction

In reviewing conventional cardiac segmentation methodologies, the previous chapter made two primary observations: (1) that clinical exponents of patient-specific 3D printing rely on a largely manual image segmentation protocol that can take hours per case; and (2) that model-driven methods developed by image processing scientists cannot provide the detailed representation necessary to visualise heart defects in a clinically meaningful fashion, nor are they adaptable to the structurally heterogeneous population with congenital heart disease (CHD).

Here, we look to the burgeoning field of deep learning for solutions, and in particular to convolutional neural networks (CNNs). This shift marks a return to primarily image-driven segmentation. Critically, however, rather than on those that are handcrafted, this family of methods rely on a data-driven set of discriminative image features, learned during network training. To understand why this might be sensible and how such learning might proceed, we firstly present a limited theoretical foundation, as necessary to comprehend our own clinical and methodological contributions in the remainder of this thesis. Subsequently, we review the literature reporting the results of CNN-based segmentation of whole heart and multi-class anatomy from 3D medical images.

4.2 Convolutional neural networks

4.2.1 Overview

In the context of medical image segmentation, we seek to model the transformation of data from a set of patient images to an anatomical or pathological label map. Whilst different mathematical formulations of this task have been presented, for illustrative purposes we consider the classification (abstractly described by the function f^*) of the N-dimensional, K-channeled input image, $\mathbf{X} : \mathbb{R}^N \rightarrow \mathbb{R}^K$, into one of the C mutually exclusive class labels of the one-hot encoded ground truth segmentation, $\mathbf{Y} : \mathbb{R}^N \rightarrow \{0, 1\}^C$:

$$\mathbf{Y} = f^*(\mathbf{X}) \tag{4.2.1}$$

A CNN approximates this transformation by a series of operators based on convolution. These are applied sequentially such that the network can be abstractly described as a directed, acyclic graph of operations arranged in layers, one that predicts the probabilistic segmentation, $\tilde{\mathbf{Y}} : \mathbb{R}^N \rightarrow [0, 1]^C$. Three properties of this approach are critical to CNNs' recent success:

Data-driven transformation
The weights of the convolutional kernel associated with each operator ($f^{(d)}$) are not determined or conceived by the image processing scientist, but form a set of free parameters ($\boldsymbol{\theta}^{(d)}$) to be optimised via data-driven training.

Non-linear modelling
The inclusion of suitable activation functions between convolutional layers admits non-linear modelling.

Deep networks
The expressive capacity of the resulting CNN is dramatically increased by repeated convolution within D sequential layers, otherwise referred to as network depth.

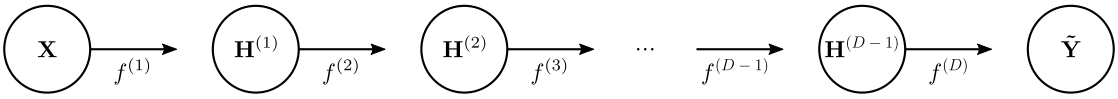


Figure 4.1: Summary mathematical operation of a feed-forward CNN. The network approximates the transformation between input image data, ($\mathbf{X}$) and its reference segmentation via a series of sequentially applied operators ($f^{(d)}$). Relying on the discrete, digital convolution for its expressive capacity, the respective weights of each operator are optimised via a data-driven training procedure, seeking an accurate and precise predicted segmentation ($\tilde{\mathbf{Y}}$).

Taken together, these attributes are characteristic of the broader family of deep learning methodologies; those which involve *deep* networks of many layers, the transformation at each of which is mediated by a parameterised function that is *learned* from exemplar data. Equipped with the CNN framework, and as depicted in Figure 4.1, the feed-forward transformation between source image and predicted segmentation can be written:

$$\tilde{\mathbf{Y}} = f^{(D)}(f^{(D-1)}(...(f^{(2)}\big(f^{(1)}(\mathbf{X};\boldsymbol{\theta}^{(1)});\boldsymbol{\theta}^{(2)})...);\boldsymbol{\theta}^{(D-1)});\boldsymbol{\theta}^{(D)}) \tag{4.2.2}$$

Together, Figure 4.1 and Equation 4.2.2 describe the so-called forward pass, the flow of information from the input ($\mathbf{X}$), which, via the hidden layers of the network ($\mathbf{H}^{(d)} = f^{(d)}(\mathbf{H}^{(d-1)};\boldsymbol{\theta}^{(d)})$), culminates in the predicted segmentation ($\tilde{\mathbf{Y}}$). By collecting the respective convolutional operations and kernels into a summary operator f, with associated parameter set $\boldsymbol{\theta}$, we can outline CNN approximation:

$$\tilde{\mathbf{Y}} = f(\mathbf{X};\boldsymbol{\theta}) \approx \mathbf{Y} \tag{4.2.3}$$

This compact description, however, neglects the constituent components on which modern, state of the art, non-linear, discriminative CNN architectures currently depend. Before presenting these building blocks, we firstly introduce the algorithms, training procedures and experimental designs that underpin CNN parameter optimisation and performance assessment.

4.2.2 Training and optimisation

Equation 4.2.3 presents the basis of an idealised and, owing to its parameter set, $\boldsymbol{\theta}$, highly flexible model. It does not specify, however, how these parameters can be chosen to optimally satisfy the approximation $\tilde{\mathbf{Y}} \approx \mathbf{Y}$, minimising the error between prediction and ground truth. We have already made passing reference to the idea that their determination is data-driven. In this section we expand on this high level description, presenting the machinery underlying the supervised training of deep CNNs.

Ambition of data-driven optimisation

Before introducing the components on which CNN training depends, we highlight the wider objective of CNN optimisation and machine learning more generally. That is: not to minimise the approximation error over the training dataset, but over unseen cases drawn from the underlying distribution that generates the real world population. Contextualising this statement for our application, we seek a CNN parameter set that can accurately predict the segmentation of 3D image data for those patients prospectively encountered in the clinic, not necessarily those historic cases that we can learn from. Due to the quantity (and prospect) of such patients and scans, neither of these sets can be explicitly gathered. Hence we rely on limited *training* and *test* sets: the former used directly within CNN parameter optimisation; the latter being held out from network training, and only used to assess performance on unseen cases. Critically, the quantity and quality of both must be carefully curated to promote the generalisability of CNN prediction and associated findings to the underlying patient population, our central aim.

Where this ambition is not met, learned parameters, whilst optimal for the training set, can be associated with inferior performance on test data. Such networks are said to overfit their associated training examples. Even in the context of seemingly challenging segmentation tasks, the possibility of overfitting is increased by the sheer capacity of state of the art CNNs. Potentially comprising millions of learnable parameters, each can be tweaked to optimise performance, prompting some to draw qualitative comparisons between overfitting and a network's capacity to "memorise" a small training dataset. In response, CNN training is most often

subject to regularisation, a topic whose techniques we will return to at the end of this Section 4.2.2. Judicious regularisation is balanced against network capacity to improve test set performance. However, where such efforts are over zealous, excessive regularisation and insufficient network capacity result in underfitting: a lack of expressive power that limits both training and test set performance.

Lastly, common practice relies upon a third, *validation* set, deployed to determine the optimal hyperparameters governing CNN configuration (rather than the model parameters, $\boldsymbol{\theta}$), including architecture, regularisation and optimisation itself. Such hyperparameters will be introduced throughout this chapter. Albeit indirectly, the features of validation cases influence network optimisation, and hence are excluded from test time performance assessment.

Supervised training

The distinction between unsupervised, weakly, semi-, and fully supervised modes of CNN optimisation is perhaps best informed by the comprehensiveness of available training data. Amongst these, and in the context of medical image segmentation, fully supervised (referred to as supervised from here on) training alone benefits from the availability of paired image-label examples (most often manually delineated by clinical experts) that fully demonstrate the task to be modelled. In contrast, the training data deployed in semi-supervised, weakly or unsupervised schemes are in some sense incomplete. Semi-supervised training typically investigates methods for learning from partially complete training data, composed by labelled and unlabelled examples; weakly supervised training more often relies on partial labels (such as scribbles (Lin et al., 2016)) or sources of prior knowledge; and unsupervised methods rely only on the features of observed image data.

These differences promote the association of each approach with different tasks. Critically, supervised training permits the formulation of CNN-based segmentation as inferring the conditional probability of a label map, given an input image: $p(\mathbf{Y}|\mathbf{X})$. Whereas, unsupervised training lends itself more naturally to feature learning and clustering, seeking to understand the underlying distribution of patient scans: $p(\mathbf{X})$. The former being better suited to predictive segmentation, supervised learning will be our primary focus for the rest of this thesis.

Gradient-based learning

Supervised training presents the distinct advantage that the approximation error between predicted and ground truth training examples can be directly computed. Most often referred to as the loss, this scalar is quantified by an associated loss function that is indicative of segmentation performance. Coupled with this supervisory signal, the parameters of a CNN define a so-called "loss landscape", associating every possible setting of $\boldsymbol{\theta}$, with an expected loss over the training set. In all practical cases, however, the high-dimensional, non-linear loss landscape is intractable, such that the space cannot be analytically sought for the globally optimal solution. Rather, our view of the loss landscape is restricted to the locality of empirically sampled parameter settings, including their initialisation.

Consequently, CNN training relies upon gradient descent, making an iterative series of incremental parameter updates, each determined to reduce the loss according to the local gradient, as sampled by feed-forward prediction. In differentiating the complex, multi-variable function described by the CNN, the gradient of the loss is proliferated throughout the network by the backpropagation algorithm (LeCun et al., 1989). This so-called backward pass returns the gradient of the loss with respect to the parameters of the model.

In this context, supervised training relies upon the following ingredients: (1) a loss function to establish the current prediction error or loss; (2) backpropagation to determine the gradient of the loss with respect to the parameters of the model; and (3) an optimiser to update the parameter values according to these gradients. We introduce these components in the following sub-sections. In the next Section 4.2.3 we introduce (4) various regularisation methods to enhance generalisation.

Loss functions

Within CNN-based medical image segmentation, a generic loss function, $L(\tilde{\mathbf{Y}}, \mathbf{Y})$, quantifies the error between a predicted labelling of input image data (determined by forward passing a training image through the network), and a ground truth reference standard. Irrespective of the formulation of a particular loss function, all such examples must be consistent with gradient-based optimisation. In other

words, they must return a scalar indicative of task performance, whose value is differentiable with respect to the parameters of the network. Informally, this requires that: (1) a small change in the model parameters and CNN output results in a small, continuous change in the loss, and (2) that contributions to the loss value itself can be determined from, and directly associated with particular voxels in the network output.

Whilst a wide variety of specialised loss functions have been designed to sensitise performance to the segmentation properties that are most salient to downstream application (El Jurdi et al., 2021), here we introduce two of the most popular choices. Both are deployed in the experimental work presented in later chapters.

Cross-entropy

If we flatten the spatial dimensions of the predicted and ground truth segmentations, and allow the index m to reference the M resulting pixels or voxels, the multi-class cross-entropy (CE) loss between the predicted and ground truth segmentations is:

$$L_{\text{CE}}(\tilde{\mathbf{Y}}, \mathbf{Y}) = -\frac{1}{M} \sum_{m} \sum_{c} Y_{c,m} \log \tilde{Y}_{c,m} \tag{4.2.4}$$

Where $Y_{c,m}$ constitute the elements of $\mathbf{Y}$, and provides an effective per voxel indicator of ground truth class membership of class label, c; and $\tilde{Y}_{c,m}$ constitute the elements of $\tilde{\mathbf{Y}}$, the probabilistic segmentation inferred by CNN. As presented in Goodfellow et al. (2016), the form of the CE loss can be naturally derived from the application of the maximum likelihood principle to the multi-classification task formulation. Optimising L_{CE} effectively minimises the distance between the one-hot encoded and continuous probability distributions defined by the ground truth and predicted segmentations, respectively. In addition to being a suitable and popular choice for multi-class segmentation, it admits straightforward schemes for weighting the contribution of different classes (perhaps to balance their frequency) and pixels (or voxels, where particular segmentation features such as geometric interfaces might confer task-specific meaning) (Ronneberger et al., 2015), and presents favourable numerical properties for gradient-based learning.

Generalised Dice loss

A popular metric of algorithmic performance, the Dice similarity coefficient (Dice, 1945) (DSC) measures the degree of overlap between predicted and ground truth segmentations. Adapting the DSC for CNN optimisation results in the generalised Dice loss:

$$L_{\text{DSC}}(\tilde{\mathbf{Y}}, \mathbf{Y}) = 1 - \frac{2\sum_c \sum_m Y_{c,m}\tilde{Y}_{c,m}}{\sum_c \sum_m Y_{c,m} + \tilde{Y}_{c,m}} \tag{4.2.5}$$

As is frequently the case, where test set performance is assessed using the DSC, directly minimising the associated loss is an attractive prospect. Moreover, others find that L_{DSC} (and other losses based on spatial overlap), can make optimisation more robust to class imbalance, even when compared against the weighted CE strategies aforementioned (Sudre et al., 2017). However, in state of the art frameworks for automated CNN optimisation, a combination of CE and generalised Dice losses is ultimately favoured (Isensee et al., 2021).

Back propagation

Whilst the value of the resulting loss evaluates the performance of the network, it does not indicate how the model parameters should be updated to reduce the prediction error. To do so, we rely on the backpropagation algorithm, a numerical framework for computing the gradient of arbitrarily complex functions, by breaking them down into simpler components (LeCun et al., 1989). This approach presents a clear symmetry with the motivations underlying CNN computation: that through the combination of simpler components, we can build up functions of arbitrary complexity. In the setting of CNN optimisation this association extends further, the two sharing computational graphs with identical topology (not computation), but proceeding in opposite direction.

At each iteration of gradient-based training, the forward pass describes the flow of information from the input ($\mathbf{X}$), via the $D-1$ hidden layers ($\mathbf{H}^{(d)}$, see Figure 4.1), to the output ($\tilde{\mathbf{Y}}$) of the network, culminating in the loss. Following, the backpropagation algorithm initiates a backwards pass, starting from the loss, and, via the recursive application of the chain rule of differentiation, determines its partial derivatives with respect to the hidden layer activations ($\mathbf{H}^{(d)}$) and the parameters ($\boldsymbol{\theta}^{(d)}$) associated with their locally determining functions ($f^{(d)}$). Col-

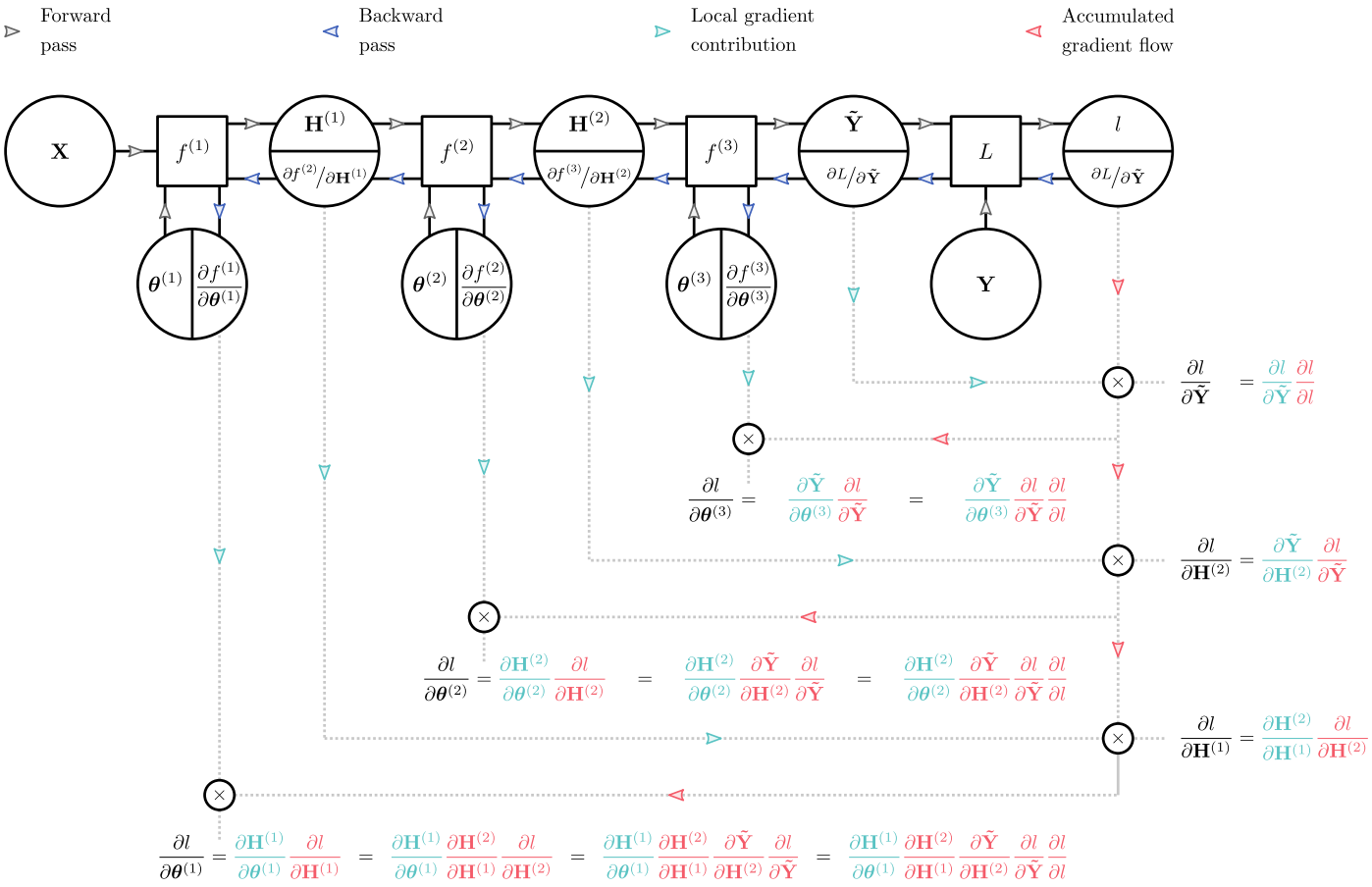


Figure 4.2: Schematic operation of the backpropagation algorithm, and its role in determining the gradient required for supervised CNN training. In the forward pass (left to right, bottom to top, grey arrows), information flows from the input ($\mathbf{X}$) to the output ($\tilde{\mathbf{Y}}$), through the parameterised transforms ($f^{(d)}(\mathbf{H}^{(d-1)}; \boldsymbol{\theta}^{(d)})$) associated with the hidden layers ($\mathbf{H}^{(d)}$) and, via the loss function (L), culminates in the loss (l). Backpropagation proceeds in the reverse direction, completing a backward pass (right to left, top to bottom, blue arrows) through a computational graph with the same connectivity. Recursively applying the chain rule of differentiation, the partial derivative with respect to each variable in the graph has two contributions: (1) the local gradient of the function transforming the variable in the forward graph (green arrows); and (2) an accumulated gradient, flowing from the upstream portions of the graph (red arrows). Even in this simple, relatively shallow neural network, by breaking a complex function down into simpler components, backpropagation simplifies the process of differentiation. The resulting partial derivatives comprise the gradient vector, $\nabla_{\boldsymbol{\theta}} = [\partial l/\partial \boldsymbol{\theta}^{(1)}, \partial l/\partial \boldsymbol{\theta}^{(2)}, \partial l/\partial \boldsymbol{\theta}^{(3)}]$, and therefore determine the local direction in parameter space that must be followed to reduce the loss and optimise the network.

Caveats:
Both the transformations ($f^{(d)}$) and the associated partial derivatives ($\partial l/\partial \boldsymbol{\theta}^{(d)}$) are more complicated than might be implied by the notation above (the latter in fact being gradient vectors themselves). Practically, both must judiciously handle >3D inputs, hidden layers and outputs, according to the convolutional and other operations applied. Whilst this perhaps could be better described using the notation of tensor calculus, our intention is to provide only an impression of the backpropagation algorithm, not its formal mathematics.

lecting the partial derivatives of the loss with respect to the parameters of the CNN constructs an associated training gradient. In the conceptual (if not mathematically rigorous) example shown in Figure 4.2, for the associated three-layer CNN, this is given by:

$$\nabla_{\boldsymbol{\theta}} l = \left[\frac{\partial l}{\partial \boldsymbol{\theta}^{(1)}}, \frac{\partial l}{\partial \boldsymbol{\theta}^{(2)}}, \frac{\partial l}{\partial \boldsymbol{\theta}^{(3)}} \right] \tag{4.2.6}$$

Via backpropagation, this vector reveals the direction through parameter space, that at least locally, most rapidly reduces the loss. We consider the formal mathematics of the algorithm to be beyond the necessary scope of this thesis, but refer the reader to Rumelhart et al. (1986, 1995) for a full account.

Optimisers

Whilst the gradient suggests the direction that should be followed to reduce the loss, it says nothing as to the magnitude of the step to be taken. Perhaps the most important hyperparameter - not just to optimisation dynamics but, once a CNN architecture has been defined (see Section 4.3) to overall performance - the learning rate (η) defines the size of each iterative update:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} l \tag{4.2.7}$$

Careful tuning of η is required just to facilitate, let alone accelerate learning. When set too high, the assumption of locality - that the gradient provides a sufficient description of the loss landscape - breaks down. In other words, for an excessive step size, the local gradient no longer suggests a trajectory to reduce the loss. The term "exploding gradients" is often used to describe the rapidly fluctuating dynamics which result and ultimately, without meaningful supervision, the explosion of the loss to infinity. On the other hand, setting η too low can preclude substantive progress.

Ultimately we seek convergence to a local minimum in the loss, or perhaps more realistically, to a location where the magnitude of the gradient is small enough to limit further gains. In the latter scenario, rather than by meeting a convergence criterion, optimisation is frequently terminated after a fixed iteration budget. This might be determined: by the quantity and structure of the training data, allow-

ing for an integer number of passes through the entire set; with reference to the expected loss over the validation set; or simply by a pragmatic assessment of the available computational resource and the time invested.

More generally than in the basic formulation presented by Equation 4.2.7, it is the job of the optimiser to shape the dynamics of loss landscape traversal. A range of optimisers and associated hyperparameters have been proposed for CNN training. Whilst all are founded on gradient descent, specialised schemes present varying levels of sophistication to adapt their operation to different training scenarios and models. In this respect, the recent popularity of CNNs has prompted the development of optimisers suited to the particular demands of deep network training, including their extremely high-dimensional feature space. These can be divided according to: (1) their stochasticity (and associated batch sampling schemes); (2) their acceleration of parameter space traversal by the accumulation of historic gradients within a velocity vector; and (3) their consideration of per parameter, adaptable learning rates. We present an example from each of these groups, choosing those which appear the most popular amongst exponents of CNN-based image segmentation, and which are deployed within the experimental work described in later chapters.

Stochastic gradient descent

In isolation, Equation 4.2.7 obscures the fact that the gradient of the loss ($\nabla_{\boldsymbol{\theta}} l$) is determined empirically by training examples. Recalling that the loss landscape describes the error expected over the training distribution, the natural approach might be to compute the gradient over all T training examples, at every iterative update. The resulting formulation describes deterministic batch optimisation:

	Deterministic gradient descent			
i)	Compute gradient	$\mathbf{g}$	$\leftarrow$	$\frac{1}{T}\sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} L(f(\mathbf{X}^{(t)}; \boldsymbol{\theta}), \mathbf{Y}^{(t)})$
ii)	Parameter update	$\boldsymbol{\theta}$	$\leftarrow$	$\boldsymbol{\theta} - \eta \mathbf{g}$

Where t indexes over the cases of the training set. This approach, however, presents challenges to the computational resources available for training, a particularly pressing obstacle for CNN-based 3D image segmentation: volumetric data being associated with a cubic increase in the memory consumed.

More often, a minibatch of $B < T$ training examples are used to estimate the gradient at each iteration. Outside of curriculum learning (Bengio et al., 2009), hard example mining (Shrivastava et al., 2016) or related schemes, typical sampling strategies seek to ensure that on average, training cases are equally represented throughout training. This can be achieved by, for example, shuffling the training set before its division into mutually independent minibatches by successive sampling without replacement. To differentiate from the deterministic case, we assume the prior use of a judicious sampling strategy, such that b indexes over each minibatch independently, and use $\tilde{\mathbf{g}}$ to indicate the resulting gradient estimate:

Stochastic gradient descent

i)	Estimate gradient	$\tilde{\mathbf{g}}$	$\leftarrow$	$\frac{1}{B}\sum_{b=1}^{B}\nabla_{\boldsymbol{\theta}} L(f(\mathbf{X}^{(b)};\boldsymbol{\theta}),\mathbf{Y}^{(b)})$
ii)	Parameter update	$\boldsymbol{\theta}$	$\leftarrow$	$\boldsymbol{\theta} - \eta\tilde{\mathbf{g}}$

Considering only a minibatch of the training set, stochastic gradient descent (SGD) accepts that parameter updates are based only on an estimate of the gradient. Goodfellow et al. (2016), however, suggest that this limitation is nullified by the following statistical arguments and outweighed by SGD's gains in computational efficiency: (1) since the standard error in this estimate scales as $1/\sqrt{B}$, improvement associated with increasing batch size is subject to diminishing returns; (2) for homogeneous training sets, minibatches avoid gradient estimation by redundant features; (3) small minibatches inject noise into learning, yielding a regularising effect that can improve generalisation.

Accordingly, different flavours of SGD and minibatch optimisation have become ubiquitous to CNN training, being preferred to their deterministic and batch counterparts. To avoid the confusing terminology used to differentiate between the two, in the remainder of this thesis we refer only to SGD, and its determination of the supervisory gradient according to batches (rather than minibatches) of training data. As per the wider literature, we call B the batch size.

(Nesterov) Momentum

Lower learning rates are sometimes required to accommodate the noisy optimisation dynamics associated with low batch sizes. Naive SGD can also limit the pace of learning in the context of small but consistent gradients. The method of

momentum (Qian, 1999) is designed to accelerate optimisation in each of these scenarios. By accumulating an exponentially decaying moving average of past updates, momentum accelerates descent in the directional components of the gradient that remain consistent across recent iterations:

SGD + Momentum

i)	Estimate gradient	$\tilde{\mathbf{g}}$	$\leftarrow$	$\frac{1}{B}\sum_{b=1}^{B}\nabla_{\boldsymbol{\theta}}L(f(\mathbf{X}^{(b)};\boldsymbol{\theta}),\mathbf{Y}^{(b)})$
ii)	Velocity update	$\mathbf{v}$	$\leftarrow$	$\mu\mathbf{v}-\eta\tilde{\mathbf{g}}$
iii)	Parameter update	$\boldsymbol{\theta}$	$\leftarrow$	$\boldsymbol{\theta}+\mathbf{v}$

Where the hyperaparameter, μ, controls the influence of the historic updates accumulated within the momentum velocity, $\mathbf{v}$ (so named for its analogy with Newtonian dynamics). More recently, Sutskever et al. (2013) adapted the momentum update to incorporate the acceleration proposed by Nesterov (2003). Their approach applies the velocity displacement to achieve an intermediate parameter set, $\boldsymbol{\theta}'$, prior to gradient estimation:

SGD + Nesterov momentum

i)	Intermediate update	$\boldsymbol{\theta}'$	$\leftarrow$	$\boldsymbol{\theta}+\mu\mathbf{v}$
ii)	Estimate gradient	$\tilde{\mathbf{g}}$	$\leftarrow$	$\frac{1}{B}\sum_{b=1}^{B}\nabla_{\boldsymbol{\theta}'}L(f(\mathbf{X}^{(b)};\boldsymbol{\theta}'),\mathbf{Y}^{(b)})$
iii)	Velocity update	$\mathbf{v}$	$\leftarrow$	$\mu\mathbf{v}-\eta\tilde{\mathbf{g}}$
iv)	Parameter update	$\boldsymbol{\theta}$	$\leftarrow$	$\boldsymbol{\theta}+\mathbf{v}$

The incremental gains made by the Nesterov formulation have led to its adoption within state of the art frameworks for CNN-based segmentation (Isensee et al., 2021).

Adam

Finally we turn our attention to optimisation strategies which deploy adaptive learning rates. Inherent within the motivations underlying momentum is the observation that traversal of particular directions within parameter space more effectively reduces the loss than others. Assuming that these directions are (or, through optimisation, can be) associated with particular elements of $\boldsymbol{\theta}$, adaptive optimisers exploit this property by assigning a learning rate to each parameter, making independent adjustments to suit their contribution to the observed optimisation dynamics.

Amongst a variety of optimisation methods that include per parameter, adaptive learning rates, for its apparent popularity we present Adam (Kingma and Ba, 2014). Through gradient reweighting, this optimiser sets out to suppress updates to those parameters presenting the largest partial derivatives of the loss. Simultaneously, its formulation allows for updates to those parameters recently associated with smaller partial derivatives to be maintained. As a result, Adam selectively descends the loss landscape, preferring directions in which the loss is reduced smoothly and predictably. Combined with Nesterov momentum, this behaviour is achieved by update rescaling, according to an accumulated history ($\mathbf{h}$) of squared gradient estimates ($\tilde{\mathbf{g}} \odot \tilde{\mathbf{g}}$):

	Adam			
i)	Intermediate update	$\boldsymbol{\theta}'$	$\leftarrow$	$\boldsymbol{\theta} + \mu\mathbf{v}$
ii)	Estimate gradient	$\tilde{\mathbf{g}}$	$\leftarrow$	$\frac{1}{B}\sum_{b=1}^{B} \nabla_{\boldsymbol{\theta}'} L(f(\mathbf{X}^{(b)}; \boldsymbol{\theta}'), \mathbf{Y}^{(b)})$
iii)	Accumulate gradient	$\mathbf{h}$	$\leftarrow$	$\rho\mathbf{h} + (1-\rho)\tilde{\mathbf{g}} \odot \tilde{\mathbf{g}}$
iv)	Velocity update	$\mathbf{v}$	$\leftarrow$	$\mu\mathbf{v} - \eta\mathbf{g} \oslash \mathbf{h}^{\circ\frac{1}{2}}$
v)	Parameter update	$\boldsymbol{\theta}$	$\leftarrow$	$\boldsymbol{\theta} + \mathbf{v}$

Where $\odot$, $\oslash$ and $\circ\frac{1}{2}$ denote the element-wise product, division and square root, respectively. Note the inclusion of a further hyperparameter, ρ, controlling the exponentially decaying contributions of previous gradients to the accumulated history, and biasing rescaling to the dynamics of the most recent parameter updates. Despite its hyperparameterisation (η, μ and ρ), Kingma and Ba (2014) promote Adam as being robust to a variety of CNN optimisation tasks, attractively advising against associated hyperparameter tuning in favour of default values.

4.2.3 Regularisation

In this section we present various techniques to regularise CNN training. Whilst all are motivated by a desire to improve the generalisability of CNN prediction to unseen cases, each is expressed through the different aspects of network optimisation previously introduced. These include methods implemented by adapting loss functions, optimisation strategies and training data, and by enhancing experimental design and task specification more generally.

At the start of Section 4.2.2, we highlighted the balance that must be struck between expressive model capacity and regularisation, if overfitting is to be avoided. Accordingly, in their investigation of the relative strengths of different approaches, Hernández-García and König (2018) further separate regularisation techniques into those placing an explicit constraint on expressive capacity versus those having only an implicit effect on the possible parameter search space. Whilst in preferring the latter approach, our experience supports their findings, for completeness we firstly present the most popular methods of explicit regularisation, followed by their implicit counterparts.

Explicit regularisation

Parameter norm penalties

Historically, constraining expressive capacity through parameter norm penalties, was a popular means of regularising machine learning. In the context of deep learning and CNNs, weight decay has been widely incorporated within associated loss functions:

$$L_{\text{WD}}(\tilde{\mathbf{Y}}, \mathbf{Y}, \boldsymbol{\theta}) = L(\tilde{\mathbf{Y}}, \mathbf{Y}) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2 \tag{4.2.8}$$

Where the hyperparameter λ controls the influence of the weight decay penalty compared with the task loss $L(\tilde{\mathbf{Y}}, \mathbf{Y})$. By effectively shrinking the weights at each gradient update (Goodfellow et al., 2016), weight decay introduces a preference for small valued parameters, constraining the expressive capacity of the CNN to regions of parameter space in the vicinity of the origin.

Dropout

Developed specifically for the training of artificial neural networks, dropout (Srivastava et al., 2014) provides an alternative form of explicit regularisation. Rather than via the loss function, dropout reduces the expressive capacity of a CNN directly, stochastically dropping channels[1] from hidden feature maps (see Section 4.2.4) such that they make no contribution to the forward or backward pass. The rate at which each is dropped is controlled by a predetermined hyperpa-

[1] Coupled with the weight sharing inherent to CNNs, the correlation of adjacent pixels of hidden feature maps renders conventional dropout (Srivastava et al., 2014), ineffective. Accordingly, spatial dropout is applied to entire channels of a hidden feature map (Tompson et al., 2015).

rameter. At each iteration of gradient-based training, dropout trains a distinct, sub-network of reduced expressive capacity. Consequently, an entire training budget can be viewed as sampling from the collection of possible sub-networks, the resulting parameter set often being presented as an approximation to the associated model ensemble. Accordingly, this interpretation of dropout's effects claims the statistical gains made through bootstrapped sampling (via minibatch methods) and aggregated prediction. An alternative rationale highlights that in the presence of a constantly shifting computational graph, dropout prevents the co-adaptation of dependent parameters, and encourages learning of independent and individually discriminative features, a property claimed as beneficial to generalisation (Hinton et al., 2012).

Implicit regularisation

Ensemble prediction

Indirectly, our explanation of dropout hinted at an implicit approach to regularisation and improved generalisability. Ensemble methods leverage the statistical gains that are achieved by aggregating the predictions made by multiple models, through various schemes such as majority voting or averaging. Key to their capacity to improve generalisation is the assumption that the errors made by each member of an ensemble are uncorrelated with each other. Where this is fulfilled, those errors made by an individual model (possibly due to it overfitting a specific feature of the training data) are not shared by the remainder (being possibly overfit to different features of the training data), such that aggregation can effectively improve the generalisation accuracy of the ensemble.

By varying the experimental setting in which each CNN is trained, the correlation between individual predictions is reduced. In some cases, the stochasticity inherent to practical CNN training (randomised minibatch sampling or parameter initialisation, for example) may be sufficient for multiple trained models to converge to different regions of parameter space. More determined schemes seek the same by optimising each network against a different subset of the training data, possibly through bagging (see random forest in Section 3.3.1) or cross-validation. In a k-fold configuration, the latter divides the training data into k evenly sized

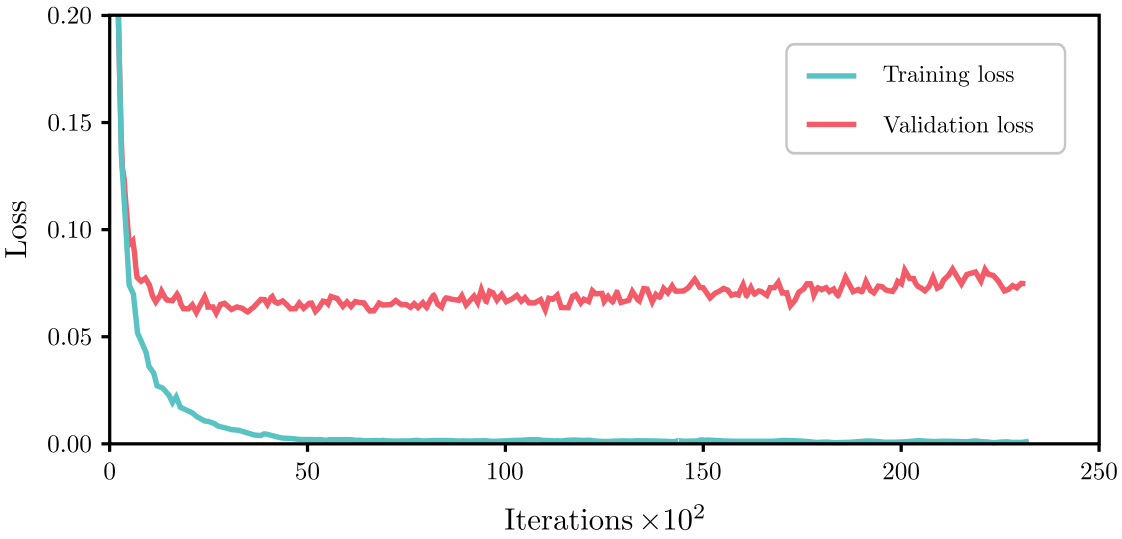


Figure 4.3: By monitoring the loss over the validation set (data which do not contribute to gradient-based parameter updates), training can be implicitly regularised through early stopping. In this example of typical loss evolution, overfitting is characterised by improvement (reduction) in the training loss that is coincident with increasing validation loss. Early stopping halts training at the point that worsening validation set performance (determined by an associated heuristic) is observed, and commonly chooses the CNN parameter set that minimises the empirical validation loss series. Adapted from Goodfellow et al. (2016).

groups (without replacement), each being respectively held out from the training of an ensemble of k networks. A final alternative is for each model to adopt a different CNN architecture (see Section 4.3). A combination of approaches is leveraged by the current state of the art for automated CNN-based segmentation: Isensee et al. (2021) amass a doubly divergent ensemble, including a range of 2D and 3D architectures, each trained using five-fold cross-validation.

Early stopping

Provided an applicable architecture, a model of sufficient expressive capacity and an appropriate hyperparameterisation, gradient-based learning reliably reduces the loss expected over the training set (the training loss). Empirical observation, however, suggests a different trajectory for the loss expected over the validation data (the validation loss). Recalling that validation examples make no direct contribution to parameter optimisation, they instead provide a means of monitoring

performance on unseen data (and where representative, generalisation to the underlying population, including the test set). Hence, as depicted in Figure 4.3, and as typically observed, after initial improvement, sustained increases in the validation loss are indicative of overfitting. They suggest that progressively, the network learns features of the training data that successfully reduce the training loss, but which do not generalise to the unseen validation examples.

In response, a straightforward solution is to terminate training prior to the observed increase in the validation loss. Practically, early stopping amounts to making periodic evaluations against the validation set, maintaining a copy of the parameters associated with the best performing model. This comprises a set of features which balance optimal performance on both training and validation data. In this idealised presentation, early stopping provides a simple and effective approach to regularisation, indirectly defining an iteration budget to preclude overfitting.

In the context of limited data, however, where the validation set itself might lack the diversity to represent the underlying patient distribution, early stopping presents a risk that the model is indirectly overfit to the validation cases. An alternative approach, which we favour within the experimental work presented in later chapters, is to use cross-validation. Rather than maintain the best performing parameter set, we use validation performance as a means of identifying an iteration budget (most often an integer number of passes through the training data) that acceptably averts overfitting across all folds of the data. In this sense, we treat the training duration as any other hyperparameter to be optimised, applying the underlying principles of early stopping to its determination.

Data augmentation

If overfitting results when learned features, rather than generalising to the unseen test set, are particular to the available training data, the most natural remedy is to curate a training set that is more representative of the underlying patient population. This is achieved by increasing the quality, diversity and quantity of training data. In many medical applications of CNN-based image segmentation

(including our own, see the results of Chapter 5), however, obtaining accompanying ground truth annotations is an expensive and time-consuming undertaking[2].

Accordingly, an attractive alternative is to increase the diversity of training data by constructing synthetic examples. Observing that ground truth segmentation is desirably equivariant to a range of transformations of the input image, data augmentation generates new cases from the manually labelled training set. For example, we anticipate that orthogonal rotation of the input should result in an equivalent transformation of its ground truth labelling. Such equivariance can be learned by simultaneous transformation of image-label pairs to form an augmented dataset.

This approach is most effective, however, when the selection of transforms is informed by domain knowledge, such that augmented representations express modes of variation exhibited by the underlying population. Modifying our prior example, since patients undergoing computed tomography (CT) or cardiac magnetic resonance (CMR) are typically scanned in a predictable position (for example, head-first-supine), rotation through small angles better captures the clinically encountered distribution of patient orientation. Other transforms that might generalise to the underlying CHD population include: translation, spatial scaling, lateral inversion (where the normal assymetry of the thoracic organs is affected by isomerism or dextrocardia) and non-rigid deformation. Augmentation need not be limited to modes of geometric variation, however, and can also be used to represent the underlying distribution of clinical image quality: such as by the introduction of random noise or by the synthesis of artefacts (Pérez-García et al., 2021).

Being only perturbed representations of the training data, it might be argued that augmentation primarily increases the diversity, rather than the quantity of training examples. In response, some have looked to generative models, sampling augmented representations from a learned distribution of clinical data (Chaitanya et al., 2019). Notwithstanding these considerations, and even in its most basic formulation, data augmentation has become a popular and successful means of regularising CNN-based segmentation.

[2]The extent of this challenge is such that some have leveraged additional unlabelled data through semi- (Bai et al., 2017) or self-supervised learning (Bai et al., 2019) (the latter also benefiting from the regularising effect of multi-task learning).

Comparison of approaches

Our experimental work relies primarily on implicit sources of regularisation, and data augmentation in particular. As per the arguments presented by Hernández-García and König (2018), we assert that it is the incorporation of domain knowledge that makes this approach an effective means of promoting the learning of generalisable features. Informally, we draw a distinction between data augmentation and the explicit methods presented: the latter limits expressive capacity by constraining parameter space; whereas, the former maintains expressive capacity, biasing convergence to a generalisable parameter set that might otherwise be precluded by explicit regularisation. This is achieved by leveraging domain knowledge to introduce features that whilst relevant to the underlying task, might be absent or poorly captured by the finite training data.

Theoretical work supports such claims. In certain experimental settings, both weight decay (Bishop, 1995) and dropout (Bouthillier et al., 2015), have been shown to be in some sense equivalent to training with examples augmented by the inclusion of random noise. It seems natural therefore, that a more sophisticated augmentation scheme, including transformations or perturbations that, with respect to the underlying patient distribution, plausibly enrich the features expressed by training data, would lead to improved generalisation. Irrespective of the perhaps disputed theoretical arguments, our experience suggests data augmentation as the most effective means of regularising training for CNN-based segmentation of medical images, particularly in the context of few training examples.

4.2.4 Building blocks

Equipped with an abstract understanding of deep networks and their training, we now make concrete the mathematical building blocks composing state of the art CNN architectures for image segmentation. Developed within a considerable body of scientific research, the following variably constitute the transformations $f^{(d)}$ learnt at each layer of the network, and as depicted in Figure 4.1.

$$\begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} \qquad \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \qquad \begin{bmatrix} \Psi_{1,1} & \Psi_{1,2} & \Psi_{1,3} \\ \Psi_{2,1} & \Psi_{2,2} & \Psi_{2,3} \\ \Psi_{3,1} & \Psi_{3,2} & \Psi_{3,3} \end{bmatrix}$$

(a) $\boldsymbol{\Upsilon}_x$ (b) $\boldsymbol{\Upsilon}_y$ (c) $\boldsymbol{\Psi}$

Figure 4.4: The horizontal (a) and vertical (b) Sobel operators have been hand-crafted to approximate the image gradient via convolution and expose directed edges within image data. Comparatively, the kernels deployed within CNN architectures (c) present a flexible basis for optimal and task-specific feature extraction and consolidation, achieved through data-driven parameter tuning.

Convolutional filters

In Chapter 3, the use of digital convolutional filters to expose discriminative image features was qualitatively introduced. More quantitatively, the value of each pixel in a derived output feature map is determined by the dot product of its corresponding, local image patch with the so-called kernel, a spatially varying but restricted set of weights. Provided a 2D input image, $\mathbf{X}$, whose elements are spatially indexed by i and j; and a 2D kernel, $\boldsymbol{\Psi}$, whose elements (or weights) are spatially indexed by u and v, the pixel-wise output of 2D convolution[3] can be written as:

$$G_{i,j} = \text{conv}_{\text{2D}}\left(\boldsymbol{\Psi}, \mathbf{X}\right)_{i,j} = \sum_{u,v} X_{i+u-1,j+v-1}\Psi_{u,v} \tag{4.2.9}$$

Since this sum is indexed over u and v, the resulting map is sensitive to features contained by the spatial extent of the convolutional kernel, which we refer to as the *receptive field* of each element of the output feature map, $G_{i,j}$. Conventionally, kernels have been hand engineered to promote the extraction of a theoretically founded feature set. Consider the role played by the Sobel operators within edge detection (Kanopoulos et al., 1988), for example. In the 2D case, these kernels are demonstrated in Figure 4.4a and Figure 4.4b. The resulting map of edge responses are subsequently used to inform downstream processing, including segmentation.

[3]Strictly, without kernel flipping, Equation 4.2.9 describes the cross-correlation operation. However, for the majority of CNN purposes this distinction is unimportant, cross-correlation even being favoured within many machine learning libraries (Goodfellow et al., 2016).

Primarily, the convolutional operations employed within CNN-based image segmentation serve the same purpose: to extract the salient features of image data that might allow the various semantic classes contained to be discriminated. Unlike conventional convolutional filters, however, those deployed within CNNs provide a flexible, parametric basis for feature extraction, see Figure 4.4c. Secondarily, and in the context of image segmentation, convolution provides a means of synthesising learned features in aid of eventual probabilistic classification. It is then the ambition of CNN training to find the optimal set of kernel weights that collectively expose the task-specific, discriminative features required to minimise the error between the predicted and ground truth segmentations: $\tilde{\mathbf{Y}}$ and $\mathbf{Y}$, respectively. Critical to this aim is an appropriate initialisation of convolutional weights. Modern approaches favour the scheme proposed by He et al. (2015).

Alone, an individual convolutional filter lacks the expressive capacity to model the complex pattern of visual features relevant to the identification of anatomy. Instead, the representational power of CNNs is dramatically increased by a number of modifications. As mentioned previously, perhaps the most significant of these is achieved by the successive operation of convolutional filters within the layers of a deep network. However, two further extensions require explanation. The first considers the extraction and interaction of multi-channel image features. Rather than a single map, it is common practice for each layer to expose a multiplicity of features. Frequently thought of as defining the *width* of the network, multi-channel features expand the capacity of the CNN to model the particular appearances that are most relevant to the data and task at hand. Extending our definition of the 2D convolutional operator to also consider the number of input and output feature channels (indexed by a and b, respectively), we write:

$$G_{b,i,j} = \text{conv}_{\text{2D}}\left(\mathbf{\Psi}, \mathbf{X}\right)_{b,i,j} = \sum_{a,u,v} X_{a,i+u-1,j+v-1}\Psi_{a,b,u,v} \tag{4.2.10}$$

Notice that the 2D convolutional kernel presented in Equation 4.2.9 has expanded to 4D in Equation 4.2.10, the additional dimensions describing the interaction of input and output channels. Despite the increased dimensionality of $\mathbf{\Psi}$, we maintain the 2D descriptor (and subscript) in order to make clear the spatial dimensionality shared by the layer, its mediating kernel and associated feature maps.

Activation functions

In the form of activation functions, the second means by which statistical capacity is extended is also the source of non-linearity within CNN-based segmentation. Thus far, our only network building block is multi-channel convolution. As defined (at least in 2D) by Equation 4.2.10, any network constructed from the sequential application of this operator alone would be limited to linear transformation, a severe constraint on model expression. Instead, non-linearity is granted by applying an activation function to the output of multi-channel convolution. Analogously to the neuronal firing that mediates the human visual system (the structure of which is often cited as inspiring the design of CNNs (LeCun et al., 2015)) this function establishes a threshold which, once surpassed, determines the activation state of learned features.

As established by Nair and Hinton (2010), the rectified linear unit (ReLU) has become the standardised approach within modern CNNs: $\text{ReLU}(x) = \max(0, x)$. Prior to non-linear activation, a learned bias shifts the ReLU activation point. These parameters are typically learned per output channel of the convolutional transformation and notationally are collected into a vector, $\boldsymbol{\delta}$.

Hidden layers

The combination of the convolutional kernel ($\boldsymbol{\Psi}$), bias vector ($\boldsymbol{\delta}$) and ReLU activation provide the basis of non-linear feature extraction and transformation. Their application generates an intermediate feature map between the input image ($\mathbf{X}$) and the output predicted segmentation ($\tilde{\mathbf{Y}}$). With reference to Figure 4.1, and in a 2D task, the pixels of the first *hidden* layer are generated as:

$$H^{(1)}_{b,i,j} = \text{ReLU}(\text{conv}_{\text{2D}}\left(\boldsymbol{\Psi}^{(1)}, \mathbf{X}\right)_{b,i,j} + \delta^{(1)}_b) \tag{4.2.11}$$

Since the hidden layers are determined in a cascade of operations within the network, their more general form is given by:

$$\begin{aligned} H^{(d)}_{b,i,j} &= f_d(\mathbf{H}^{(d-1)}; \boldsymbol{\theta}^{(d)})_{b,i,j} \\ &= \text{ReLU}(\text{conv}_{\text{2D}}\left(\boldsymbol{\Psi}^{(d)}, \mathbf{H}^{(d-1)}\right)_{b,i,j} + \delta^{(d)}_b) \end{aligned} \tag{4.2.12}$$

Where d is a descriptive index of the depth at which the hidden layer is determined, with $\mathbf{H}^{(0)} = \mathbf{X}$ and $\mathbf{H}^{(D)} = \tilde{\mathbf{Y}}$ representing the special cases of the input and output[4], respectively (refer back to Figure 4.1); and the set of learnable parameters associated with the hidden transformation $f^{(d)}$, is $\boldsymbol{\theta}^{(d)} = \{\boldsymbol{\Psi}^{(d)}, \boldsymbol{\delta}^{(d)}\}$. Notationally, and unless illustrative of a wider point of theory, for clarity we omit the index d when referring to a hidden layer at arbitrary depth.

Hidden transformations of this form make up the bulk of a CNN's capacity to extract and synthesise discriminative features. However, to increase their efficiency and performance, dedicated network architectures have been developed for image segmentation, including a range of extended feature operations. Hence, whilst the form implied by Equation 4.2.12 is useful to demonstrate the theoretical mechanisms underlying CNN performance, it does not provide a generalised and exhaustive description of feed-forward models, as they are practically deployed. Real world architectures include other operations that might comprise the intermediate feature transformations, $f^{(d)}$. We now introduce those that are most relevant to modern CNN-based image segmentation.

Pooling layers

Provided suitable padding (the options being thoroughly discussed by Dumoulin and Visin (2016)), the hidden layers of the CNN maintain the spatial extent of the input image and its intermediate feature maps. In contrast, by summarising hidden features within spatially limited local regions, pooling layers serve to reduce their size (see Figure 4.5). Most commonly based on mean or maximal statistics, their design ensures that irrespective of the local feature configuration, each region is represented by the most discriminative feature contained.

Window regions with sides of an integral number of pixels (w) are commonly used to provide a sliding window view (as per the convolution operator) over the input, intermediate feature map. To reduce the spatial size of the feature map, $\mathbf{H}$, however, windows must be sampled without maximal (and normally without any) spatial overlap. The spacing between sampled regions is referred to as the stride (s) of the pooling layer. When the stride is equal to the window side, there

[4]Note that in the majority of CNN architectures, a specialised output activation is applied in the place of the ReLU (see later in this section).

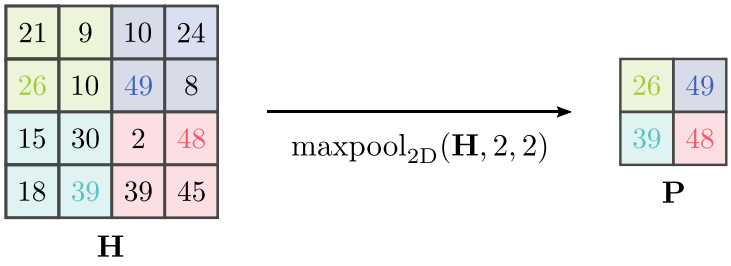


Figure 4.5: By spatial summary, max pooling is used to spatially abstract the features of an input feature map, $\mathbf{H}$. In this example, elements of the pooled feature map, $\mathbf{P}$, are determined within square regions of width and stride of two, $w = s = 2$ (shaded in contrasting colours). Note the resulting reduction in the spatial extent of learned features.

is no overlap between sampled regions, and the size of the output feature map is reduced by a factor of s. In the case of so-called max pooling, elements of the pooled feature map, $\mathbf{P}$, are given by:

$$P_{b,i,j} = \text{maxpool}_{\text{2D}}\left(\mathbf{H}, w, s\right)_{b,i,j} = \max_{u,v \in \{1,\ldots,w\}} H_{b,s(i-1)+u,s(j-1)+v} \tag{4.2.13}$$

This formulation implies: that the channels (indexed by b) of the feature map to be summarised are pooled independently; that each pooling region is square; and that the stride is isotropic in the spatial dimensions indexed by i and j. Whilst these reflect the most common design choices for the inclusion of a max pooling layer, each can be adapted to the problem at hand. Lastly, note that by selecting an alternative summary operation (say the mean rather than the max) Equation 4.2.13 can be modified to achieve different flavours of feature pooling.

This approach grants a network response that is invariant to small translations of the input image. Given that state of the art architectures for CNN-based segmentation have their roots within image-level classification (predicting a single label per image rather than per pixel), such invariance to small translations of the input was an attractive property. Being concerned with the semantic content of the data in totality, these and related tasks are less troubled by the localisation of relevant features and structures to a particular portion of the image. Whether the same motives justify the use of pooling layers within CNN-based segmentation is questionable, their primary motivation appearing at odds with our task objectives.

Nevertheless, and as demonstrated by their incorporation within ubiquitous, state of the art architectures, pooling layers present other advantages. By reducing the spatial extent of hidden feature maps, the network is made more computationally efficient. These gains can be leveraged to increase the representational capacity of the CNN, rapidly expanding its width and depth, albeit at low spatial resolution. In Section 4.3, we shall see that the competition for computational resource has driven the development of specialised segmentation CNNs and how their use of pooled features can be reconciled with the requirements of pixel-wise labelling.

Strided convolution

Strided convolution provides an alternative means of reducing the spatial extent of learned features. This approach leverages the expressive capacity of the convolutional operator to learn a spatial synthesis of the most discriminative features, whilst simultaneously downsampling the input feature map (Springenberg et al., 2014). This is in contrast with more restrictive pooling operations based on the statistical properties of the local feature distribution (such as the maximum or mean). As per these operations, when the size of the kernel equals the stride, the spatial extent of the input feature map is reduced by a factor of s:

$$G_{b,i,j} = \text{conv}_{\text{2D}}\left(\boldsymbol{\Psi}, \mathbf{H}, s\right)_{b,i,j} = \sum_{a,u,v} H_{a,s(i-1)+u,s(j-1)+v}\Psi_{a,b,u,v} \tag{4.2.14}$$

Transposed convolution

Having introduced transformations capable of reducing the spatial extent of learned feature maps, in the context of image segmentation, it is natural to anticipate the need for operations that can reverse this downsampling, and "upsample" the hidden activations to their original resolution. This could be straightforwardly achieved by conventional image resampling, including nearest neighbour or linear interpolation. Historically, others have stored the spatial locations associated with max pooled features in an attempt to approximate a reverse "unpooling" operation (Zeiler et al., 2011). A modern approach, however, is achieved by transposed convolution.

To better understand transposed convolution, and its role in restoring the spatial extent of learned feature maps to the resolution of their associated source images, it is informative to express convolutional transformation using matrix multiplication. Consider the 2D convolution of a particular input image or learned feature map, represented by the matrix $\mathbf{H}' \in \mathbb{R}^{4\times4}$, with a particular convolutional kernel, $\mathbf{\Psi}' \in \mathbb{R}^{3\times3}$ (whose elements are Ψ'_{ij}, see Figure 4.4):

$$\mathbf{G}' = \text{conv}_{\text{2D}}\left(\mathbf{\Psi}', \mathbf{H}'\right) \tag{4.2.15}$$

In the absence of any padding of the input image, $\mathbf{H}'$, there are four valid placements of $\mathbf{\Psi}'$, two per spatial dimension. The result, $\mathbf{G}'$ is therefore a 2×2 matrix. In other words, convolution has reduced the spatial extent of $\mathbf{H}'$.

Rather than by Equation 4.2.9, we can formulate this convolutional transformation using matrix multiplication. By spatially unrolling (left to right along rows and then top to bottom down columns) $\mathbf{H}'$ and $\mathbf{G}'$ into the column vectors $\mathbf{h} \in \mathbb{R}^{16}$ and $\mathbf{g} \in \mathbb{R}^4$, we can rewrite Equation 4.2.15 as a matrix-vector multiplication:

$$\mathbf{g} = \mathbf{\Xi}\mathbf{h} \tag{4.2.16}$$

Where the sparse matrix $\mathbf{\Xi}$ is:

$$\begin{bmatrix}
\Psi'_{1,1} & \Psi'_{1,2} & \Psi'_{1,3} & 0 & \Psi'_{2,1} & \Psi'_{2,2} & \Psi'_{2,3} & 0 & \Psi'_{3,1} & \Psi'_{3,2} & \Psi'_{3,3} & 0 & 0 & 0 & 0 & 0 \\
0 & \Psi'_{1,1} & \Psi'_{1,2} & \Psi'_{1,3} & 0 & \Psi'_{2,1} & \Psi'_{2,2} & \Psi'_{2,3} & 0 & \Psi'_{3,1} & \Psi'_{3,2} & \Psi'_{3,3} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \Psi'_{1,1} & \Psi'_{1,2} & \Psi'_{1,3} & 0 & \Psi'_{2,1} & \Psi'_{2,2} & \Psi'_{2,3} & 0 & \Psi'_{3,1} & \Psi'_{3,2} & \Psi'_{3,3} & 0 \\
0 & 0 & 0 & 0 & 0 & \Psi'_{1,1} & \Psi'_{1,2} & \Psi'_{1,3} & 0 & \Psi'_{2,1} & \Psi'_{2,2} & \Psi'_{2,3} & 0 & \Psi'_{3,1} & \Psi'_{3,2} & \Psi'_{3,3}
\end{bmatrix},$$

and represents the convolutional transformation, each row accounting for a valid location of the convolutional kernel.

This formulation prompts the realisation that we can recover a vector of equal length to that of the input, $\mathbf{h}$, from the output, $\mathbf{g}$, by multiplication with the transpose of the matrix representing the initial convolution:

$$\mathbf{\Xi}^T\mathbf{g} = \mathbf{r} \tag{4.2.17}$$

Note that since we consider the transpose $\mathbf{\Xi}^T$, and not the inverse $\mathbf{\Xi}^{-1}$ (where that might exist), in all practical cases the recovered vector, $\mathbf{r} \neq \mathbf{h}$. Rather, it

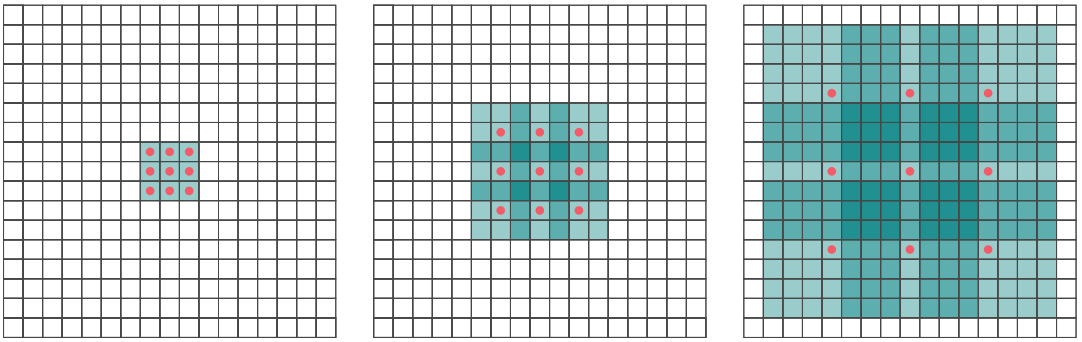
is only relevant the two have the same dimensionality. Furthermore, in addition to sharing the same length, the relationship between forward and reverse convolutional matrices, one being the transpose of the other ($\boldsymbol{\Xi}$ and $\boldsymbol{\Xi}^T$, respectively), guarantees that the "elemental connectivity" of these operations is mirrored. By this we mean that the elements of $\mathbf{h}$ which determine a particular element g_i, share indices with those elements of $\mathbf{r}$ which are derived from g_i. In either case the vectors $\mathbf{g}$ and $\mathbf{r}$ must be reshaped to assume the shapes of their 2D spatially structured counterparts, $\mathbf{G}' \in \mathbb{R}^{2\times2}$ and $\mathbf{R}' \in \mathbb{R}^{4\times4}$.

Hence, the transposed convolutional operator can be said to reverse the change in feature space dimensionality, prescribed by an associated convolutional transformation. In our example, and only for illustrative purposes, the two share the same set of kernel parameters, Ψ'_{ij}. In practice, both convolution and transpose convolution are deployed to downsample and upsample feature maps, independently and as required: there is no sense in which the counterparts need be paired. Critically, each new convolutional layer, irrespective of whether it might involve transposed operation, presents its own set of learnable parameters. As a result, transposed convolution presents a more expressive means of upsampling hidden features, than those based on conventional image resampling or unpooling.

Whilst we have developed this theoretical basis in the context of convolution with unit stride, the same approach can be applied to strided outputs, accelerating the rate of spatial upsampling. Finally, we acknowledge that there are different approaches to thinking about transposed convolution, including through intricate input padding schemes (Zeiler et al., 2011), or esoteric arithmetic. Rather than within an explicit and complex closed-form expression, however, we favour the implicit presentation through matrix multiplication: at the very least it best explains why we refer to this operation as *transposed* convolution.

Dilated convolution

As will be explained in Section 4.3, pooling, and strided and transposed convolutional layers have all been incorporated within CNN architectures as a means of rapidly increasing and controlling the receptive field of learned features. An alternative approach is provided by dilated convolution (Yu and Koltun, 2015). Rather



(a) $\text{conv}_{2\text{D}}(\boldsymbol{\Psi}, \mathbf{H}, r = 1)$ (b) $\text{conv}_{2\text{D}}(\boldsymbol{\Psi}, \mathbf{H}, r = 2)$ (c) $\text{conv}_{2\text{D}}(\boldsymbol{\Psi}, \mathbf{H}, r = 4)$

Figure 4.6: By spatially distributing the associated kernel (shown by red markers composing the elements of $\boldsymbol{\Psi}$) using a rate, r, dilated convolution increases the receptive field. When applied sequentially (as in this example: (a) $\rightarrow$ (b) $\rightarrow$ (c)), the resultant receptive field of the network (shown shaded in green for the central pixel of each feature map, $\mathbf{H}$) can be grown exponentially. Note that since the dilated elements of each kernel realise a sparse sampling on the input, pixels within the resultant receptive field are not equally represented: those more frequently sampled are indicated by increasing the saturation of their shading. Adapted from Yu and Koltun (2015).

than via a dense rectangular kernel, this operation relies on a spatially distributed sampling of input pixels (see Figure 4.6), per output element of the derived feature map. The dilation rate, r, controls the expansion of sampled locations according to:

$$G_{b,i,j} = \text{conv}_{2\text{D}}(\boldsymbol{\Psi}, \mathbf{H}, r)_{b,i,j} = \sum_{a,u,v} H_{a,i+r(u-1),j+r(v-1)} \Psi_{a,b,u,v} \tag{4.2.18}$$

This sampling of input or hidden layers expands the receptive field of the associated convolution, allowing spatially extended features to be learnt. Whilst the same can be nominally achieved by the action of convolutional layers after pooling, it should be noted that features extracted by dilated convolution are determined at the full resolution of its input. The sparse sampling of the receptive field (and the more complex network connectivity patterns established by repeated dilated convolution, see Figure 4.6b and Figure 4.6c), however, means that not all spatial locations make equal contribution to learned features.

Generalised convolution

Motivated by a desire to handle and manipulate the spatial extent of learned features, thus far we have presented two modifications of the basic, 2D multi-channel convolutional operator, including strided and dilated variants. We can write a function defining a generalised convolutional operator that, by considering different values of the stride (s) and dilation rate (r), can express each of these modes:

$$\begin{aligned} G_{b,i,j} &= \text{conv}_{\text{2D}}\left(\boldsymbol{\Psi}, \mathbf{H}, s, r\right)_{b,i,j} \\ &= \sum_{a,u,v} H_{a,s(i-1)+r(u-1)+1,s(j-1)+r(v-1)+1} \Psi_{a,b,u,v} \end{aligned} \tag{4.2.19}$$

Note that when $s = r = 1$ we recover the basic multi-channel convolution defined by Equation 4.2.10. To avoid excessively cluttered index notation this formalism has been developed in the context of 2D medical image segmentation. However, and as relevant to our work on 3D image segmentation, the same approach can be applied to 3D inputs, the equivalent generalised result being given by:

$$\begin{aligned} G_{b,i,j,k} &= \text{conv}_{\text{3D}}\left(\boldsymbol{\Psi}, \mathbf{H}, s, r\right)_{b,i,j,k} \\ &= \sum_{a,u,v,w} H_{a,s(i-1)+r(u-1)+1,s(j-1)+r(v-1)+1,s(k-1)+r(w-1)+1} \Psi_{a,b,u,v,w} \end{aligned} \tag{4.2.20}$$

Output layers

After feature extraction and transformation via the application of successive convolutional and pooling operators, the forward pass results in an array of unnormalised scores or logits that in some way, characterise CNN prediction. In the context of classification, a meaningful output is achieved by the selection of an appropriate normalisation scheme, converting the raw outputs into an indication of probabilistic class membership. So-called output layers include specialised activation functions to squash the dynamically ranged CNN output into the interval $[0, 1]$. They are applied to the logits returned by the final convolutional (or classification) layer, in a spatially pixel-wise fashion. Selection of an appropriate output

activation depends on the formulation of the classification task at hand, and in particular, the number of segmentation labels considered.

Those using a CNN to model a binary segmentation task frequently use the logistic sigmoid to indicate the predicted probability that a voxel is associated with the positive class:

$$\sigma(x) = \frac{\exp(x)}{\exp(x) + 1} \tag{4.2.21}$$

Where x is a scalar input. Note that in such tasks, the number of channels in the CNN output is normally set as one, indicating the score or, once normalised, the probability of membership within the positive class. In this scheme, the probability of the negative class is left implicit, only being inferred as $1 - \sigma(x)$.

In multi-class segmentation, and where $\mathbf{x}$ is a per pixel vector of raw network scores over the C class labels, the softmax output provides a differentiable analogue to the max operator:

$$\text{softmax}(\mathbf{x})_b = \frac{\exp(x_b)}{\sum_{c=1}^{C} \exp(x_c)} \tag{4.2.22}$$

The softmax generalises the sigmoid function, allowing the CNN output to approximate a valid probability distribution over C segmentation classes. Its form enforces competition between the semantic classes of the considered segmentation task, via the relative magnitude of the raw scores contained in $\mathbf{x}$. Where any element (x_b) of the raw output increases, the probability of the union of the remaining classes $(\sum_{c \neq b} \text{softmax}(\mathbf{x})_c)$ must fall. This property aligns with the ambition of a multi-class (rather than a multi-label) task specification, in which each voxel is associated with a single label, without partial or hierarchical membership of multiple classes.

4.3 Network architectures

Section 4.2.4 introduced the various building blocks from which CNNs are constructed. A network's architecture describes the selection, combination and connectivity of these components to suit the motivations and technical requirements of the task at hand. Amongst the multitude of design choices governing CNN construction and parameter optimisation, network architecture is possibly the most

Table 4.1: A history of the developmental contributions (as most relevant to the design of modern segmentation CNN architectures) made by key network architectures for image classification. Illustrative of the rapid progress that has been made in this field in just the last decade, according to Google Scholar these works have been collectively referenced over 350,000 times.

Year	Architecture	Authors	Contribution
2012	AlexNet	Krizhevsky et al.	Deeper networks can be trained and exceed the state of the art by a large margin.
2014	VGG	Simonyan and Zisserman	Deeper networks comprising convolution with small kernels is superior to shallower networks of large kernels.
2015	GoogLeNet	Szegedy et al.	Spatially multi-scale processing improves the accuracy of performance.
2016	ResNet	He et al.	Deeper networks can be trained more easily by learning residual functions.
2017	DenseNet	Huang et al.	Dense connectivity between layers promotes efficient feature learning.

critical. At minimum, it must accommodate the respective structures of the input data and anticipated outputs or predictions. However, and as alluded to previously, judicious architectural design also admits a means of manipulating the spatial dimensionality of learned features, expanding the receptive field and boosting computational efficiency: desirable attributes in the context of 3D image segmentation.

Today's state of the art architectures for CNN-based segmentation are the culmination of the rapid scientific progress that has been observed over the last decade. To understand the key features of modern segmentation CNNs, we firstly present a historical review of the most pivotal contributions made in the published literature. These, and their associated architectures, are summarised in Table 4.1.

4.3.1 Classification

Modern CNN-based segmentation architectures owe their foundation to methodological developments motivated by the task of image classification. Rather than a dense array of labels per pixel or voxel, such applications seek to assign a single class that summarises the content of an entire image. In this context, Krizhevsky et al. (2017) presented AlexNet, a seminal piece of work in the image processing

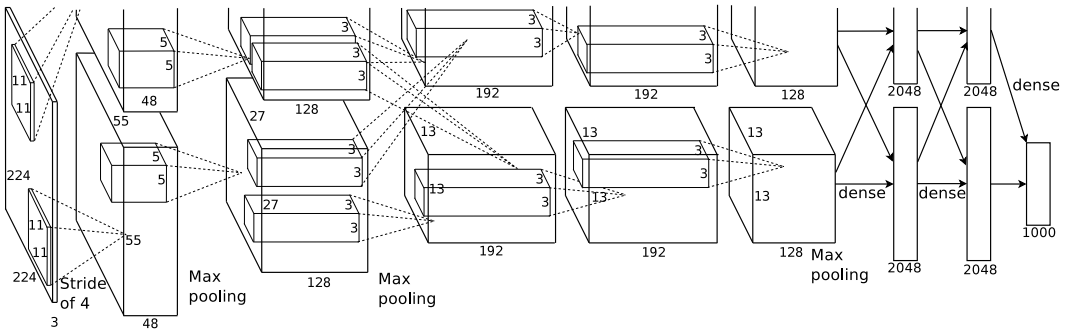


Figure 4.7: The AlexNet architecture: implemented via two identical processing streams (upper and lower), each assigned to a separate graphics processing unit (GPU), this CNN learns a spatially abstracted representation of image content in aid of global classification.

field. Their architecture included the basic building blocks of today's state of the art segmentation CNNs, including five convolutional, and three pooling layers.

Through the successive application of these operations, AlexNet (in common with other architectures for CNN-based classification) bridges the gap between the spatially structured input, and the non-spatial, semantic class prediction. This is demonstrated in Figure 4.7, in which each multi-channel feature map is described by a cuboidal box. Through the forward pass, the growth in the number of feature channels, and the increasingly rich description of semantic content, is represented by the broadening horizontal width of each hidden layer. Semantic enrichment is accompanied by occasional max pooling, each application halving the spatial extent of the hidden layers. By extracting the most discriminative features from each portion of the image, the semantic content of the image is abstracted from its spatial configuration. In Figure 4.7, spatial abstraction is indicated by the narrowing of each feature map cross-section (projected into the plane of the page).

Finally, after the third max pooling operation, the image dimensions of the resulting hidden layer (just 6×6) are flattened to realise a vector, devoid of any spatial structure, but containing the semantic features from which the global content of the input can be determined. After, the final predicted probability distribution over the 1000 labels of the ImageNet database (the task to which AlexNet was first applied), is returned by three learned "classification" layers. Otherwise known as fully connected layers, these are achieved by the multiplication

of a parameterised weight matrix with the input feature vector and subsequent non-linear activation. Accordingly, and in the associated computational graph, every element of the intermediate 2048-dimensional feature vectors, and the final class distribution, is densely connected with their respective inputs.

Practically, Krizhevsky et al. (2017) showed that by leveraging graphics processing units (GPUs), such deep networks, though theoretically developed decades before (Fukushima and Miyake, 1982; LeCun et al., 1989), could now be practicably optimised. These computational and operational gains admitted the application of CNNs to tasks comprising high-dimensional, complex and heterogeneous image data, the like of which had only been idealised to that point. Whilst Krizhevsky et al. (2017) may not have been the first to present such an approach, perhaps due to the superiority of their results (exceeding pre-existing state of the art methods by a large margin) their paper has been highly influential. In large part, the success of AlexNet stimulated the research effort that has culminated in the popularity and variety of today's CNN-based, image processing methods.

Their advances also allowed for architectural investigation. Critically, they observed that performance was degraded by the removal of even a single convolutional layer, prompting their conclusion that network depth was key to improving performance, a foundational tenet of today's deep CNNs. Network depth admits the capacity to learn a hierarchy of features at increasing levels of abstraction from the raw content of pixel data (Litjens et al., 2017). By combining the low level image cues extracted within shallower portions of the architecture, deeper layers learn a high-level, abstract representation of the anticipated classification outputs (see Figure 4.8). Where these are anatomical targets, the synthesis of low-level features builds a high-level representation of shape, relative size and spatial arrangement. The result is that increasing network depth delivers a highly expressive model of the task at hand, one in which a wide variety of outputs can be represented by a CNN with high statistical capacity and efficiency[5].

Figure 4.8 also reflects the growth of the receptive field through the successive layers of the deep, AlexNet architecture. We highlight that where Section 4.2.4

[5]For further discussion of the advantages presented by deep CNN architectures, including their extraction of *distributed*, *sparse* and *hierarchical* representations of data, we direct the reader to Bengio et al. (2013).

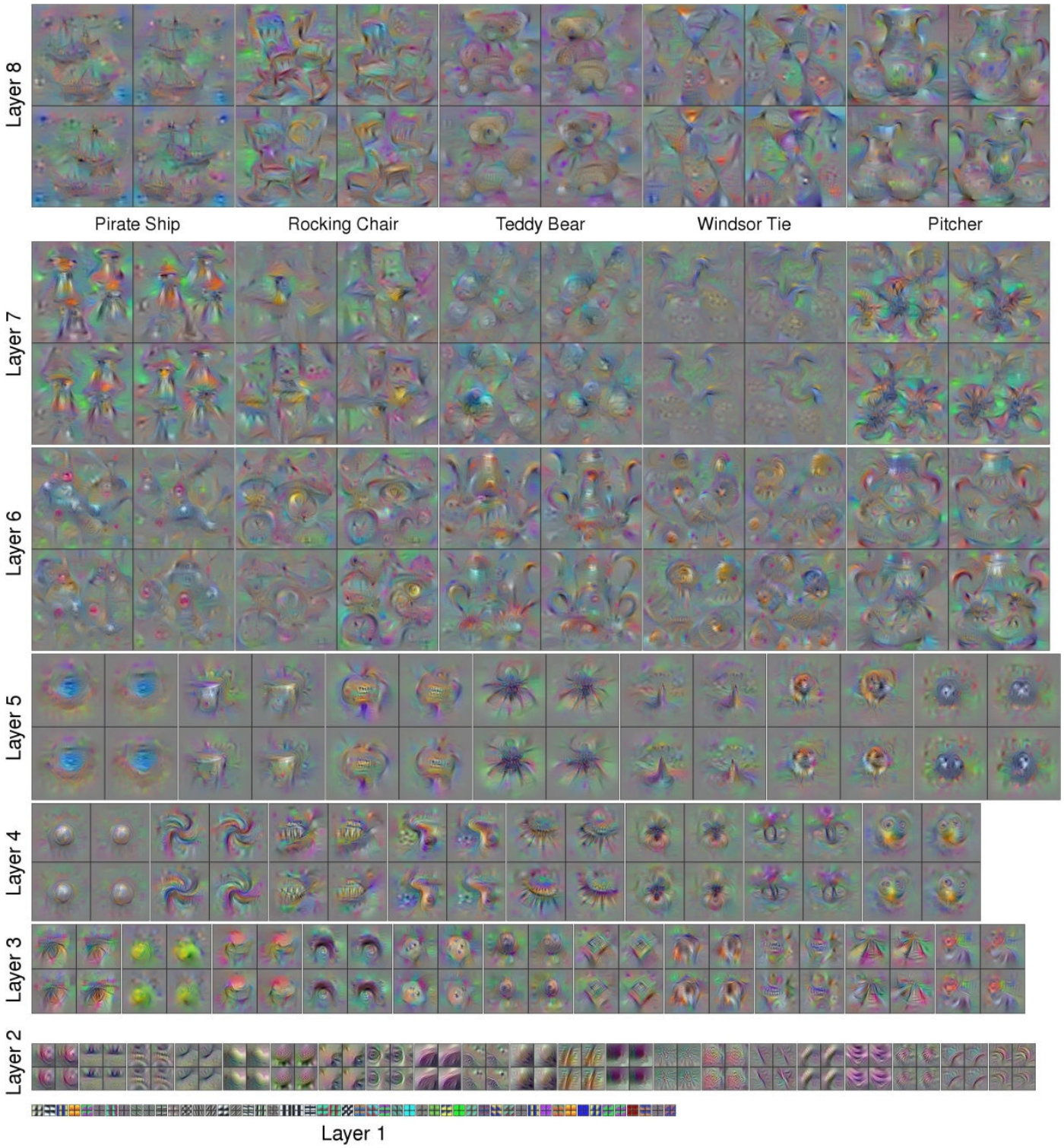
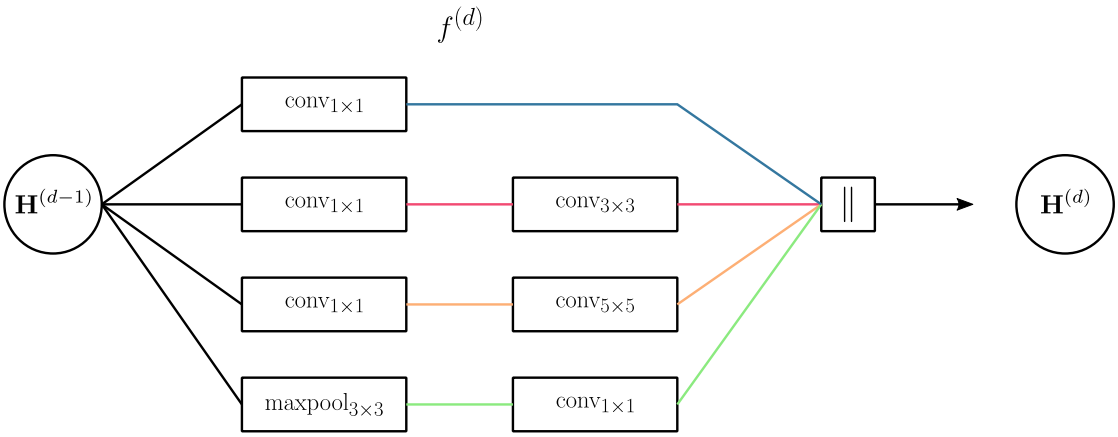


Figure 4.8: Developing an approach based on regularised image space optimisation, Yosinski et al. (2015) (from where this example is taken) sought to visualise the features learned at each layer of an AlexNet-like CNN architecture. Their results demonstrate that successively deeper layers of a CNN learn a higher-level and increasingly abstract representation of image content. Where the first layer is sensitive to only low-level image features such as edges and basic texture, the eighth combines these, resolving the high-level semantic structure of associated objects. Note that the relative size of each feature map reflects the layer-wise growth of the receptive field, and that the colours indicate the true red-green-blue (RGB) space of natural images.

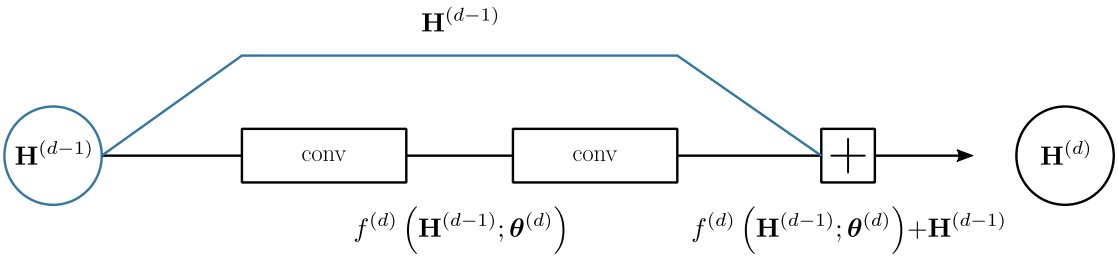
introduced the receptive field of a single convolutional operation, we now extend this notion. Rather than against its immediate hidden input, here we characterise the receptive field of a given convolutional layer with respect to the input *image* ($\mathbf{X}$). Via the relatively shallow ($d < d'$) portion of the computational graph established by the forward pass, the receptive field is governed by the set of pixels in $\mathbf{X}$ which determine each hidden element in $\mathbf{H}^{(d')}$. Accordingly, the receptive field of AlexNet, is most rapidly increased by max pooling after the second and third convolutional layers.

To a lesser extent, the receptive field is also expanded by successive convolution with unit stride and dilation rate. In such cases, its growth is largely governed by the spatial extent of associated convolutional kernels. Where the AlexNet architecture (Krizhevsky et al., 2017) employed convolutional kernels of varying spatial extent (as large as 11×11), within their Visual Geometry Group (VGG) network, Simonyan and Zisserman (2014) preferred successive feature extraction and transformation via smaller, 3×3 filters. Their experiments found that replacing a 5×5 with two 3×3 kernels, not only achieved an identical receptive field, but also, by increasing the network depth (from one to two non-linear layers), classification accuracy was improved. This result was observed despite the associated reduction in the number of convolutional parameters: $2 \times 3^2 < 5^2$. For its efficiency and ultimately superior performance, VGG's use of multiple small convolutional kernels has been adopted by modern CNN architectures for image segmentation.

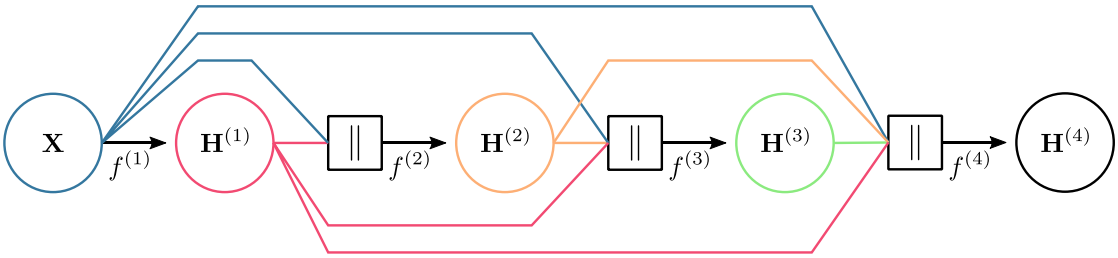
The association between increasing CNN capacity (either through depth or width) and improved performance, however, increases computational demand, and presents challenges to gradient-based optimisation. Addressing the first of these, Szegedy et al. (2015) observed that for a linear increase in the number of input ($a \leq A$) and output ($b \leq B$) features, the overall operational burden of the associated convolutional layer increases quadratically ($A \times B$). Such growth consumes substantial computational resource, expending GPU memory in the storage of hidden feature maps and multiplicative operations required by convolution. In pursuit of computational and statistical efficiency, they sought to approximate an idealised architecture which, rather than via the application of dense convolutional filters, made use of only sparse, optimal feature connectivity between hidden layers. Their resulting Inception module is shown in Figure 4.9a. Key to its operation is the



(a) Inception block, adapted from Szegedy et al. (2015).



(b) Residual block, adapted from He et al. (2016).



(c) Dense block, adapted from Huang et al. (2017).

Figure 4.9: Advanced computational blocks developed within image classification, and deployed within CNN-based segmentation architectures. The || and + symbols imply channel-wise concatenation and element-wise summation respectively. Note that for clarity, we omit explicit indication of where non-linear activation and batch normalisation (Ioffe and Szegedy, 2015) are included, their presence only implied by the notation representing layer-wise operation by $f^{(d)}$. In any case, the inclusion and order of such operations, as well as the number of convolutional operations applied per block, are free design choices. Moreover, arguments as to whether each block increments the network depth by one, or by the number of convolutional layers contained are largely inconsequential and, so long as they are applied consistently, only a matter of nomenclature.

use of 1×1 convolutional layers to project hidden activations to a lower dimensional feature space ($B < A$), saving computational resources. They also employ convolutional kernels of varying size, extracting and then concatenating features learned at different spatial scales. Incorporating the Inception module within their GoogLeNet architecture, they outperformed existing CNN architectures that were both shallower and less wide, with only a modest increase in parameter and computational resources. Whilst the Inception module has not proved as popular in novel architectures for image segmentation, its underlying intentions have informed related developments: (1) the drive to achieve deep networks that are both parameter- and computationally efficient; (2) the use of 1×1 convolution to control feature dimensionality; and (3) the interest in feature extraction and combination at multiple spatial scales.

Addressing the second challenge posed by the pursuit of ever deeper networks, He et al. (2016) observed that such architectures were more difficult to train. Moreover, previous experiments suggested that the associated degradation in performance (even compared with their shallower counterparts) was not a product of gradient-vanishing (an obstacle largely solved by the inclusion of intermediate batch normalisation layers (Ioffe and Szegedy, 2015)) nor over-fitting. Instead, their assessment pointed to limitations in gradient-based optimisation schemes, and their inability to find locally optimal locations in the very high-dimensional parameter space of deeper networks. In contrast, they demonstrated that such solutions (which at least maintain, rather than degrade, the level of performance achieved by shallower networks) can be analytically constructed at arbitrary network depth by successive identity mapping of shallow outputs.

Equipped with this insight, they formulated deep residual learning. By recasting the underlying mapping sought by the layers of a conventional CNN, $f(\mathbf{H}; \theta)$, they hypothesised that the resulting residual functions, $f(\mathbf{H}; \theta) + \mathbf{H}$, would be easier to train. As shown by Figure 4.9b, these residual units were achieved by the inclusion of skip connections within the CNN architecture. At the time, this approach allowed some of the deepest networks presented to be successfully trained. ResNet models with hundreds of sequential convolutional layers demonstrated performance gains over shallower architectures. As a means of training deeper net-

works, residual connectivity can and has been incorporated within architectures for CNN-based segmentation.

Amongst other works, ResNets tackled the challenge of deep network optimisation by introducing short cut connections, effectively bypassing layers of the network to reduce the path length between shallower and deeper feature maps. Taking this principle to its ultimate extreme, Huang et al. (2017) presented densely connected CNNs, in which every hidden layer receives as input, the output of all preceding feature maps (see Figure 4.9c). They argue that such dense, direct connectivity improves the flow of gradient information from the scalar loss to all network layers. Furthermore, by virtue of the fact that shallow, low-level features are concatenated and reused as input at all depths, DenseNets can be made parameter efficient by reducing the number of output channels realised by each hidden layer. These advantages were borne out in their experimental results, achieving state of the art classification performance with relatively few parameters. As per its residual counterpart, dense connectivity has also been tested within modern architectures for CNN-based segmentation.

4.3.2 2D segmentation

Unsurprisingly, initial attempts at CNN-based image segmentation started in 2D and built on the principles developed for classification (those listed in Table 4.1). Differences between the two tasks, however, preclude the immediate application of associated architectures, at least without suitable adaptation to rationalise their operation with the segmentation objective: the global summary of image content sought by CNN-based classification is achieved by spatial abstraction; whereas, segmentation seeks to localise the constitutive structures through a pixel-wise labelling.

Such a high level analysis might reject those sources of spatial abstraction, which though critical to classification architectures, may compromise the localisation necessary to segmentation. In particular max pooling appears at odds with the latter's objectives. However, in the context of segmentation - where we recognise that in common with classification, a pixel-wise labelling depends not only on local cues, but also on spatially extended and even global features of the im-

Table 4.2: A history of the developmental contributions made by key network architectures for image segmentation.

Year	Architecture	Authors	Contribution
2012	CNN	Ciresan et al.	Applied in a sliding window across the input, CNN-based pixel classifiers can build up a dense segmentation.
2015	FCN	Long et al.	Features can be spatially reconstructed from classification nets by 1×1 and transposed convolution; skip connection.
2015	Deconvolutional network	Noh et al.	Mirroring the contracting path used in classification with an expanding, deconvolutional path improves reconstruction.
2015	U-Net	Ronneberger et al.	Combining skip connections with incremental expansion balances reconstruction and semantic feature extraction.

age[6] - pooling layers provide a means of rapidly expanding the network's receptive field. Hence, rather than solely as a means of spatial abstraction, pooling allows for multi-scale features to be learned efficiently. By attempting to balance the two, CNN-based segmentation faces an inherent tension between the semantics of image features, and their spatial localisation to particular regions of the data (Long et al., 2015). The succession of segmentation architectures presented in this section (and listed in Table 4.2) are specially adapted to meet this expectation.

Historically, attempts to extend classification architectures had first to contend with the disparity between the non-spatial class vector returned by prior CNNs and the highly structured output anticipated by segmentation. Rather than architecturally, this gap was initially bridged operationally. Instead of predicting a dense probabilistic labelling per forward pass, Ciresan et al. (2012) inferred the class membership of a single pixel centred at the network input. An associated pixel-wise labelling was then built up by multiple passes through the network within a sliding window framework. Whilst this approach was perhaps reasonable in the domain of low resolution, natural images, its inefficiency rendered it impractical in most real world use cases. In this scheme, 3D medical imaging comprising millions of voxels would demand the same number of forward computations.

More efficient solutions rely on architectural modification to rationalise CNN output with the dense, structured labelling demanded by segmentation. In their

[6]In the segmentation of cardiac anatomy from medical images, this reflects the fact that individual sub-structures can be more easily identified with reference to their gross position, relative to the predictable arrangement of the remaining components of the thorax.

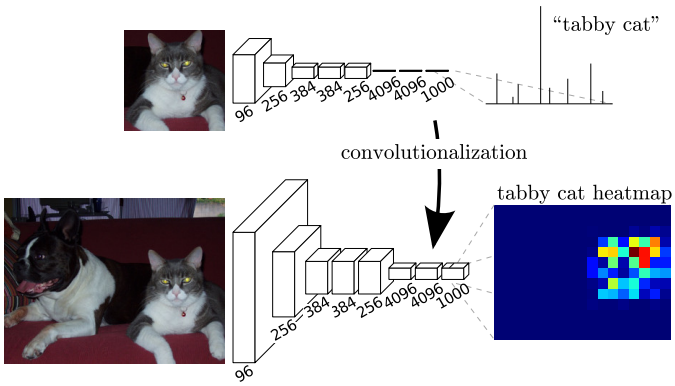


Figure 4.10: Fully convolutional networks (FCNs) replace fully connected layers with 1×1 convolutions and use upsampling to recover pixel-wise labels from the spatially abstract feature map learned by a VGG network. Reproduced from Long et al. (2015).

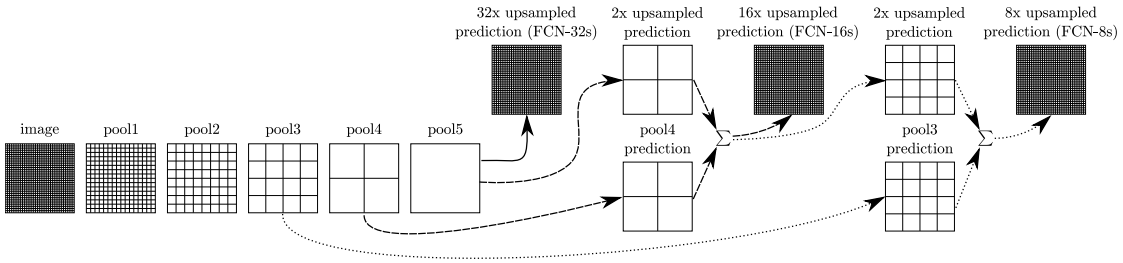


Figure 4.11: Fully convolutional network (FCN) architectures improve the accuracy with which the fine details of segmentation targets are predicted, by introducing skip connections to fuse multi-scale and multi-resolution features. In this example, the resolution of each feature map is indicated by interposed gridlines. On the left hand side, the features extracted at each pooled stage of a VGG network are indicated. Depth increasing from left to right, these are labelled "image", "pool1", "pool2" and so on. On the right hand side, arrows indicate the skip connections defining the fusion of various feature maps, prior to multi-scale prediction. Reproduced from Long et al. (2015).

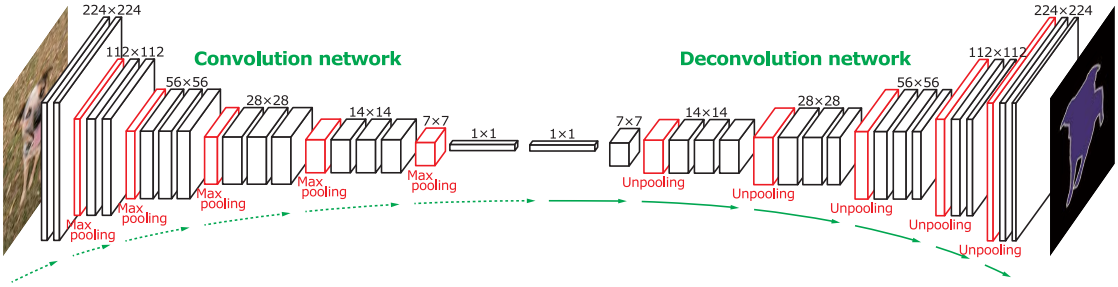


Figure 4.12: Compared with a conventional VGG network for classification, the deconvolutional network presents a symmetric expanding path in which the dense labelling demanding by segmentation is recovered by incremental unpooling and transposed convolution. Reproduced from Noh et al. (2015).

seminal work on fully convolutional networks (FCNs), Long et al. (2015) made four adaptations to the VGG network (Simonyan and Zisserman, 2014). Firstly, rather than collapse the pooled features obtained by the final convolutional layer into a vector, their FCNs maintain spatial structure throughout (albeit at a coarse resolution, see the heatmap in Figure 4.10). Secondly, rather than by fully connected transformation, their classification layers are implemented as 1×1 convolutions, allowing the FCNs to accommodate inputs of variable size. Thirdly, they recover the full spatial resolution of the input by learned upsampling, implemented as transposed convolution. Fourthly (and given the current state of the art in CNN-based segmentation, perhaps most importantly), they improve the granular accuracy of predicted segmentation by the summation of feature maps learned at different spatial scales and resolutions. As shown in Figure 4.11, via skip connections and learned upsampling, shallow but high resolution features are combined with those which though semantically rich, are extracted by deeper layers of the network, at coarse resolution. The resulting FCN-16s and FCN-8s architectures made significant improvements compared with their FCN-32s counterpart; the latter being based on direct upsampling of the single, most spatially abstracted, but lowest resolution VGG feature map.

The idea of incremental upsampling was taken to its extreme by Noh et al. (2015). In their account of deconvolutional networks (see Figure 4.12), they investigate a symmetric architecture in which the spatial abstraction associated with the convolutional and pooling layers of a conventional VGG network, is mirrored by un-

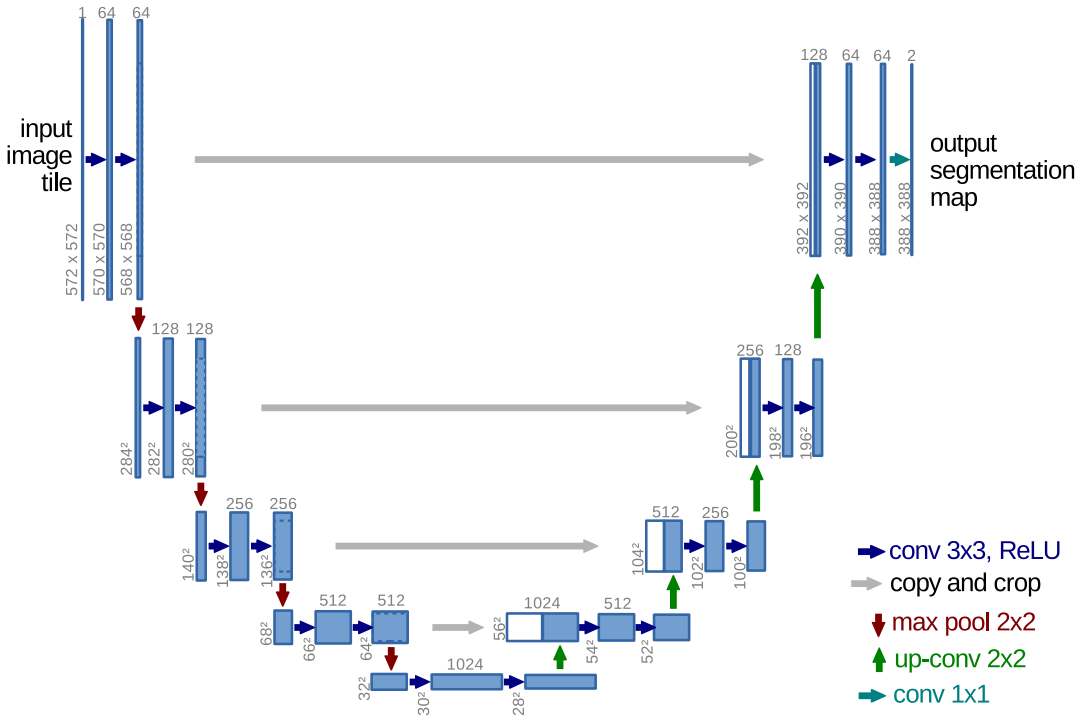


Figure 4.13: Based on symmetric contracting and expanding pathways, linked by skip connections between shared resolution levels, the U-Net architecture has come to dominate the state of the art for CNN-based segmentation. Reproduced from Ronneberger et al. (2015).

pooling (Zeiler et al., 2011) and deconvolution (or transposed convolution) within an opposing path for spatial reconstruction. In the contracting path, this architecture efficiently expands the receptive field and learns a rich semantic description of image content at low spatial resolution. In the expanding path, compact features are incrementally unpacked in a coarse-to-fine framework and localised to specific pixels of the image. Their approach does not, however, include skip connections for the synthesis of multi-scale features.

Combining the architectural modifications described previously, Ronneberger et al. (2015) established the U-Net architecture shown in Figure 4.13. Critical to its success, to the the symmetric form described by Noh et al. (2015), they added the skip connections introduced by Long et al. (2015). At each pooling level, these allowed for the high resolution features learned in the contracting path to be directly incorporated within spatial reconstruction. Rather than by

summation, features forwarded from the contracting path are concatenated as additional channels of the symmetric expansion. In aid of spatial reconstruction, this allows for the synthesis of the two to be learned by convolution.

As well as their architectural developments, Ronneberger et al. (2015) made use of intensive data augmentation based on non-rigid deformation (a mode of transformation that has since been incorporated within dedicated open source libraries (Pérez-García et al., 2021). Furthermore, they employed a geometric weighting schemes to bias the CE loss in the vicinity of the background interfaces dividing foreground targets and balance class frequencies. Collectively, their contributions achieved a level of performance that, though dependent on only a small number of training examples, exceeded the state of the art in the binary segmentation of cellular microscopy images by a large margin. Its superiority was measured not only by metrics of spatial overlap, but also by the warping error (Jain and Farrokhnia, 1991), indicating that U-Net's incremental, reconstruction of fine details realised a topologically meaningful delineation of cellular interfaces.

Since this time, U-Net has solidified its position as the state of the art architectural solution to CNN-based 2D segmentation across a number of domains (Isensee et al., 2021). Success has bred popularity, its original paper having received over 40,000 citations. Arguably, this status is a product of the balance struck between the extraction of semantic features in the contracting path (leveraging the efficiency of max pooling to expedite multi-scale learning), and incremental spatial reconstruction (incorporating transposed convolution and skip connections at multiple scales) in the following expansion. In so doing (and until a more optimal balance is found), U-Net represents our best attempt to overcome the tension between the spatial abstraction of semantic features and their localisation to the dense structure of the image (Long et al., 2015).

4.3.3 3D segmentation

Despite the success of the 2D U-Net architecture, its natural extension to 3D segmentation by the inclusion of volumetric convolutional kernels (first demonstrated by Çiçek et al. (2016)), was not established as an equivalent gold standard until recently. Before ultimately returning to the state of the art nnU-Net framework

(Isensee et al., 2021) in Section 4.3.4, here we describe the historical development of competing approaches and architectures, those motivated by the particular challenges of 3D segmentation. Due to the cubic increase in the number of voxels, in addition to carrying a higher computational burden, volumetric segmentation depends on a more complex set of visual patterns than might be found in 2D (Li et al., 2017c). Coupled with the equivalently scaled increase in the number of network parameters, the typical paucity of medical image training data raises the possibility of overfitting in 3D tasks. These observations have led some to suggest that training a 3D CNN from scratch is infeasible where training data are insufficient and computer memory limited (Mortazi et al., 2018).

Accordingly, the following architectures and associated formalisms have sought to address ways to: efficiently balance the intensive demands of 3D representation with the computing resources available; facilitate learning of complex 3D appearances; or both. These gains have been achieved either through: (1) adapting existing methods to 3D segmentation (such as via novel task formulations, or schemes for training established 2D, 2.5D or 3D architectures based on U-Net or FCNs); or (2) through the conception of novel architectures, tailored to the demands of 3D CNN-based segmentation.

Adapting training schemes to 3D segmentation

A first, and perhaps most obvious approach to overcoming the computational challenge of 3D CNNs, is to abandon any attempt to learn the 3D geometry of anatomy, instead focusing on its appearance in 2D cross-section. At test time, 3D structure is recovered by the slice-wise combination of 2D predictions (and possible domain- and task-specific refinement). This approach admits the use of the 2D U-Net, or other architectures explored in the previous Section 4.3.2. In the context of data acquired on a spatially anisotropic grid, where the spacing between slices might far exceed the in plane resolution, this approach may even be preferable. For example, within conventional 2D short axis cine images, 3D appearances are compromised. Accordingly, 2D architectures remain the gold standard approach, having achieved a level performance consistent with the inter-observer variation between clinical experts (Bai et al., 2018).

Where acquired with isotropic 3D spatial resolution, however, lone 2D CNNs do not leverage volumetric features. A compromise between the two, the so-called 2.5D approach seeks to incorporate this context without incurring the computational costs of an equivalent 3D model. Typically it involves training three independent 2D CNNs, one per orthogonal plane of the 3D volume, and fusing their predictions at test time (Hesamian et al., 2019). The 2.5D moniker, however, has also been applied to a range of different configurations, each setting out to augment 2D CNNs (or their inputs) to leverage volumetric spatial context without applying 3D kernels. These were recently compared by Minnema et al. (2021) and include: triplanar majority voting; concatenation of slices from the third spatial dimension as additional channels of the input (or which consider adjacent slices more generally); and training on randomly oriented 2D cross-sections. Finally, a more advanced triplanar CNN is presented by Prasoon et al. (2013) who, through a shared softmax output layer realise the dependent training of three 2D networks, respectively focused on axial, sagittal and coronal planes.

Ultimately however, 2D and 2.5D approaches lack the expressive capacity of 3D CNNs. Accepting the increased parameterisation associated with 3D convolution, a popular approach to reduce computational demand is to separate task complexity into a cascade of operations, typically: (1) 3D region of interest (ROI) localisation at low resolution; followed by (2) high spatial resolution segmentation. In the first step, spatially downsampled input allows for the incorporation of 3D context, without dramatically increasing the memory required to store hidden activations. This infers a focal ROI that is cropped from the high resolution data, and fed to the second step, reducing the size of downstream feature maps. Moreover, by isolating the two components of the overall task, each is simplified, and demands a lower capacity network (with curtailed depth or width) than might be required were they addressed simultaneously. To segment white matter lesions from magnetic resonance imaging (MRI), Valverde et al. (2017b) implement a cascade of independently trained 3D CNNs. Alternatively, to label multiple organs within abdominal CT, Roth et al. (2018) train a cascade of 3D U-Nets in an end-to-end fashion (extending the 2D approach of Christ et al. (2016)). Zhou et al. (2017) combine multi-planar 2.5D U-Nets with spatially cascaded processing, achieving an iterative approach to segment the pancreas from CT data.

Table 4.3: The range of CNN architectures that have been specially developed to address the demands of 3D image segmentation.

Year	Architecture	Authors	Contribution
2016	3D U-Net	Çiçek et al.	The 2D U-Net can be extended to 3D segmentation via the incorporation of 3D convolutional kernels.
2016	V-Net	Milletari et al.	Dense inference via an end-to-end trained 3D U-Net incorporating residual learning may improve performance.
2017	DeepMedic	Kamnitsas et al.	Comparatively, by restricting the number of spatial scales, parallel stream architectures are more parameter-efficient.
2017	HighResNet	Li et al.	Similar gains in efficiency are made by using dilated convolutions to increase the receptive field at high resolution.
2018	Dense 3D CNN	Chen et al.	Densely connected convolutional blocks may also facilitate learning of the complex features of 3D anatomy.
2019	TeTrIS	Lee et al.	Alternatively, these features can be instilled indirectly, via the learned transformation of an anatomical prior.
2021	nnU-Net	Isensee et al.	Despite this wealth of architectural contributions, few make reliable improvements compared with the 3D U-Net.

CNN architectures dedicated to end-to-end 3D segmentation

Works cited in the previous section met the computational and representational challenges of 3D CNN-based segmentation by making adaptations to simplify training. Rather than match the full 3D structure of volumetric data and their segmentation with an architecture composing 3D convolutional layers, they instead developed schemes for reconstructing labels from predictions made by 2D or 2.5D networks. Where 3D kernels were introduced, others broke down the complex segmentation task into simpler facets. Here we address the natural alternative, providing an account of architectures exclusively including 3D convolution: those which have been specially developed to tackle volumetric segmentation, and trained in an end-to-end fashion. These are summarised in Table 4.3.

After Çiçek et al. (2016) first demonstrated the 3D equivalent of U-Net, Milletari et al. (2016) presented the modified V-Net architecture. This work remains notable, being an early example of an end-to-end CNN-based solution to volumetric segmentation, one that also applied the Dice loss later generalised by Sudre et al. (2017). Additionally, and as shown in Figure 4.14, V-Net learns residual features (He et al., 2015) at each level of the multi-scale expanding and contracting

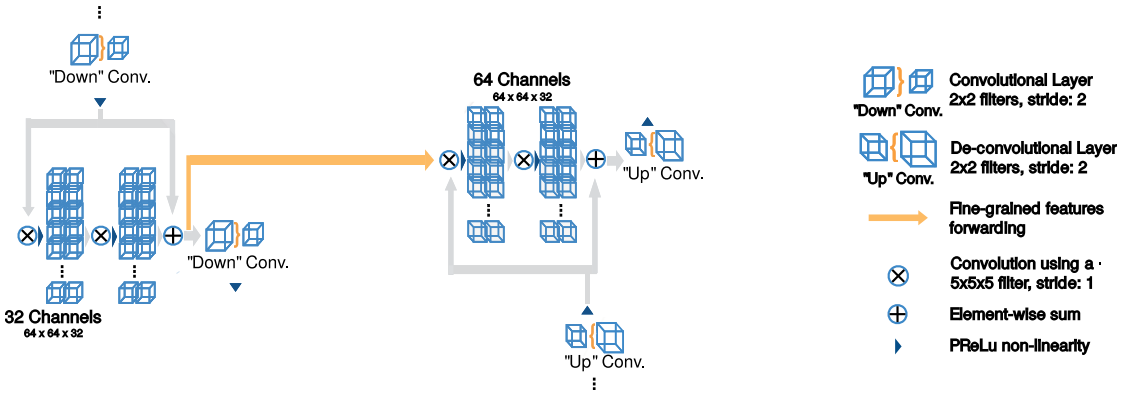


Figure 4.14: The V-Net architecture for dense volumetric segmentation includes two types of skip connection: the first (in yellow) incorporates high resolution features within the expanding path for spatial reconstruction; the second (in grey) admit residual feature learning at each spatial scale. This figure has been extracted from the schematic found in Milletari et al. (2016), and reflects a single spatial scale of the entire V-Net architecture.

paths. The authors suggested this modification was beneficial to both performance (perhaps addressing the complex appearances of 3D anatomy) and the speed of convergence during training. A final architectural subtlety, rather than by channel-wise concatenation, high resolution features forwarded from the expanding path were incorporated within spatial reconstruction via element-wise summation: a design choice perhaps seeking to reduce the memory demands of storing large, 3D multi-channel features maps. Despite a limited training set of only fifty examples, Milletari et al. (2016) achieved impressive spatial overlap and surface localisation performance when segmentating the prostate gland from volumetric MRI.

The success of U-Net-based, 3D architectures (including V-Net), comes with the computational and statistical cost associated with the millions of parameters contained. In response, Kamnitsas et al. (2017) presented the DeepMedic architecture, boasting increased efficiency by limiting the number of spatial scales interrogated (at least compared with the majority of U-Net-based networks). Perhaps inspired by cascaded CNN processing, including prior ROI localisation, DeepMedic accommodates two, mutually centred inputs. The first accepts a large 3D patch resampled at low resolution, allowing the extended spatial context of segmentation targets to be learnt. The second, a smaller 3D patch maintaining high spatial

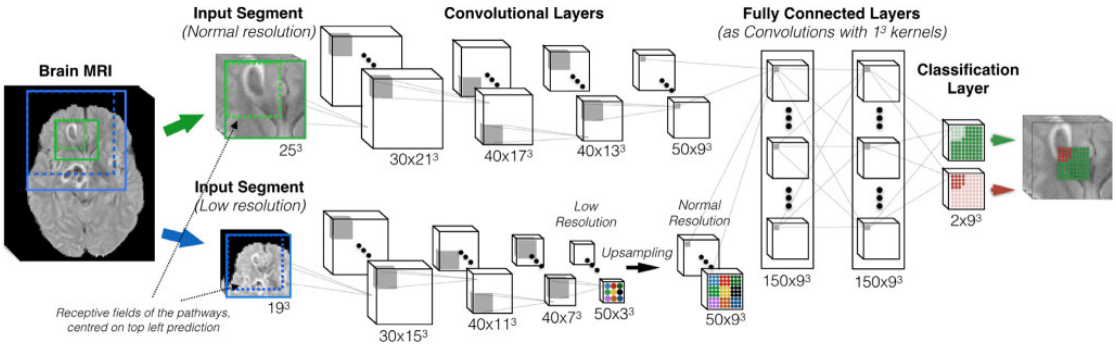


Figure 4.15: The DeepMedic architecture for dense volumetric segmentation, including dual streams for feature extraction at varying spatial scales. These are learnt in parallel, prior to their eventual fusion via convolutional classification layers. Reproduced from the work of Kamnitsas et al. (2017).

resolution, admits the extraction of features describing the fine details of these structures. Unlike cascaded networks, and as shown in Figure 4.15, each of these streams are learned in parallel, such that their associated features are latterly fused prior to prediction. This is achieved by their synthesis within classification layers implemented as $1 \times 1 \times 1$ convolutions for dense prediction. In addition to their architectural contribution, Kamnitsas et al. (2017) also explored different sampling strategies to counter the class imbalance within their challenging task of segmenting brain lesions from neuroradiological MRI.

Bearing similarity with the structure of the DeepMedic architecture, Chen et al. (2018a) also presented a dual stream 3D CNN. However, rather than combine features learned at multiple scales, their network isolated the semantics captured by different neuroradiological MRI acquisitions (including fluid attenuated inversion recovery and T2-weighted imaging in one stream; and T1-weighted and contrast-enhanced image series in another). After fusion with the first stream, the second is trained to classify the lesion sub-structures relevant to diagnosis. Specific to their particular clinical setting, we are less interested in the semantic formulation of their dual stream architecture. Perhaps of more general interest, Chen et al. (2018a) employ densely connected convolutional blocks (Huang et al., 2017) to tackle the challenge of learning complex 3D appearances.

As per Kamnitsas et al. (2017), Li et al. (2017c) also sought an efficient solution to CNN-based volumetric segmentation. They presented their HighResNet

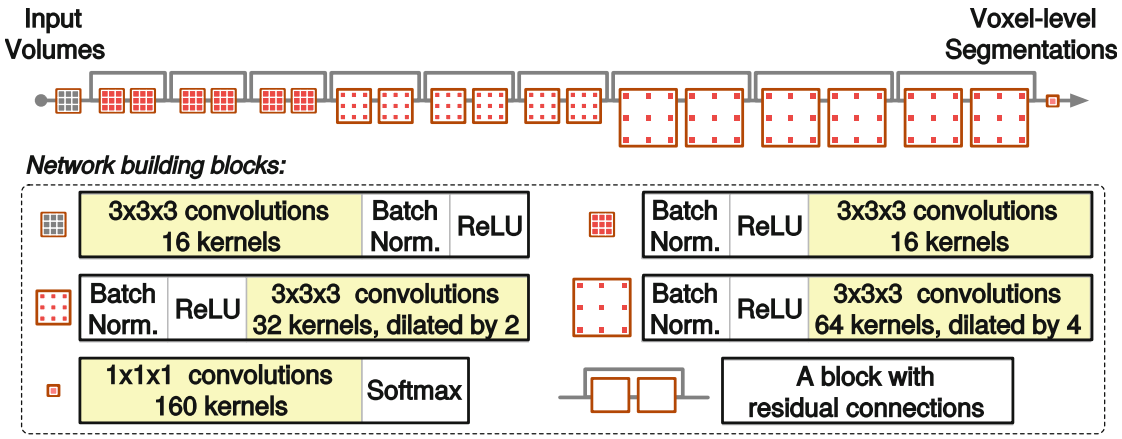


Figure 4.16: The HighResNet architecture combines $3 \times 3 \times 3$ dilated convolutions and residual connectivity to achieve efficient CNN-based, volumetric segmentation, learning features across a distribution of spatial scales.

architecture for the parcellation of brain anatomy from MRI (see Figure 4.16). Rather than through pooling, this learns features at multiple spatial scales by successive dilated convolution (Yu and Koltun, 2015), rapidly expanding the receptive field whist maintaining high spatial resolution. Moreover, as per its modern counterparts (and unlike the preceding V-Net (Milletari et al., 2016) and DeepMedic (Kamnitsas et al., 2017) networks), it employed small $3 \times 3 \times 3$ kernels. Together, these design choices reduced the number of network parameters by an order of magnitude when compared with U-Net-like architectures. Additionally, since HighResNet extracts semantic features within residual blocks, Li et al. (2017c) argue that their network considers a multitude of computational paths between input and prediction, and learns a feature set across a distribution of receptive fields.

Finally, we present the work of Lee et al. (2019), who met the challenge of learning complex 3D geometry by leveraging prior information. Relying on spatial transformer networks (Jaderberg et al., 2015), their template transformer networks for image segmentation (TeTrIS) established a new approach, one incorporating the principles of atlas-based segmentation. Rather than through a CNN's forward computation, TeTrIS make predictions by learning an optimal spatial transformation of a task-specific geometric prior (see Figure 4.17). Critically (and though the output of their CNN is limited by the parameters of its mediating spatial transform), their entire framework is trained end-to-end, supervised by a conventional

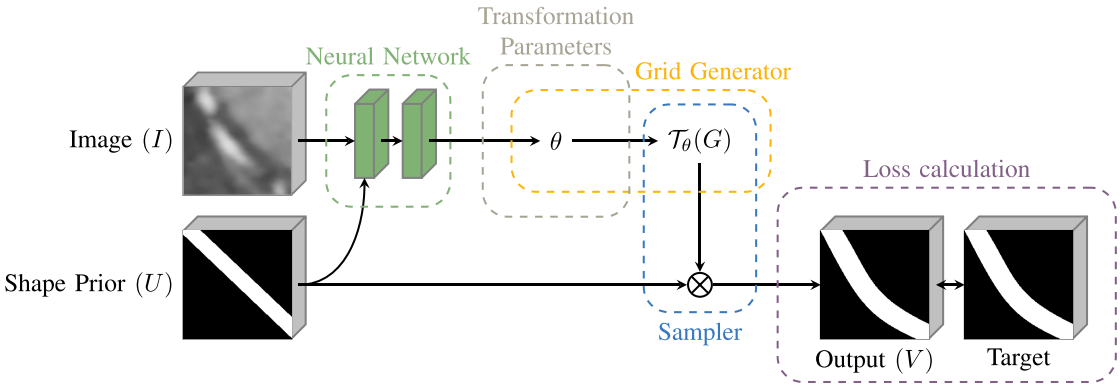


Figure 4.17: Leaning on the principles of atlas-based segmentation, template transformer networks for image segmentation (TeTrIS) present a novel approach to CNN-based segmentation of 3D medical image data. Rather than a direct mapping between image (I) and predicted segmentation (V), their CNN learns the parameters (θ) of a free form transformation $(\mathcal{T}_\theta)$. Predictions are made according to the learned transformation of a prior description of expected anatomy (U).

segmentation loss between ground truth and predicted label maps. Constraining the free form deformation between prior and prediction, TeTrIS benefit from the advantages of their atlas-based methodological forerunners: that inferred label maps conform to the morphological and topological properties described by the prior. In so doing, Lee et al. (2019) mitigate the challenge of learning the complex 3D features of anatomy from training data. Of peripheral interest to our focus on CHD, they demonstrate their approach to segment the coronary arteries from contrast-enhanced CT.

4.3.4 State of the art

The previous Section 4.3.3 presented a wealth of research seeking to improve the efficiency or performance of what were at the time, state of the art architectures for the CNN-based segmentation of 3D medical images. Provided this array of contributions (spanning network dimensionality; task formulation; approaches to promote efficient feature learning at multiple scales; dilated convolution; advanced network connectivity patterns to leverage residual and dense learning; and the incorporation of priors), determining which combination might optimally address a given task is far from trivial.

This observation underlies the work of Isensee et al. (2021) who, in presenting their nnU-Net ("no new net"), establish that a plain U-Net, incorporating only a select few of these modifications, is at least competitive with the majority of more involved configurations. Moreover, they find those design choices that might otherwise be relegated to minor concerns of low level implementation (for example, resampling and normalisation of training data, and ensemble prediction) can be more influential than higher level, architectural considerations. This allows their nnU-Net to outperform a wide array of competing architectures across an equally broad array of 23 segmentation tasks, spread between 2D and 3D applications. Accordingly, and in the experimental work contained in the rest of this thesis, we rely exclusively on 2D and 3D U-Net architectures; both as a baseline, and as the architectural foundation to our novel contributions.

4.4 Applications in cardiac image segmentation

To this point, our review of CNN-based segmentation methodologies has been largely application-agnostic. We now turn our attention to their use within cardiac imaging, considering practical applications across CT and MRI primarily, and echocardiography where informative. Despite the main finding of the previous Section 4.3 - that though performance may be enhanced by bells and whistles, the U-Net architecture is representative of the current state of the art in CNN-based segmentation - a review of the literature focused on cardiac applications provides critical context to our experimental contributions. This topic is served by an incredible amount of research literature. Amongst the various clinical motivations and associated acquisitions (and in common with the body of research concerning conventional segmentation methodologies, see Chapter 3), no clinical segmentation task has received as much attention as ventricular volumetry from 2D short axis cine MRI. In contrast, our focus is on 3D applications, and in particular those for which segmentation returns a representation of multi-class (to at least include the four cardiac chambers) or whole heart anatomy (see Figure 2.10 and Figure 3.11). Accordingly, in Section 4.4.1, we primarily limit our review to studies concerned by the segmentation of these targets (and their constituents) in aid of a downstream application reliant on a 3D model of anatomy. Most often,

therefore, we are concerned with works seeking to label 3D images of isotropically high spatial resolution, largely dispensing with those dealing with short or long axis data acquired at significant inter-slice spacing. Given that our own motives surround the translation of such segmented data into patient-specific models of CHD, in Section 4.4.2 we refine our search further, reviewing only those studies investigating the application of CNN methodologies to delineate the structures of the congenitally malformed heart. For a review of the CNN-based segmentation of all other cardiac applications, we refer the reader to the article written by Chen et al. (2020)[7].

In each of these sections, and in light of the findings of Section 4.3.3 we group citations according to their methodological approach. We compare those that: (1) seek to simplify their implementation by 2D or 2.5D networks, or by task reformulation; with (2) those that employ a fully 3D CNN. In each, we make special mention of reports which include or introduce novel aspects to their network, or that have a strong focus on downstream clinical application.

4.4.1 3D Whole heart and multi-class segmentation

Adapting training schemes to 3D segmentation

Several published reports have presented solutions to 3D cardiovascular segmentation based on 2D CNNs. Primarily, these differ in their approach to fusing the slice-wise predictions returned. Mortazi et al. (2017) and Mortazi et al. (2018) present an approach based on connected component analysis, weighting each isolated component according to its contribution to the total ground truth volume. Respectively, they applied this scheme to the 2013 Left Atrium Segmentation Challenge (Tobon-Gomez et al., 2015) (LASC) and to the multi-class cardiac anatomy demanded by the 2017 Multi-Modality Whole Heart Segmentation (Zhuang et al., 2019) (MM-WHS) Challenge label specification. Addressing the latter task, Wang and Smedby (2018) developed a more involved scheme incorporating statistical shape modelling (see Figure 4.18). To do so, they divided the training data pro-

[7]Where our later experimental work requires, we provide a review of short axis segmentation for ventricular volumetry, focused on the incorporation of shape priors within the introductory sections of Chapter 7.

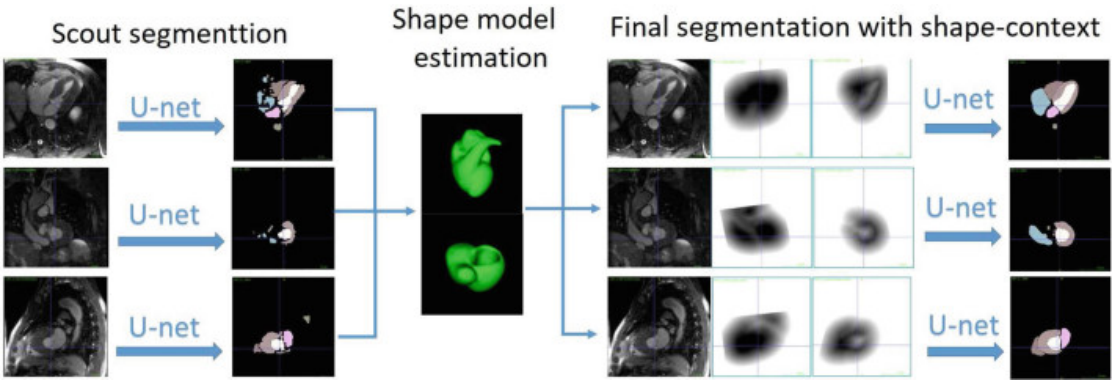


Figure 4.18: An overview of the joint CNN-active shape model framework presented by Wang and Smedby (2018).

vided by the MM-WHS Challenge into two sets: from the twenty CT cases they constructed an active shape model; leaving the twenty isotropic CMR volumes for experimentation. Their pipeline proceeds in a two-step cascade. Firstly, an initial segmentation of CMR data was made by triplanar U-Net architectures. To this, they fit their active shape model, appending the result as additional channels of the input to a second, identical set of U-Nets. Both intermediate and final predictions are determined by averaging the probabilistic outputs for each view.

A more conventional CNN-based segmentation cascade appeals to a coarse-to-fine strategy, training dedicated networks for localisation (at low spatial resolution) prior to labelling (at high resolution). Such a cascade was applied to the CT component of the MM-WHS dataset by Sundgaard et al. (2020). Their implementation proceeded via two sets of 2D U-Nets (each 2.5D set being composed of three networks trained by exclusive orthogonal views), the first acting on a spatially downsampled input to isolate the cardiac ROI. After cropping the data accordingly, the second performed segmentation at high resolution. In both cases, a straightforward fusion scheme was achieved by pixel-wise non-maximal suppression across all classes and imaging planes.

Though applying the same coarse-to-fine strategy, Sharobeem et al. (2022) adopt an alternative formulation of the preceding localisation task, using a 2D VGG network to regress a dense distance map predicting pixel-wise separation from the aortic valve. After extracting a centred ROI, they segment a rich description of 3D cardiac anatomy including ten multi-class labels. Their work is

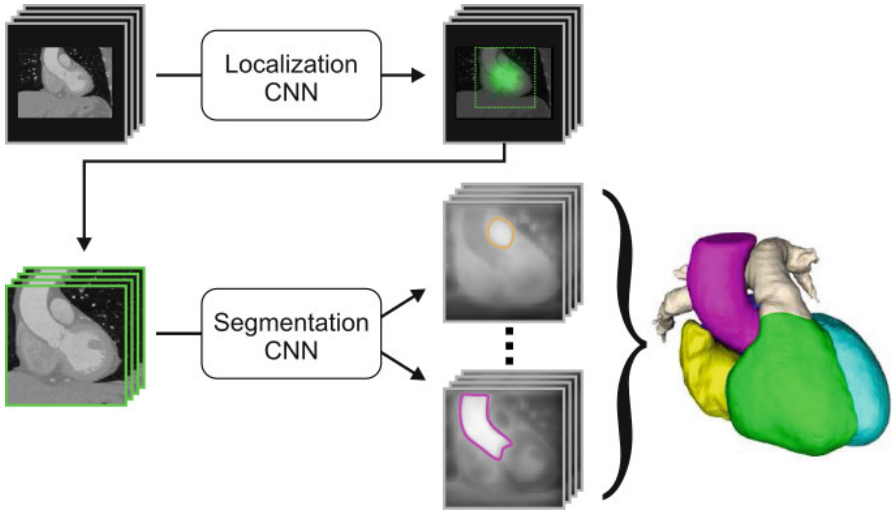


Figure 4.19: An overview of the coarse-to-fine 3D U-Net cascade of localisation and then segmentation, presented by Payer et al. (2018).

particularly impressive for its dependence on an in house curated dataset of CT scans of 71 patients with acquired heart disease; each of which was scanned prior to transcatheter aortic valve implantation. Amongst their technical findings, it is noteworthy that spatial overlap performance was markedly reduced for the predicted segmentation of small or variable structures including: the coronary sinus, pulmonary veins and pulmonary arteries (including pathology-induced changes in morphology).

The attraction of the localisation-segmentation cascade extends to workflows entirely dependent on 3D CNNs. Moreover, its application allowed the approach presented by Payer et al. (2018) to (at least with respect to the segmentation of multi-class anatomy from 3D CT) lead the original submissions made to the MM-WHS Challenge (see Figure 4.19). It is noteworthy, however, that even to their high resolution 3D U-Net (the second in the cascade and that needed to predict the labels of the cropped ROI), input data remained relatively coarse, having a pixel spacing of 4 mm^3. Though consistent with the relatively featureless ground truth segmentations provided by the MM-WHS Challenge, it is questionable whether such data could resolve the septal defects characteristic of CHD, being often defined at the limit of acquired spatial resolution. In the same year, Xia et al. (2019) applied a closely related coarse-to-fine cascade to the gadolinium-enhanced MRI

data provided by the 2018 Atrial Segmentation Challenge (Xiong et al., 2021) (ASC). Applying the V-Net architecture at each stage, by reducing the batch size to one and given their restricted focus on only the left atrium, they achieved predictions at full spatial resolution.

Lastly within the collection of works seeking to simplify 3D CNN-based cardiac segmentation through its decomposition into a localisation-segmentation cascade, we review those formulations depending on region-based convolutional neural networks (R-CNNs). The R-CNN framework promotes the application of specialised CNNs to (or via) proposed regions rather than the entire image. Whilst in its earlier presentations (Girshick et al., 2014), R-CNN relied on region proposals returned by a prior Selective Search (Uijlings et al., 2013) of the input, its modern descendent, Faster R-CNN (Ren et al., 2015), integrates ROI extraction and subsequent analysis within a single CNN.

The Faster R-CNN framework lends itself naturally to the cascade discussed. For example, Xu et al. (2018) use a region proposal network to predict the box bounding the cardiac ROI within the MM-WHS CT data, prior to segmentation by 3D U-Net. In addition, they describe an auxiliary edge extraction network, its output used to construct a boundary sensitive loss that improves spatial overlap with the ground truth. Later, Liu et al. (2019) used a similar R-CNN approach to extract the region bounding the left atrium. In their consideration of the data provided by the LASC, rather than by subsequent CNN, predictions were refined by Otsu thresholding and evolution of a 3D active contour. Finally, Harms et al. (2021) investigated a reversed scheme in which the image features extracted by 3D U-Net were subsequently interrogated by an R-CNN, including semantic segmentation. They curated their own dataset of 55 CT scans (all patients being treated for lung cancer), including an impressively rich multi-class task specification according to fifteen labels.

CNN architectures dedicated to end-to-end 3D segmentation

Compared with those presented previously, the following works all rely on a single 3D CNN as an end-to-end segmentation solution. Perhaps the most straightforward approach involves the use of a single 3D U-Net. Borra et al. (2019) trained

such a model in their submission to the ASC. As has become commonplace within modern CNN-based segmentation, they used connected component analysis to eliminate small, spurious components from their predicted labelling of the left atrium. Following this work, and within their respective segmentation tasks, others have made incremental adaptions, including: an advanced scheme for data augmentation based on the principal components of variation expressed by the MM-WHS training set of CT images (Habijan et al., 2019); dense connections between the hidden feature maps learned from the same dataset (Kanakatte et al., 2021); and the application of residual learning to the LASC data (Kausar et al., 2021). Though all made valuable contributions, the performance of each was largely consistent with the state of the art.

More significant modifications were made by Yang et al. (2018b). In their consideration of both MRI and CT portions of the MM-WHS data, they presented a hybrid loss function combining both frequency balanced CE, and generalised Dice score, within a deep supervision framework (Lee et al., 2015). However, perhaps more interestingly, they leveraged transfer learning (Weiss et al., 2016), initialising their network with a parameter set pre-trained on a video recognition task. Transfer learning from the 2D domain of natural images to 3D medical imaging is normally precluded by the inherent mismatch in dimensionality between the two. Overcoming this barrier by depending on the features associated with video recognition (with its two spatial (2D) and single temporal (T) dimensions) makes for a simultaneously sensible but intriguing choice. Unfortunately, however, they fail to perform the ablation experiments that might expose the performance gains associated with their approach. Instead, they demonstrate the improvement conferred by their hybrid loss (its formulation shared by the state of the art nnU-Net framework (Isensee et al., 2021)), comparing against the Dice loss alone. In closely related work (albeit limited to the CT partition of the MM-WHS data), similar findings were made by Ye et al. (2019) who, in extending the focal loss (Lin et al., 2017) to the multi-class setting, demonstrated its marginal superiority over the hybrid loss presented by Yang et al. (2018b). This conclusion, however, is obfuscated by their coincident modification of the 3D U-Net architecture, including dense connectivity (similar to the fractal expansion described by Larsson et al. (2016)) at each level of the expanding and contracting paths.

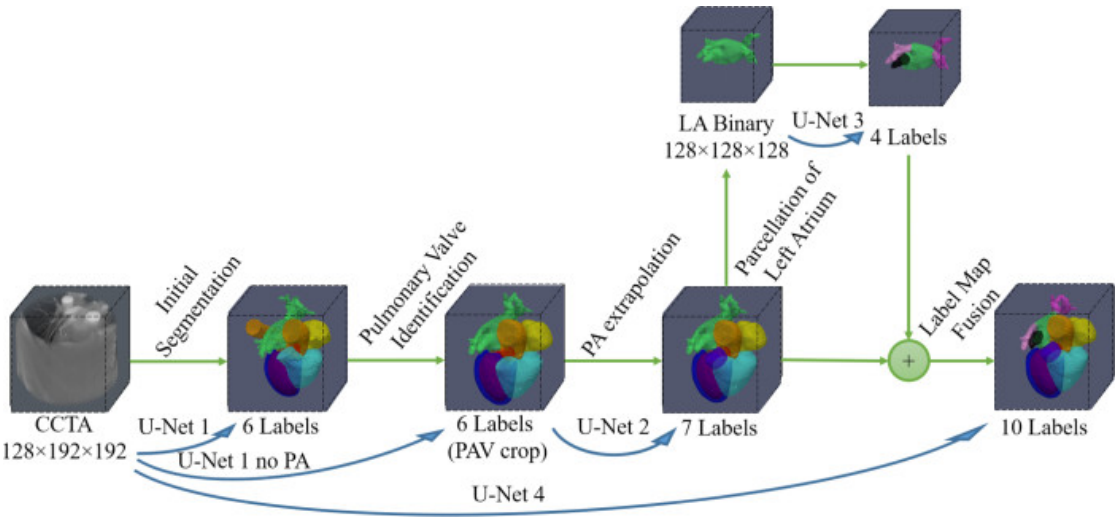


Figure 4.20: An overview of the segmentation and incremental enrichment of the associated multi-class description of the Scottish Computed Tomography of the Heart (The SCOT-Heart Investigators, 2015) (SCOT-HEART) dataset, achieved by Xu et al. (2021).

Illustrated in Figure 4.20, Xu et al. (2021) present an interesting application in which a series of 3D U-Nets are used to initially label, and then enrich the multi-class segmentation of the 1770 cases comprising the Scottish Computed Tomography of the Heart (The SCOT-Heart Investigators, 2015) (SCOT-HEART) CT database. Starting without any labels at all, they first apply the method developed by Zheng et al. (2008), based on marginal space learning and steerable features (see Section 3.3.3), to segment six structures (four cardiac chambers, the left ventricular myocardium and ascending aorta) from every volume. After refinement, including manual identification of the pulmonary valve to terminate the right ventricular outflow, these were used to train an image-to-label (conventional) 3D U-Net. In smaller subsets of the SCOT-HEART data, manual segmentations were then made to enrich the description of anatomy returned by Zheng et al. (2008) so as to: (1) include the main pulmonary artery (distal to the valve); and (2) isolate the pulmonary veins from the body of the left atrium. Each new label set was used to train a label-to-label 3D U-Net, the resulting models being applied to the remaining cases of the database. Finally, the results from a manually verified subset of 260 of these predictions were used to train a 3D image-to-label U-Net

capable of delineating not only the classes catered for by Zheng et al. (2008), but also the enriched description incrementally achieved through their careful manual curation. This work is a great example of human-in-the-loop machine learning, as reviewed by (Budd et al., 2021).

Next we review a pair of publications, sharing a distinctive interest in the segmentation of multi-class anatomy from dual energy CT scanning. Conventional cardiac CT is typically reliant on exogenous, iodinated agents to generate contrast between the cardiovascular blood pool and associated tissues. In the absence of pharmacological contrast, non-contrast-enhanced studies remain a highly challenging target for segmentation. However, through its sensitivity to photon energy and from a single, contrast-enhanced scan, dual energy CT admits the synthesis of virtual non-contrast images.

Bruns et al. (2020) leveraged this quality in pursuit of a solution to the segmentation of conventional non-contrast-enhanced data. They curated a manually labelled training set of eighteen dual energy, contrast-enhanced CT datasets. For each, the resulting segmentation could be propagated to its associated virtual, and conventional non-contrast-enhanced counterparts by identity mapping, establishing a dataset for supervised training. They assessed performance using a six-fold cross-validation, quantitatively determining that a 3D U-Net could be optimised to successfully segment both virtual and conventional non-contrast-enhanced scans. In a qualitative analysis of a further 290 conventional non-contrast-enhanced scans, subjective assessment found predicted segmentations to be mostly accurate. Lartaud et al. (2021) leveraged the same principle within a controlled experiment comparing different augmentation schemes.

Finally, we review a series of works that, inspired by conventional atlas-based segmentation, rely on prior representations of anticipated cardiac anatomy to constrain prediction. Work presented by Dong et al. (2020) is closely related to TeTrIS (Lee et al., 2019). However, where TeTrIS train a CNN to learn the parameters of a single freeform deformation, Dong et al. (2020) decouple atlas transformation into affine and non-rigid components. They combine their approach with a single atlas to segment the left ventricle from 3D echocardiography. Comparatively, Ding et al. (2020) integrate CNN operation within a 3D patch-based multi-atlas framework (see Figure 4.21), training networks to learn: (1) the non-rigid trans-

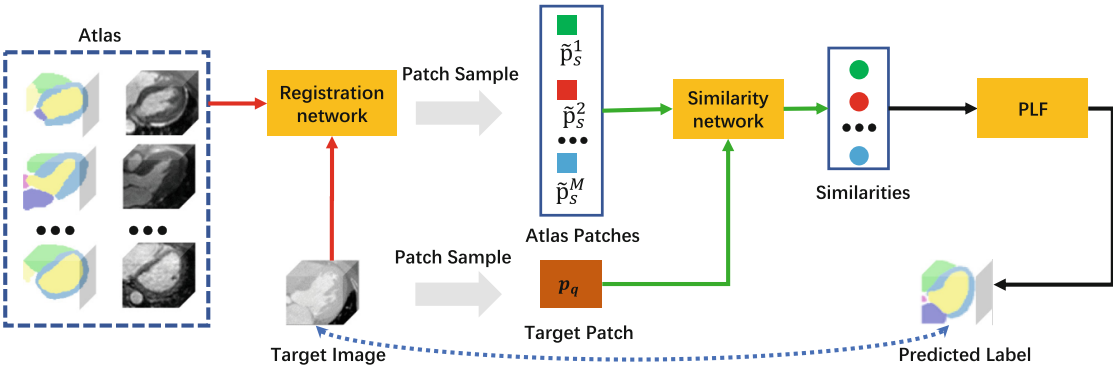


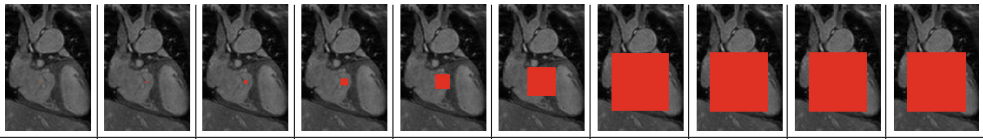
Figure 4.21: An overview of the CNN-based multi-atlas segmentation framework (including patch-based label fusion), presented by Ding et al. (2020).

formation between input and atlas patches; and (2) patch similarity according to separation within a compact latent space. Using a cross-over design in which the twenty CT (CMR) images of the MM-WHS data serve as multi-atlas for the remaining twenty CMR (CT) volumes, they assess their methods on myocardial segmentation. Whilst far superior to conventional multi-atlas segmentation, their approach remained consistent with the CNN-based state of the art; although its dependence on an atlas prior reduced the number of anatomically spurious predicted features. Lastly, Sinclair et al. (2022) present an involved framework for joint atlas construction, registration and segmentation. They rely on 3D U-Nets for semantic segmentation and to regress the parameters of atlas transformation, leveraging a large but commercially protected dataset.

4.4.2 Patient-specific modelling of CHD

Adapting training schemes to 3D segmentation

Owing to its dependence on the availability of training examples, accounts of the CNN-based segmentation of 3D CHD anatomy have been dominated by works considering the Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease (Pace et al., 2015) (HVSMR) dataset (Pace et al., 2015). Its total of just twenty cases (ten for which the labels are made public and ten for which the ground truth is concealed), labelled with only two classes (the whole heart blood pool and biventricular myocardium), might appear to preclude



Layer	1	2	3	4	5	6	7	8	9	10
Convolution	3×3	3×3	3×3	3×3	3×3	3×3	3×3	3×3	1×1	1×1
Dilation	1	1	2	4	8	16	32	1	1	1
Field	3×3	5×5	9×9	17×17	33×33	65×65	129×129	131×131	131×131	131×131
Channels	32	32	32	32	32	32	32	32	192	3
Parameters	320	9248	9248	9248	9248	9248	9248	9344	6912	579

Figure 4.22: The ten layer, dilated CNN presented as a solution to the HVSMR Challenge by Wolterink et al. (2017). Red shading indicates the growing receptive field.

substantive methodological investigation. Despite these limitations, a wealth of research has been conducted. We have already met such examples: with reference to Table 3.5, the following citations can be compared.

Of those approaches rejecting a single 3D CNN solution in favour of an adapted or simplified training scheme (at least compared with studies of normal anatomy or that affected by acquired disease) we find a reduced dependence on the decoupled learning of localisation and then segmentation in a coarse-to-fine cascade[8]. This is likely a consequence of the way data are provided by the HVSMR Challenge organisers. In addition to full volumetric data, they provide cropped images, determining a ROI tightly bounding the heart and great vessels. Hence, a prior localisation step is redundant.

Instead, exponents of adapted learning schemes have focused on the application of 2D CNNs or their combination with 3D networks. In their submission to the original HVSMR Challenge, Wolterink et al. (2017) train a ten layer CNN relying on 2D dilated convolutional operators for feature extraction (see Figure 4.22). As argued, these realise an exponential growth in the network's receptive field (ultimately covering a square region of 131 × 131 voxels), for only a linear growth in the number of parameters. They train their network for each of axial, sagittal and coronal planes, making a triplanar prediction by averaging pixel-wise probabilistic predictions. Compared with their contemporary competitors, they achieve

[8]Deploying a coarse-to-fine cascade the approach described Han et al. (2020) presents a notable, but flawed counter example. Unfortunately the authors misconceive the HVSMR segmentation task, training a CNN-based solution to label the whole heart blood pool and myocardium as a single, binary foreground. We are not inclined to discuss this work further.

excellent quantitative performance, leading the standings when ranked by spatial overlap for the whole heart blood pool and myocardium classes. They are only beaten into second place due to the perceived superiority of the winner's fully 3D convolutional implementation, a relative novelty at the time. Amongst their qualitative findings, they observe the network's capacity to learn extended features capturing long range spatial coherences, presumably by virtue of its expansive receptive field. Perhaps with the benefit of hindsight, unlike the HVSMR Challenge organisers, we consider the relative simplicity of their approach a strength of this work.

In contrast, Du et al. (2020) present a far more complicated 2D architecture with a focus on multi-scale feature learning and integration. It successfully includes: an inception module; a dilated residual network including a spatially unrolled convolutional long short-term memory (LSTM) (Shi et al., 2015) module[9]; and a hybrid pyramid pooling network. In a series of sound arguments, they motivate and justify this level of complexity in response to the challenges of CHD segmentation: including the delineation of complex and variable anatomy, possibly defined in relation to blurred tissue interfaces. However, it is unclear to us how each of their network components contribute to this aim. In particular, whilst the principle of leveraging a spatially unrolled recurrent neural network (RNN) to learn global features is sound, we are left uncertain as to whether this can be adequately achieved within a 2D network. We feel that such global features of anatomy are meaningfully defined in 3D, rather than their appearance in 2D cross-section. Though we raise this concern, their approach achieves strong empirical performance, albeit limited to a leave-one-out cross-validation (train on nine, test on one) on the HVSMR training set. Comparison of their spatial overlap results with those of Wolterink et al. (2017) are made challenging by divergent experimental designs, the latter performing a five-fold cross-validation to train on eight and test on two. Allowing for this difference, the two achieve similar spatial overlap performance as measured by the DSC, (Wolterink et al. (2017), Du et al. (2020)): (0.92,0.946) and (0.80,0.824), for the blood pool and myocardium, respectively.

[9]The LSTM cell constitutes the basis of a recurrent neural network (RNN), a variety of neural network designed to model sequential data. Due to its infrequent application within image segmentation and for brevity, we have not dwelt on in this thesis. For an indicative work, we direct the reader to Chen et al. (2016).

Given the small dataset, Wolterink et al. (2017) trained with an eighth fewer examples than Du et al. (2020). Hence, the complexity of the approach presented by the latter, though undoubtedly making a novel contribution, may not be as significant as might first appear.

More generally, both these works highlight a weakness shared by the majority of works presented in this application-focused section of our review. That is, technical metrics of CNN-based segmentation performance (for example, the DSC and Hausdorff distance (Huttenlocher et al., 1993)) are rarely sensitive to the features most pertinent to the clinical applications of 3D models of patient-specific anatomy. In part, this constraint is imposed by the HVSMR Challenge submission system, in which the ground truth labellings of the test set are kept private, preventing researchers from considering their own, perhaps more clinically focused metrics of performance (unless they are willing to experiment with only the ten public cases made available for training).

A clear solution to this challenge is to leverage alternate data for training and testing. In commendable work, Nurmaini et al. (2020) curate a collection of exemplar 2D foetal ultrasound images. Though they fail to detail the number of patients represented, by considering the temporal frames of the resulting videos separately, they achieve a total of 764 raw images, evenly distributed between foetal patients exhibiting a range of septal defects (atrial septal defect, ventricular septal defect, atrioventricular septal defect and normal septal isolation). Leveraging these data, they train a 2D mask R-CNN, (He et al., 2017a) for simultaneous cardiac ROI detection and multi-class instance segmentation of the four cardiac chambers, aorta and any septal defects. Critically, their evaluation metrics consider the spatial overlap performance for predicted segmentations of each variety of septal defect. Their results are encouraging, supporting the extension of their methods to incorporate a range of standard views.

Similarly clinically focused work presented by Nova et al. (2021) is also highly promising. Albeit within a binary segmentation of the four cardiac chambers from 2D echocardiography, they infer the presence of septal defects through the pixel adjacency relationships of the predicted labels. Despite the attractive principles of their approach, their experiment is let down by its small dataset (including two patients with each of atrial septal defect, ventricular septal defect, atrioventricular

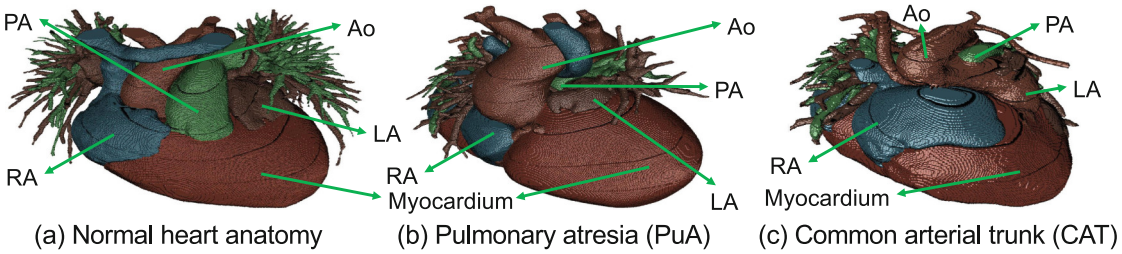


Figure 4.23: Example segmentations from the ImageCHD dataset. Reproduced from Xu et al. (2019b).

septal defect and normal septal isolation; for a total of just eight cases) and the fact that they make no attempt to understand how or whether their pixel-wise metrics translate into the clinically relevant detection of defects.

Another alternative dataset supporting a string of publications is provided by the CT volumes collected within ImageCHD (see Figure 4.23). Highly relevant to our interests, patient-specific 3D printing motivated the publication of its first 68 cases and associated experimental work (Xu et al., 2019b). Distinct from the HVSMR task, ImageCHD includes a multi-class ground truth, separating a wide range of CHD anatomy into seven different labels. Since this time, this highly impressive publicly available dataset has grown to include 110 scans (Xu et al., 2019a). These examples have been leveraged in a series of related works, all proposing adapted training strategies based on 2D CNNs and variably their combination with 3D architectures. This group of publications shares a common interest in the incorporation of spatially extended or global features via graph reasoning.

Adopting a purely 2D solution, Liu et al. (2020b) deploy a U-Net with asymmetric convolution (Ding et al., 2019). At the low-resolution bottleneck, they introduce a graphical CNN to learn features relevant to the global spatial coherence of anatomy. Though intrigued by their base architecture, we feel that the incremental value conferred by their graphical CNN might be limited by attempting to learn the global features of 3D anatomy by their representation in 2D slices. Finally, though Liu et al. (2020b) claim to leverage prior information concerning anatomical shape, this knowledge is explicitly captured within a ground truth segmentation of the data and hence unavailable at test time. Hence we feel a better description might characterise their, no less valuable contribution as a spe-

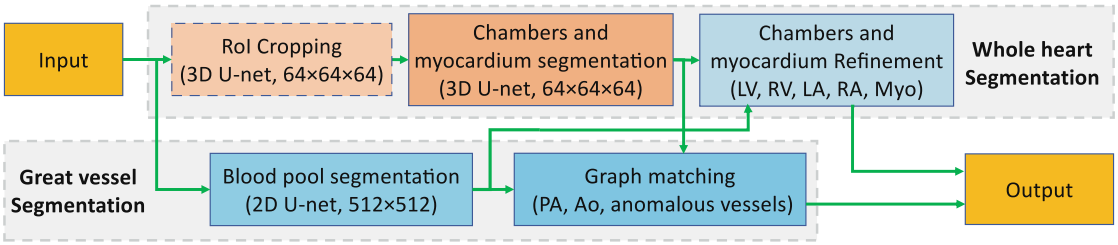


Figure 4.24: Overview of the segmentation framework first presented by Xu et al. (2019b).

cialised loss function. More problematically, their experiments are made difficult to understand, the authors neglecting to describe their test set.

This uncertainty is reduced in the work of Xu et al. (2019a) and Xu et al. (2020). In each publication, the curators of the ImageCHD dataset rely on the same methodology, combining both 2D and 3D CNNs within a highly engineered workflow, incorporating heuristics based on domain knowledge and graph reasoning (see Figure 4.24). Their approach isolates multi-class and blood pool segmentations to low resolution 3D and high resolution 2D U-Nets, respectively. Subsequently, the interfaces of the whole heart blood pool class, learned at high 2D resolution, are used to refine the coarse 3D segmentation of multi-class anatomy. This "connection analysis" step takes place according to relatively rudimentary heuristics based on morphology operations and pixel adjacency. Their separate "shape analysis" receives inputs from the respective U-Nets, applying Boolean operations to extract, and then skeletonisation (Lobregt et al., 1980) to graphically represent, the great vessels. Following, they compare the resulting vessel graph with a template library (determined from the training data). Though a little unclear, our assumption is that subsequently, pertinent features of the template are propagated to the predicted segmentation.

In their former work Xu et al. (2019a) apply this pipeline in aid of segmentation, achieving qualitatively impressive results. Quantitatively, their assessment is limited to spatial overlap. Though this makes it difficult to understand the extent to which clinically meaningful features are represented (for example the delineation of congenital defects), they make the highly relevant observation that performance degraded in the presence of severe structural malformation. In their later work

Xu et al. (2020) deploy their pipeline in aid of diagnostic classification, inferring the presence of defects according to the adjacency of predicted multi-class labels (for example, a ventricular septal defect is characterised by neighbouring voxels being classified as each of left and right ventricles). We feel that coupled with the engineered heuristics inherent within their approach (for example, they require great vessels be disconnected, presumably even in the context of aortopulmonary window), their diagnoses are susceptible to segmentation error, even at the level of an individual pixel. Moreover, we are not convinced by the principle that segmentation should precede and inform diagnostic classification. At least in our local practice, by echocardiogram and related clinical examination, congenital diagnoses are determined far in advance of tomographic CT or CMR acquisition. Hence, we suggest that a more appropriate semantic hierarchy understands segmentation as a means of describing and localising the structural characteristics of a more fundamental diagnosis. The latter might even serve as prior information to inform anatomical labelling (see Chapter 7).

CNN architectures dedicated to end-to-end 3D segmentation

Before a more detailed discussion of end-to-end, 3D CNN-based solutions to the segmentation of CHD anatomy, we first list a group of publications which, though performing experiments using the HVSMR dataset, are not motivated by its clinical context. Rather, these reports seek generalised improvements in CNN methodology without specific focus on CHD or even medical imaging. In such cases, the publicly available HVSMR data merely provide a convenient basis for assessing the performance of: dense networks incorporating spatial and channel attention modules for semi-supervised learning (Min et al., 2019, 2020); a universal decoder for CNN-based segmentation (Liang et al., 2019); novel ensembling strategies (Zheng et al., 2019b; Ma et al., 2021); sparse image annotation and associated learning (Zheng et al., 2020); and contrastive learning for semi-supervised training and transfer learning (Zeng et al., 2021). Whilst some of these achieve strong spatial overlap performance, none stand out from Table 3.5 to exceed the state of the art. Nor do they contain discussion dedicated to the segmentation of CHD anatomy.

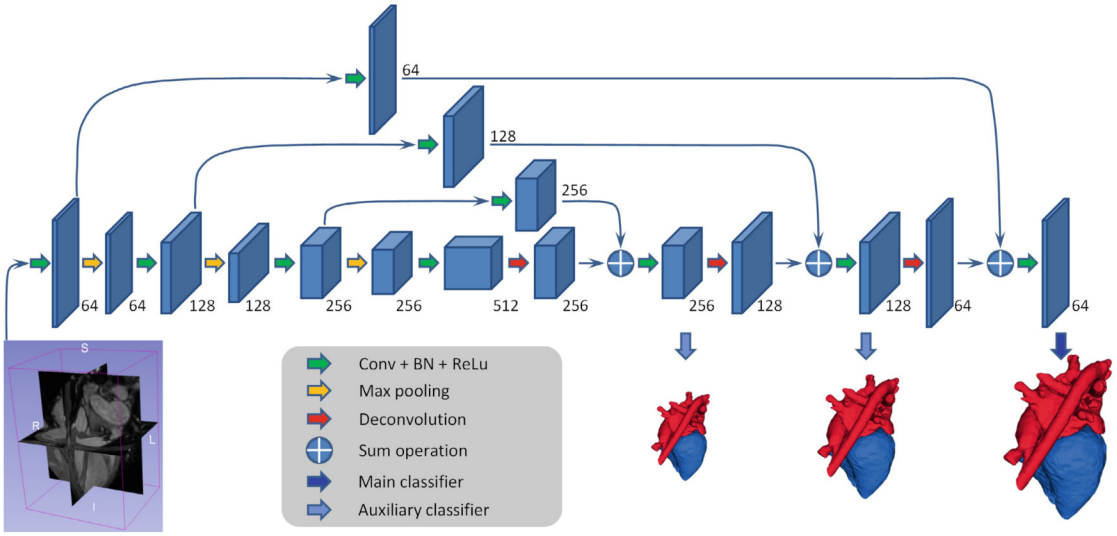


Figure 4.25: The U-Net-like architecture including fractal connectivity submitted to the HVSMR Challenge by Yu et al. (2017b).

Two submissions to the HVSMR Challenge made use of 3D CNNs trained end-to-end. Chief amongst these, Yu et al. (2017b) deployed a U-Net-like architecture, including fractal connectivity (Larsson et al., 2016) (see Figure 4.25). Ultimate winners of the Challenge, they achieved Dice scores of 0.931 and 0.786 for the segmentation of the whole heart blood pool and myocardium, respectively. Preferring an FCN architecture, Li et al. (2017b) presented a more complex network that also included 3D dilated convolution. Their approach, however was outperformed by the remaining CNN-based solutions, and two submissions that deployed conventional atlas-based segmentation. Lastly, though not a submission to the original Challenge, Dou et al. (2017) offered an alternative 3D FCN, achieving results to rival Yu et al. (2017b). All three leveraged deep supervision, directly injecting gradient information at the various spatial scales probed by their networks.

The record HVSMR performance (at least for a 3D CNN) achieved by Yu et al. (2017b) did not stand long. In the same year, the same authors employed a densely connected 3D architecture that they called DenseVoxNet, including max pooling for feature learning at multiple spatial scales (Yu et al., 2017a). They attribute their superior performance to the advantages of dense connectivity, including improved gradient flow and parameter-efficient feature reuse.

Densely connected 3D architectures also proved attractive to: Ran et al. (2018), within their dense U-Net; and to Zhang et al. (2019b), who trained two DenseVoxNets for each HVSMR class, before a third was used to combine label-wise predictions. Perhaps notably, and at least when ranked by spatial overlap, the currently best-performing 3D CNN solution also includes dense connectivity. After parallel asymmetric convolution, Zheng et al. (2019a) aggregated triplanar features using a dense module. Our final exponent of densely connected CNNs, we cite work recently published by Nainamalai et al. (2022). At least outwardly, this article appears particularly pertinent to our clinically focused interests, promising an account of the clinical integration of CNN-based segmentation for 3D visualisation of CHD anatomy. Disappointingly, in assessing their dense V-Net architecture, their test set is limited to just six local cases. Though they provide a diagnostic and demographic description of each, it seems unlikely that such a limited sample could be representative of the structurally heterogeneous CHD population. Allowing for the fact that their manuscript remains a pre-print, we also find it problematic that they do not attribute the publicly available training dataset relied upon (including 66 cardiac CT scans) to a specific source. Though tempting to speculate that these cases might be related to the 68 labelled volumes provided in the first tranche of the ImageCHD dataset (Xu et al., 2019b), we cannot be certain.

Another work depending on alternative training data to those provided by the HVSMR Challenge, Giannakidis et al. (2016) curated a labelled set of isotropic CMR volumes, segmenting the right ventricle in each. Given the patient population (all patients exhibiting tetralogy of Fallot), their focus represents a clinically motivated choice, changes in the volume and structure of the right ventricle being critical to ongoing management. Moreover, substantial variation in morphology, including that associated with pathological remodelling or previous interventional modification, makes for a particularly challenging task. They train a DeepMedic architecture in a two-fold cross-validation, achieving a sizeable average absolute volume difference of 12%. In a faithful analysis, they stress their ambitions to improve their quantity of training data.

To conclude our review, we present three papers, each made distinctive by their unconventional, but clinically focused approach. In the first of these, Gilliland et al. (2022) address the segmentation of the congenitally malformed foetal circulation

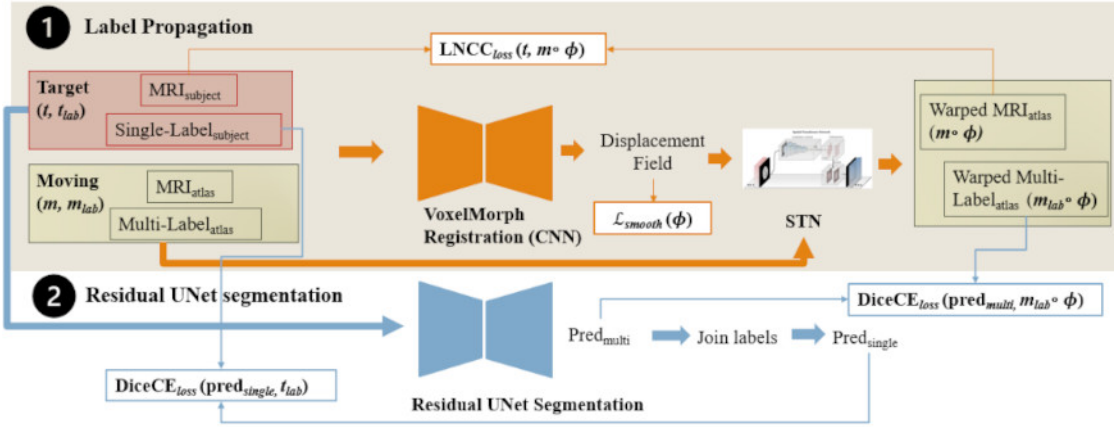


Figure 4.26: The combined framework for CMR registration, segmentation and label enrichment of a multi-class representation of the foetal great vessels. Reproduced from Gilliland et al. (2022).

from 3D T2-weighted CMR data. Its staggering application aside, this work elegantly combines atlas-based segmentation and label propagation to maximise the information returned by clinically routine image post-processing (see Figure 4.26). In a first step, they train the U-Net-like 3D VoxelMorph architecture (Balakrishnan et al., 2019) to infer the displacement field necessary to warp a multi-class foetal atlas of the great vessels and associated head and neck vasculature to a target CMR volume. Leveraging the paired CMR-multi-class labels that result, they subsequently train a 3D residual U-Net to learn the associated mapping, optimising against two CE losses between: (1) multi-class, U-Net prediction and warped atlas; and (2) the union of predicted multi-class labels, and a binary ground truth, delineated within the clinical routine. In so doing, they propagate the information contained within a *single* multi-class atlas, enriching the features learned by 3D CNN; at the same time they ensure that predictions are optimised against a *patient-specific* representation of geometry, captured within the data to which they have access, the clinically prepared, binary ground truth.

In our own work (Byrne et al., 2019), we associate congenital defects with changes in the topology of the cardiac blood pool. Conventional schemes for data augmentation rely on spatial transformation and image resampling. We observed that when these operations are applied to discrete label maps, changes in ground truth topology, including splits and mergers of the cardiac blood pool, can result.

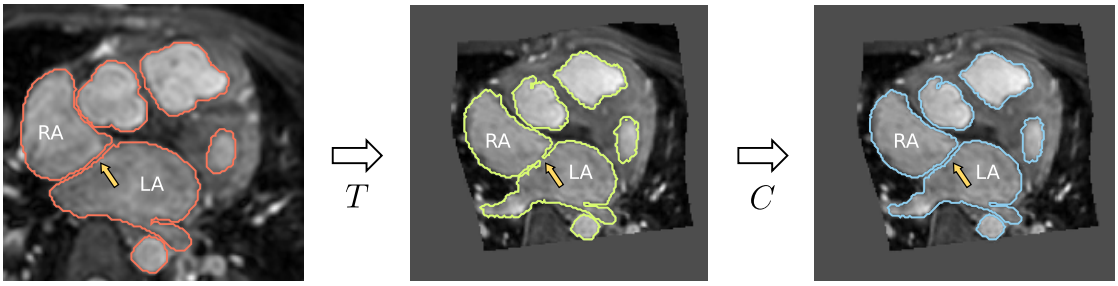


Figure 4.27: Adapted from Byrne et al. (2019), in which we implemented an augmentation pipeline to preserve the topology of manually segmented training labels provided by the HVSMR Challenge. This was composed of two steps: (1) the original images and labels are spatially transformed (T); and (2) the labels are subsequently topologically corrected (C). In this case topological arguments preserve the appearance of the atrial septum, where conventional spatial transformation and resampling falsely indicate the presence of an ASD (see arrows).

Though the result of image resampling, such errors might be mistaken for clinically relevant defects: splits being topologically equivalent to discontinuities in the circulation (such as atretic defects); mergers being topologically identical to associative, extra-anatomical communication (such as septal defects). Engineering a spatial transformation pipeline to constrain digital topology (Kong and Rosenfeld, 1989), we ensured that the congenital defects exhibited by the HVSMR training labels could be faithfully transmitted to their augmented representations (see Figure 4.27). Compared with conventional schemes, we found that when trained using our topology-preserving augmentation scheme, a V-Net architecture made fewer topological warping errors (Jain et al., 2010) in a five-fold cross-validation. Though having little to no impact on spatial overlap performance, this approach suggested an alternative, topological metric of performance, one relevant to the most clinically pertinent features of CHD anatomy: the defects.

Sadly, we were unable to reproduce these findings in a study of larger data. The precise cause of this discrepancy remains uncertain. However, we can speculate as to the range of factors that might offer a partial explanation. A clear difference between the experimental setting of our conference paper (Byrne et al., 2019), and our attempts to reproduce the same findings within a larger study, concerns the data available for experiment. Our previous work employed the HVSMR dataset of just ten cases for each of training and testing. In contrast, our follow up experiment

enjoyed access to the Evelina London Children's Hospital (ELCH) dataset of 150 examples (a contribution of this thesis presented in Chapter 5), including fifty for testing. We therefore speculate that the findings made within the smaller HVSMR test set were compromised by their limited statistical power.

Differences in data extend to the quality of the training sets. Later, Figure 5.7b will indicate that the ELCH cohort contains a broader range of structural variation than the HVSMR dataset. Hence it is possible that despite the topologically accurate augmented representations returned by our pipeline, the structural heterogeneity of the ELCH dataset hampered associated feature learning. In contrast, and in the context of the relatively homogeneous HVSMR training labels, the consistent appearances (perhaps in respect of their size, morphology and localisation to regions of the heart or image volume) of salient topological features could firstly be learned more readily, and then more reliably, where their accuracy was ensured by our augmentation pipeline. These differences might explain our failure to reproduce previous findings (Byrne et al., 2019) within the large ELCH dataset.

Finally, we cite work published by (Pace et al., 2018), the organisers of the original HVSMR Challenge. Having access to all twenty CMR volumes (ten for each of training and testing), they modified the associated labels to isolate the aorta and left ventricle separately. Methodologically, rather than through direct CNN prediction, they reformulated respective binary segmentation tasks within a framework for recursive refinement. Given an incomplete segmentation, they trained a 3D U-Net to infer the next most likely binary label, using a recurrent relationship to incrementally evolve prediction. Motivating this approach, they appeal to the advantages of conventional active contours, including topological constraint. Interestingly, they find their framework to be more robust to severe congenital malformation than existing methods for direct estimation.

4.5 Conclusion

This chapter has provided the theoretical and practical foundations necessary to comprehend the remainder of this thesis. By understanding the research findings that led to the development of the nnU-Net framework (Isensee et al., 2021) and reviewing its competitors, we have established its underlying U-Net architecture as

both baseline (for performance comparison) and state of the art solution to CNN-based segmentation of medical images. In the experimental work presented in the following chapters, we seek to understand the compatibility of this approach with our motivating clinical interest: the construction of patient-specific, 3D models of anatomy, so as to inform the personalised care of patients with CHD.

The literature review presented in Section 4.4.1 found that 3D U-Net-like architectures performed strongly across a number of whole heart and multi-class cardiac segmentation tasks, and endorses our line of enquiry. Our more specialised (and sometimes critical) review (see Section 4.4.2) of the CNN-based segmentation of CHD anatomy was not so positive. In particular, though the majority of cited works claim clinical motivations including advanced 3D visualisation for treatment planning, rarely did they conduct experiments capable of understanding whether their methods might meet these ambitions. We identify three ways in which the clinical generalisability of reviewed findings might be improved, addressing each in the following chapters:

Data

The majority of works concerned with the CNN-based segmentation of 3D CHD anatomy rely on the HVSMR dataset (Pace et al., 2015). Given the wealth of research it has admitted, its contribution cannot be in doubt. However, we do recognise that it is severely limited in its *quantity*, including only ten labelled CMR volumes for training and a further ten unlabelled, for testing. We do not imagine that such a small sample could represent the underlying and structurally heterogeneous CHD population. This limits the generalisability of both learned features, and of performance assessment.

Applications of the HVSMR training data (and any representation of anatomy that can be learned from them) are also limited by the semantic *quality* of the provided labels. Namely, the task specification presented by the HVSMR Challenge seeks the segmentation of data into a unified blood pool class. Neglecting multi-class labels limits our description of pathology-induced changes that might be localised to individual cardiac sub-structures. Perhaps even more importantly, being that they are often defined by the relationship between two or more car-

diac segments (consider a ventricular septal defect, for example, characterised by the communication or isolation of left and right ventricles), descriptions of the whole heart blood pool are ignorant of congenital defects without more careful consideration.

In these respects, the ImageCHD dataset is far superior to HVSMR. However, it contains only CT volumes. Free from ionising radiation, isotropic 3D CMR is a particularly effective imaging modality for the paediatric CHD population (Ntsinjana et al., 2011), one that presently, has no publicly available source of segmented, multi-class training data. Unfortunately, amassing annotated CMR volumes remains both a significant challenge (having been established in Chapter 3) and motivation for academic research.

In response others have considered alternatives to manual curation. Koehler et al. (2020) tested whether U-Net models trained with data reflective of acquired heart disease might faithfully segment the short axis scans of patients with repaired tetralogy of Fallot. They found that learned features did not generalise well between pathologies. Alternatively, Karimi-Bidhendi et al. (2020) took a generative approach to data augmentation, synthesising training examples from a learned distribution of short axis data acquired from 64 paediatric patients with complex CHD. Though their approach was associated with improved performance, their augmented inputs remain closely tied to their initial sample.

1. These arguments outline the need for training data of improved *quantity* and *quality*, as required to support CNN-based solutions to the patient-specific, multi-class segmentation of CHD anatomy from 3D CMR.

Quantitative performance assessment

Across all works cited in Section 4.4.1 and Section 4.4.2, performance is primarily assessed by spatial overlap and in particular by the DSC. Though an effective and easily interpreted metric of segmentation performance, inspection of Table 3.5 suggests that according to such metrics, the broad array of existing CNN method-

ologies are largely indistinguishable. Moreover, the DSC does little to characterise whether a predicted label set captures the clinically relevant features of data. Spatial overlap makes a pixel-wise assessment of segmentation agreement, all elements making an equal contribution to the statistic returned. However, in the presence of CHD, we are primarily concerned by the size, shape and location of defects, defined in relation to a smaller, localised portion of the image. As such, a predicted segmentation might share a strong spatial overlap with the ground truth, whilst failing to convey or describe the most clinically important features of the data.

Until performance metrics are better aligned with the downstream clinical requirements that motivate the underlying segmentation task, continuing to compare CNN-based methodologies using measures of spatial overlap does not appear an efficient way to gauge nor improve the state of the art. Assessed against a clinically relevant alternative, the ranking presented by Table 3.5 may be transformed.

These arguments motivate the design of:

2. performance metrics, and
3. loss functions

which expose the most clinically relevant features of image data, reflecting the extent to which congenital defects are captured and faithfully described by an accompanying segmentation.

Chapter 5

The ELCH dataset

5.1 Introduction

Chapter 2 motivated the design of automated solutions for the segmentation of congenital heart disease (CHD) anatomy from 3D cardiac magnetic resonance (CMR) data. Drawing on insights from the broader field of cardiovascular image analysis, Chapter 3 highlighted the recent success of data-driven methodologies, and their superiority over conventional approaches. The statistical models associated with deep learning, and most pertinently convolutional neural networks (CNNs), automatically learn a set of discriminative features through training (Litjens et al., 2017). The depth of these non-linear models admits the extraction of primitive image cues and their aggregation within abstract representations of data, such as anatomical geometry. Exploiting these high capacity models, however, assumes a large and representative set of training data, capturing the task at hand.

Many datasets (including the comparatively vast UK Biobank (Fry et al., 2017)), have been both established and successfully applied to CNN-based segmentation of short axis cine data (Chen et al., 2020). In contrast, equivalent resources relevant to the segmentation of multi-class anatomy from high resolution 3D CMR are scarce. Moreover, Chapter 4 suggested that training data germane to CHD anatomy and morphology are fewer still, being limited to a single sample of just ten training, and ten test CMR volumes (Pace et al., 2015). This number pales in comparison with the structural heterogeneity exhibited by the CHD population.

In this chapter, we seek to rectify this deficit, curating the unique Evelina London Children's Hospital (ELCH) dataset. We recount our methods of data collection and manual segmentation, and assess the demographic, imaging, diagnostic and anatomical characteristics of our patient cohort. Each is informed by a detailed consideration of the clinical and technical requirements of patient-specific anatomical modelling, for individuals with CHD.

5.2 Contributions

In this chapter we make the following contributions:

1. We present a dataset of 150, clinically acquired, 3D, isotropically high resolution CMR volumes:
   (a) Each is segmented into instances of eighteen different class labels relevant to cardiac anatomy, CHD morphology and associated structural intervention.
   (b) At the pixel level, every segmentation is clinically meaningful, indicating the communication or isolation of cardiac anatomy as relevant to the presence of congenital defects or structural intervention.
   (c) All cases are accompanied by the additional inclusion of gadolinium-enhanced, 4D time-resolved magnetic resonance angiography (TR-MRA).
2. We quantitatively characterise the operator burden of manual image segmentation associated with establishing this dataset, including duration and task complexity.
3. We characterise the heterogeneity of our dataset, reflecting the clinical status, demographics and anatomical features of each case, comparing these properties against existing, publicly available CMR training data.

5.3 Methods

5.3.1 Information governance

Our use of patient data was reviewed and approved by the UK Health Research Authority (HRA) (Integrated Research Application System (IRAS) ID: 273807, Research Ethics Committee (REC) reference: 19/HRA/6918).

5.3.2 Data collection

Data were collected retrospectively from ELCH and Guy's and St Thomas' NHS Foundation Trust electronic archives. All cases underwent clinically indicated CMR investigation for the assessment of CHD at ELCH between 2013 and 2020, including isotropically high spatial resolution, 3D steady state free precession (SSFP) and gadolinium-enhanced, 4D TR-MRA[1]. Eligibility was considered at two stages: (1) CMR report review; (2) Qualitative imaging review.

Suitable paediatric cases (less than eighteen years of age at scan) were initially identified from their clinical CMR report, building a sample of 150 cases stratified by predominant congenital diagnosis. Five equally sized diagnostic groups were considered: double outlet right ventricle (DORV), transposition of the great arteries (TGA), ventricular septal defect (VSD), hypoplastic left heart syndrome (HLHS) and hypoplastic right heart syndrome (HRHS). Under the lesser used label of HRHS, we collect patients exhibiting hypoplastic, stenotic or atretic defects of the right heart, including tricuspid valve or pulmonary valve stenosis or atresia and right ventricular hypoplasia. Patients co-exhibiting the following diagnoses were excluded: atrioventricular discordance or criss cross heart, atrial isomerisms or heterotaxy syndrome and dextrocardia or apex-to-the-right[2].

The design of this sample (including its size and diagnostic stratification) sought to balance two ambitions: (1) to collect a clinically representative group

[1]Strictly, angiography refers to a radiological study seeking to visualise (components of) the vascular system, most often via the administration of an exogenous contrast agent. Throughout this thesis, by using this term within our characterisation of 4D time-resolved magnetic resonance angiography (TR-MRA), we extend this definition to include visualisation of both blood vessels, and the blood pool contained by the heart's chambers.

[2]Whilst these cases have been excluded from this study, some have already been manually segmented as part of ongoing efforts to grow and broaden our dataset.

of patients; and (2) to limit the operator burden of manually segmenting all cases carefully and with high fidelity. Given the structural and demographic heterogeneity of the CHD population, the first of these aims is best met by increasing the sample size. For every additional patient considered, however, time invested in image segmentation increases, reducing the time available for experimental investigation. Accordingly, in balancing these competing demands, we settled on a pragmatic sample of 150 patients across five representative diagnostic groups.

For all groups: scan date, weight at scan, associated diagnoses and structural cardiac interventions were noted and subsequently corroborated at imaging review. Given current applications of patient-specific 3D modelling, those cases who had not undergone previous structural intervention were prioritised. However, after exhausting this purpose, patients from each group were randomly sampled, including those to have undergone biventricular repair, or partial or complete conversion to a palliative univentricular circulation.

Cases deemed eligible with respect to CMR report review were used to query the picture archive and communications system (PACS). Image quality was subjectively assessed for its consistency with the extraction of a clinically meaningful patient-specific model. This assessment was made by a clinical scientist with more than five years' experience in the segmentation, fabrication and clinical application of patient-specific 3D printed models of CHD anatomy. Judgement was made jointly, considering the 3D and 4D acquisitions of interest. In the event that dual phase 3D data were acquired, the diastolic volume was preferred. Accepted cases were finally anonymised and downloaded from PACS in Digital Imaging and Communications in Medicine (DICOM) format.

Prior to manual segmentation, and to accommodate spatial misalignments associated with respiratory motion, 4D data were registered to the coordinate space of the 3D SSFP volume. Registration was performed in two steps: (1) a cascade of temporally sequential non-linear transforms were used to co-register each temporal dynamic from the 4D series to the coordinate space of the final time step; and (2) further non-linear transformation was used to spatially register the aligned temporal dynamics with the 3D SSFP volume. All registrations were completed using NiftyReg (Modat et al., 2010).

5.3.3 Guiding principle of manual image segmentation

Understanding and representing the continuity of cardiac anatomy is critical to the utility of a patient-specific 3D model of CHD. In particular, an ideal model reflects the clinically meaningful communication between all chambers and associated vasculature. Throughout this work, we rely on voxel adjacency to infer the haemodynamic continuity between the sub-components of cardiac anatomy.

As illustrated by Figure 5.1, this principle permits the definition of clinically meaningful anatomical features. Adjacency of labelled voxels may indicate the continuity of blood flow across anatomical junctions (consider the atrioventricular valves, for example); or might specify the presence of a shunt or septal defect. Equivalently, non-adjacency excludes communication between anatomical structures, defining thin tissue interfaces such as the atrial septum; or implying atretic defects. It is through this principle that the segmented image gains clinically relevant meaning. Moreover, holding our ground truth segmentations to this standard guarantees the accurate delineation of defects, providing a faithful representation of their spatial extent, morphology and location.

Accordingly, manual segmentation required that communication between subcomponents of the cardiovascular blood pool be reflected at the pixel level, considering a 26-connected image grid[3]. This allows our segmented images to demonstrate the presence of congenital defects and interventional modifications, as specified by clinical CMR report. Whilst such a highly detailed segmentation may not be required by all the applications reviewed in Section 2.2.1, adopting this demanding formulation (and multi-class specification, see Section 5.3.4) makes our approach generalisable to the entire spectrum of use cases.

[3]Considering the 3D image as a close-packed lattice of cuboids, 26-connectivity associates all voxels that share a face, edge or vertex. This description admits the 6-connected grid of voxels that share faces; and the 18-connected grid of voxels that share faces and edges. For segmented objects to have meaningful digital topology (Kong and Rosenfeld, 1989), complementary (and reversible) connectivity patterns are defined (foreground, background): (26, 6) and (18, 6).

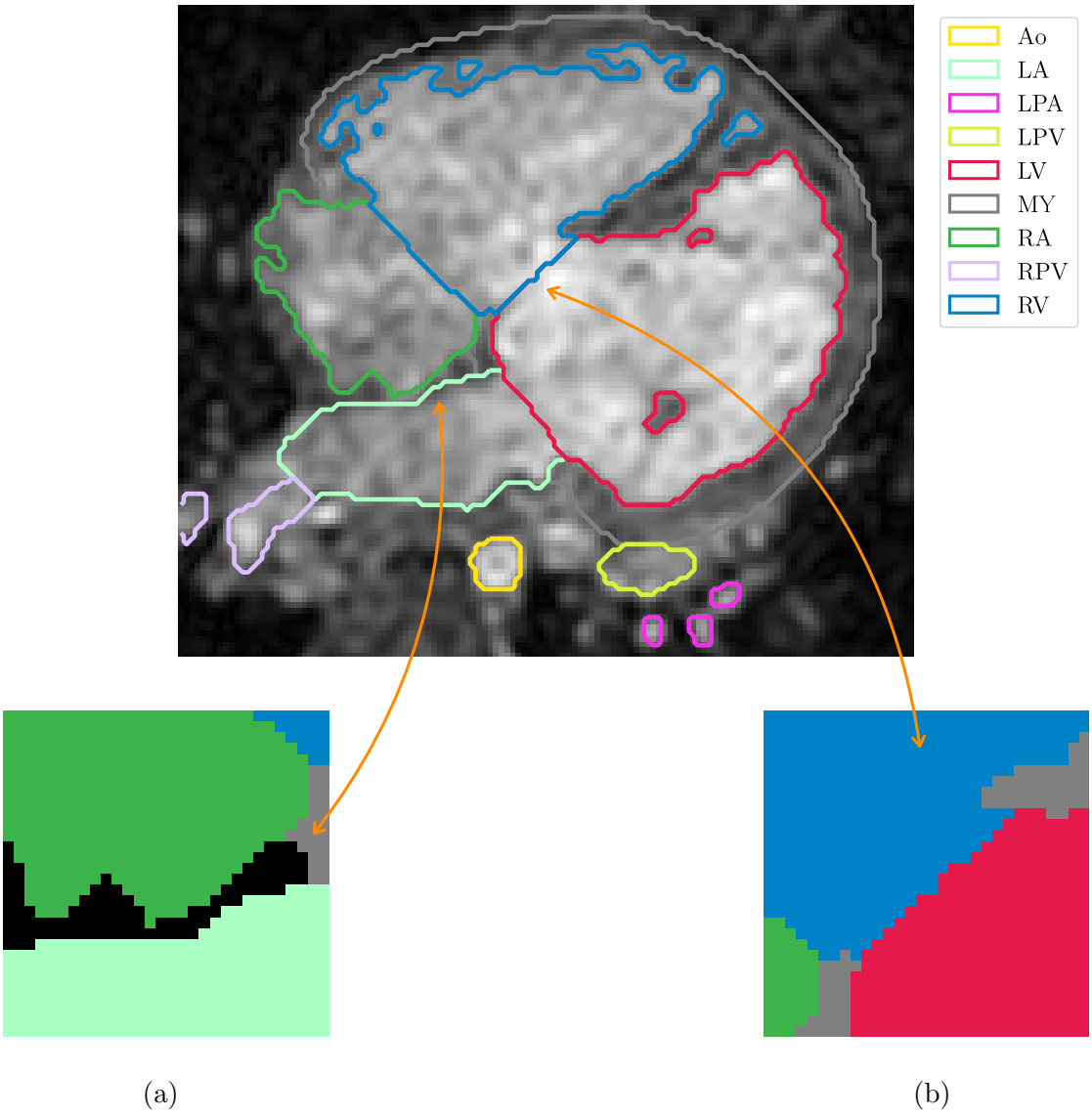


Figure 5.1: Throughout this work, we use voxel adjacency to indicate haemodynamic continuity. (a) Non-adjacency indicates the isolation of cardiac segments, such as the septum separating left and right atria. (b) Adjacency indicates the communication of segments, such as is found at valvular junctions, vascular confluence, or - as in this case - VSD.

5.3.4 Manual image segmentation protocol

All segmentations were completed using Mimics Medical Software (v18.0, Materialise NV, Leuven, Belgium) by the same clinical scientist referred to in Section 5.3.2. Segmentation sought a highly detailed representation of anatomy, being informed primarily by high resolution 3D data, but also considering 4D TR-MRA.

The imaging volume was separated into instances of eighteen different foreground classes, including: aorta (Ao), Blalock-Taussig (BT), Damus-Kaye-Stansel (DKS) connection, inferior vena cava (IVC), left atrium (LA), left pulmonary artery (LPA), left pulmonary vein (LPV), left superior vena cava (LSVC), left ventricle (LV), main pulmonary artery (MPA), ventricular myocardium (MY), patent ductus arteriosus (PDA), right atrium (RA), right pulmonary artery (RPA), right pulmonary vein (RPV), right superior vena cava (RSVC), right ventricle (RV), and total cavopulmonary connection (TCPC). Note that not all of these classes are present in all cases. In previous work, Pace et al. (2015) opted to terminate branching vascular structures proximally, citing a sensible desire to avoid obscuring cardiac anatomy. There, however, data were segmented into a combined blood pool class. In contrast, our multi-class formulation affords control over the visualisation of individual structures relative to one another. Therefore, we elect to segment the major branches of the pulmonary vasculature as can be confidently delineated within 3D and 4D data[4].

Practically, segmentation proceeded as follows. Initially, a crude representation of cardiac anatomy was achieved by semi-automated methods including thresholding, cropping and region selection. Ultimately, however, the standard described necessitated meticulous manual adjustment using multi-slice editing. More generally, the variety in both structural anatomy and image quality required a considered treatment, using the majority of editing tools made available in software.

During the course of segmentation, human-computer interaction (including mouse clicks and key presses) was logged using Mousotron activity monitor (Blacksun Software, 2021). Duration was also recorded. Due to the extended nature of the task, timing was automatically paused after idle periods of more than 10 s, allowing for accurate accounting in the event of distraction.

[4]TR-MRA often provides differential visualisation of branching pulmonary arteries and veins.

Unfortunately, given the operator burden of manual segmentation (see Section 5.4.4), repeated segmentation for the purpose of intra- or inter-observer variation was not practical. Each segmentation, however, was reviewed by an expert paediatric cardiologist with ten years' experience in the care of patients with CHD, including within CMR and the fabrication of patient-specific models of anatomy (Valverde et al., 2017a). Changes were made where indicated.

5.3.5 Comparative analysis

To contextualise the structural heterogeneity expressed by the ELCH dataset against the array of publicly available training data for cardiac image analysis, we complete a comparative analysis. Ideally this would contrast our cohort with publicly available data that were comparable in: patient population (CHD); image acquisition (isotropically high resolution 3D data); and segmentation task specification (multi-class anatomy). Though there exist two notable examples dedicated to CHD, neither is an entirely suitable reference standard.

The Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease (Pace et al., 2015) (HVSMR) Challenge data include only ten cases and segment 3D CMR into collective blood pool and myocardial classes (Pace et al., 2015); though larger ($n = 110$), ImageCHD computed tomography (CT) data are made available without physical pixel spacing, precluding any but dimensionless analyses (Xu et al., 2020). Whilst the 2017 Multi-Modality Whole Heart Segmentation (Zhuang et al., 2019) (MM-WHS) dataset also includes patients with tetralogy of Fallot, these are limited to seven of the twenty available CMR scans, without representation in the twenty CT acquisitions (Zhuang et al., 2019). More generally, quantifying the discrete structural changes associated with defective anatomy remains a significant challenge.

Accordingly and primarily, we frame our analysis in terms of ventricular volume. This presents the distinct advantage of broadening the pool of reference datasets, including those associated with short axis cine CMR. Thanks to its position as the gold standard approach for ventricular volumetry (Ruijsink et al., 2020), the segmentation of such images has received more machine and deep learning attention than any other cardiac image analysis task (Chen et al., 2020). Though

having different imaging characteristics, both short axis cine and spatially isotropic 3D SSFP seek to represent the same underlying ventricular anatomy. As a result, in addition to those outlined above, we also compare the anatomy of our ELCH cohort against: Sunnybrook (Radau et al., 2009), 2012 Right Ventricle Segmentation Challenge (Petitjean et al., 2015) (RVSC), Kaggle (Kaggle, 2016) Automatic Cardiac Diagnosis Challenge (Bernard et al., 2018) (ACDC) and 2020 Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (Campello et al., 2021) (M&Ms), short axis datasets.

5.3.6 Statistical analysis

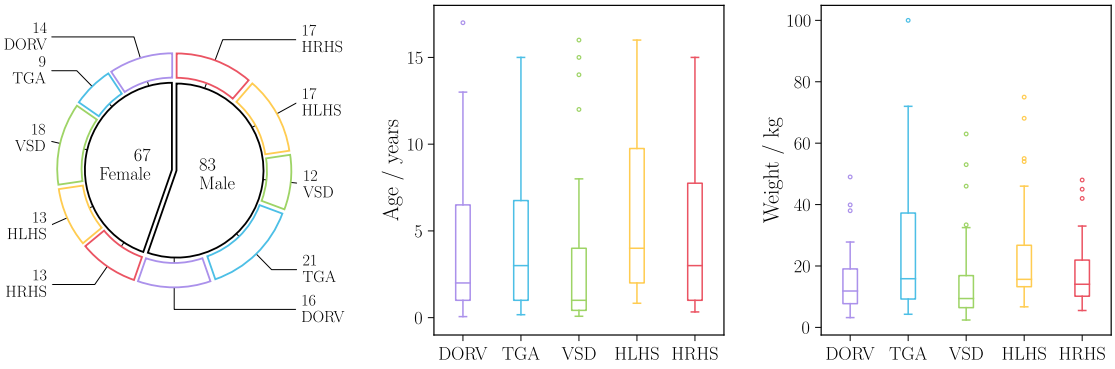
The previous sections invite statistical analyses to compare and contrast: the ELCH dataset with publicly available data; or its sub-groups. In Section 5.4.1, we check for significant demographic differences between diagnostic sub-groups using an omnibus and post-hoc testing procedure, using Kruskal-Wallis H-test, followed by Dunn test with Bonferroni correction. We check for gender disproportion across the ELCH dataset using a one-proportion z-test. To compare the anatomical characteristics of the ELCH cohort with those of publicly available data, in Section 5.4.5 we use Mann-Whitney U-tests with Bonferroni correction for multiple comparison.

5.4 Results

The ELCH dataset constitutes a cohort of 150 patients, reflecting a range of congenital diagnoses. Its ambition is to provide a snapshot of those cared for between 2013 and 2020. The following sections seek to characterise these patients: their disease; relevant interventional histories; CMR investigations; and their cardiovascular anatomy. In so doing, we reflect the heterogeneity of the CHD population.

5.4.1 Demographics

Figure 5.2 demonstrates the demographic characteristics of the ELCH dataset. By one-proportion z-test, there is no statistically significant evidence for an uneven



(a) (b) (c)

Figure 5.2: Demographic characterisation of the patients within the ELCH dataset, reflecting: (a) sex, (b) age at scan, and (c) weight at scan. Results are reported for each diagnostic group. Please see Section 5.3.2 for acronym definitions.

ratio of female to male patients (67:83, $p = 0.1914$, see Figure 5.2a). Given the small sample size ($n = 30$), we do not attempt sub-group analysis of categorical sex differences within each diagnostic category. In any case, we do not expect such differences to be clinically important.

More significantly, the ELCH dataset demonstrates a wide range in age (between three weeks and seventeen years) and weight (between 2.4 kg and 100 kg) at the point of scan. These intervals illustrate the heterogeneity of the paediatric CHD population. Figure 5.2b and Figure 5.2c suggest that the distributions of both age and weight are positively skewed, with the majority of included CMR investigations taking place during infancy.

Kruskal-Wallis H-test suggests statistically significant differences within the omnibus of diagnostic groups ($p = 0.0403$). Post-hoc Dunn test with Bonferroni correction reveals that the distribution of patient age significantly differs only between VSD and HLHS samples ($p = 0.0186$). At the time of scan, the median age of the HLHS patients exceeds that of the VSD group by three years. This reflects the different patient pathways followed in the management of these conditions. HLHS more frequently includes defects which would otherwise preclude life, necessitating urgent surgical, or hybrid intervention in the neonatal period and prior to CMR investigation. Diagnostic imaging is subsequently used to consider the timing of staged palliation or alternative intervention. On the other hand, the

spectrum of VSDs exhibited by this cohort allows for comparatively conservative management, and can be informed by CMR earlier in life.

This observation also accounts for the significant difference between the distribution of weight between HLHS and VSD patients ($p = 0.0123$ by the same omnibus and post-hoc analyses). As illustrated in Figure 5.2c, growth between the first and fourth years of life engenders a discrepancy in median patient weight of 6.25 kg: being 15.65 kg and 9.40 kg in HLHS and VSD samples, respectively. We are reassured that these demographic characteristics of the ELCH dataset reflect the distribution of patients encountered in the clinic and their management.

5.4.2 Congenital diagnoses, defects and interventions

In addition to considerable variation in age and weight, the heterogeneity of the CHD population is extended by an expansive set of structural defects, provoking deviation from expected cardiovascular anatomy. Whilst the presentation of such lesions is itself highly variable (see Section 5.4.5 for the impact upon anatomical distribution), defects are nominally reported for diagnostic and radiological descriptive purposes.

As determined from clinical CMR report, Table 5.1 breaks down the 600 cardiac malformations present at the birth of patients making up the ELCH dataset. By far the most common, ventricular and atrial septal defects account for almost one third of this total. Distinct from nearly all the remaining defects listed, septal defects occur in all five diagnostic groups. Conversely and to some extent, the frequency of most other defects correlates with the separation of cases into diagnostic subgroups. Consider the lateralisation of defects associated with HLHS and HRHS, for example.

The structural interventions sustained by the ELCH dataset are presented in Table 5.2. This suggests that at the time of scan, only 22 patients had no history of cardiac intervention. A further 35 cases had undergone intracardiac (such as atrial septostomy or septectomy) or extracardiac (such as MPA banding or BT shunt) intervention consistent with consideration for biventricular repair. Another 28 patients had undergone intracardiac (such as Rastelli or VSD patching) or extracar-

Table 5.1: Number of cardiac defects exhibited by the ELCH cohort at birth.

	DORV	TGA	VSD	HLHS	HRHS	**Total**
Septal defects						
Ventricular septal defect	31	24	26	6	24	**111**
Atrial septal defect	13	4	13	24	15	**69**
Membranous septal defect	2	3	7	0	6	**18**
Atrioventricular septal defect	2	0	1	3	0	**6**
Tetralogy of Fallot	1	0	2	0	1	**4**
Coronary sinus defect	1	0	0	0	0	**1**
Stenotic defects						
Sub-pulmonary stenosis	8	12	1	0	4	**25**
Pulmonary valve stenosis	4	8	1	0	5	**18**
Coarctation of the aorta	3	1	6	4	4	**18**
Mitral valve stenosis	2	0	2	10	0	**14**
Left pulmonary artery stenosis	5	4	2	2	1	**14**
Right pulmonary artery stenosis	5	2	4	0	2	**13**
Sub-aortic stenosis	3	3	1	0	0	**7**
Aortic valve stenosis	0	0	1	5	0	**6**
Main pulmonary artery stenosis	2	0	0	0	0	**2**
Pulmonary vein stenosis	1	0	1	0	0	**2**
Tricuspid valve stenosis	1	0	0	0	0	**1**
Atretic defects						
Mitral valve atresia	8	0	0	8	0	**16**
Aortic valve atresia	0	0	0	15	0	**15**
Pulmonary valve atresia	1	1	0	0	7	**9**
Tricuspid valve atresia	0	0	0	0	6	**6**
Interrupted aortic arch	1	0	0	0	1	**2**
Hypoplastic defects						
Hypoplastic left ventricle	10	0	0	22	0	**32**
Hypoplastic right ventricle	1	2	2	0	24	**29**
Hypoplastic aortic arch	4	2	4	1	2	**13**
Hypoplastic left pulmonary artery	2	0	0	1	2	**5**
Hypoplastic right pulmonary artery	1	0	1	1	0	**3**
Discordant defects						
Transposition of the great arteries	0	30	0	0	17	**47**
Double outlet right ventricle	30	0	0	0	0	**30**
Double inlet left ventricle	0	1	0	0	16	**17**
Double inlet right ventricle	1	0	0	0	0	**1**
Valvular defects						
Straddling mitral valve	3	1	1	0	0	**5**
Bicuspid aortic valve	0	0	3	1	0	**4**
Straddling tricuspid valve	0	2	1	0	0	**3**
Bicuspid pulmonary valve	0	1	0	0	0	**1**
Mitral valve cleft	0	1	0	0	0	**1**
Vascular defects						
Bilateral superior vena cavae	5	0	4	2	3	**14**
Patent ductus arteriosus	0	2	4	0	0	**6**
Aberrant right subclavian artery	0	1	2	0	1	**4**
Partial anomalous pulmonary venous drainage	2	0	0	0	0	**2**
Left superior vena cava	1	0	0	0	0	**1**
Total anomalous pulmonary venous drainage	0	0	0	0	1	**1**
Aortopulmonary window	0	1	0	0	0	**1**
Left pulmonary arteriovenous malformation	0	0	0	1	0	**1**
Anomalous left pulmonary artery	0	0	0	0	1	**1**
Right aortic arch	1	0	0	0	0	**1**

Table 5.2: Number of interventional modifications exhibited by the ELCH cohort at CMR acquisition.

	DORV	TGA	VSD	HLHS	HRHS	**Total**
No intervention	4	1	14	0	3	**22**
Biventricular surgical repair						
Arterial switch	1	12	0	0	0	**13**
Ventricular septal defect closure	0	8	2	1	0	**11**
Atrial septal defect closure	0	7	1	0	0	**8**
Rastelli	4	1	0	0	0	**5**
Rastelli with pulmonary conduit	0	2	0	0	1	**3**
Atrial switch	0	2	0	0	0	**2**
Arterial switch with pulmonary conduit	0	1	0	0	0	**1**
REV	0	1	0	0	0	**1**
Other surgical repairs						
Left ventricular outflow tract resection	0	4	0	1	1	**6**
Right ventricular outflow tract resection	1	1	0	0	1	**3**
Vascular surgery						
Main pulmonary artery band	11	3	10	0	5	**39**
Coarctation repair	3	2	4	3	4	**16**
Ligation of patent ductus arteriosus	3	1	0	0	1	**5**
Aortic arch repair	4	1	0	0	0	**5**
Main pulmonary artery repair	4	0	0	0	0	**4**
Right pulmonary artery repair	0	1	0	2	0	**3**
Left pulmonary artery repair	0	0	0	0	2	**2**
Bilateral branch pulmonary artery bands	0	0	2	0	0	**2**
Main pulmonary artery plication	1	0	0	0	0	**1**
Repair of partial anomalous pulmonary venous drainage	1	0	0	0	0	**1**
Repair of total anomalous pulmonary venous drainage	0	0	0	0	1	**1**
Valvular surgery and surgical replacement						
Tricuspid valve repair	1	0	0	1	0	**2**
Mitral valve repair	1	0	0	1	0	**2**
Aortic valve replacement	1	1	0	0	0	**2**
Oversewing of the tricuspid valve	0	0	0	0	2	**2**
Oversewing of the pulmonary valve	0	0	0	0	2	**2**
Mitral valve cleft closure	0	1	0	0	0	**1**
Pulmonary valve replacement	0	1	0	0	0	**1**
Surgical palliation						
Right hemi Fontan cavopulmonary connection	8	2	1	23	19	**53**
Damus-Kaye-Stansel connection	4	0	0	28	5	**37**
Fontan total cavopulmonary connection	3	1	0	14	6	**24**
Atrial septectomy	5	2	0	2	6	**15**
Right Glenn cavopulmonary connection	3	1	0	5	3	**12**
Right Blalock-Taussig shunt	1	2	2	0	2	**7**
Left Glenn cavopulmonary connection	5	0	0	0	2	**7**
Atrial septal defect enlargement	1	0	0	1	1	**3**
Left Blalock-Taussig shunt	0	0	1	0	1	**2**
Ventricular septal defect enlargement	1	0	0	0	0	**1**
intracardiac catheter intervention						
Balloon atrial septostomy	2	11	0	4	0	**17**
Balloon pulmonary valvotomy	1	0	0	0	0	**1**
Extracardiac catheter intervention						
Patent ductus arteriosus stent	2	2	2	7	3	**15**
Left pulmonary artery stent	0	0	0	6	0	**6**
Balloon dilatation of coarctation of the aorta	0	0	0	2	0	**2**
Blalock-Taussig shunt stent	0	1	0	0	0	**1**
Right pulmonary artery stent	1	0	0	0	0	**1**
Patent ductus arteriosus closure	0	0	1	0	0	**1**
Balloon dilatation of the branch pulmonary arteries	0	0	1	0	0	**1**
Balloon dilatation of the aortic valve	0	0	0	1	0	**1**
Balloon dilatation of the right pulmonary artery	0	0	0	1	0	**1**
Balloon dilatation of the left pulmonary artery	1	0	0	0	0	**1**

Table 5.3: Number of cardiac defects exhibited by the ELCH cohort at CMR acquisition.

	DORV	TGA	VSD	HLHS	HRHS	**Total**
Septal defects						
Ventricular septal defect	27	12	24	5	23	**91**
Atrial septal defect	18	9	12	28	21	**88**
Membranous septal defect	2	3	7	0	6	**18**
Residual ventricular septal defect	3	0	3	1	0	**7**
Atrioventricular septal defect	2	0	1	3	0	**6**
Tetralogy of Fallot	1	0	2	0	1	**4**
Coronary sinus defect	1	0	0	0	0	**1**
Stenotic defects						
Sub-pulmonary stenosis	7	10	1	0	3	**21**
Pulmonary valve stenosis	4	8	1	0	5	**18**
Mitral valve stenosis	2	0	2	10	0	**14**
Right pulmonary artery stenosis	4	2	3	0	2	**11**
Left pulmonary artery stenosis	3	4	1	1	1	**10**
Sub-aortic stenosis	3	2	1	0	0	**6**
Aortic valve stenosis	0	0	1	5	0	**6**
Coarctation of the aorta	0	0	2	0	0	**2**
Main pulmonary artery stenosis	2	0	0	0	0	**2**
Tricuspid valve stenosis	1	0	0	0	0	**1**
Pulmonary vein stenosis	0	0	1	0	0	**1**
Atretic defects						
Mitral valve atresia	8	0	0	8	0	**16**
Aortic valve atresia	0	0	0	15	0	**15**
Pulmonary valve atresia	1	1	0	0	7	**9**
Tricuspid valve atresia	0	0	0	0	8	**8**
Interrupted aortic arch	0	0	0	0	0	**0**
Hypoplastic defects						
Hypoplastic left ventricle	10	0	0	22	0	**32**
Hypoplastic right ventricle	1	2	2	0	24	**29**
Hypoplastic left pulmonary artery	2	0	0	1	2	**5**
Hypoplastic aortic arch	0	0	4	0	0	**4**
Hypoplastic right pulmonary artery	1	0	1	1	0	**3**
Discordant defects						
Transposition of the great arteries	0	13	0	0	17	**30**
Double outlet right ventricle	26	0	0	0	0	**26**
Double inlet left ventricle	0	1	0	0	14	**15**
Double inlet right ventricle	1	0	0	0	0	**1**
Valvular defects						
Straddling mitral valve	3	1	1	0	0	**5**
Bicuspid aortic valve	0	0	3	1	0	**4**
Straddling tricuspid valve	0	2	1	0	0	**3**
Bicuspid pulmonary valve	0	1	0	0	0	**1**
Mitral valve cleft	0	0	0	0	0	**0**
Vascular defects						
Bilateral superior vena cavae	5	0	4	2	3	**14**
Patent ductus arteriosus	0	1	3	0	1	**5**
Aberrant right subclavian artery	0	1	2	0	1	**4**
Left superior vena cava	1	0	0	0	0	**1**
Aortopulmonary window	0	1	0	0	0	**1**
Left pulmonary arteriovenous malformation	0	0	0	1	0	**1**
Anomalous left pulmonary artery	0	0	0	0	1	**1**
Right aortic arch	1	0	0	0	0	**1**
Partial anomalous pulmonary venous drainage	0	0	0	0	0	**0**
Total anomalous pulmonary venous drainage	0	0	0	0	0	**0**
Hypertrophic defects						
Hypertrophic right ventricle	3	5	7	1	2	**18**
Hypertrophic left ventricle	1	2	1	0	0	**4**

diac (such as arterial switch) surgical repair maintaining biventricular circulation[5]. Lastly, 65 patients had histories including staged or anticipated univentricular palliation: a single patient had undergone hybrid palliation, awaiting the next operative stage; 19 had been palliated with superior cavopulmonary connection (via either hemi Fontan or Glenn anastomosis) without neo-aortic reconstruction; 21 cases had progressed through the combined stage I and II Norwood procedure, including the creation of DKS and superior cavopulmonary connections; and 24 patients had fully transitioned to a univentricular circulation, including Fontan completion (TCPC).

The combination of congenital diagnoses and interventional histories described, culminates in the patient-specific anatomy and disease morphology of the ELCH dataset. In addition to the circulatory modifications outlined in Table 5.2, the 563 defects resultant at the point of CMR examination are tallied in Table 5.3.

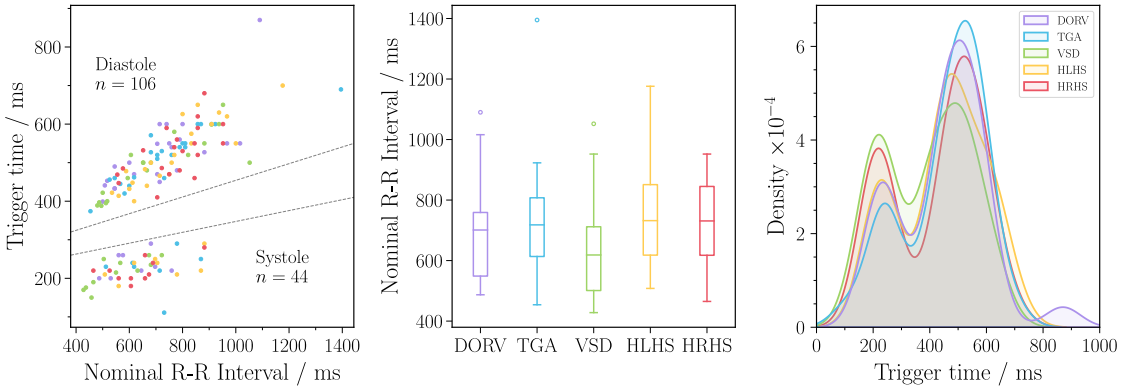
5.4.3 Imaging characteristics

All image data were acquired at 1.5 T using Philips Achieva magnetic resonance imaging (MRI), spanning three scanner software releases. In an effort to secure patient compliance, and to avoid the distress of extended CMR investigation, local practice allowed for acquisition under general anaesthesia for 125 out of 150 patients.

High resolution 3D SSFP was achieved via electrocardiogram (ECG) gating and respiratory navigation. Where the cardiac cycle could accommodate dual phase acquisition, the diastolic volume was maintained for segmentation, maximising the apparent size of septal defects. This scheme results in the bimodal distribution demonstrated in Figure 5.3a, including 44 systolic, and 106 diastolic volumes. Qualitatively, and given our preference for diastolic segmentation, this plot also suggests a subtle bias for single phase, systolic acquisition at high heart rate. Given that there are no statistically significant differences in heart rate between diagnostic groups[6], Figure 5.3c confirms that the bimodal distribution is independent of diagnosis.

[5]However, it should be noted that residual VSDs remained in seven such cases.

[6]Although Kruskal-Wallis H-test points to significant difference within the omnibus, this is not borne out by post-hoc Dunn testing with Bonferroni correction: $p > 0.06$ in all comparisons.



(a) (b) (c)

Figure 5.3: The interaction between electrocardiogram (ECG) gating and heart rate for high resolution 3D imaging. (a) The ELCH dataset is bimodally distributed between systolic and diastolic volumes. (b) There are no statistically significant differences in heart rate between patients with different diagnoses. (c) Accordingly, kernel density estimation suggests the same bimodal distribution is largely consistent across all diagnostic sub-groups.

TR-MRA relied upon the administration of gadolinium-based contrast agent in all cases, including: Dotarem ($n = 145$, 0.2 mLkg^{-1}); Gadovist ($n = 4$, 0.1 mLkg^{-1}); and Multihance ($n = 1$, 0.2 mLkg^{-1}). Other representative imaging characteristics are summarised in Table 5.4.

5.4.4 Burden of manual image segmentation

Our results confirm that high fidelity (including the expression of circulatory continuity at the pixel level) 3D manual image segmentation is a laborious task. After adjusting for idle periods, the median segmentation time was 3 hours 17 minutes, and up to 5 hours 31 minutes in the worst case. Moreover, this task demands intensive human-computer interaction, requiring a median number of 11,848 individual mouse clicks and 6429 key presses. In mediating these actions, the mouse pointer moved a median distance of 831 m across a 23 inch computer monitor[7], and the scroll wheel completed 814 full revolutions. Although results for comparative medical image processing tasks are scarce, recent publication suggests that

[7]In segmentation of the ELCH data, the mouse pointer covered a total of 130 km, or around 60 single journeys between Guy's and St Thomas' Hospitals.

Table 5.4: Imaging characteristics of the ELCH dataset. Unless otherwise indicated, all results, $P_{50}\ \ (P_i, P_j)_{i,j}$, report the median and ranges between the i^{th} and j^{th} percentiles. References to x, y and z reflect the DICOM standard coordinate space given by left-posterior-superior convention.

	Imaging characteristics			
	3D SSFP		4D TR-MRA	
Image				
Size / voxels				
x	150	$(140, 169)_{25,75}$	140	$(130, 160)_{25,75}$
y	384	$(336, 432)_{25,75}$	320	$(320, 348)_{25,75}$
z	400	$(352, 432)_{25,75}$	320	$(320, 352)_{25,75}$
Spacing / mm				
x	0.72	$(0.65, 0.80)_{25,75}$	0.82	$(0.65, 0.88)_{25,75}$
y	0.71	$(0.66, 0.78)_{25,75}$	0.79	$(0.71, 0.85)_{25,75}$
z	0.72	$(0.66, 0.78)_{25,75}$	0.79	$(0.71, 0.85)_{25,75}$
Slice thickness / mm	1.44	$(1.30, 1.50)_{25,75}$	1.54	$(1.30, 1.76)_{25,75}$
Over contiguity*	2.00	$(1.00, 2.00)_{0,100}$	2.00	$(1.00, 2.00)_{0,100}$
Timing				
Scan duration / s	158.5	$(131.3, 188.8)_{25,75}$	18.5	$(14.8, 22.2)_{25,75}$
Phases	2	$(1, 2)_{25,75}$	N/A	
Dynamics	N/A		11	$(9, 11)_{25,75}$
Spacing / s	N/A		1.7	$(1.36, 2.10)_{25,75}$
Magnetic resonance				
Echo time / ms	2.36	$(2.24, 2.43)_{25,75}$	1.33	$(1.20, 1.45)_{25,75}$
Repetition time / ms	4.72	$(4.48, 4.87)_{25,75}$	4.47	$(3.97, 4.96)_{25,75}$
Flip angle / °	90	$(70, 90)_{0,100}$	25	$(25, 30)_{0,100}$
Pixel bandwidth / Hz	542	$(541, 969)_{25,75}$	289	$(248, 332)_{25,75}$

*To improve the signal-to-noise ratio our local imaging protocol acquires data on an isotropic grid composed of over contiguous slices, the slice thickness typically exceeding the spacing between slices by a factor of two.

this level of interaction is at least comparable with the computer operation of a radiologist completing a clinical reporting shift.

Vosshenrich and Breit (2021) found that a single radiologist moved the mouse 4.6 m and initiated 23 key strokes every minute. In segmentation of the ELCH data, equivalent median results are: 4.1 m and 30 strokes per minute. Whilst the distances traversed are difficult to compare due to differences in screen size, the elevated rate of keyboard input indicates the complexity and intensity of manual segmentation. This is further increased by another 58 mouse clicks, adding up to a total of almost 90 discrete inputs per minute.

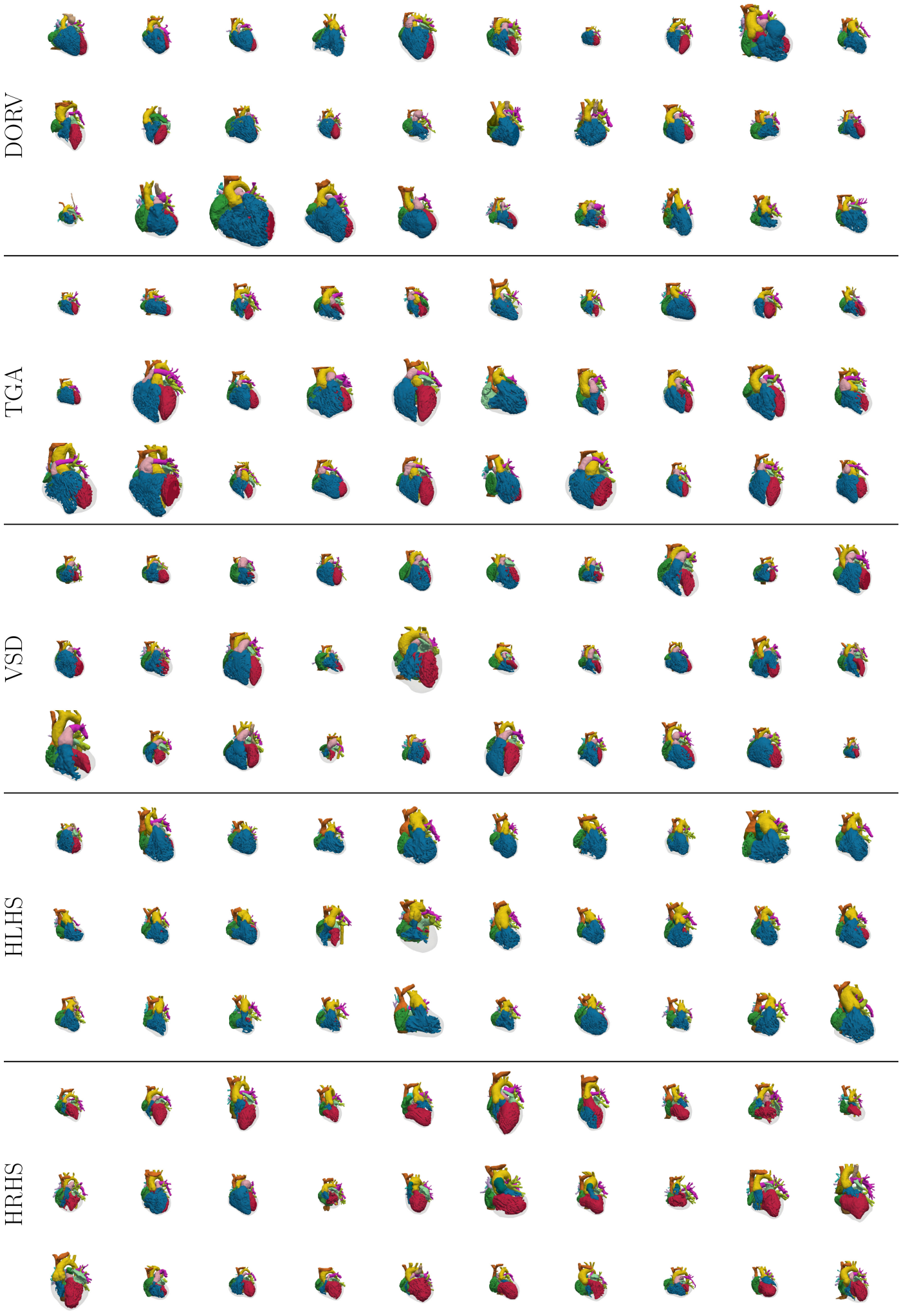
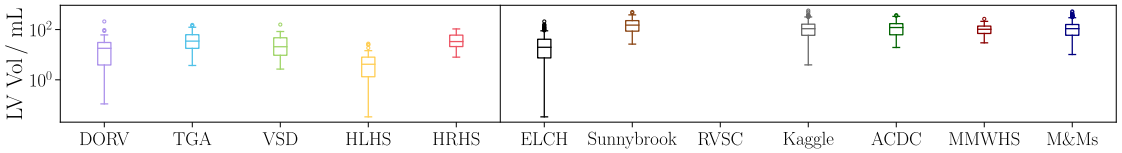
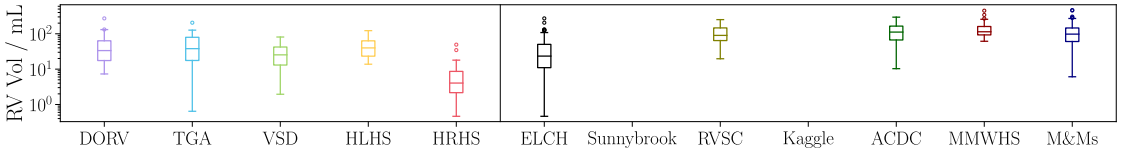


Figure 5.4: The ELCH dataset (see digital version for "zoomable" view). All examples share a common coordinate space such that relative differences in anatomical size are visually meaningful.



(a) LV Volume



(b) RV Volume

Figure 5.5: Disparities in ventricular volume are characteristic of CHD diagnosis. LV and RV volume are severely reduced in HLHS and HRHS, in (a) and (b) respectively. The ELCH dataset includes smaller patients with reduced left and right ventricular volume than those represented in publicly available data.

The results of manual image segmentation are visualised by surface-rendered representation in Figure 5.4. Note that the perspective (or camera position) of each rendering takes a constant position with respect to the anatomical centre of mass and DICOM coordinate space. This allows visual differences in spatial scale and orientation to reflect the variation in CHD anatomy encountered within clinical image data. Figure 5.4 provides an immediate impression of CHD heterogeneity including the distribution of patient sizes discussed in Section 5.4.1. In addition, it presents a qualitative impression of certain defects, including ventricular imbalance or hypoplasia, and muscular hypertrophy.

5.4.5 Anatomical characteristics and comparative analysis

The anatomical heterogeneity of the ELCH dataset is quantitatively borne out by the following comparative analysis. Given its paediatric membership, it is not surprising that LV and RV volumes expressed in the ELCH dataset are lower than those represented in publicly available data (see Figure 5.5). Mann-Whitney U-test after Bonferroni correction suggests that the median ELCH LV and RV volumes of 19.1 mL and 23.4 mL, respectively, are significantly less than equivalent results drawn from publicly available data ($p < 10^{-16}$ in all cases). Perhaps more

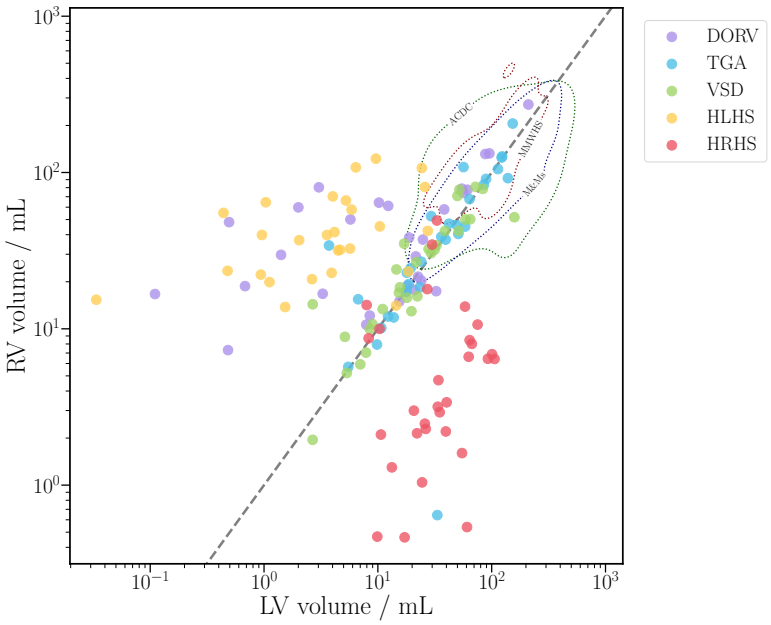
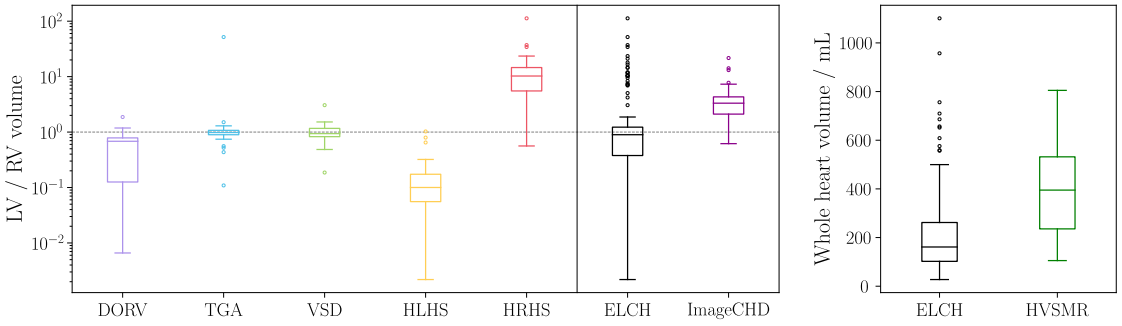


Figure 5.6: The paediatric CHD population predictably exhibits lower ventricular volumes than those expressed by healthy adults, or those with acquired cardiac disease (ACDC, M&Ms, MM-WHS). Moreover, patients in the ELCH dataset (particularly those with HLHS and HRHS) exhibit significant ventricular imbalance, manifesting as deviation from the dashed, grey line of equality.

interestingly, Figure 5.5 also confirms that LV and RV volume are characteristic of CHD, particularly for those diagnostic groups that include lateralised defects. As would be expected, the ELCH dataset demonstrates reduced LV and RV volume in HLHS and HRHS groups, respectively.

In Figure 5.6 we consider imbalances between LV and RV volume, possibly associated with CHD, and whether such effects are captured by the ELCH cohort. In those publicly available datasets for which both LV and RV cavity labels are made available (ACDC, MM-WHS, M&Ms), cases are tightly grouped either side of the line indicating equal ventricular volume. In contrast, patients from the ELCH dataset show substantial ventricular imbalance. This is particularly clear in diagnostic groups involving ventricular hypoplasia (HLHS and HRHS) and to a lesser extent within the DORV category. Although a well established and fully anticipated observation (ventricular imbalance being a definitive characteristic of these diagnoses), this plot makes clear that the cardiovascular anatomy of the



(a) (b)

Figure 5.7: (a, *left*) CHD diagnoses associated with lateralised defects demonstrate substantial deviation from ventricular volume balance (see HLHS and HRHS) (a, *right*) and (b) Even when compared against other CHD-specific examples, the entire ELCH cohort ($n = 150$) captures a far broader range of ventricular anatomy than those making up the ImageCHD ($n = 110$; dark purple boxplot) or HVSMR datasets ($n = 10$).

paediatric cohort cannot be well approximated by a spatially affine transformation of structurally normal image data, or those reflective of acquired heart disease. Moreover, the anatomical heterogeneity of the ELCH cohort is not limited to the discrete circulatory changes associated with the defects list in Table 5.3, but is also informed by increased continuous variation in anatomical scale and geometry. We are pleased that our segmentation of the ELCH dataset captures these core features of CHD morphology.

The structural heterogeneity of the ELCH cohort is reinforced by comparison with publicly available datasets for the segmentation of CHD anatomy. Given that the ImageCHD dataset is provided without meaningful pixel spacing, in Figure 5.7a we consider the dimensionless quantity of LV to RV ratio. As per previous results, this demonstrates the association between ventricular disproportion and congenital diagnosis. More interestingly, it also suggests that at least by this metric, the ELCH dataset captures a wider range of ventricular imbalance (and perhaps anatomical variation) than the ImageCHD cohort. In particular, the ImageCHD dataset includes fewer examples in which the RV is dominant over a small, hypoplastic or rudimentary LV. Numerous patients matching this anatomical configuration are represented in DORV and HLHS sub-groups of the ELCH dataset.

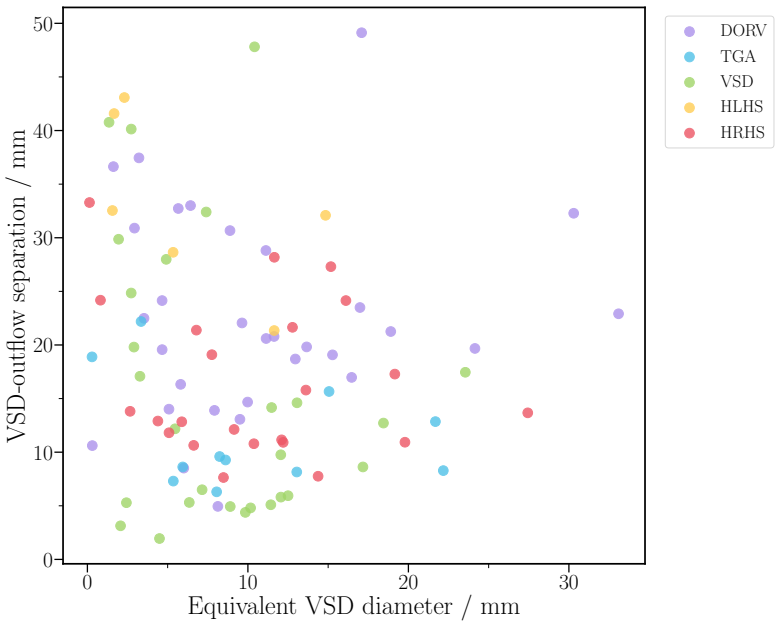


Figure 5.8: The bivariate distribution characterising the structural appearance of ventricular septal defects according to: defect size (or equivalent circular diameter) and separation from ventricular outflow. The ELCH dataset captures a wide range of different defects, reflecting the heterogeneity of the CHD population.

This somewhat surprising disparity (given the approximate equivalence between the sizes of these two datasets) may expose local differences in clinical practice and outcomes for various patient groups.

The HVSMR dataset separates 3D SSFP data into a collective blood pool, and a myocardial class. As such, Figure 5.7b is framed in terms of whole heart blood pool volume. As might be anticipated from a dataset of just ten cases, the HVSMR data span a narrower range than the more heterogeneous ELCH dataset. There is also evidence that the ELCH patients are substantially smaller than their HVSMR counterparts. The median HVSMR whole heart volume (395 mL) is almost 2.5 times that of the ELCH examples (161 mL).

Finally, we briefly depart from the examination of ventricular volume and turn our attention to the characteristics of discrete defects. Unfortunately, there are no appropriate public datasets against which we can compare the following results. The most frequent lesion within the ELCH cohort, we limit our analysis to VSDs. Figure 5.8 characterises the distribution of these 98 examples according to their

size and proximity to the nearest ventricular outflow. These characteristics are relevant to the consideration of possible surgical closure.

Defects were measured according to surface-rendered representation of relevant anatomical interfaces (VSD: between LV and RV; aortic valve: between either ventricle and the (neo-) aorta; pulmonary valve: between either ventricle and the MPA), as presented by manual segmentation. Subsequently, individual interfaces were localised by their centre of mass; and VSD size was assessed as the diameter of the circle with equivalent surface area. Figure 5.8 shows that the majority of VSDs are located within 35 mm of the nearest ventricular outflow and have an equivalent diameter of less than 20 mm.

In general, however, these results demonstrate the wide range of septal defects included within the ELCH dataset. They confirm that the heterogeneity of CHD is not limited to an accounting of defects by their discrete presence or absence (as per Table 5.3), or to continuous changes in anatomy (such as ventricular volume). Rather, variability also extends to the continuous characteristics governing the appearance of discrete, structural defects themselves. Though these observations are well established properties of CHD, we are encouraged that they are not only expressed by patients from the ELCH cohort, but that they can be derived automatically from our manual segmentation of associated image data. This confers clinical meaning to our ground truth label maps.

5.5 Discussion

5.5.1 Context

Through the lens of CMR, the ELCH dataset aims to reflect the range of anatomy encountered by those caring for patients with CHD. This variety is captured by the segmentation of high resolution, spatially isotropic image data. Within this context, and to the best of our knowledge, the ELCH dataset is unique. The results of this chapter have demonstrated that compared with publicly available datasets it offers distinct advantages in both quantity and quality.

In raw patient numbers, our total of 150 cases is comparable with the largest publicly available, cardiovascular datasets[8]. Of those considered, only the M&Ms dataset includes a greater number of patients, 375, once training and test sets are combined (Campello et al., 2021). The size of the ELCH cohort is on par with the popular ACDC dataset (Bernard et al., 2018), also comprising 150 patients, and exceeds that of: Sunnybrook, 45 (Radau et al., 2009); RVSC, 48 (Petitjean et al., 2015); HVSMR, 20 (Pace et al., 2015); MM-WHS, 60 CMR and 60 CT (Zhuang et al., 2019); and ImageCHD, 110 (Xu et al., 2020)[9].

In line with the majority of research activity (Chen et al., 2020), most publicly available data concern the segmentation of short axis cine CMR for the purpose of quantitative ventricular volumetry. In contrast, segmentation of the ELCH dataset was primarily motivated by qualitative downstream application, such as 3D visualisation of patient-specific anatomy. As such, and with respect to both image data and segmentation task, the ELCH dataset diverges from the majority of publicly available data.

In terms of image acquisition, the ELCH dataset is most comparable with MM-WHS and HVSMR resources. Each depicts cardiovascular anatomy within high resolution, isotropic 3D SSFP data. However, and apart from by their quantity, the HVSMR labels are limited by the formulation of their segmentation task, separating pixels into whole heart blood pool, myocardium and background classes. Although this may be useful to downstream applications such as patient-specific 3D printing, whole heart segmentation otherwise limits wider utility. A more flexible partitioning of data separately delineates anatomical sub-structures of the heart. MM-WHS and ImageCHD share a formulation that separates image data into seven foreground classes. Whilst the two diverge in their treatment of the great arteries, both reflect the anatomy of the four cardiac chambers, the two cardiac outflows and the ventricular myocardium. The ELCH formulation is more fine-grained still, including a total of eighteen anatomical components and surgical connections.

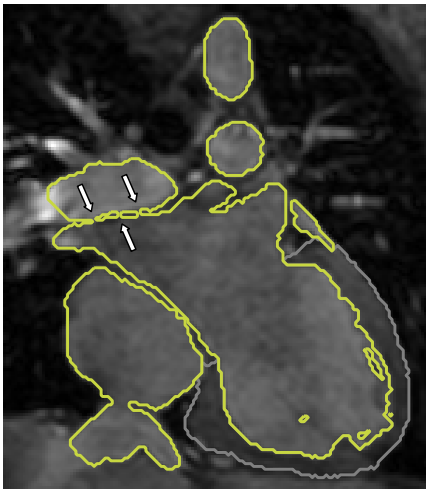
[8]Note that whilst far larger, at the time of writing, the UK Biobank is not publicly available (Fry et al., 2017).

[9]Whilst the combined training and test sets associated with the Kaggle data totals 700 patients, images are provided without ground truth segmentation (Kaggle, 2016).

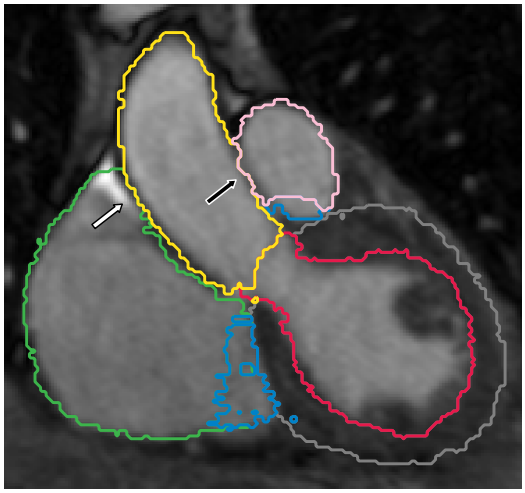
Critically, we also point out limitations in the quality of segmentations made available by each of HVSMR, MM-WHS and ImageCHD datasets. We would not go so far as to consider these complaints shortcomings. Rather, the following critique reflects the ways in which our approach might differ from practice elsewhere. Having compiled the ELCH segmentations, we are fully aware of the challenges of curating such resources and do not wish to devalue any of the publicly available datasets considered. Each will have had their own set of priorities and motivations, likely differing from our own.

Firstly, whilst ostensibly derived from 3D image data, we do not consider all MM-WHS labels as true 3D segmentations. Figure 5.9 demonstrates the irregular anatomical boundaries which result when volumetric data are manually segmented in a slice-wise fashion. This appearance is shared by more than one quarter of the MM-WHS training cases. Secondly, where motivated by treatment or surgical planning, we consider it important that segmented data reflect continuity of the blood pool. In our work, we confer clinical meaning by reflecting haemodynamic communication by pixel adjacency (see Figure 5.1). Figure 5.9 shows that none of MM-WHS, HVSMR or ImageCHD take the same approach, presumably being more concerned by spatial overlap and surface localisation. Finally, although a rich resource of 110 patients, ImageCHD data are made publicly available without meaningful physical pixel spacings. This is problematic in the case that the features characteristic of cardiac anatomy correlate with patient size. Moreover, inspection of these data suggests that images were neither acquired nor reconstructed at (near) isotropic spatial resolution, possibly resulting in distortions in aspect ratio.

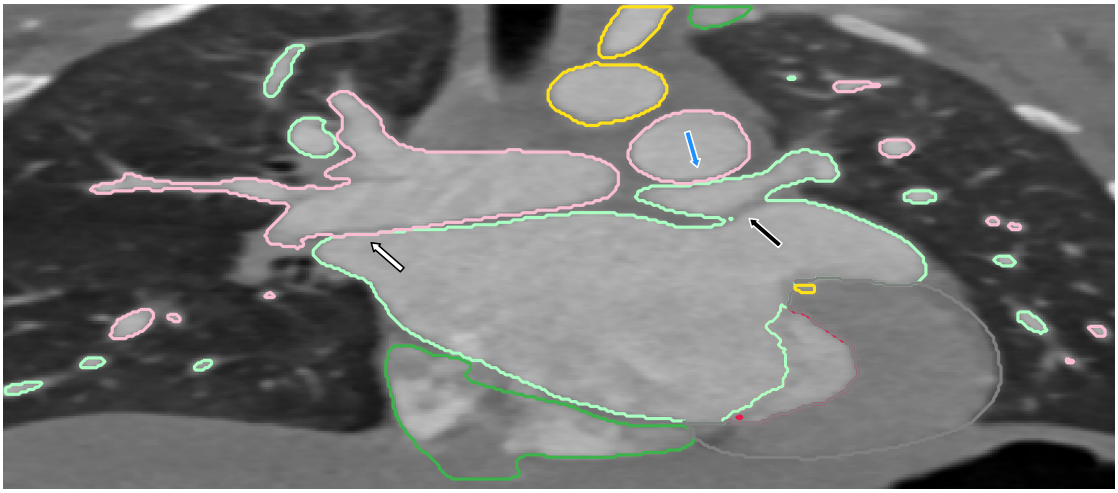
Diagnostically, ImageCHD patients are the closest matched to the ELCH cohort, both reflecting a wide range in CHD. In this respect the two are highly complementary, capturing CHD anatomy through the segmentation of CT and CMR data, respectively. Accumulated across our dataset, detailed reporting identifies over 600 congenital defects (see Table 5.1), 563 of which are resultant at the point of diagnostic examination (see Table 5.3). The gap between the two is partially reconciled by endogenous resolution (for example, spontaneous VSD closure), but more substantively, is explained by the imposition of discrete structural intervention (see those listed in Table 5.2). This picture, is further complicated by the interaction between congenital diagnosis and clinical management.



(a) HVSMR



(b) MM-WHS



(c) ImageCHD

Figure 5.9: Representative segmentations from publicly available datasets. None of the HVSMR, MM-WHS or ImageCHD datasets consistently reflect anatomically meaningful boundaries. (a) The HVSMR labels include small non-physiological connections between the superior margin of the LA and the inferior RPA (white arrows). (b) A significant proportion of the MM-WHS data include jagged artefacts associated with 2D slice-wise segmentation of volumetric data. By pixel adjacency, this example also includes large haemodynamic windows between the right atrial appendage and aorta (white arrow) and the aorta and MPA (black arrow). (c) This example from the ImageCHD dataset falsely implies connection between the LA and RPA (white arrow); LA and LPV (black arrow); and LPV and RPA (blue arrow). Note that the CT data composing the ImageCHD dataset are made available without physical pixel spacing, resulting in aspect ratio distortion, as per this example.

As alluded to in Section 5.4.1, patient pathways, whether culminating in biventricular repair or staged univentricular palliation, are disease- and patient-specific. Despite this variation, we observe that fundamentally, both surgical and catheter-based structural intervention modify the geometry of congenital anatomy in some way. Such changes are highly relevant to the segmentation and visualisation of patient-specific disease, and in particular the representation of defects. Successful surgical VSD closure, for example, represents a discrete change in the communication of left and right ventricles. Perhaps more significantly, staged univentricular palliation induces dramatic circulatory modification, including DKS and cavopulmonary connections. Critically, in both its labelled image data, and in its characterisation, the ELCH reflects these features of structural anatomy.

Our analysis of anatomical characteristics lays bare the extreme differences in structural anatomy exhibited by the paediatric CHD population. If anything, measurements of ventricular volume suggest greater structural heterogeneity within the ELCH cohort than between ImageCHD patients. Unlike publicly available data, the variety captured in the ELCH dataset is further demonstrated by a detailed reporting of patient demographics, diagnoses and medical histories. Finally, comparative analysis highlights the stark differences between congenitally malformed or surgically modified anatomy, and that of healthy volunteers or those affected by acquired disease. Whilst commonly understood clinically, these results provide quantitative evidence, compelling to medical and scientific researchers alike.

Our results rigorously establish the operator burden associated with 3D manual image segmentation. A median duration of 3 hours 17 minutes is approaching half a working day for a whole time equivalent member of NHS staff. Perhaps more problematically, the wide distribution of segmentation times (being over 5 hours in the worst case) hampers meaningful prediction of the workload associated with a given case. Comparison with published estimates is made challenging by infrequent and imprecise reporting of segmentation time, as well as differences in task formulation and motivation (see Section 3.2.3). Acknowledging these difficulties, previous systematic review made a best estimate of between 2 and 3 hours (Byrne et al., 2016), with which our findings are comparable.

Whilst more difficult to characterise, at least with respect to the intensity of interaction, our results confirm the manual complexity of this task. If nothing

else, they highlight the control exerted by the operator, and their influence in determining the form and extent of segmented anatomy. In this regard, the sheer volume of inputs likely explains the inter- and intra-observer variation previously reported (Meier et al., 2017). Beyond these observations, drawing more substantive conclusions may necessitate a comprehensive assessment of human-computer interaction. This line of enquiry may provide a means for assessing and considering the introduction of semi-automated tools for cardiac segmentation in the future.

We observe that the faithful understanding of anatomy depends on significant expertise in image interpretation and protracted experience inspecting and analysing CHD morphology. Experientially, however, we suggest that these skills are insufficient: appreciating anatomical structure is one thing; capturing such an understanding (including the accurate representation of defects and haemodynamic continuity) within a segmented image, is another. Manual segmentation of the ELCH dataset also demanded a knowledge of image processing methodologies, their implementation in software, and *above all*, practice in their application. Problematically, these traits are rarely expressed by an individual, tending to be siloed within medical (radiology or cardiology specialties) and scientific or engineering professions. In combination, our findings confirm the challenges of this task and to some extent, explain the limited expansion of patient-specific 3D modelling outside of all but the largest teaching hospitals and research centres of excellence.

5.5.2 Limitations and future work

The previous Section 5.5.1 presented the unique qualities of the ELCH dataset, endorsing its potential to inform research and education across a number of domains (including within image processing and cardiac morphology). To deliver on this promise (and provided we can overcome current obstacles associated with: information governance, including the ethics of sharing patient data; the imposition of suitable commercial restrictions; and identification of sustainable funding for long term storage and access) we intend to make this resource publicly available in the near future. Leveraging the dataset as the basis of our own segmentation challenge (similar to HVSMR), may afford an opportunity to maximise exposure and hence the impact of our contribution.

For all the strengths of the ELCH dataset, however, we also acknowledge notable limitations. Firstly, the size of the cohort is not sufficient to appreciably represent all possible congenital defects. Several rare lesions are characterised by lone example, including right aortic arch, coronary sinus defect and aortopulmonary window. In future, increasing the size of the dataset would undoubtedly make it more representative of the CHD population.

Adopting such an approach, however, may be to oversimplify the challenge and its motives. Once surgical modifications are taken into account, the number of structural permutations is colossal. This is not to mention the heterogeneity in the presentation of defects as illustrated by Figure 5.8. Hence, it is unlikely that an expanded ELCH dataset will ever truly represent the underlying distribution of patients. Therefore, whilst increasing the size of the ELCH cohort remains vital to our future work, we are pragmatic as to how best to go about this.

In particular, we recognise that the primary motivation in establishing the dataset was to provide a resource informing efforts to automate the segmentation of patient-specific models from clinical data. To expedite curation, we opted to gather historic cases. In future, and if ethical requirements can be satisfied, expanding the dataset by the inclusion of patients conteporaneously with the delivery of care may be preferable. Simultaneously, this approach allows for the expansion of the ELCH dataset; the development of methodologies for automated segmentation; and investigation the effectiveness of patient-specific 3D models within care, education and patient communication. This is not to mention the direct clinical benefits conferred by patient-specific 3D printing (reviewed in Section 2.2.1).

The second major limitation of the ELCH dataset concerns its specificity. Whilst features of the data are undoubtedly generalisable to the wider CHD population, they are expressed by patients treated at a single centre, scanned using MRI technology provided by a single manufacturer and segmented by a single operator. In these respects, the ELCH data define their own domain. Given our wider ambition to exploit this dataset, driving feature learning relevant to automated segmentation, this may prove a limitation. More specifically, due to the domain specificity of CNNs (Kamnitsas et al., 2017), models trained using our dataset may not be peformant when presented with images acquired at different hospitals, using different MRI scanners. Despite this, we consider the ELCH data

a valuable foundation, to which data from other centres could be contributed. Multi-centre collaboration, and rigorous assessment of inter- and intra-observer variability (possibly including ELCH examples) is fertile ground for future work.

5.6 Conclusion

In the curation of the unique ELCH dataset, we have established the motivation for, and resources necessary to investigate, methods to extract patient-specific CHD anatomy from 3D CMR. It includes 150 distinct cases, each segmented into one of eighteen anatomical classes. Vitally, our label maps reflect haemodynamic continuity by pixel adjacency, conferring measurable clinical meaning. Through a careful diagnostic and anatomical analysis, we have shown that our dataset captures the significant structural variation associated with this population. In this respect, and compared with existing datasets, including those representative of CHD, ELCH cases demonstrate greater variety. Accordingly, we find that this task, and the clinically meaningful formulation presented, embody a fundamentally different proposition to the delineation of ventricular anatomy within short axis cine data. In this context, the ELCH cases represent the first training dataset appropriate for CNN-based segmentation, an application we will explore in the following chapters. More generally, however, we hope that the ELCH dataset will prove an invaluable starting point for those seeking to understand, and improve the care of patients with, CHD.

Chapter 6

CNN segmentation of congenital heart defects

6.1 Introduction

Chapter 5 presented the Evelina London Children's Hospital (ELCH) dataset, demonstrating its unique quantity and quality. Equipped with this resource, this chapter establishes baseline performance for the convolutional neural network (CNN)-based segmentation of multi-class congenital heart disease (CHD) anatomy from 3D cardiac magnetic resonance (CMR). As per our curation of these training data, we attend closely to the clinical aspects of this task. Recognising that a typical CMR study comprises the acquisition of complementary image series, we leverage the combination of 3D steady state free precession (SSFP) and 4D time-resolved magnetic resonance angiography (TR-MRA). We anticipate that spatio-temporal features describing the dynamics of blood flow inform the structure of the heart and circulation, including the presence of defects. To ensure the clinical relevance of our work, CNN predictions are assessed not only against widely accepted metrics of performance, but also against bespoke measures sensitive to the representation of congenital defects in an associated patient-specific 3D model.

Prior to our account of baseline performance, we motivate and briefly review the application of CNNs to high-dimensional data, with particular focus on 4D and spatio-temporal deep learning.

6.2 Learning from 4D data

6.2.1 Rationale

Each 3D SSFP image in the ELCH dataset is accompanied by dynamic, TR-MRA. This acquisition reveals the passage of a bolus injection of gadolinium contrast agent. As observed by Shin et al. (2012), differential enhancement informs the haemodynamic relationships between anatomical structures. In relation to the normally configured left and right heart, this is illustrated by Figure 6.1a. Within such patients, we observe: (1) early enhancement of the right heart following venous injection (blue curve); (2) enhancement of the left heart (red curve); and (3) final[1] enhancement of the cardiovascular blood pool (convergence of the blue and red curves). In this description, the patterns of enhancement associated with the blue and red curves are characteristic of the right and left heart, respectively. Taken together, the two can be said to indicate differential enhancement, informing the separation of anatomy.

The presence of a congenital heart defect, however, may disrupt this picture. Consider an inter-atrial communication, placing the left and right chambers in continuity. The passage of contrast agent between the two will be reflected in TR-MRA data, with pixels describing each atrium (and certainly those in close proximity to the defect) being more likely to share enhancement characteristics. Specific to CHD, distinctive patterns of enhancement can reflect the presence of septal defects or atretic junctions. More generally, characteristic enhancement is an expression of haemodynamic continuity; differential enhancement is an expression of haemodynamic isolation. In this way, clinically meaningful, anatomical structure can be inferred from inspection of these data. Notice that these conclusions are drawn from the temporal domain, revealing the presence and integrity of thin tissue interfaces (such as the atrial septum) which otherwise are a challenge to resolve spatially, including within 3D SSFP (see Figure 6.2a and Figure 6.2b).

This is not the only way in which TR-MRA and 3D SSFP data are complementary. Owing to the administration of exogenous contrast agent, TR-MRA visualises cardiovascular anatomy with high sensitivity. This improves the faithful

[1] Note that we refer to final rather than late enhancement to avoid confusion with late gadolinium enhancement for the visualisation of ventricular scar.

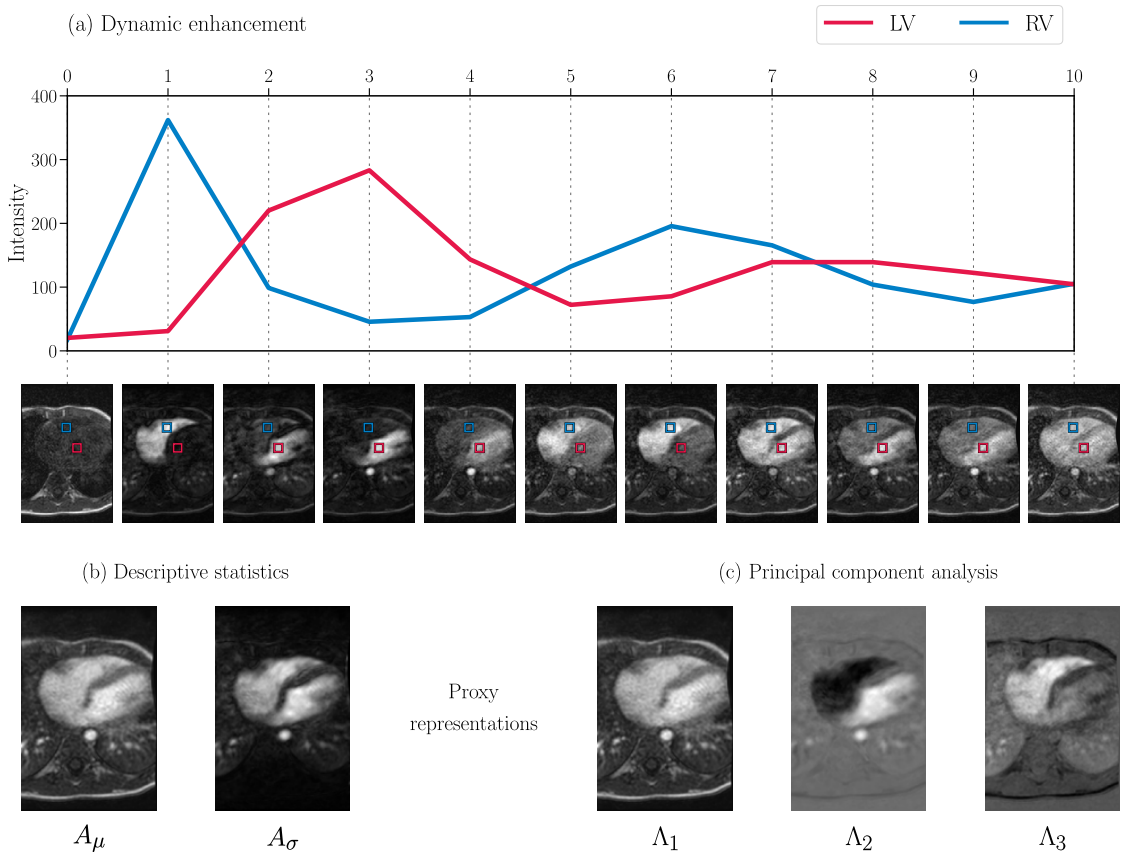


Figure 6.1: Characteristic enhancement patterns within contrast-enhanced, time-resolved magnetic resonance angiography (TR-MRA) reveal the configuration of cardiovascular anatomy and can imply the presence of congenital defects. (a) Temporal variations in signal intensity averaged within regions of interest positioned in the left (red) and right (blue) ventricles (LV and RV, respectively). We construct proxy representations of the volumetric time series. Those based on: (b) Descriptive statistics such as the pixel-wise, temporal mean (A_μ) and standard deviation (A_σ) reflect global enhancement; (c) Dimensionality reduction by principal component analysis (PCA) (Λ_1, Λ_2, Λ_3), summarises temporally differential enhancement in a compact form. All images demonstrate anatomy in a single axial plane from the volume time series.

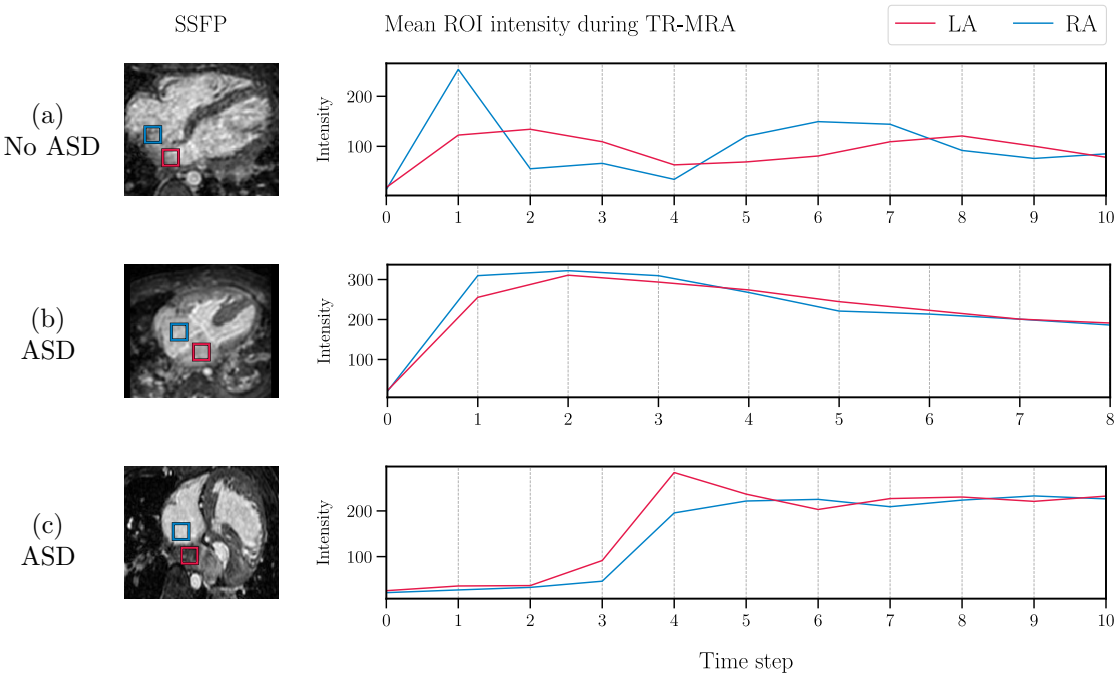


Figure 6.2: TR-MRA can inform the configuration of the heart where inspection of SSFP data remains equivocal. In (a) and (b), the limited spatial resolution of SSFP data provides an ambiguous impression of the atrial septum, possibly including an atrial septal defect (ASD). (a) Differential enhancement of the left atrium (LA) and right atrium (RA) rule out an ASD. Whereas in (b), the two atria share enhancement characteristics, indicating a communication. Finally in (c), the appearance of the LA is compromised by dephasing artefacts associated with blood flowing from the right pulmonary vein. Enhancement within TR-MRA data firstly reveals the approximate shape, size and location of the LA. Secondly, the common enhancement of both atria is indicative of ASD.

appearance of pulmonary vasculature. In SSFP data, these vessels are frequently affected by inhomogeneity in signal intensity and dephasing artefacts associated with flow acceleration (particularly at the confluence of pulmonary veins and left atrium, see Figure 6.2c). Similarly, TR-MRA data can aid in the visualisation of anatomy in the presence of the susceptibility artefacts associated with metallic devices. Relevant to the CHD population, this includes vascular stents (see Table 5.2). For all these strengths, however, it must be remembered that TR-MRA acquisition is not electrocardiogram-gated. Whilst it provides valuable insight as to the structure of the heart and its arrangement, it does so via a spatially coarse

representation. Given this configuration, we rely on 3D SSFP for the delineation of anatomy at high resolution. This is critical to the segmentation of intracardiac structures and in particular to septal defects. Therefore, and throughout this chapter, we consider TR-MRA data as an adjunct to 3D SSFP data, not the subject of segmentation in its own right.

6.2.2 Challenges of 4D deep learning

The deep learning methodologies which support state of the art performance across a range of medical image processing challenges, were first developed within the natural image domain (Litjens et al., 2017). For the most part, this family of tasks considers lone 2D images or their sequential combination as frames in a conventional video. In contrast and within the medical domain, 3D volumetric acquisition via any of CMR, computed tomography (CT), ultrasound or nuclear imaging (including positron emission tomography (PET)) are commonplace. Whilst volumetric time series can also be acquired, in the broader field of image processing, tasks addressing 4D data are rare (Gessert et al., 2020).

Accordingly, where established methods for 2D and 3D image segmentation are ubiquitous (consider the popularity of U-Net-like architectures (Ronneberger et al., 2015)), the same cannot be said for 4D data. Moreover, whilst most deep learning software libraries, such as PyTorch (Paszke et al., 2019) and TensorFlow (Abadi et al., 2016), provide functionality for the construction of CNNs based on 2D and 3D convolution, their 4D equivalent is notably absent. The difficulty of 4D deep learning is exacerbated by the significant memory demands associated with an exponential increase in the size of the data, challenging the resources relied upon for graphics processing unit (GPU)-accelerated training (Gessert et al., 2020). Methodologically, the same explosion applies to the number of model parameters, hampering effective, regularised optimisation: avoiding overfitting is particularly challenging considering the dearth of labelled 4D training data.

Not all of these challenges are unique to 4D segmentation tasks. Software availability aside, each obstacle posed by the extension of deep learning from 3D to 4D, equally characterises the difference between 2D and 3D applications. As a result, previous work has found that the theoretical advantages of 3D application

cannot always overcome the practical difficulties of 3D CNN training. In such cases, pragmatic 2D treatments of 3D data often return superior performance (Choy et al., 2019). Given these challenges, accepted schemes for the application of deep learning methods to 4D image processing tasks are yet to be established and remain the subject of open investigation.

6.2.3 Previous work in spatio-temporal deep learning

To inform our review of the relevant literature, we note that the 4D image data made up by TR-MRA reflect a volumetric time series. That is, the data are defined in a 3D+T spatio-temporal coordinate space. Correspondingly, our application is informed not only by previous investigation of 4D image processing tasks, but also by those concerning lower dimensional spatio-temporal image data. Far more commonplace, conventional 2D+T video has received greater research interest than its 3D+T equivalent. We also recognise the ambiguity in references to 2D, 3D and 4D data. More precisely, we will review the literature relevant to the segmentation of spatio-temporal 4D data, making explicit references to 2D+T and 3D+T where appropriate.

Previous works can be differentiated by their approach to spatio-temporal feature learning: whether spatial and temporal features are learned separately in series or in parallel; or whether truly spatio-temporal features are learned jointly. Each approach further divides into those based solely on CNNs, and those which also employ recurrent neural networks (RNNs) to model sequential data.

Separately learned spatial and temporal features

The first family of methods extracts spatial and temporal features separately and in parallel. Simonyan and Zisserman (2014) used optical flow to construct a temporally sensitive input of stacked displacement fields. Combined with a static single frame, they used a dual stream 2D CNN to extract spatial and temporal features. Subsequent late fusion was used to classify human actions within 2D+T video. A related approach has been presented by Kang et al. (2019). Rather than by CNN, they use long short-term memory (LSTM) to extract temporal features per voxel of 3D+T dynamic contrast-enhanced magnetic resonance imaging (MRI) of the

prostate. They subsequently fuse spatial and temporal features using a fully connected multi-layer perceptron, demonstrating improved segmentation performance compared with the application of either stream in isolation.

Alternatively, others have taken a serial approach to spatio-temporal feature learning. Sun et al. (2015) factorise the 2D+T convolutional operator: features extracted using a 2D CNN are subsequently transformed using learned 1D convolutional operators, modelling inter-frame dependencies within a video classification task. In the medical domain, a similarly staged approach has been used to identify cognitive networks from 3D+T functional MRI data using CNNs (Zhao et al., 2018). The serial combination of CNN and RNN features has also been explored. Across a variety of 2D+T natural video processing tasks, Donahue et al. (2015) use stacked LSTM modules to interrogate the spatial features extracted by 2D CNN. In the medical domain, this approach was applied to the detection of deficits in a cohort of 396 patients undergoing 3D+T CT perfusion examination (Vargas et al., 2019).

Jointly learned spatial and temporal features

Compared with those presented previously, the following works extract truly spatio-temporal features, leveraging operators (whether convolutional or recurrent) matching the dimensionality of the data considered. The first family rely solely on high-dimensional convolution, without differentiating between spatial and temporal axes. This approach has been deployed within 2D+T processing tasks: 3D CNNs have been applied to human action recognition (Ji et al., 2012), video classification (Tran et al., 2015) and, pertinently to CMR image analysis, promoting temporally consistent myocardial segmentation from short axis cine (Yang et al., 2018b). Within 3D+T processing tasks, 4D CNNs have been used to label 3D video data capturing rooms or street scenes (Choy et al., 2019); to reconstruct 4D cardiac CT data (Clark and Badea, 2019); and within the segmentation of the left ventricular cavity and myocardium from 4D cardiac CT (Myronenko et al., 2020). In the last case, however, spatial overlap performance was only comparable with, and did not exceed the application of a 3D CNN.

In contrast, the following works deploy convolutional recurrent operators for feature extraction. Following the work of Shi et al. (2015), these methods combine the advantages of CNNs and their treatment of spatially structured data, with those of temporally unrolled RNNs. Several works have investigated the benefit conferred by the introduction of convolutional LSTM cells to the U-Net architecture. Gao et al. (2018) employ these layers within the upsampling path, improving the segmentation of white and grey matter from longitudinal (3D+T) paediatric MRI. Meanwhile, Dong et al. (2020) and Fehling et al. (2020) investigate performance differences when convolutional LSTM cells are included in both or either of the encoder and decoder. Their findings suggest that the most significant performance gains are made by the introduction of these cells to the expanding portion of the U-Net; but that small incremental gains can be made by additionally including recurrent layers within the encoder. They demonstrate improved segmentation of the fruit fly heart from 2D+T optical coherence tomography, and laryngeal anatomy from high speed 2D+T video, respectively. Finally, Van De Leemput et al. (2019) use stacked, bi-directional convolutional LSTMs to synthesise non-contrast and contrast-enhanced volumes from a 3D+T CT perfusion study, as relevant to the assessment of stroke. This approach is closely related to our segmentation task, modelling a "volume-series-to-volume" transformation.

Comparison of approaches

The preceding review presented different approaches to spatio-temporal feature learning. However, relatively few authors have rigorously compared the performance of different schemes for handling high-dimensional and spatio-temporal data. This number dwindles significantly when our scope is limited to volume-series-to-volume transformation or segmentation. Notably, Van De Leemput et al. (2019) find their approach outperforms 3D U-Net. However, questionable design choices may prevent this baseline from being considered state of the art: they train for only 1500 iterations; opt to recover high resolution information by the use of nearest neighbour interpolation rather than learned deconvolution; and fail to concretely describe their strategy for passing sequential volumes to the 3D U-Net.

In contrast, the related works of Bengs et al. (2020) and Gessert et al. (2020) present a rigorous comparison of different schemes for the treatment of 3D+T spatio-temporal data. In the classification of autism spectrum disorder from fMRI, Bengs et al. (2020) consider spatio-temporal features learned: (1) jointly via 4D convolution; (2) sequentially by gated recurrent unit (GRU) and then by spatially 3D CNN; and (3) by adding frames as additional input channels to the network. Related to the final of these, they also consider (4) a formulation of their task conditioned on a multi-channel, spatially 3D input. Each channel reflects the pixel-wise mean and standard deviation, as assessed over the temporal dimension. Whilst this "temporal proxy" captures a relatively meagre representation of sequential information, this approach outperformed the trivial inclusion of frames as additional channels and was comparable to 4D convolution, only being surpassed by sequential temporal and spatial feature learning. In keeping with Section 6.2.2, we speculate that the challenges of faithfully optimising a high-dimensional and high capacity neural network (such as a 4D CNN) with limited medical training data, allow for this relatively basic approach to remain competitive.

Within the regression task of force estimation from 3D+T optical coherence tomography data, Gessert et al. (2020) compare spatio-temporal feature learning by: (1) 4D convolution within a residual architecture; (2) by factorised convolution (3D spatial and 1D temporal components); and (3) sequentially by 3D CNN followed by temporally unrolled GRU and vice versa. In common with Bengs et al. (2020), their best performing architecture preceded spatial CNN with recurrent feature extraction. However, 4D convolution was far more competitive within this work. We speculate that the improved performance of a 4D CNN can be explained by the availability of a larger training dataset: Gessert et al. (2020) had access to on the order of 10^4 samples, whereas Bengs et al. (2020) trained with only 134 subjects. It is plausible that these training data allowed for regularised optimisation of the high capacity 4D CNN, whilst avoiding overfitting.

Summary

The preceding review demonstrates the potential performance gains that can be achieved through a considered treatment of spatio-temporal image data. How-

ever, coupled with the challenges of 4D deep learning, the lack of an accepted architecture for 3D+T feature extraction makes any application of the reviewed methodologies an exploratory and open investigation. Given the limited training data and significant variation associated with the clinical CHD population, we leave the application of advanced neural architectures, possibly including high-dimensional or recurrent, convolutional operators, for future work. Instead, we are encouraged by the relative success of pragmatic efforts to straightforwardly incorporate time-series information by the construction of temporal proxies, and their input to CNNs as additional channels (Bengs et al., 2020). Here we investigate two such proxies, including pixel-wise, summary statistics of dynamic constrast enhancement (the mean and standard deviation, see Figure 6.1b); and dimensionality reduction via principal component analysis (PCA) (see Figure 6.1c).

6.3 Contributions

In this chapter we make the following contributions:

1. We conduct experiments using the ELCH dataset, establishing a new CNN-based state of the art for the segmentation of CHD anatomy and disease morphology from 3D CMR acquisition.

2. Our challenging task formulation separates image data into sixteen anatomical labels relevant to the description of congenital disease morphology, exceeding the seven-class description favoured by previous work.

3. We present schemes to efficiently leverage 4D CMR data within this task, demonstrating associated performance gains.

4. In addition to commonly used metrics of performance, we assess the clinical suitability of predicted segmentations, constructing metrics relevant to the quantity and quality of congenital heart defects.

5. In so doing, we expose limitations of conventionally trained CNN-based segmentation, demonstrating its inadequacy for clinical practice.

6.4 Methods

6.4.1 Task formulation

We assess the performance of CNN-based segmentation of CHD anatomy from 3D and 4D CMR data. Our task seeks to divide these images into clinically meaningful anatomical segments. We consider the input image $\mathbf{X} : \mathbb{R}^3 \to \mathbb{R}^C$, as the combination of C channels $X_1, X_2, ..., X_C$, each constructed from two types of data:

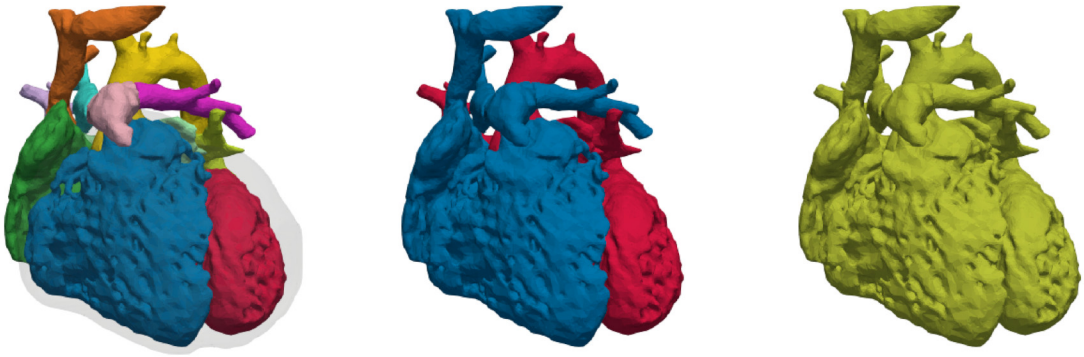
1. 3D SSFP, the morphological acquisition denoted $\mathbf{M} : \mathbb{R}^3 \to \mathbb{R}$, and being made up by the single volume M.

2. 4D TR-MRA, the angiographic acquisition denoted $\mathbf{A} : \mathbb{R}^3 \to \mathbb{R}^T$, and being made up by a time series of T volumes $A_1, A_2, ..., A_T$.

Our baseline considers the case in which segmentation probability is conditioned only on 3D SSFP data: $\mathbf{X}_M = (M)$. Experimentally, we consider the concatenation of M with the proxy temporal representations described in Section 6.4.2, achieving pairings with the pixel-wise temporal mean $\mathbf{X}_\mu = (M, A_\mu)$; standard deviation $\mathbf{X}_\sigma = (M, A_\sigma)$; and first principal mode of variation $\mathbf{X}_{\Lambda_1} = (M, \Lambda_1)$. We also construct three and four channel inputs by further concatenation of the second and third principal components, achieving: $\mathbf{X}_{\Lambda_2} = (M, \Lambda_1, \Lambda_2)$ and $\mathbf{X}_{\Lambda_3} = (M, \Lambda_1, \Lambda_2, \Lambda_3)$, respectively.

In all cases, we train a CNN to infer the probabilistic segmentation $\tilde{\mathbf{Y}} : \mathbb{R}^3 \to [0, 1]^{17}$, being made up by seventeen mutually exclusive class label maps: $\tilde{Y}_1, \tilde{Y}_2, ..., \tilde{Y}_{17}$, including sixteen foreground classes[2]. Metrics are computed with respect to the discrete, predicted segmentation which maximises conditional pixel-wise probability inferred by CNN. We denote this as $\hat{\mathbf{Y}} : \mathbb{R}^3 \to \{0, 1\}^{17}$. We also allow subscripts to identify anatomical segments by their abbreviation: for example, $\hat{Y}_{\text{LV}}$ and $\hat{Y}_{\text{RV}}$ are the class label maps of the left and right ventricles respectively.

[2]Note that due to low frequency of appearance in ELCH data, we subsume labels for both patent ductus arteriosus and Blalock-Taussig shunt into the aorta class.



$\mathbf{Y}$ Y_{LH},Y_{RH} Y_{WH}

Figure 6.3: We assess performance of different sets of cardiac labels, each relevant to a variety of downstream applications of segmented data. Left, $\mathbf{Y}$ is the multi-class blood pool and myocardium. Central, Y_{LH} (red) Y_{RH} (blue) are the left and right heart. Right, Y_{WH} is the whole heart blood pool.

In addition to these individual labels, we consider groupings of the blood pool components of the left and right heart anatomy: $\hat{Y}_{\text{LH}}$ and $\hat{Y}_{\text{RH}}$. Each is formed by the union of pixels from all associated anatomical structures, with the sixteen foreground classes divided as follows. Left heart structures include: aorta, Damus-Kaye-Stansel (DKS) connection, left atrium (LA), left pulmonary veins (LPVs), left ventricle (LV) and right pulmonary veins (RPVs). Right heart structures include: inferior vena cava (IVC), left pulmonary artery (LPA), left superior vena cava (LSVC), main pulmonary artery (MPA), right atrium (RA), right pulmonary artery (RPA), right superior vena cava (RSVC), right ventricle (RV) and total cavopulmonary connection (TCPC). The final foreground class describes the myocardium bounding the union of ventricular blood pools. Further, we refer to the combination of $\hat{Y}_{\text{LH}}$ and $\hat{Y}_{\text{RH}}$ as the whole heart blood pool, $\hat{Y}_{\text{WH}}$ (in essence, the union of all foreground classes except the myocardium, see Figure 6.3). This form and its associated notation are mirrored by the ground truth segmentation, $\mathbf{Y}$; its mutually exclusive class label maps, Y_i; and left, right and whole heart groupings, Y_{LH}, Y_{RH} and Y_{WH}.

6.4.2 Representation of 4D data

We intend to feed proxy representations of TR-MRA data as input to the CNN. We consider these in two groups:

The first group is based on the pixel-wise statistics describing the temporal distribution of intensities, the mean and standard deviation[3]:

$$A_\mu = \frac{\sum_t A_t}{T} \tag{6.4.1}$$

$$A_\sigma = \sqrt{\frac{\sum_t (A_t - A_\mu)^2}{T}} \tag{6.4.2}$$

For the most part, A_μ and A_σ reflect the (temporally) *global* enhancement characteristics of the cardiovascular blood pool. All temporal information is reduced to a single volume, returning a coarse representation of cardiac anatomy with high sensitivity: enhancing structures appear bright, the non-enhancing background remains dark (see Figure 6.1). We anticipate that these characteristics provide a source of features relevant to anatomical segmentation. Firstly, A_μ and A_σ approximate the whole heart cardiac blood pool at low spatial resolution, capturing its localisation and crude geometry. Secondly, these images (and TR-MRA data more generally) complement the deficits in image quality commonly associated with 3D SSFP data. Where the latter is frequently compromised by flow artefacts, A_μ and A_σ visualise pulmonary vasculature with high contrast. Similarly, thanks to the introduction of exogenous contrast agent, TR-MRA often appears less affected by metallic susceptibility artefacts. Relevant to the ELCH dataset, this can improve the visualisation of anatomy within close proximity to vascular stents.

The second group of TR-MRA proxy representations exploits PCA to reduce the temporal dimensionality of each angiographic sequence (typically containing between nine and eleven volumes, see Table 5.4). Where N is the number of pixels in each A_t, we consider the $N \times T$ design matrix, $\mathbf{A}^\dagger$, achieved by reshaping $\mathbf{A}$. In essence, $\mathbf{A}^\dagger$ lists the enhancement curves of every pixel in $\mathbf{A}$. In this context PCA seeks a change of basis (mediated by linear transformation $\mathbf{P}$), that maximally

[3]More precisely, this should be referred to as the uncorrected, or sample standard deviation. However, for brevity we refer simply to the standard deviation throughout this chapter.

explains the different modes of enhancement expressed by the pixels of $\mathbf{A}$:

$$\mathbf{\Lambda}^{\dagger} = \mathbf{P}\mathbf{A}^{\dagger} \tag{6.4.3}$$

We recover the transformed volume series $\mathbf{\Lambda} : \mathbb{R}^3 \rightarrow \mathbb{R}^T$, by the reverse of the reshaping operation associating $\mathbf{A}$ with $\mathbf{A}^{\dagger}$. This approach hopes to achieve a transformed space in which each of the T successive volumes $\Lambda_1, \Lambda_2, ..., \Lambda_T$, reflect a different, characteristic mode of enhancement. Compared with $\mathbf{A}$, we reduce dimensionality by considering the first three modes of variation only: $\Lambda_1, \Lambda_2, \Lambda_3$. This is a principled choice, observing the modes of enhancement discussed in Section 6.2.1. On average, these explain over 95% of the total variance in dynamic enhancement.

Consequently, where A_μ and A_σ reflect *global* enhancement (whether a pixel describes an enhancing structure or not), $\Lambda_1, \Lambda_2, \Lambda_3$ capture the *differential* enhancement of cardiac substructures. By the arguments presented in Section 6.2.1, we anticipate that this may aid the discrimination of anatomical classes, and indicate the presence of defects. Our investigation will test whether transformation by PCA, a pragmatic, low dimensional treatment of 4D data, returns a source of features consistent with these aims.

6.4.3 Experiments

All experiments in this chapter make use of the ELCH dataset. Ensuring equal representation of each diagnostic group, the dataset of 150 pairs of images and labels were randomly split between training and test sets in the ratio 2:1. This achieved a training set of 100 examples, exceeding the quantity of data made available for the Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease (Pace et al., 2015) (HVSMR) Challenge by a factor of ten. Correspondingly, our held out test set of fifty examples was a factor of five larger than the number of cases used for HVSMR evaluation. Prior to experimentation, all data were resampled to an isotropic spatial resolution of 1.00 mm, and normalised to have zero mean and unit variance.

Inspired by the nnU-Net framework for self-configuring deep learning (Isensee et al., 2021)[4], we make segmentation predictions using an ensemble of models, trained by five-fold cross-validation. At each fold, a validation set of twenty training cases was withheld. Training cases were randomly divided between validation sets without replacement, and after stratifying by diagnostic group. Final predictions were made by averaging individually inferred probabilistic segmentations.

For its ubiquity and state of the art performance, we adopt the 3D U-Net architecture (Çiçek et al., 2016) in all experiments. Each model was trained using an equally weighted combination of cross-entropy and generalised Dice losses (Sudre et al., 2017) for 24,000 iterations. Auxiliary classification of all but the lowest resolution features learned in the expanding portion of the U-Net, allowed deep supervision (Lee et al., 2015). Each contribution to the loss was weighted as per the nnU-Net formalism (Isensee et al., 2021). Optimisation was mediated by stochastic gradient descent including Nesterov momentum ($\mu = 0.99$). An initial learning rate of 0.01 was decayed according to the the poly learning rate policy (Chen et al., 2017). Each minibatch contained two large volumetric patches of dimension $128 \times 128 \times 128$. We made use of intensive data augmentation, precomputing 200 examples per case. Transformations included: rotation about all three spatial axes ($[-10°, 10°]$), scaling ($[0.9, 1.1]$) and non-rigid deformation, using Torchio (Pérez-García et al., 2021). All CNN experiments were implemented using PyTorch (Paszke et al., 2019).

6.4.4 Metrics

The following metrics can broadly be grouped into those used widely throughout the image processing literature, and those which we have adapted specifically for our application. We refer to these as *technical* and *anatomical* metrics respectively.

Technical metrics

Being task agnostic, the following technical metrics are commonly used throughout the image segmentation literature, assessing performance against the ground truth

[4]Whilst design choices and hyperparameter settings were inspired by the nnU-Net framework (Isensee et al., 2021), we favoured a bespoke implementation for its flexibility and efficiency.

reference standard. They concern well-established technical aspects of performance including spatial overlap and surface localisation. We measure spatial overlap performance between ground truth and predicted image segments, Y_i and $\hat{Y}_i$, using the Dice similarity coefficient (Dice, 1945) (DSC):

$$\mathrm{DSC}(Y_i, \hat{Y}_i) = \frac{2|Y_i \cap \hat{Y}_i|}{|Y_i| + |\hat{Y}_i|} \tag{6.4.4}$$

Where $\cap$ indicates pixel-wise set intersection and $|\cdot|$ is the cardinality of the set contained. A summary measure for multi-class problems, we rely on the generalised Dice similarity coefficient (Crum et al., 2006) (GDSC) as an objective measure of spatial overlap performance across foreground anatomical segments:

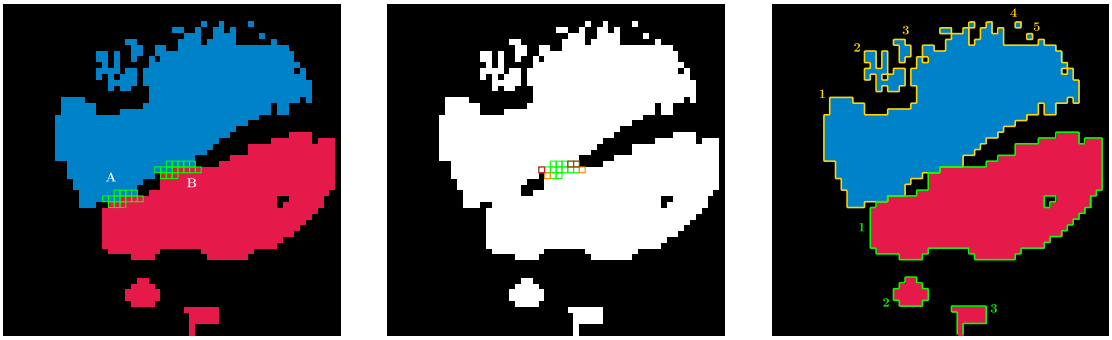
$$\mathrm{GDSC}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{2\sum_i |Y_i \cap \hat{Y}_i|}{\sum_i |Y_i| + |\hat{Y}_i|} \tag{6.4.5}$$

To align our results with previous work (Pace et al., 2015; Yu et al., 2017a,b; Wolterink et al., 2017; Du et al., 2020), we also report a whole heart Dice similarity coefficient (Dice, 1945): $\mathrm{DSC}_{\mathrm{WH}} = \mathrm{DSC}(Y_{\mathrm{WH}}, \hat{Y}_{\mathrm{WH}})$.

Defect visualisation (including extent, morphology and locale) depends on the accurate localisation of the anatomical surfaces composing the heart. We assess surface localisation by the bidirectional Hausdorff distance (Huttenlocher et al., 1993) (HDD) between ground truth and predicted image segments, Y_i and $\hat{Y}_i$:

$$\begin{aligned} \mathrm{HDD}(Y_i, \hat{Y}_i) &= \mathrm{HDD}(\hat{Y}_i, Y_i) && (6.4.6)\\ &= \max(h(Y_i, \hat{Y}_i), h(\hat{Y}_i, Y_i)) && (6.4.7)\\ h(Y_i, \hat{Y}_i) &= \max_{y_i \in Y_i} \min_{\hat{y}_i \in \hat{Y}_i} \delta(y_i, \hat{y}_i) && (6.4.8) \end{aligned}$$

Where $h(Y, \hat{Y})$ is the directed HDD and $\delta(y, \hat{y})$ is the Euclidean distance between the centre of the voxels y_i and $\hat{y}_i$ contained by Y_i and $\hat{Y}_i$, respectively. As per spatial overlap, we are also concerned with the HDD for the union of blood pool classes, $\mathrm{HDD}_{\mathrm{WH}}(Y_{\mathrm{WH}}, \hat{Y}_{\mathrm{WH}})$.



(a) (b) (c)

Figure 6.4: Schematic representation of clinically relevant, anatomical metrics. (a) The shunt error (SE) is calculated with respect to the number of connected components of pixels, adjacent between the left (red) and right (blue) heart. In this case $\hat{S} = 2$, in including: (A) atrial septal defect (ASD) and (B) ventricular septal defect (VSD). (b) The VSD boundary intersection (VBI) finds the association between the ground truth VSD (highlighted in green orange and red pixels) and the boundary of the predicted whole heart blood pool (white). Whilst the ground truth whole heart blood pool is not shown explicitly, its form can be inferred by consideration of the terms of Equation 6.4.13: Y_{VSD} is the union of green, orange and red pixels; $Y_{\text{VSD}} \cap \beta(Y_{\text{WH}})$ is the union of orange and red pixels; and $Y_{\text{VSD}} \cap \beta(Y_{\text{WH}}) \cap \beta(\hat{Y}_{\text{WH}})$ is set of red pixels only. In this case, VBI = 0.6. (c) Discontinuities in the left (red) and right (blue) heart are identified by connected component analysis, being two and four, respectively in this example.
Note that all examples depict a single, 2D axial slice from the CMR volume, downsampled to ease visualisation. In reality all metrics are computed throughout the full resolution, 3D volume, considering adjacency and component connectivity on a 26-connected grid.

Anatomical metrics

Visually demonstrated in Figure 6.4, the following metrics are designed to capture the extent to which predicted segmentation returns a clinically meaningful representation of CHD anatomy. In particular, as per the ground truth (see Section 5.3.3), our ambition is that automated segmentation reflects haemodynamic continuity via pixel adjacency. This is key to the accurate delineation or exclusion of a variety of defects.

Accordingly, our first anatomical metric is simply the difference between the number of connections between the left and right heart exhibited by predicted and ground truth segmentations. This metric is sensitive to the presence of defects such as inter-chamber communication, double inlet ventricle and ventriculo-arterial discordance; and to surgical, structural modification, such as septal defect closure or Rastelli repair. Often referred to as shunts, the shunt error (SE) associated with these connections is constructed as follows:

$$\mathrm{SE} = \hat{S} - S \tag{6.4.9}$$

Where S is the number of shunts demonstrated by the ground truth segmentation (and also known *a priori* from the CMR report and reflected in Table 5.2 and Table 5.3), and $\hat{S}$ is the number shunts presented by the predicted segmentation. $\hat{S}$ is automatically determined by: finding all voxels within the left heart ($\hat{Y}_{\mathrm{LH}}$) that are adjacent to a voxel in the right heart ($\hat{Y}_{\mathrm{RH}}$), and vice versa; separating the resulting set into 26-connected components; and counting the number of distinct clusters which result:

$$S = \kappa(\phi(\hat{Y}_{\mathrm{LH}}, \hat{Y}_{\mathrm{RH}})) \tag{6.4.10}$$

Where the operator $\phi(U, V)$ returns the voxels that are adjacent between sets U and V; and $\kappa(W)$ returns the number of connected components of W. A schematic representation of S is given in Figure 6.4a.

The most frequent shunt within the ELCH cohort, we subject ventricular septal defects (VSDs) to further scrutiny. We anticipate that the insight gained from the associated metrics will be representative of the capacity of CNNs to capture anatomical interfaces more generally. Firstly, defining the VSD as the union of pixels that are adjacent between left and right ventricles (see pixel set B in Figure 6.4a), $Y_{\mathrm{VSD}} = \phi(Y_{\mathrm{LV}}, Y_{\mathrm{RV}})$ we consider metrics of spatial overlap and surface localisation, specific to these defects:

$$\mathrm{DSC}_{\mathrm{VSD}} = \mathrm{DSC}(Y_{\mathrm{VSD}}, \hat{Y}_{\mathrm{VSD}}) \tag{6.4.11}$$

$$\mathrm{HDD}_{\mathrm{VSD}} = \mathrm{HDD}(Y_{\mathrm{VSD}}, \hat{Y}_{\mathrm{VSD}}) \tag{6.4.12}$$

Secondly, we consider a metric relevant to the particular downstream application of patient-specific 3D printing. In this segmentation use case, cardiac anatomy may be judiciously demonstrated as the union of blood pool sub-components. Once this combination is made, it is not so important that the individual left and right ventricles are labelled with high spatial overlap and accurate surface localisation. Rather, it is important that their union has these properties. Accordingly and in this scenario, defining the predicted VSD as the set of pixels that are adjacent between predicted LV and RV segmentations might be overly punitive. Instead, we compare the ground truth VSD, with the surface of the predicted blood pool union. The VSD boundary intersection (VBI) metric finds the fraction of the pixels on the boundary of Y_{VSD} shared by the boundary of $\hat{Y}_{\text{WH}}$:

$$\text{VBI}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{|Y_{\text{VSD}} \cap \beta(Y_{\text{WH}}) \cap \beta(\hat{Y}_{\text{WH}})|}{|Y_{\text{VSD}} \cap \beta(Y_{\text{WH}})|} \tag{6.4.13}$$

Where $\beta(U)$ returns the set of pixels on the boundary of the binary segmentation U. In essence, the VSD boundary intersection relaxes the condition that LV and RV be segmented faithfully, instead assessing the accuracy with which the predicted, whole heart blood pool delineates VSDs. A schematic representation is given in Figure 6.4b.

Where the preceding metrics relate to the inter-connectivity of the left and right heart, the final metric considered concerns the continuity of each of the great and small circulations separately. Here, continuity describes the flow of blood between successive anatomical components. In the normal heart, each of the left (pulmonary veins, LA, LV, aorta) and right (systemic veins, RA, RV, pulmonary arteries) circulations are respectively continuous: each is composed by a single connected component. Discontinuities isolate these flows and increase the number of connected components. Practically, discontinuities are associated with the presence of congenital defects such as valvular atresia, and surgical modifications including superior and total cavopulmonary connections. Our final anatomical metric is the difference between the number of discontinuities in ground truth and predicted circulations. The discontinuity error (DE) is constructed as follows:

$$\text{DE}_{N \in \{\text{LH,RH}\}} = \hat{D}_N - D_N \tag{6.4.14}$$

Where $\mathrm{DE}_{\mathrm{LH}}$ and $\mathrm{DE}_{\mathrm{RH}}$ are the discontinuity differences within the left and right heart circulations respectively; and D_N and $\hat{D}_N$ are the number of discontinuities in the relevant circulations of ground truth and predicted segmentations, respectively. Whilst the number of discontinuities in ground truth circulations (D_{LH} and D_{RH}) are known *a priori*, the number of discontinuities in the predicted cardiac anatomy are found as follows (see Figure 6.4c):

$$\hat{D}_{N\in\{\mathrm{LH,RH}\}} = \kappa(\hat{Y}_N) - 1 \tag{6.4.15}$$

6.4.5 Clinically relevant covariables

Chapter 5 demonstrated the heterogeneity of the CHD population, including significant demographic and structural variation. In response, we seek to understand the dependence of CNN-based segmentation performance on clinically relevant covariables. Perhaps most obviously, we report technical performance per diagnostic category of the ELCH dataset. However, we also consider the following dependencies on continuous covariables.

Section 5.4.5 demonstrated that the ratio of LV to RV volume is characteristic of the defects associated with CHD. Therefore, to crudely characterise structural heterogeneity, we define ventricular imbalance as:

$$\delta V = \left| 1 - \frac{V_{\mathrm{LV}}}{V_{\mathrm{RV}}} \right| \tag{6.4.16}$$

Where V_{LV} and V_{RV} are the left and right ventricular volumes respectively, δV measures the deviation from ventricular balance ($V_{\mathrm{LV}} = V_{\mathrm{RV}}$).

Section 5.4 demonstrated that the heterogeneity of the CHD population is not limited to differences in diagnosis, anatomical structure and disease morphology, but also extends to fundamental patient characteristics, such as age and weight. Moreover, the challenges of CMR acquisition in small children (Ntsinjana et al., 2011) are well established. Hence, we consider patient weight at the point of scan, as an explanatory covariable of segmentation performance.

Finally, Table 5.4 highlighted the variety of acquisition parameters associated with the CMR data composing the ELCH dataset. More generally, we anticipate

that the distribution of 3D SSFP image quality (Q_M) exhibited by these data will affect CNN performance. As a measure of image quality, we consider the standard deviation of pixel intensities captured under the whole heart blood pool of each ground truth segmentation, σ_{WH}.

$$Q_M = 1 - \sigma_{\text{WH}} \tag{6.4.17}$$

To allow comparison between images, this is calculated with respect to the normalised image data with zero mean and unit variance. By this measure, we sensitise our analysis to spatial inhomogeneities in signal intensity and the presence of artefacts, both of which manifest as deviations from the mean blood pool intensity.

6.4.6 Statistical analysis

Across different CNN inputs and clinically relevant covariables, our experimental setting invites a number of comparisons. We assess these statistically using the following tests. To compare performance for different CNN inputs (that including only morphological, with those also admitting TR-MRA data), we use Wilcoxon signed rank test, after Bonferroni correction. As shall become clear in Section 6.5.2, we adopt the same test statistic to compare performance before and after segmentation post-processing.

Making the same comparison between the diagnostic groups of the ELCH dataset, we use the Kruskal-Wallis H-test for omnibus analysis, prior to post-hoc Dunn tests with the same correction for multiple comparison. Where categorical covariables of performance are limited to two groups (comparing performance for patients on biventricular and univentricular pathways), we again use the Wilcoxon signed rank test. To understand the relationship between CNN performance and clinically relavant covariables we use linear regression, assessing the normality of residuals to inform our reporting of correlation coefficients. In all cases, we adopt $p = 0.05$ as our threshold for statistical significance.

Table 6.1: Technical metrics summarising the multi-class and whole heart blood pool segmentation performance of CNN prediction. Spatial overlap is characterised by the generalised and whole heart Dice similarity coefficients (GDSC and DSC_{WH}); surface localisation is conveyed using the whole heart Hausdorff distance (HDD_{WH}). All results reflect performance on the held out ELCH test set ($n = 50$) and are presented as $P_{50}(P_{25}, P_{75})$ where P_i indicates the ith percentile.

Input data	GDSC		DSC_{WH}		HDD_{WH} / mm	
$\mathbf{X}_M = (M)$	0.854	(0.825, 0.884)	0.899	(0.872, 0.919)	16.4	(13.7, 19.1)
$\mathbf{X}_\mu = (M, A_\mu)$	0.861	(0.834, 0.888)	0.911	(0.891, 0.925)	15.1	(13.1, 17.0)
$\mathbf{X}_\sigma = (M, A_\sigma)$	0.865	(0.826, 0.887)	0.911	(0.893, 0.925)	14.7	(13.2, 17.7)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1)$	0.859	(0.826, 0.888)	0.911	(0.887, 0.925)	14.6	(13.2, 16.4)
$\mathbf{X}_{\Lambda_2} = (M, \Lambda_1, \Lambda_2)$	0.853	(0.830, 0.887)	0.912	(0.885, 0.927)	13.8	(12.1, 16.7)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1, \Lambda_2, \Lambda_3)$	0.853	(0.826, 0.883)	0.914	(0.883, 0.927)	13.8	(12.3, 16.8)

6.5 Results

6.5.1 Technical performance

Baseline performance

At least with respect to spatial overlap Table 6.1 demonstrates the strong performance of CNN-based segmentation on the ELCH test set. Though broadly reflecting normal anatomy or that affected by acquired disease, leading submissions to the 2017 Multi-Modality Whole Heart Segmentation (Zhuang et al., 2019) (MM-WHS) Challenge test set (of similar size: $n = 40$; image data: 3D SSFP; and task: multi-class cardiac segmentation) achieved a GDSC of around 0.87 (Zhuang et al., 2019). Comparatively, the ELCH dataset exhibits significant structural variation and imposes the demanding requirement that extracardiac vasculature be delineated. Given these differences, a median GDSC of 0.854 for the segmentation of CHD anatomy from 3D SSFP data appears a reasonable baseline.

Methodologically comparable submissions to, and since the HVSMR Challenge have achieved whole heart Dice scores in excess of 0.92 (Pace et al., 2015; Yu et al., 2017a; Du et al., 2020). Comparatively, our results for the ELCH dataset lag behind, achieving a median DSC_{WH} of only 0.899. Whilst both datasets nominally reflect the CHD population and consider identical acquisitions, Section 5.4.5

demonstrated the greater structural variation captured by the ELCH data. As previously mentioned, the challenge of segmenting the distal pulmonary vasculature likely also accounts for this gap. In any case, we find it unlikely that algorithmic performance could adequately be assessed in the HVSMR test set of just ten cases, especially when drawn from such a heterogeneous patient population. When coupled with differences in task specification, image quality and anatomical heterogeneity, the dependence of the HDD on extreme points of disagreement precludes meaningful comparison with previous works. In general, however, we are satisfied that the first row of Table 6.1 reflects a strong baseline for further comparison.

Performance when exploiting 4D data

Overall, predicting segmentation probability conditioned on both 3D SSFP and 4D TR-MRA data improves performance. The second and third rows of Table 6.1 suggest that the inclusion of either the pixel-wise, temporal mean ($\mathbf{X}_\mu$) or standard deviation ($\mathbf{X}_\sigma$) of TR-MRA aids prediction, improving all metrics compared with $\mathbf{X}_M$. By Wilcoxon signed rank test, and after Bonferroni correction for multiple comparison ($n = 15$), this improvement is statistically significant for both inputs and across all metrics ($p < 0.003$ in all tests).

The benefit of a temporal proxy based on the pixel-wise principal components of variation, however, is less clear. Row four suggests that as per $\mathbf{X}_\mu$ and $\mathbf{X}_\sigma$, concatenating the first principal component of temporal variation ($\mathbf{X}_{\Lambda_1}$) to the CNN input realises statistically significant improvements when compared with the segmentation of 3D SSFP alone: $p < 0.001$ for comparisons of all metrics. For $n > 1$, however, and albeit without statistical significance ($p > 0.1$ in both cases), $\mathbf{X}_{\Lambda_n}$ actually degrades the median GDSC. On the other hand, the inclusion of $\mathbf{X}_{\Lambda_2}$ and $\mathbf{X}_{\Lambda_3}$ yield statistically significant improvement in $\mathrm{DSC}_{\mathrm{WH}}$ and $\mathrm{HDD}_{\mathrm{WH}}$ ($p < 0.0005$ in all cases). Moreover, these inputs return the strongest performance according to metrics relevant to the whole heart blood pool: achieving median Dice scores in excess of 0.912 and an HDD of 13.8 mm. In the latter case, the gains of $\mathbf{X}_{\Lambda_2}$ over $\mathbf{X}_\mu$ and $\mathbf{X}_\sigma$ are even statistically significant: $p < 0.004$ in both cases. Taken together, these results suggest that whilst $\mathbf{X}_{\Lambda_2}$ and $\mathbf{X}_{\Lambda_3}$ might provide more salient features for distinguishing the blood pool from background, the same are

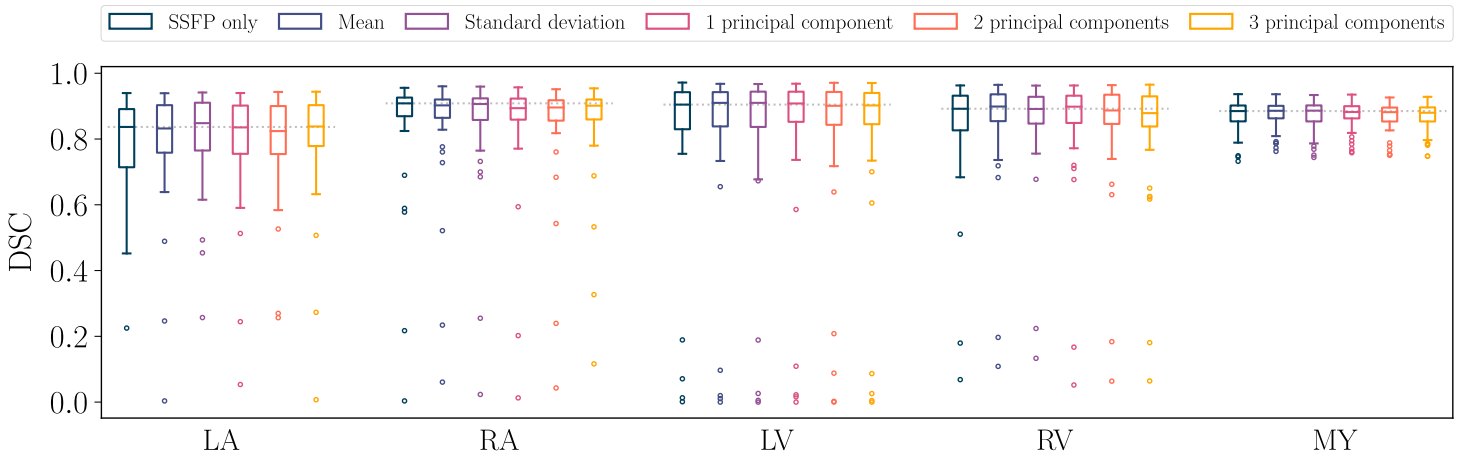
not as relevant to the discrimination of one cardiac sub-structure from another. This is a direct contradiction of our motivation for their inclusion, and discussed further in Section 6.6.2.

Performance across cardiac sub-structures

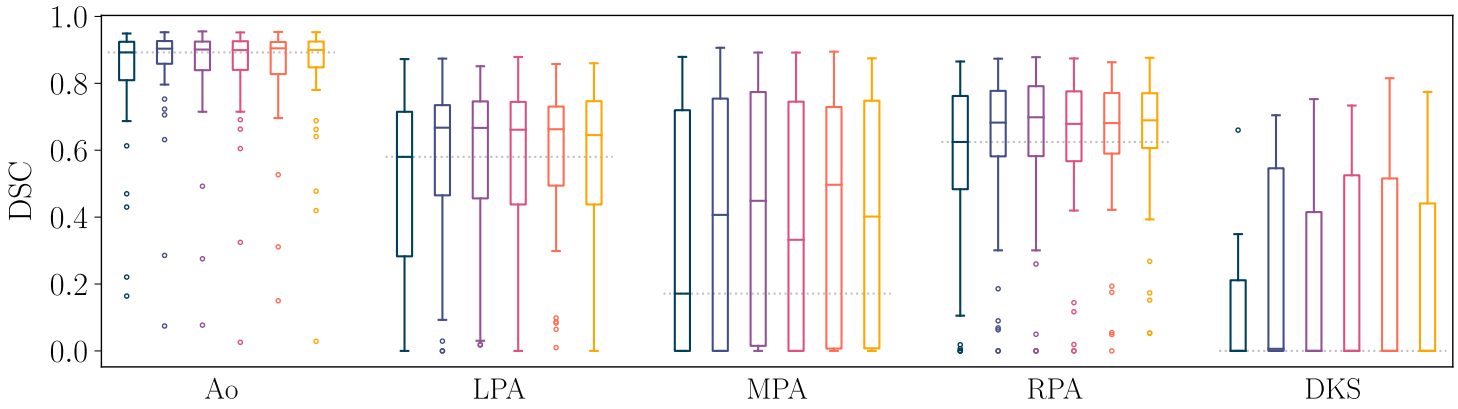
Figure 6.5 breaks down the GDSC results of Table 6.1 into cardiac sub-classes. Individual segmentation labels are grouped into: (a) chambers and ventricular myocardium; (b) arterial vasculature; and (c) venous vasculature. In so doing, this grouping suggests that the improvement associated with the inclusion of TR-MRA within CNN input is most substantial in the segmentation of vascular sub-components. This can be observed in the qualitative results presented in Figure 6.6b: in particular, note the improved segmentation of the right ventricular outflow tract (segmented as part of the RV) and main pulmonary artery at the 25^{th} percentile of performance (P_{25}). This supports our rationale for the inclusion of TR-MRA data, given their superior visualisation of the pulmonary vasculature compared with 3D SSFP.

More generally and irrespective of CNN input, Figure 6.5 makes clear the superior spatial overlap performance achieved in the delineation of cardiac chambers, when compared with their vascular counterparts. Undoubtedly, this result is influenced by the association between Dice score and surface area to volume ratio of the target; the latter tending to be lower for chambers, inflating apparent spatial overlap. However, performance also depends on the representation of anatomical structures within the ELCH dataset (training and test), in terms of both frequency and geometrical complexity.

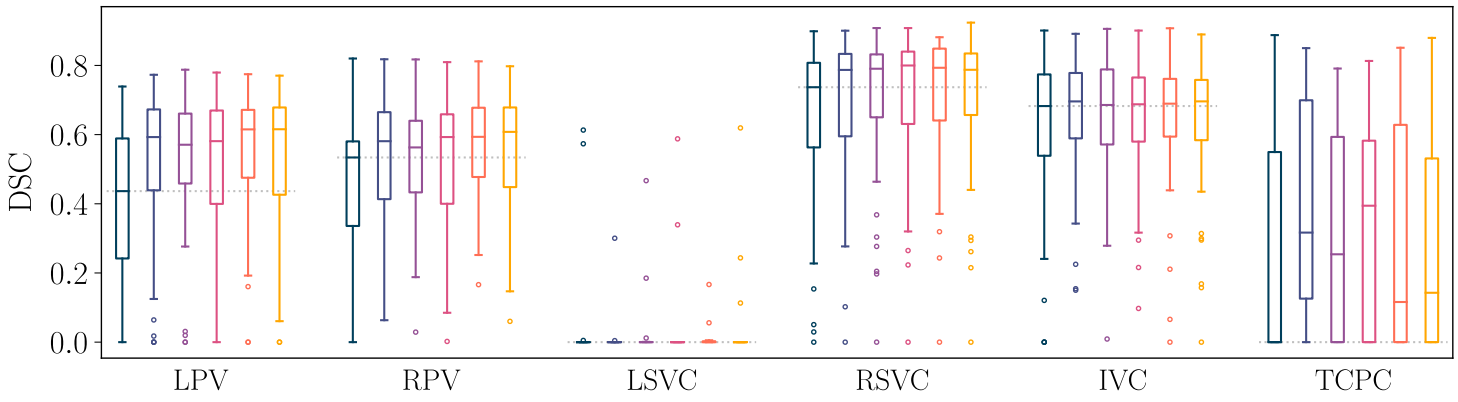
Predictably, classes which appear with higher frequency are better segmented than those characterised by fewer examples. All atrial and ventricular anatomy are present in 99.6% of cases, with the LV absent in just three. Comparatively, no case includes all vascular components: for example, just 10% of patients from the ELCH data exhibit left superior vena cava or bilateral superior vena cavas. This difference contributes to the superior segmentation performance of chamber anatomy when compared with the vasculature.



(a) Cardiac chambers and ventricular myocardium



(b) Extracardiac arterial vasculature



(c) Extracardiac venous vasculature

Figure 6.5: Per class spatial overlap performance for CNN-based segmentation of the ELCH test set.

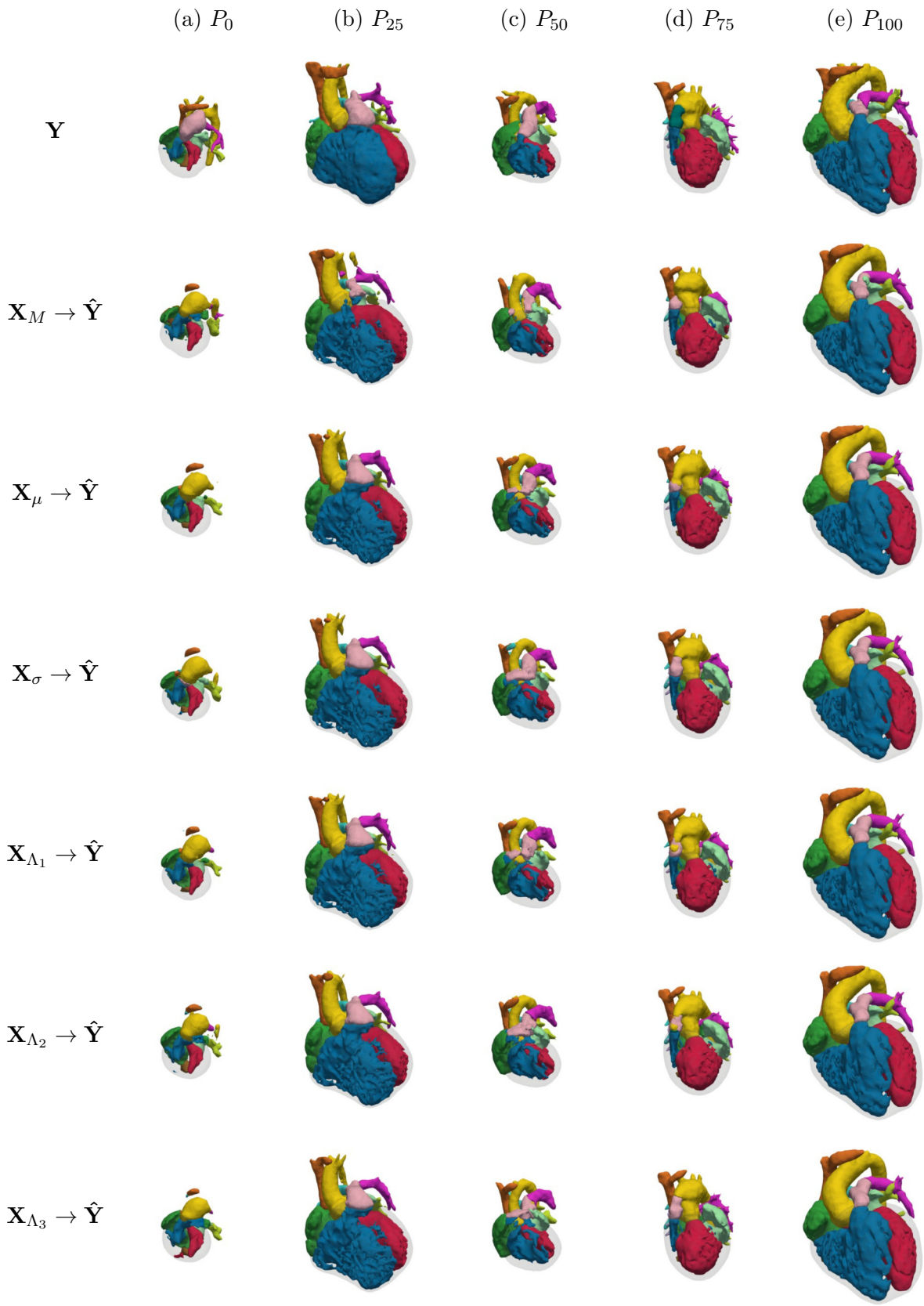


Figure 6.6: Qualitative results of CNN-based segmentation of the ELCH test set. Cases are selected to represent the ith percentile in GDSC for each input, as indicated by P_i.

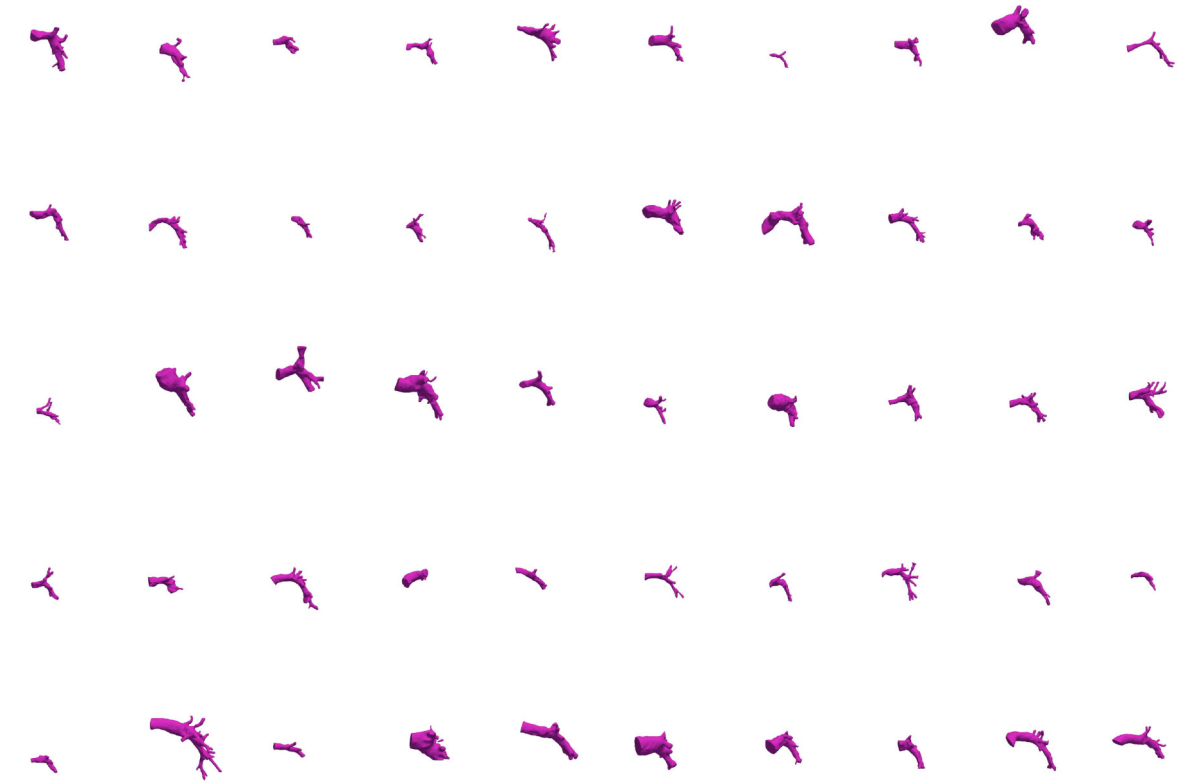
This effect is also apparent *within* anatomical groupings. In Figure 6.5b, the abstract representation of the aorta is learned from 100 training cases, whereas the DKS connection appears only 23 times. Qualitatively, this is reflected in Figure 6.6d: the DKS is correctly included as part of the blood pool, but the CNN cannot determine which vascular label it should be assigned. Depending on CNN inputs, labelling as the aorta, main pulmonary artery, or some combination of the two are all apparent. A more extreme case, in Figure 6.5c, the right superior vena cava is characterised in 100 ground truth segmentations; whereas its left-sided analogue is present in only nine training, and six test cases.

We suggest that other disparities in DSC, including those between structures equally represented in training data, are associated with the relative per class geometrical complexity and heterogeneity. Consider segmentation of the aorta and LPA from 3D SSFP, for example. Both are labelled in all 100 training cases, however the median DSC for the predicted segmentation of the former is substantially higher (0.892 compared with 0.580). Qualitatively, this disparity is explained by the complex branching structure of the LPA, compared with the smooth ascending, arched and descending portions of the aorta. Moreover, allowing for spatially affine transformation, the aorta has a relatively uniform appearance throughout the ELCH dataset. Whereas, variations in both anatomy and image quality make the geometry of the segmented LPA far more heterogeneous, including highly variable branching structures (see Figure 6.7).

We suspect that consistency and complexity of appearance also favours the segmentation of cardiac chambers. Whilst trabeculation of the ventricular cavities introduces a highly irregular interface with the myocardium, this complexity is largely reflected by pixel intensity[5]. Moreover, this complexity accounts for a relatively small fraction of ventricular volume, which otherwise appears bounded by a convex and approximately conical envelope. Finally, heterogeneity in the appearance of ventricular anatomy is largely limited to affine changes in spatial scale and orientation, modes of variation which are well represented in the ELCH dataset (see Figure 6.8), and enhanced by data augmentation by spatial transformation.

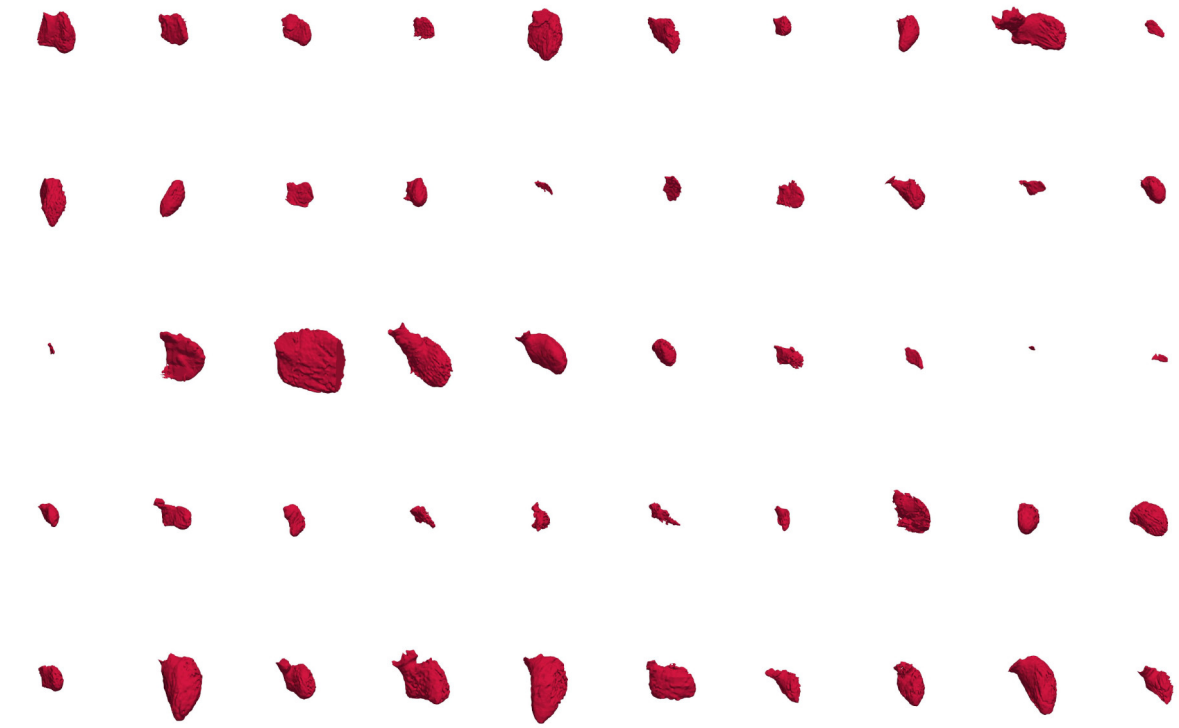
[5]This is apparent to anyone who has manually segmented these data. Where branching vasculature is often defined by manual editing, the trabeculated surface of the ventricular cavities is achieved simply by intensity thresholding.

(a)

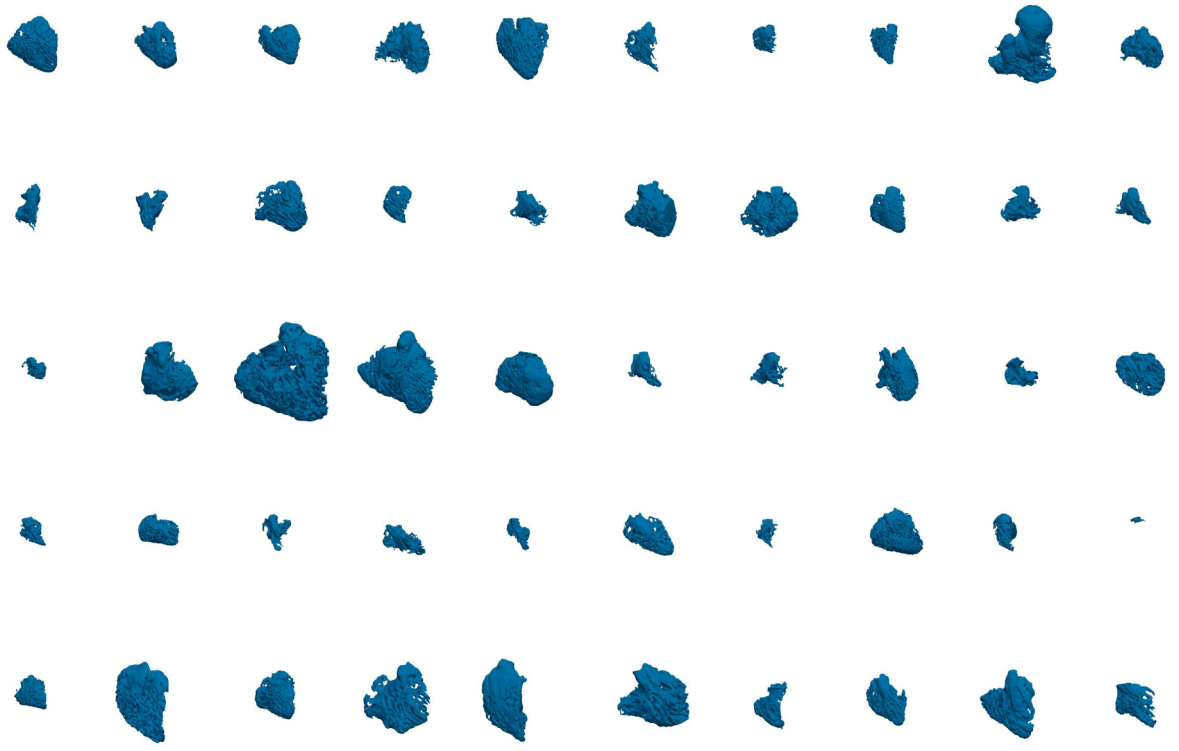


(b)

Figure 6.7: Surface-rendered representation of aorta ((a), yellow) and left pulmonary artery (LPA) ((b), pink) anatomy in the ELCH test set. Notice that while both anatomical components demonstrate large variation in size and orientation, the morphology of the LPA is more heterogeneous. In particular, its branching structure is highly variable compared with the smooth consistent arch of the aorta.



(a)



(b)

Figure 6.8: Surface-rendered representation of the anatomy of the left ((a), red) and right ((b), blue) ventricular cavities in the ELCH test set. Both structures have trabeculated epicardial surfaces (being more pronounced around the RV cavity) and exhibit significant variation in size. However, notice that for the most part, the envelope of ventricular anatomy takes a relatively simple, conical geometry. This allows for improved abstract representation by CNN.

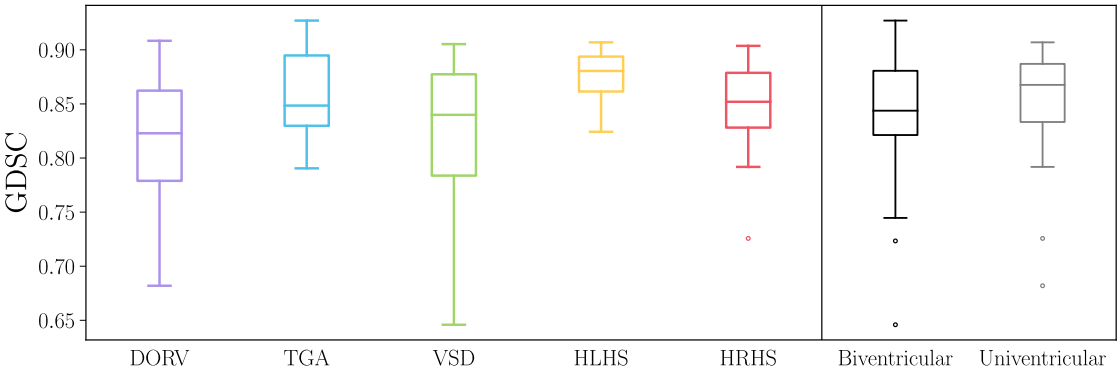


Figure 6.9: A comparison of predicted segmentation performance between patient groups: (*left*) Performance on different diagnostic groups (see Section 5.3.2). (*right*) Performance on different interventional groups: those on a care pathway culminating in (or which has culminated in) biventricular repair, and those receiving univentricular palliation.

Other performance dependencies

Here we consider the dependence of CNN performance on clinically relevant covariables. With respect to the segmentation of 3D SSFP data only ($\mathbf{X}_M$)[6], Figure 6.9 reflects differences in CNN performance between diagnostic groups. However, Kruskal-Wallis H-test indicates no significant differences between the omnibus of diagnostic classes ($p = 0.149$). Whilst statistically insignificant, qualitative differences are apparent. In particular, we note the superior segmentation of cases drawn from the HLHS group compared with those with DORV or VSD: median overlap results are 0.880, 0.823 and 0.840, respectively. For some, this might present a surprising result, especially when it is considered that a greater proportion of the HLHS patients exhibit anatomy consistent with staged univentricular palliation. We may have expected associated complex anatomical modifications (including DKS and total cavopulmonary connection connection), to hinder the accurate delineation of anatomy. Examination of the training set, however, reveals that the ratio of patients progressing through a care pathway culminating (or which has culminated) in biventricular repair, to those receiving univentricular palliation is almost evenly split. At 51:49, the optimised CNNs had approximately

[6]Qualitatively, predicted segmentations of $\mathbf{X}_\mu$, $\mathbf{X}_\sigma$, $\mathbf{X}_{\Lambda_1}$, $\mathbf{X}_{\Lambda_2}$ and $\mathbf{X}_{\Lambda_3}$, exhibit the same differences in performance between diagnostic and interventional groups. Quantitatively, statistical testing results in the same conclusions, irrespective of CNN input.

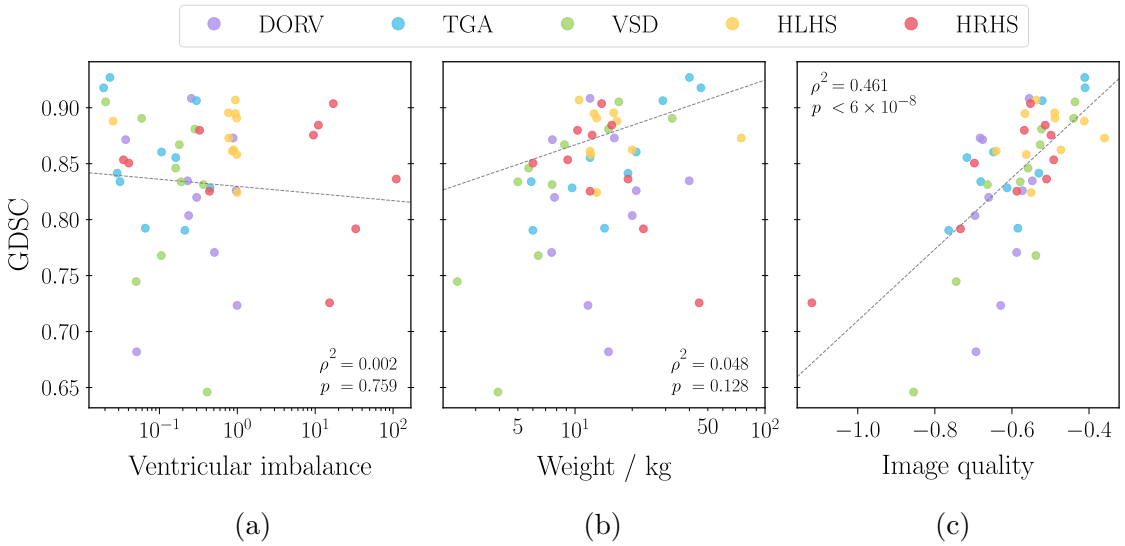


Figure 6.10: Dependence of CNN performance on: (a) ventricular imbalance; (b) patient weight; and (c) image quality. Note that in each case, the dashed grey line indicates the result of linear regression (ρ^2 capturing the proportion of shared variance between the ranked variables as the square of Spearman's ρ; p being the statistical significance), and that spatial overlap performance is measured for predictions conditioned on 3D SSFP data only ($\mathbf{X}_M$). Perhaps predictably, image quality is a strong determinant of CNN performance; perhaps surprisingly, the structural variation associated with CHD (and represented by ventricular imbalance) is not a strong predictor of CNN performance. Please see Section 6.4.5 for a description of how ventricular imbalance (δV) and image quality (Q_M) are determined.

equal exposure to each mode of clinical management, explaining the comparable performance observed in their automated segmentation (see Figure 6.9). This is confirmed by Mann-Whitney U-test: $p = 0.188$.

The apparent insensitivity of CNN performance to diagnostic classification is reinforced by considering the relationship between spatial overlap and ventricular imbalance (see Figure 6.10a). Whilst this plot visualises the linear regression of these data, residuals to the grey dashed line are not normally distributed. Hence, we reflect monotonicity by Spearman's correlation coefficient: $\rho = -0.042$, indicating a very weak association between worsening segmentation performance and ventricular imbalance. Moreover, the apparent insensitivity to structural het-

erogeneity endorses the ELCH dataset as being representative of the underlying distribution of CHD anatomy.

Given the demographic heterogeneity of the CHD population, and as relevant to the segmentation of $\mathbf{X}_M$, Figure 6.10b demonstrates the relationship between spatial overlap performance and patient weight. This indicates a weak association ($\rho = 0.218$) between improved spatial overlap performance and weight. Interestingly, the monotonicity of this relationship is reduced by the addition of TR-MRA data to CNN input (ranging between 0.140 and 0.173 for $\mathbf{X}_\mu$ and $\mathbf{X}_{\Lambda_2}$, respectively). This may confirm our expectation that by virtue of their contrast sensitivity, TR-MRA data provide a source of features relevant to the localisation of the heart, irrespective of patient size.

Figure 6.10c demonstrates a strong correlation between GDSC and image quality: Spearman's $\rho = 0.676$. It is perhaps surprising that despite the significant structural variation associated with CHD (as reflected by diagnostic grouping and patient weight), image quality remains the most influential covariable informing CNN performance. As per its relationship with patient weight, the correlation between spatial overlap performance and 3D SSFP image quality falls after the introduction of TR-MRA data (ranging between 0.588 and 0.649 for $\mathbf{X}_\mu$ and $\mathbf{X}_\sigma$, respectively). This supports our expectation that TR-MRA data complement deficits in 3D SSFP image quality, including the artefacts and spatial inhomogeneities aforementioned.

6.5.2 Clinical performance

Anatomical coherence

Compared with the technical performance of CNN-based segmentation of congenital heart defects, the following clinical assessment paints a very different picture. The anatomical metrics presented in Table 6.2 illustrate the tendency of CNN prediction to include spurious communications between the left and right heart (or shunts, SE), and discontinuities within each of the great ($\mathrm{DE}_{\mathrm{LH}}$) and small circulations ($\mathrm{DE}_{\mathrm{RH}}$), respectively. Depending on task formulation (or equivalently mode of presentation, see Figure 6.3), the left and right heart can be combined in a binary representation of the whole heart blood pool. Given this association, it is

Table 6.2: Clinically relevant, anatomical metrics summarising the segmentation of 3D CMR by CNN. With respect to the configuration of the left and right circulations inferred from clinical CMR report: the shunt error (SE) reflects the inclusion of spurious shunts between the left and right heart; and the discontinuity errors (DEs), $\mathrm{DE}_{\mathrm{LH}}$ and $\mathrm{DE}_{\mathrm{RH}}$, count the number of anomalous discontinuities in the right and left circulation, respectively. All results reflect performance on the held out ELCH test set ($n = 50$) and are presented as $P_{50}(P_{25}, P_{75})$ where P_i indicates the i^{th} percentile.

Input data	SE		$\mathrm{DE}_{\mathrm{LH}}$		$\mathrm{DE}_{\mathrm{RH}}$	
$\mathbf{X}_M = (M)$	6.0	(4.0, 9.0)	4.0	(3.0, 6.0)	8.5	(6.0, 13.8)
$\mathbf{X}_\mu = (M, A_\mu)$	7.0	(5.0, 9.0)	3.0	(2.0, 6.0)	7.0	(5.0, 11.0)
$\mathbf{X}_\sigma = (M, A_\sigma)$	6.5	(5.0, 10.0)	3.0	(2.0, 5.0)	5.0	(4.0, 8.8)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1)$	8.0	(6.0, 9.0)	3.5	(2.3, 6.0)	8.0	(5.0, 10.0)
$\mathbf{X}_{\Lambda_2} = (M, \Lambda_1, \Lambda_2)$	7.5	(5.0, 10.8)	4.0	(2.3, 6.8)	9.0	(6.3, 12.0)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1, \Lambda_2, \Lambda_3)$	7.0	(4.3, 10.0)	4.0	(2.0, 6.8)	8.5	(6.0, 11.0)

perhaps surprising that both these modes of error present simultaneously: $\mathrm{SE} > 0$ being associated with false positive, and $\mathrm{DE}_{\mathrm{LH}}, \mathrm{DE}_{\mathrm{RH}} > 0$, with false negative classification. The coincidence of such errors speaks to a lack of spatial coherence within CNN prediction, and the inference of results which lack anatomically and clinically relevant meaning.

The inclusion of TR-MRA data within CNN input does little to improve clinical performance compared with the SSFP baseline. Amongst these results, Wilcoxon signed rank test (after Bonferroni correction for multiple comparison, $n = 15$) suggests that the only significant improvements are conferred by the additional inclusion of the temporal mean and standard deviation. Compared with $\mathbf{X}_M$, $\mathrm{DE}_{\mathrm{RH}}$ is significantly improved (reduced) by conditioning prediction on $\mathbf{X}_\sigma$ ($p < 10^{-5}$); $\mathrm{DE}_{\mathrm{LH}}$ is reduced by $\mathbf{X}_\mu$ and $\mathbf{X}_\sigma$ ($p < 0.02$ for both). Whilst these differences are statistically significant, associated improvements remain relatively small compared with the total number of circulatory discontinuities. Moreover, the inclusion of the first and second modes of temporal variation actually serve to increase the number of spurious shunts: $p < 0.02$ when $\mathbf{X}_{\Lambda_1}$ and $\mathbf{X}_{\Lambda_2}$ are compared with $\mathbf{X}_M$.

Qualitative performance for the segmentation of SSFP data is illustrated in Figure 6.11. Frequently isolated by thin tissue interfaces (such as the atrial sep-

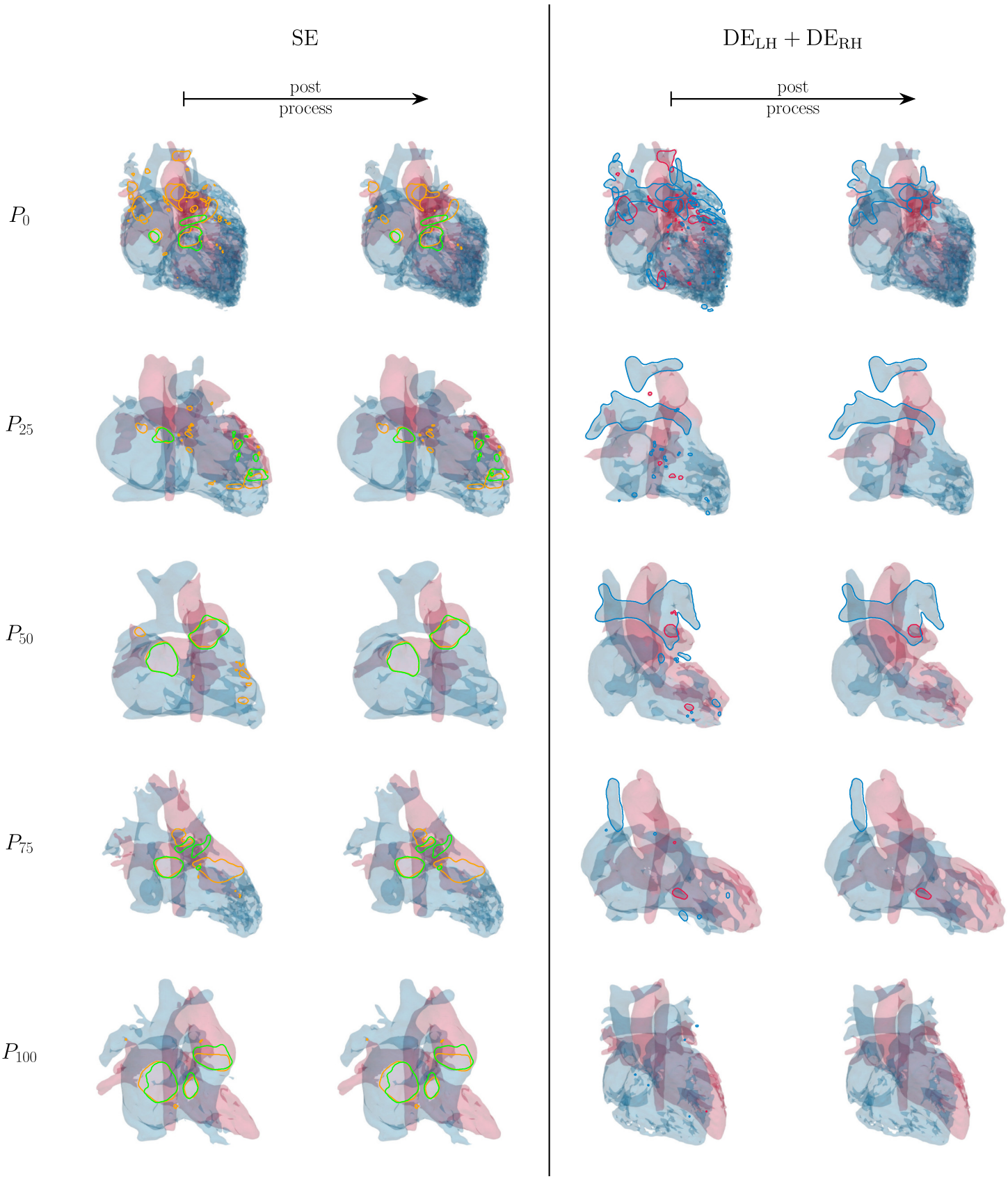


Figure 6.11: Representation of clinically relevant, anatomical metrics of segmentation performance relating: (*left*) the shunts between left and right heart (transparent red and blue, respectively) within the ground truth (orange loops) and predicted (green loops) segmentations; and (*right*) the discontinuous components of the left (red loops) and right (blue loops) circulations of the predicted segmentation. Note that performance percentiles, P_i, apply to the CNN prediction prior to post-processing.

tum), spurious shunts can be introduced by the inclusion of only a small number of false positive pixels. At the 0^{th} percentile (P_0) of performance, this culminates in a large number of anomalous connections between left and right heart, and demonstrates the challenge of delineating morphologically complex anatomy in a clinically meaningful fashion. Even in the median case, six spurious shunts remain.

Turning our attention to circulatory discontinuity, Figure 6.11 demonstrates that such errors can isolate components of variable volume. All examples presented illustrate the anomalous inclusion of trivial extra-anatomical components. However, within examples up until P_{75}, continuity errors can also isolate components of substantial volume, and which otherwise faithfully represent anatomy (sharing significant overlap with the ground truth). Consider the isolation of the pulmonary arteries from the RV at the median level of performance: there, the apparent absence of the right ventricular outflow tract limits the anatomical and clinical coherence of predicted segmentation.

We are far from the first to observe the inclusion of spurious connected components, holes and discontinuous gaps within CNN-based segmentation (Painchaud et al., 2020). Accordingly, previous work has leveraged rudimentary post-processing. This has included connected component analysis (CCA), suppressing all but the largest component per foreground class of $\hat{\mathbf{Y}}$ (Isensee et al., 2021). With the exception of the pulmonary veins, for which we anticipate several components per side (typically but not reliably two on the left; three on the right), such a prior is appropriate to our task. Previous works have also chosen to fill small holes (or voids) predicted within each class. In our case, post-processing via the application of CCA and hole-filling modifies anatomical performance as per Table 6.3[7].

Compared with Table 6.2, post-processing most clearly resolves errors in the continuity of the cardiac blood pool, with particularly pronounced improvement (reduction) in DE_{RH}. We speculate that such gains are more pronounced in the right heart due to the highly trabeculated endocardial surface of the RV. Its complex morphology, including a multiplicity of small, muscular loops, admits many anomalous connected components to CNN prediction, and elevates DE_{RH} above DE_{LH} in Table 6.2. These are subsequently resolved by CCA.

[7]Post-processing has a clinically insignificant deleterious effect on spatial overlap performance, reducing the median GDSC by an amount $\leq 10^{-4}$, irrespective of CNN input.

Table 6.3: Anatomical metrics summarising the segmentation of CMR by CNN after post-processing by connected component analysis (CCA) and hole-filling. With respect to the configuration of the left and right circulations inferred from clinical CMR report: SE reflects the inclusion of spurious shunts between the left and right heart; and DE_{LH} and DE_{RH} count the number of anomalous discontinuities in the right and left circulation, respectively. All results reflect performance on the held out ELCH test set ($n = 50$) and are presented as $P_{50}^{\delta}(P_{25}, P_{75})$ where P_i indicates the ith percentile, and δ is the difference in performance after post-processing, compared with raw CNN prediction.

Input data	SE		DE_{LH}		DE_{RH}	
$\mathbf{X}_M = (M)$	$3.0^{-3.0}$	(2.0, 5.0)	$0.0^{-4.0}$	(0.0, 1.0)	$0.0^{-8.0}$	(0.0, 2.0)
$\mathbf{X}_\mu = (M, A_\mu)$	$4.0^{-2.0}$	(2.3, 6.0)	$0.0^{-3.0}$	(0.0, 1.0)	$0.0^{-5.0}$	(0.0, 1.0)
$\mathbf{X}_\sigma = (M, A_\sigma)$	$3.5^{-2.0}$	(2.0, 5.8)	$0.0^{-3.0}$	(0.0, 1.0)	$0.0^{-5.0}$	(0.0, 1.0)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1)$	$4.0^{-3.0}$	(2.0, 6.0)	$0.0^{-4.0}$	(0.0, 1.0)	$0.0^{-7.0}$	(0.0, 1.8)
$\mathbf{X}_{\Lambda_2} = (M, \Lambda_1, \Lambda_2)$	$4.5^{-3.0}$	(2.3, 7.8)	$0.0^{-3.5}$	(0.0, 0.8)	$0.0^{-8.5}$	(0.0, 1.0)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1, \Lambda_2, \Lambda_3)$	$4.0^{-3.0}$	(2.0, 5.8)	$0.0^{-4.0}$	(0.0, 0.8)	$0.0^{-8.0}$	(0.0, 1.0)

It is striking that whilst median performance is characterised by $DE_{LH}, DE_{RH} = 0$, examples in Figure 6.11 demonstrate cases for which discontinuity errors remain after post-processing. Closer inspection suggests that this scheme lends itself to the elimination of errors associated with the prediction of extra-anatomical components of trivial volume. After removing small, anomalous components, only the discontinuity errors associated with more sizeable objects remain (see Figure 6.12). Figure 6.11 suggests that the latter can be associated with anatomically and clinically meaningful errors in circulatory continuity. Despite this, and irrespective of CNN input, Wilcoxon signed rank test suggests that post-processing significantly improves the meaningful continuity of the cardiac blood pool: $p < 10^{-8}$ for all comparisons of DE_{LH} or DE_{RH}, before and after post-processing.

Whilst the same analysis suggests that rudimentary post-processing also significantly reduces the number of anomalous shunts ($p < 10^{-7}$ for all CNN inputs), the median SE in Table 6.3 remains non-zero and clinically problematic.

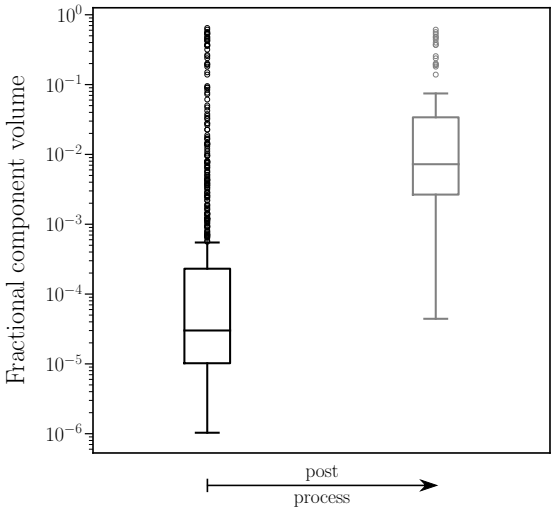


Figure 6.12: The effect of rudimentary post-processing, including CCA, on the volume of extra-anatomical components: those separated from the predicted whole heart blood pool by discontinuity errors with the ground truth cardiac circulation. Note that to allow comparison, volume is presented as a fraction of the ground truth blood pool. These boxplots highlight the ability of CCA to remove extra-anatomical components of small or trivial volume. The distribution of those which remain has a significantly larger fractional volume ($p < 2 \times 10^{-29}$ by Mann-Whitney U-test) compared with raw CNN prediction. As echoed by Figure 6.11, however, its resolution of discontinuous errors involving larger components is limited.

Defect representation

Given the success of CCA in rectifying discontinuity errors in the left and right heart, extending this approach to the interfaces between $\hat{Y}_{\mathrm{LH}}$ and $\hat{Y}_{\mathrm{RH}}$ appears a natural choice. Such a procedure firstly involves the identification of shunts according to pixel adjacency (as per Figure 6.4a), with each considered a connected component. Subsequently, and in a semi-automated framework, shunts are included or excluded from the post-processed prediction according to a prior specifying the number anticipated, S. In the simplest case and as per the description of CCA supplied previously, this scheme selects the S largest shunts predicted by CNN. It assumes that the prediction of spurious shunts is associated with small groups of pixels receiving false positive classification within any of the blood pool classes. Moreover, an extension of this scheme might consider, not only the prediction of shunts between the left and right heart, but further specify the communication of

particular cardiac sub-structures, considering all anatomical classes predicted in $\hat{\mathbf{Y}}$. Leveraging multi-class prediction, this grants sensitivity to different defects, differentiating between atrial and ventricular septal defects, for example.

Irrespective of the class of shunt considered, any attempt to enforce an anticipated configuration of cardiac anatomy upon discrete CNN prediction requires that: (1) candidate shunts can be isolated from one another by pixel adjacency; (2) that the most credible candidates can be differentiated by their size; and (3) that within these candidates, there exists an accurate representation of the ground truth defect or defects concerned. Put another way, it is one thing to manipulate CNN prediction such that it contains the correct *number* of shunts, as we have considered up until now; ensuring that the resulting segmentation meaningfully represents the structure of the underlying anatomy, including the size, morphology and location of any defects, is another. To this end, we now extend our analysis to consider these facets of defect representation. Whilst we limit our investigation to VSDs, being the most frequent defect within the ELCH dataset, we expect our findings to be relevant to the wider array of congenital lesions. As per the previous discussion, and to make best use of the multi-class prediction $\hat{\mathbf{Y}}$, we use CCA to select the V largest VSDs from the predicted segmentation, where V is the number of defects anticipated *a priori* and provided in a semi-automated framework.

Table 6.4 reflects these aspects of performance for such post-processed CNN predictions. It suggests that the deficiencies of CNN-based segmentation extend to the accuracy of defect representation. For the segmentation of 3D SSFP data only, median spatial overlap of predicted and ground truth segmentations does not reach 0.2. Whilst DSC_{VSD} is improved by the inclusion of TR-MRA data within CNN input, Wilcoxon signed rank test (after Bonferroni correction, $n = 15$) does not return a significant result ($p > 0.24$ in all cases). Such poor performance is reinforced by Hausdorff distance (Huttenlocher et al., 1993)$_{\text{VSD}}$ in excess of 7 mm.

Taken together, DSC_{VSD} and HDD_{VSD} reveal the extent to which CNN prediction satisfies the requirements (1-3) of the effective post-processing scheme outlined above. Table 6.4 associates a clear improvement in HDD_{VSD} with the selection of defects from $\hat{\mathbf{Y}}$. This suggests that inferred segmentations satisfy requirements (1) and (2), shortening the HDD by the removal of smaller, false positive defects that are remote and disconnected from ground truth VSDs. Moreover and apart from

Table 6.4: Clinically relevant, anatomical metrics summarising the segmentation of CMR by CNN as relevant to the representation of ventricular septal defect (VSD) after rudimentary post-processing by CCA and hole filling. After identifying VSD as the set of pixels adjacent between left and right ventricles, the $\mathrm{DSC}_{\mathrm{VSD}}$ and $\mathrm{HDD}_{\mathrm{VSD}}$ reflect the Dice Similarity Coefficient and Hausdorff Distance respectively. Whereas, VSD boundary intersection (VBI) assesses the intersection of the boundary of the predicted whole heart blood pool with the rim of any VSD captured by the ground truth segmentation. All results reflect performance on the held out ELCH test set ($n = 50$) and are presented as $P_{50}^{\delta}(P_{25}, P_{75})$ where P_i indicates the ith percentile and δ is the difference in performance after post-processing, compared with raw CNN prediction.

Input data	$\mathrm{DSC}_{\mathrm{VSD}}$	$\mathrm{HDD}_{\mathrm{VSD}}$/mm	VBI
$\mathbf{X}_M = (M)$	$0.193^{+.000}$ (0.009, 0.297)	$7.9^{-6.7}$ (6.4, 22.3)	$0.356^{+.004}$ (0.263, 0.457)
$\mathbf{X}_\mu = (M, A_\mu)$	$0.234^{+.009}$ (0.056, 0.407)	$7.0^{-5.9}$ (5.5, 21.1)	$0.410^{-.001}$ (0.285, 0.472)
$\mathbf{X}_\sigma = (M, A_\sigma)$	$0.276^{+.041}$ (0.091, 0.391)	$7.3^{-7.7}$ (5.5, 20.4)	$0.367^{+.000}$ (0.263, 0.473)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1)$	$0.229^{+.010}$ (0.013, 0.339)	$8.1^{-7.7}$ (5.7, 20.1)	$0.356^{-.047}$ (0.237, 0.447)
$\mathbf{X}_{\Lambda_2} = (M, \Lambda_1, \Lambda_2)$	$0.214^{+.013}$ (0.002, 0.390)	$7.0^{-13.4}$ (6.4, 20.5)	$0.373^{-.002}$ (0.233, 0.473)
$\mathbf{X}_{\Lambda_1} = (M, \Lambda_1, \Lambda_2, \Lambda_3)$	$0.236^{+.012}$ (0.097, 0.357)	$7.2^{-15.5}$ (5.0, 20.5)	$0.393^{+.000}$ (0.275, 0.484)

for $\mathbf{X}_M$, this improvement is statistically significant: $p < 0.02$ for all other inputs. In contrast, post-processing of defects by CCA results in only a meager increase in $\mathrm{DSC}_{\mathrm{VSD}}$, unlikely to be of clinical significance. This implies that CNN prediction is not consistent with requirement (3).

Both $\mathrm{DSC}_{\mathrm{VSD}}$ and $\mathrm{HDD}_{\mathrm{VSD}}$ infer VSD location and morphology from the adjacency of pixels predicted within the left and right ventricular classes of $\hat{\mathbf{Y}}$. Consequently, these metrics demand not only that the rim of the VSD (the interface between the ventricular septum and blood pool) be accurately delineated, but also that the effective interface between left and right ventricles be localised. They are sensitive to the misclassification of pixels between these classes. If, however, our segmentation is motivated by a desire to visualise the whole heart blood pool, such stringent requirement may be unnecessary, and these metrics, overly punitive. Instead, VBI compares the effective delineation of VSDs within ground truth and predicted whole heart blood pools. It is sensitive only to the overlap between the VSD rim in $\mathbf{Y}$ and the blood pool surface of $\hat{Y}_{\mathrm{WH}}$. Predictably, therefore, median VBI exceeds $\mathrm{DSC}_{\mathrm{VSD}}$ for all CNN inputs. As per $\mathrm{DSC}_{\mathrm{VSD}}$, the inclusion of TR-

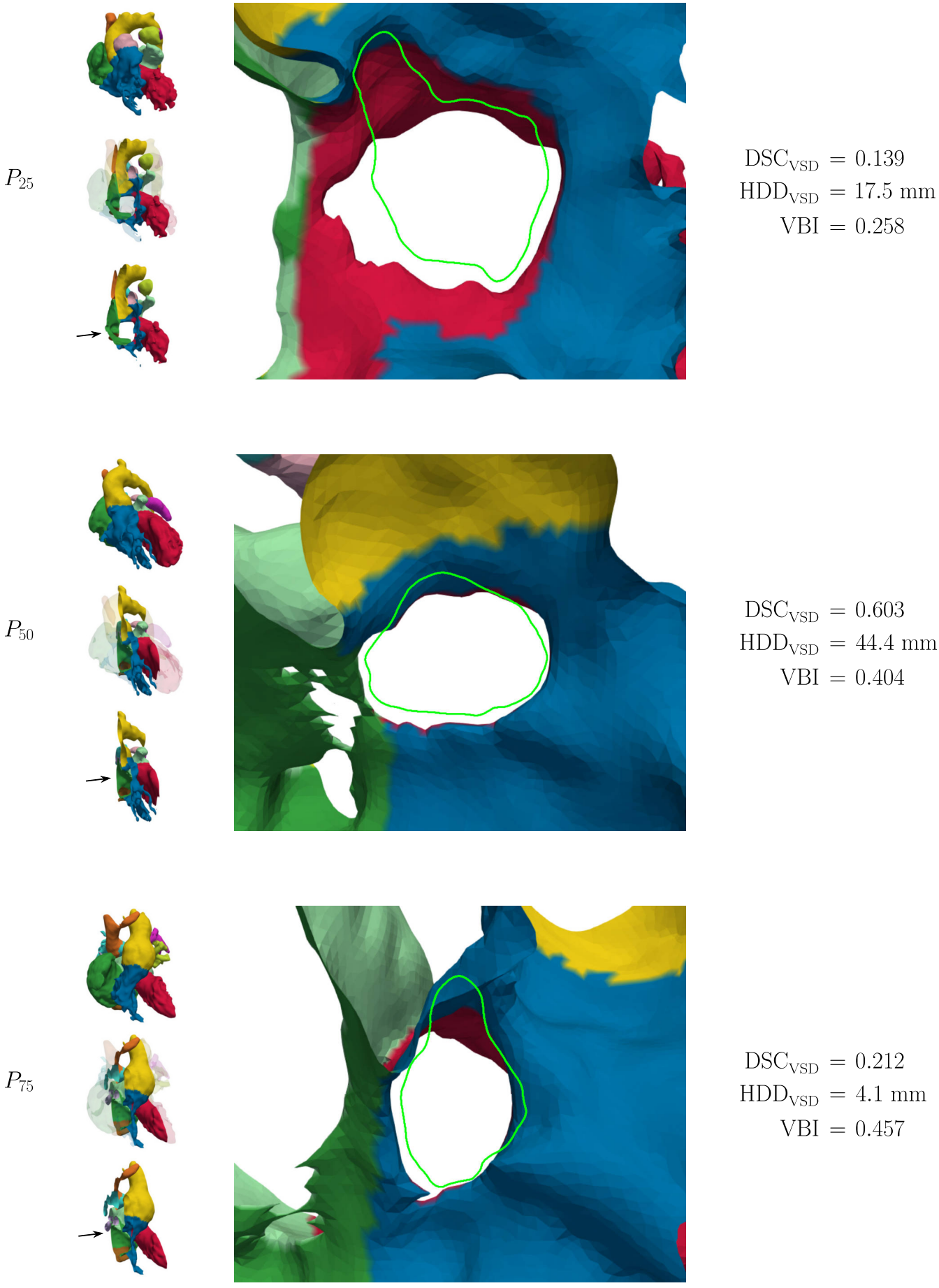


Figure 6.13: Interpretation of anatomical metrics associated with the representation of ventricular septal defect (VSD). The percentiles of performance (P_i) relate to the VSD boundary intersection (VBI) with corresponding Dice similarity coefficients (DSC_{VSD}) and Hausdorff distances (HDD_{VSD}) provided. Each view demonstrates the VSD from the perspective indicated by the respective black arrows, with anterior and posterior sections of the heart removed. Note that the ground truth VSD is indicated by the lime green contour.

MRA data within CNN input appears to improve VBI. Again, however, this is found to be statistically insignificant ($p > 0.1$ for all comparisons).

To aid in the interpretation of DSC_{VSD}, HDD_{VSD} and VBI results, Figure 6.13 visualises the prediction of VSDs in the ELCH test set. The disagreement between ground truth and predicted defects makes raw CNN prediction inadequate for clinical application without manual adjustment.

6.6 Discussion

6.6.1 Context

This chapter spans many aspects of the task of segmenting CHD anatomy from CMR data. These include different task formulations (multi-class and whole heart specifications); CNN inputs (3D SSFP data and various representations of TR-MRA); and performance metrics (including those developed to assess the suitability of segmentation for clinical application). In summarising these developments, we firstly highlight the novelty of our work, it being the first dedicated investigation of multi-class, CNN-based segmentation of 3D CMR data. Whilst previous works have examined the application of deep learning methodologies to congenital CMR, these have either: focused on 2D short axis segmentation for ventricular volumetry (Karimi-Bidhendi et al., 2020), including investigation limited to patients with tetralogy of Fallot (Backhaus et al., 2019; Koehler et al., 2020) or segmentation of the RV only (Giannakidis et al., 2016); deployed the HVSMR training and test sets, totalling just twenty cases (Yu et al., 2017a,b; Wolterink et al., 2017; Li et al., 2017b; Zheng et al., 2019a; Rezaei et al., 2020; Han et al., 2020; Du et al., 2020) (limiting task specification to the segmentation of the whole heart blood pool as a single class); or examined the 3D segmentation of individual anatomical sub-components only (Pace et al., 2018). Experiments involving larger data, including their segmentation into multiple anatomical classes have been limited to X-ray CT imaging (Xu et al., 2019a; Liu et al., 2020a), and have employed the ImageCHD dataset of 110 cases (or its precursors), provided by Xu et al. (2020). Irrespective of imaging modality, none of the previous works cited consider the

segmentation of data into more than seven foreground classes. In contrast, our ambitious experiments seek the separation of anatomy into sixteen classes.

In the body of work outlined above, little attention has been paid to aspects of segmentation performance outside of spatial overlap (most often measured by the DSC with the ground truth) and surface localisation (most often measured by the HDD or mean surface separation). Such assessments are likely sufficient in studies of 2D short axis segmentation. However, for those motivated by the extraction of 3D patient-specific anatomical models for treatment or surgical planning (a majority of those cited above), we argue that a closer examination of anatomical coherence and defect representation is critical (see Section 6.6.3). This is implicitly recognised by Pace et al. (2018); Liu et al. (2020a) and Xu et al. (2020), all of whom reflect on the topological changes associated with defective anatomy, and the challenge presented by their accurate segmentation in the presence of poorly defined or "blurry" image boundaries. Some claim improvements in the delineation of such interfaces, but provide no quantitative foundation to such findings, relying only on purposively sampled qualitative examples (Du et al., 2020).

We wish not to discredit the contribution of these works more generally. All conform to the most popular and widely recognised approaches for the *technical* assessment of novel segmentation methodology. We also acknowledge that the limited size of the HVSMR dataset likely precludes more detailed analyses, particularly for those submitting to the original Challenge, being bound by its associated assessment protocol. However, we consider that the work presented in this chapter provides a comprehensive and unique investigation of CNN performance relevant to the *clinical* application of segmented, 3D CMR data.

6.6.2 Successes and failures

The breadth of our analysis benefits greatly from our curation and use of the ELCH dataset, permitting the investigation of multi-class anatomical segmentation from 3D and 4D CMR data for the first time. Our task formulation, in combination with the use of bespoke metrics, makes it difficult to compare performance with previous work. However, once allowances for the difficulty of the task posed are made (considering the diversity of congenital defects and surgical modifications

to anatomy; and the challenge of delineating distal pulmonary vasculature), we speculate that our spatial overlap results are at worst consistent with the state of the art. Experimenting on the ELCH dataset also allowed us to investigate the extent to which clinically relevant covariables influence segmentation. This suggested image quality as the strongest predictor of technical performance, better explaining spatial overlap than the heterogeneity of the CHD cohort (expressed through their weight and anatomical diversity), than might have been expected.

Our novel introduction of 4D TR-MRA data via temporal proxies, however, delivered mixed performance. The rationale for their inclusion was two fold: (1) that 4D TR-MRA data complement 3D SSFP acquisition, providing highly specific contrast between the cardiovascular structures (and relevant segmentation targets) and the background; and (2) that differential enhancement of cardiac sub-structures might inform the continuity of the cardiac blood pool, including the presence of congenital defects. For the most part, our technical findings endorsed the first of these hypotheses, with the most reliable improvements associated with the inclusion of those proxies based on pixel-wise descriptive statistics (the mean and standard deviation). Analysing the relationship between performance and clinically relevant covariables evidenced the complementarity of 3D SSFP and 4D TR-MRA, the latter accounting for deficiencies in 3D image quality.

In contrast, our attempts to capture and exploit differential modes of enhancement in aid of the second hypothesis were unsuccessful. Figure 6.1 demonstrates the sensitivity of PCA to differential modes of dynamic contrast enhancement. We had hoped that the CNN might leverage this representation as a source of discriminative features, improving the isolation and association of cardiac sub-components. In particular, we were keen to understand whether differential enhancement might improve the delineation of congenital defects (see the example provided in Figure 6.2, in which enhancement is characteristic of atrial septal defect). At least within our current approach based on PCA, the results in Table 6.1 do not support this expectation.

Due to the heterogeneity of dynamic enhancement, and its dependence on patient size and heart rate, we elected to compute the principal components of temporal variation, *independently* expressed by each case. Accordingly, high-dimensional (Λ_n, $n > 1$) modes of variation captured were not consistent across the ELCH co-

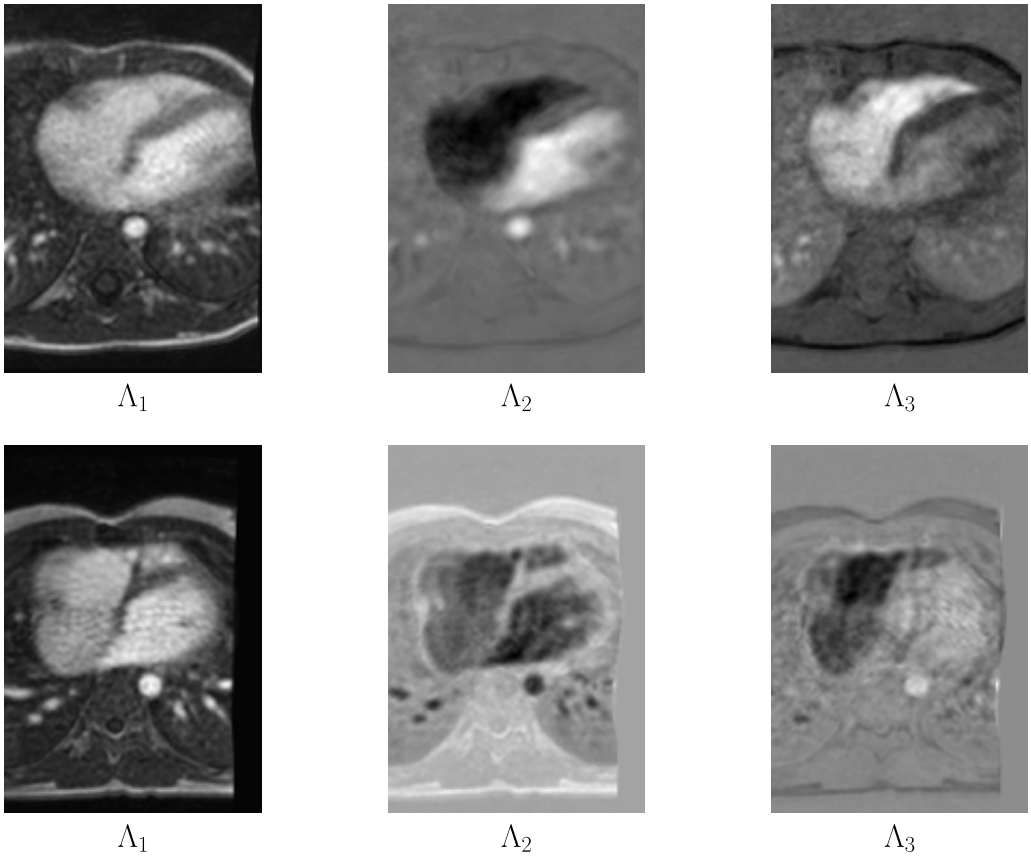


Figure 6.14: Computed *per patient* the principal components of temporal enhancement do not reflect consistent modes of variation *between* cases of the ELCH dataset. In both (a) and (b), differential enhancement of the left and right heart is captured by $\mathbf{X}_{\Lambda_2}$ and $\mathbf{X}_{\Lambda_3}$. However, as the transformed space is not shared between (a) and (b), the same pattern of enhancement presents differently.

hort. Figure 6.14 illustrates this phenomenon: note that whilst Λ_1 has a similar appearance in both (a) and (b), modes of characteristic enhancement are distributed differently across the remaining principal components. This effect may explain the discrepant performance of $\mathbf{X}_{\Lambda_1}$ compared with $\mathbf{X}_{\Lambda_2}$ and $\mathbf{X}_{\Lambda_3}$.

Despite these challenges, we still believe that dynamic enhancement remains an untapped source of features relevant to the configuration of cardiac anatomy. However, and as described, per case PCA did not facilitate effective learning. It is possible that the alternative, computing principal components across the entire training set, may provide a representation of dynamic enhancement that is common

between cases, fostering effective feature learning. Canonical Correlation Analysis (or other forms of manifold alignment) may also offer a solution (Hardoon et al., 2004). Another option might consider augmentation by shuffling input channels. Perhaps a more credible conclusion, however, is that the heterogeneity in dynamic enhancement cannot be sufficiently modelled by low capacity models based on handcrafted features, such as PCA. Feature extraction by high capacity, 4D deep CNN, possibly involving recurrent architectures, might present a more realistic solution.

The most significant failing in CNN performance concerns the limited clinical applicability of associated segmentations. Though fruitless in this respect, it is a strength of our analysis and its bespoke metrics that they draw attention to clinically relevant aspects of performance. These include failings in both the number of defects implied by multi-class or whole heart segmentation, and the accurate representation of their location, size and shape. Yet to be comprehensively addressed in the wider literature, we now expand on these facets of performance and discuss their importance to clinical applications within CHD.

6.6.3 Relevance to patient-specific anatomical modelling

Section 6.5.2 assessed various aspects of clinical performance across a range of metrics. To best understand the importance of our findings, it is worth briefly retreating from the details of this section and recalling the motivations underlying 3D segmentation of CHD imaging. In this task, we seek a patient-specific representation of anatomy, including the presence of defects and surgical modifications. These models find application in each of communication and education; patient consultation and consenting; and clinical management.

With respect to spatial overlap, CNN-based segmentation performed strongly. In some cases, this level of performance may be consistent with certain applications, especially those which predominantly provide a holistic and qualitative description of patient-specific disease. For example, some of the cases illustrated in Figure 6.6 may enhance the training of medical students (where patient-specificity is less critical than at the point of care) or patient consultation. More general applications might also be enhanced by the presentation of multi-class segmen-

tation within a unified whole heart blood pool. Such models are likely to be of greatest value where extracardiac anatomy and its malformation motivate inspection, rendering the accurate representation of intracardiac features and thin tissue interfaces incidental.

Perhaps our primary motivation, however, seeks the wider introduction of 3D models as a means of understanding, considering and managing CHD and the particular defects exhibited by the specific patient. Via an advanced form of 3D rendering (such as 3D printing (Giannopoulos et al., 2016) or virtual reality (Ong et al., 2018)), segmented images can be used to inform the management of patients with CHD, including planning for catheter-based or surgical intervention. To this end, the accurate representation of defects is vital. For example, in considering biventricular repair of double outlet right ventricle with sub-aortic VSD, the surgeon must plan the course of the intra-ventricular tunnel between VSD and systemic outflow. This relies on a faithful appreciation of defect size and shape, its muscular rim, and its displacement from the aortic valve.

In common with other researchers (Painchaud et al., 2020), however, we found CNN-based segmentations to lack spatial coherence, including spurious connected components, holes and discontinuities. Given that a majority of congenital defects concern deviations from the expected configuration of the heart's chambers, this deficiency is problematic. When coupled with the observation that a major subset of malformations (and in particular intracardiac defects) are defined in relation to (septal defects), or by (atretic defects), thin tissue interfaces, this culminates in predicted segmentations which lack clinical meaning.

To a limited extent, we were able to rectify such errors by leveraging our multi-class formulation (relying on the detailed labelling provided by our ELCH dataset). Whilst rudimentary post-processing reduced the *number* of errors in the continuity of the blood pool and eliminated some spurious shunts, the structural representation (including *size*, *shape* and *location*) of remaining defects was poor. Without further manual editing, we encountered a level of segmentation performance that was inadequate for the idealised precision decision-making and interventional planning described.

6.6.4 Limitations and future work

At the outset of this project, our ambition was to assess the suitability of CNN-based segmentation for clinical deployment within the care of patients with CHD. We had hoped to complete a pragmatic investigation by extending our performance metrics beyond the quantitative outcomes presented, and asking clinical imaging experts to rate clinical acceptability. However, the work presented in this chapter, and in particular our clinically facing metrics, preclude such an experiment at this stage. There are technical deficiencies within our current CNN methodology that must be resolved, and performance gains made before any such assessment is warranted. We touch on some of these developments in the following.

In any segmentation task, the frequency with which targets appear in the underlying distribution of image data influence data-driven solutions. Where frequency imbalances are significant, previous work has introduced loss weighting schemes (Ronneberger et al., 2015) or judicious batch sampling strategies (Kamnitsas et al., 2017) to ensure an optimal representation of each class. Such methods are highly applicable to the multi-class formulation of our task, and would likely improve the segmentation of rare anatomical components (such as left superior vena cava) or surgical modifications (such as DKS connection). However, and whilst appropriate to any clinical application, given that these approaches are established in literature, they are less fertile ground for novel investigation.

In contrast, there are a host of developing methods that might be of interest. These include: high capacity and high-dimensional neural networks for learning 4D features of TR-MRA data, possibly including recurrent operation or temporal convolution; and the application of spatial transformer networks (Jaderberg et al., 2015) to handle the diversity of CHD patient size and anatomy. More closely related to the specifics of our task, dataset and patient population, it would also be of interest to examine possible performance differences associated with the isolation of diagnostic groups from the ELCH dataset, and their treatment as individual training sets. Lastly, given the dependence of performance on image quality, methods such as: bias field compensation to resolve spatial inhomogeneity in CMR signal intensity; and the possible synthesis of relevant image artefacts within data augmentation (possibly via a generative approach); are both of interest. Whilst each

of these address the limitations in current clinical image quality, we also recognise the benefits that improvements in acquisition and reconstruction methods might confer in future.

6.7 Conclusion

To the best of our knowledge, this chapter presents the first concerted and comprehensive effort to investigate the CNN-based segmentation of CHD anatomy from 3D CMR. Through our ELCH cohort, we achieved a training dataset that was: an order of magnitude larger than that used in previous works (the HVSMR dataset being the most comparable); admitted the multi-class segmentation of data into sixteen anatomical classes; and allowed CNN inference to be conditioned on the combination of 3D SSFP and 4D TR-MRA acquisitions. In so doing, we established the state of the art in this task, including strong spatial overlap performance. Our results suggest that segmentation can be straightforwardly enhanced by the introduction of proxy representations of 4D data based on statistics describing contrast enhancement. However, our attempts to leverage differential enhancement, using PCA to extract features relevant to the circulatory configuration of the heart were unsuccessful. Perhaps most importantly, our investigation sheds light on the deficiencies within CNN-based segmentation. The use of bespoke metrics relevant to the number and representation of congenital heart defects revealed a lack of anatomical coherence within predicted results, sufficient to preclude the majority of clinical applications without prior manual editing. Being our primary motivation for the segmentation of these image data, improving the anatomically and clinically meaningful delineation of such defects will be central to the remainder of this thesis.

Chapter 7

Topological loss functions

7.1 Introduction

The preceding chapters made every effort to apply, and consider the suitability of fundamental convolutional neural network (CNN) methodology for the segmentation of congenital heart disease (CHD) anatomy from 3D cardiac magnetic resonance (CMR) images. Chapter 5 curated the Evelina London Children's Hospital (ELCH) dataset, its ground truth segmentations describing a clinically relevant cohort of 150 patients with CHD. Despite this resource, Chapter 6 demonstrated the limitations of conventional CNN-based segmentation, and its failure to delineate anatomy in a spatially and clinically coherent fashion. Predictions distorted the continuity of the blood pool, including errors violating the configuration of cardiac chambers, and the presence of defects.

In response, this chapter attends to CNN parameter optimisation leveraging topology as a means of understanding anatomical configuration, and builds associated loss functions to promote clinically plausible segmentation. Moreover, we substantively exploit the multi-class formulation of our data and task, allowing optimisation against a diagnostically relevant, prior description of CHD. Before presenting our findings, we review literature relevant to the incorporation of prior information within cardiac segmentation; and introduce the theory of persistent homology (PH), on which our loss functions are based.

7.1.1 Limitations of pixel-wise optimisation

The methodological focus of this work, deep learning, and in particular CNNs, have fostered significant performance gains across an array of cardiac image segmentation tasks (Chen et al., 2020). One key to their success has been the design of specialised architectures dedicated to image segmentation (the ubiquitous U-Net model, being the prime example (Ronneberger et al., 2015)). Implicit within this multi-scale architecture is an acknowledgement that image segments are discriminated by each of pixel, local and global image features. This observation is reflected by the introduction of operations such as pooling, and strided or dilated convolution for the extraction of multi-scale features; and skip connections for their synthesis. In combination, these serve to expand the receptive field of a model's constituent neurons faster than the repeated application of convolution alone. Theoretically, such approaches permit the learning of image features with extended spatial context, such as anatomical morphology and topology.

Whilst considerable effort has been devoted to methods for the extraction of multi-scale image features, less attention has been paid to their role in network optimisation (Duan et al., 2019). For the most part, segmentation CNNs have been trained using pixel-wise loss functions such as cross-entropy (CE) or the Dice similarity coefficient (Dice, 1945) (DSC). Whilst easily implemented and possessing favourable numerical properties, their treatment of pixels as independent from one another renders them insensitive to higher order features of the data such as morphology and topology. This is in stark contrast with the spatially extended features that we anticipate are learned during training.

Ignorant of such features, CNN optimisation against pixel-wise losses can result in predicted segmentations which lack spatial coherence. Not limited to the findings of our own work (see Chapter 6), CNN-based predictions are frequently reported as presenting with unrealistic properties such as spurious connected components or holes (Painchaud et al., 2020). To the operator, such errors can appear nonsensical, even violating fundamental properties of anatomy. Since these errors are frequently small and necessarily constrained to the boundaries of predicted anatomy, their associated segmentation can remain suitable for the assessment of ventricular volume and certain clinical indices (Ruijsink et al., 2020). For a wider

array of downstream applications, however, including patient-specific visualisation of CHD, it is crucial to represent such features faithfully (Byrne et al., 2016).

Across a range of cardiac image segmentation tasks, prior knowledge has been used to inform the expected configuration of anatomical components (see Chapter 3). In fact historically, and before the advent of deep learning methodologies, state of the art cardiac segmentation methods were dependent on the use of strong prior information. Consider atlas-based approaches or those relying on statistical models of anatomical shape or appearance, for example. In each, prior information concerning the plausible arrangement of anatomy is implicitly characterised by the training cases within the atlas or statistical model. Ideally, however, such prior information should be abstract and adaptable to the variety of cases encountered in the clinic: not dependent on its appearance within exemplar training data.

In the simplest case, priors might specify a healthy configuration of the heart's various chambers, valves and associated vasculature. For example, in short axis CMR images the right ventricular cavity appears bound to the left ventricular myocardium, which in turn surrounds the left ventricular blood pool. However, such a description can also be adapted and extended to characterise the structural defects associated with CHD. Advantageously to our application, the details of patient-specific diagnosis are frequently known from previous examination, most often echocardiography. Hence, at the point of care, prior knowledge can specify disease morphology, explicitly indicating defect presence and number, and implicitly reflecting size, shape and locale.

Whether describing normal or pathological anatomy, and in contrast to the pixel-wise loss functions aforementioned, such anatomical priors provide a global description of segmentation coherence. However, whilst these constraints are simple to express qualitatively, their effective quantitative statement is not trivial. Furthermore, the opaque nature of CNNs has meant that up until recently, it has proved difficult to explicitly exploit such prior information in model optimisation. Addressing these limitations, we present a topological loss function for multi-class image segmentation. Our approach not only makes use of global features, but also leverages an explicit prior description of anticipated label map topology that is independent of exemplar training data.

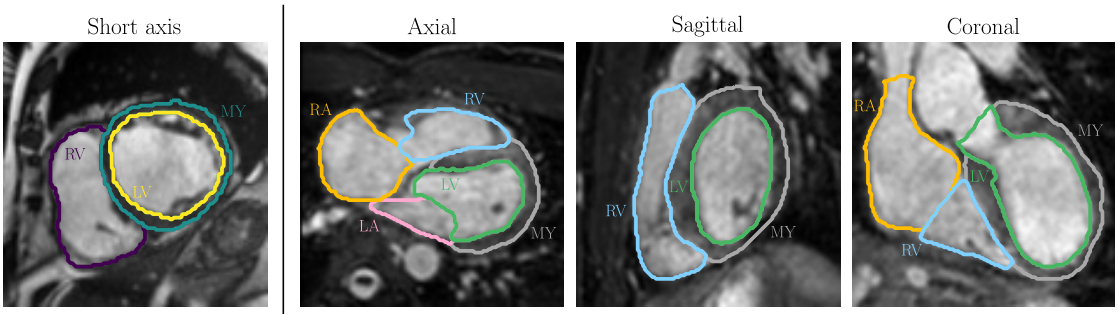


Figure 7.1: The CMR segmentation tasks considered: (a) segmentation of 2D short axis data into left ventricle (LV) and right ventricle (RV) blood pool cavities and the left ventricular myocardium (MY); and (b) whole heart segmentation of 3D data into left atrium (LA) and right atrium (RA) classes in addition to those presented in (a).

7.1.2 Anatomical priors in cardiac segmentation

Given that a comprehensive review of cardiac segmentation methodologies was provided in Chapter 3 and Chapter 4, here we limit our focus to those works whose primary contribution leverages prior information. Moreover, in developing and validating the topological loss functions presented, we explore their application to tasks of increasing difficulty, incrementing image dimension (moving from 2D to 3D) and pathological variety (moving from normal to congenitally malformed anatomy). Hence, the following review examines both 2D short axis and 3D whole heart or multi-class segmentation (see Figure 7.1).

2D Short axis segmentation

Thanks to its position as the gold standard approach for ventricular volumetry, CMR short axis segmentation has received the most significant research interest of these applications. At least in studies of healthy patients, including large training data, this effort has culminated in a level of performance consistent with inter-observer variation (Bai et al., 2018). However, for studies involving cardiovascular disease (for which training sets are typically smaller and morphologically more variable), a performance deficit remains. This gap is in part characterised by the sort of anatomically implausible error described in Section 7.1.1 and observed in

Chapter 6. To address these modes of failure, authors have sought to introduce prior information to CNN-based segmentation.

One avenue of research has combined conventional methods such as atlas-based segmentation (Dong et al., 2018; Duan et al., 2019; Zotti et al., 2018) and active contour refinement (Avendi et al., 2017; Ngo et al., 2017; Rupprecht et al., 2016) with CNNs. Whilst these extensions admit improvement, their capacity to model the variation of pathological cases is limited by the make up of the atlas or hand-crafted features employed. CNNs have also been used in conjunction with statistical shape models: Tóthová et al. (2018) used a CNN to infer the coefficients of variation at test time; Milletari et al. (2017) include a principal component analysis layer to learn the distribution of left ventricular shape. The ability of CNNs to learn shape priors from multi-view CMR data has also been shown to improve segmentation performance (Chen et al., 2019).

Another area of work has attempted to inject prior information into CNN optimisation directly via a learned, latent representation of anatomically plausible shapes (Degel et al., 2018; Oktay et al., 2017; Yue et al., 2019). Their implicit embedding, however, makes it difficult to understand the extent to which such priors are related to morphology or topology as claimed. Bridging this gap, Painchaud et al. (2020) augmented the latent space via a rejection sampling procedure, maintaining only those cases satisfying sixteen criteria related to anatomical plausibility. At test time, searching the augmented space for a case's nearest neighbour guarantees a credible segmentation result.

3D Whole heart segmentation

Before the shift to deep learning methods, whole heart segmentation made use of strong prior knowledge. Statistical shape models (Wierzbicki et al., 2008), and in particular (multi) atlas-based segmentation (Zuluaga et al., 2013; Zhuang, 2016) made up the state of the art. However, rather than make similar use of priors, exponents of CNN-based segmentation have been necessarily preoccupied with solutions for best handling large 3D volumes. To make best use of limited graphics processing unit (GPU) memory, relevant work has focused on architectural modifications including cascaded processing at multiple spatial scales (Payer et al., 2018;

Isensee et al., 2018), patch-based inference (Yang et al., 2018b) and slice-by-slice or 2.5D segmentation (Wolterink et al., 2017).

There are exceptions to this trend, however. In the original 2017 Multi-Modality Whole Heart Segmentation (Zhuang et al., 2019) (MM-WHS) challenge, Wang et al. (2017a) incorporated the results of statistical shape modelling as an additional CNN input channel. In the context of congenital computed tomography (CT), Xu et al. (2020) used a graphical representation of the great vessels to improve both diagnostic classification and segmentation within a large training dataset. Works combining multi atlas-based registration and CNNs have also been presented (Luo and Zhuang, 2020; Ding et al., 2020; Dong et al., 2020; Sinclair et al., 2022). Most recently, Habijan et al. (2021) employed a latent representation of whole heart anatomy (akin to Oktay et al. (2017)) as a means of assessing spatial coherence within an associated loss function. Wang et al. (2021b) used the same approach to determine reliable pseudo-labels in a few shot learning framework, achieving impressive results.

7.1.3 Topological priors in image segmentation

Here, we touch on key references from the wider image processing literature relevant to segmentation topology. In alternative clinical settings, priors specifying the adjacency and hierarchical containment of anatomical components have been used to build associated loss functions. Examples have considered features related to the anticipated adjacency of brain regions (Ganaye et al., 2018) and the hierarchy of cellular structures (BenTaieb and Hamarneh, 2016). Alternatively, reformulating the segmentation task as a layer-wise regression has allowed the segmentation of optical coherence tomography images, with an anatomical ordering of retinal layers (He et al., 2019). Inspired by the work of Jaderberg et al. (2015) and taking a lead from atlas-based registration, Lee et al. (2019) used spatial transformer networks to learn an optimal mapping of a coronary artery prior to the coordinate space of the test image.

We observe that across applications (BenTaieb and Hamarneh, 2016; Ganaye et al., 2018; Painchaud et al., 2020), criteria used to define anatomical plausibility can frequently be summarised by segmentation topology. Though a global

property, researchers have adapted well-established topological metrics such as the Rand Index (Briggman et al., 2009) and Warping Error (Jain et al., 2010), for use in gradient-based, CNN training.

More recently, PH, an emerging mathematical tool for topological data analysis, has been combined with machine learning methodology. In addition to practical applications, the theoretical basis for such approaches is developed by Hofer et al. (2017) and Gabrielsson et al. (2020). Topological features returned by PH have been employed within image classification and segmentation by K-means clustering (Assaf et al., 2016) and K-nearest neighbours (Qaiser et al., 2016) classifiers. They have also been leveraged to distill dermoscopic images (Vandaele et al., 2020). Specially constructed input layers have been used to feed PH features to CNNs, improving electroencephalogram classification (Hofer et al., 2019), and the detection of coronavirus disease (2019) (COVID-19) (Hajij et al., 2021). In cardiac imaging, PH was used in the identification and restoration of papillary muscles to a myocardial segmentation of CT data (Gao et al., 2013).

Pertinently to our motives, PH has been employed, not only as a source of features to be passed to a downstream classifier, but also to extract a supervisory signal for learned feature optimisation. Moreover, PH has been used to build topological loss functions for CNN-based image segmentation. Hu et al. (2019), established a topological loss according to the Wasserstein distance between the persistence barcodes of predicted and ground truth segmentations, applying their approach to cellular microscopy and within the natural image domain. They have also extended their approach to generative adversarial networks (Wang et al., 2020b).

Other exponents of PH-based losses have instead optimised segmentation topology against an explicit topological description known *a priori*. Example losses have penalised segmentations of the murine neurovasculature which deviate from its anticipated tree-like topology (Haft-Javaherian et al., 2020). Shin et al. (2020) used a similar scheme at training time, optimising a set of CNN parameters to infer the cylindrical topology of small bowel segmentation. Relevant to our tasks, a similarly constructed loss has been used to encourage toroidal appearance of the myocardium in short axis view (Clough et al., 2019, 2020). In contrast to those losses built on a latent representation of plausible shape, PH provides a mechanism

for segmentation evaluation against an explicit topological prior. The benefits to interpretability aside, this affords the opportunity to decouple relevant topological, prior information from its representation within training data, a statistical shape model or atlas.

7.2 Contributions

To the best of our knowledge and outside of our preliminary work (Byrne et al., 2021), the application of PH-based loss functions to the task of multi-class segmentation is yet to be explored. Compared with the binary case, extension to this setting permits consideration of a richer set of topological priors, including hierarchical class containment and adjacency. In the following, we explore the application of PH-based loss functions to various CMR segmentation tasks, leveraging priors related to the topological relationships between the chambers of the heart. Our work makes the following contributions:

1. We present a formalism for the construction of PH-based loss functions that can be used to optimise CNN-based, multi-class segmentation topology.

2. We employ these losses in a CNN post-processing framework, demonstrating significant improvements in topological performance across 2D and 3D problems, considering a range of CMR segmentation tasks (see Figure 7.1).

3. We introduce an efficient implementation based on cubical complexes and parallel computation, making significant performance gains, and admitting 3D application at full spatial resolution for the first time.

4. In addition to within structurally normal anatomy, we apply our approach to the segmentation of patient-specific CHD anatomy, improving the topology of predicted segmentations.

5. Throughout we present a faithful and detailed evaluation of our approach, speculating as to the limits of its generalisability.

7.3 Theory

Here we provide a practical and largely qualitative introduction to homology and PH as necessary to inform our construction of cubical complexes and topological loss functions. For a mathematical background, we direct the reader to Kaczynski et al. (2006), Edelsbrunner et al. (2008) and Otter et al. (2017).

7.3.1 Betti numbers, homology and cubical complexes

In ND, objects with differing topology can be distinguished by the first N Betti numbers: $\mathbf{b} = (b_0, b_1)$ in 2D and $\mathbf{b} = (b_0, b_1, b_2)$ in 3D. Intuitively, b_0 counts the number of connected components which make up an object, b_1, the number of 2D holes or loops present and b_2, the number of 3D holes or voids contained (Otter et al., 2017). For the purposes of our work, homology is a branch of algebraic topology concerned with procedures to compute the Betti numbers of objects. Rather than interrogating an arbitrary topological space, such computation is developed in conjunction with the representation of objects by combinatorial structures called simplicial and cubical complexes. Cubical complexes are well suited to data structured on a rectangular lattice, constructing 3D objects as the combination of points (0-*cells*), and unit line intervals (1-*cells*), squares (2-*cells*) and cubes (3-*cells*). Note that the edges of high-dimensional cells ($n >$ 0-*cells*) align with the cardinal directions of the image space. In the 2D case, this is illustrated in Figure 7.2. Provided its representation as a valid cubical complex, homology returns the Betti numbers of a binary segmentation by linear algebra alone.

In the context of image data, this representation is not only more computationally efficient and elegant than its simplicial equivalent (as employed in our previous work (Clough et al., 2019, 2020; Byrne et al., 2021)), but also affords precise control over the connectivity of pixels. As exemplified by Garin et al. (2020), there are two approaches for the construction of cubical complexes from ND image data. These differ in their treatment of image pixels as either 0-*cells* or N-*cells* (in other words, *"top dimensional"-cells*) of the resulting complex.

The repercussions of this choice are illustrated in Figure 7.2, and best understood in the context of classical pixel connectivity, as it bears on the formalism

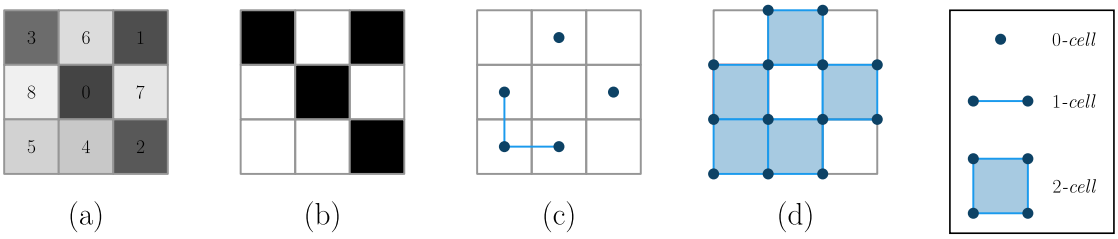


Figure 7.2: Construction of a cubical complex from 2D data. Pixel intensities in (a) exceeding an arbitrary threshold of three appear white in the binary image (b). In (c), these white pixels are considered 0-*cells*, representing a 4-*connected* foreground, including three components. In (d) these are considered 2-*cells*, representing an 8-*connected* foreground component, containing a hole.

of digital topology (Kong and Rosenfeld, 1989). In the 2D case: treating pixels as 0-*cells* results in a 4-*connected* representation of foreground objects; treating pixels as 2-*cells* results in an 8-*connected* representation of foreground objects. We will refer to these cases as the 0-*construction* and N-*construction* respectively. It should be noted that the same argument applies in 3D, with correspondences drawn between the 0-*construction* and the 6-*connected* lattice; and between the N-*construction* and the 26-*connected* lattice. In Section 7.5.3, we perform experiments comparing these approaches.

7.3.2 Persistent homology

We intend to interrogate and optimise the topological features of segmented images, relying on tools from algebraic topology for their extraction. The previous section demonstrated the construction of a cubical complex for a discrete, binary image. This approach can be applied equally to the ground truth label map $S : \mathbb{R}^N \to \{0, 1\}$, and to its associated predicted segmentation $\tilde{S} : \mathbb{R}^N \to [0, 1]$, once binarised at a probability threshold, $\tilde{S}_p = \tilde{S} \geq p$. The homology of such an object, whilst easily exposed via the mechanics of algebraic topology, is not immediately compatible with gradient-based, CNN optimisation. In more specific terms, a loss function comparing the homology of S with $\tilde{S}_p$ does not return a differentiable supervisory signal. This is demonstrated by Figure 7.2: modulating the central pixel incurs a discontinuous change in the object's Betti numbers. To overcome this hurdle, previous work (Clough et al., 2019, 2020; Shin et al.,

2020) has exploited PH, acquiring a representation of the topological features of the probabilistic segmentation, $\tilde{S}$.

In general, PH describes a scheme for the particular application of algebraic topology which exposes the topological features of data (or equivalently their Betti numbers) at multiple scales (Otter et al., 2017). We use the 2D example of Figure 7.3 to illustrate the result returned by such a procedure (the so-called persistence barcode), providing a practical explanation as it relates to our application. Critically, rather than consider topology at a single probability threshold (as might be exposed at test time), PH captures the topology of $\tilde{S}$ binarised at all possible thresholds. Descending from the maximal threshold $p \geq 1$, this amounts to computing the homology of a nested sequence of binarised segmentations: $\tilde{S}_1 \subset ... \subset \tilde{S}_p \subset ... \subset \tilde{S}_{-\infty}$. The barcode, therefore, is a dynamic characterisation of the way that probabilistic segmentation topology evolves as a function of this threshold, p.

Example segmentations from this sequence are demonstrated in Figure 7.3. At each value of the threshold, the number of vertically arranged bars indicates the number of topological features presented by $\tilde{S}_p$. Moreover, the presentation of each bar reveals the dimensionality of the feature described: solid bars indicate the presence of connected components; open bars indicate the presence of loops. In this way, the Betti numbers of $\tilde{S}_p$ are returned by counting the number of each type of bar vertically intersected.

A bar extends horizontally for the probability interval over which its associated feature is maintained. Critical values of p admit changes in the topological features of $\tilde{S}_p$. In Figure 7.3, such values are indicated by the endpoints of each bar. Accordingly, the persistence of a topological feature Δp is the horizontal length of its associated bar. Persistent bars are considered robust to small perturbations, suggesting that they are true topological features of $\tilde{S}$. Hence, in Figure 7.3, we arrange bars in order of descending persistence after grouping by topological dimension. From the persistence barcode of the probabilistic segmentation $\tilde{S}$, we write the lifetime of the l^{th} most persistent feature of dimension d as $\Delta p_{d,l}(\tilde{S})$.

Compared with the Betti numbers of a discrete, binary segmentation, topological persistence provides a differentiable quantity that is consistent with gradient-based learning (see Gabrielsson et al. (2020)). To appreciate this, we present two

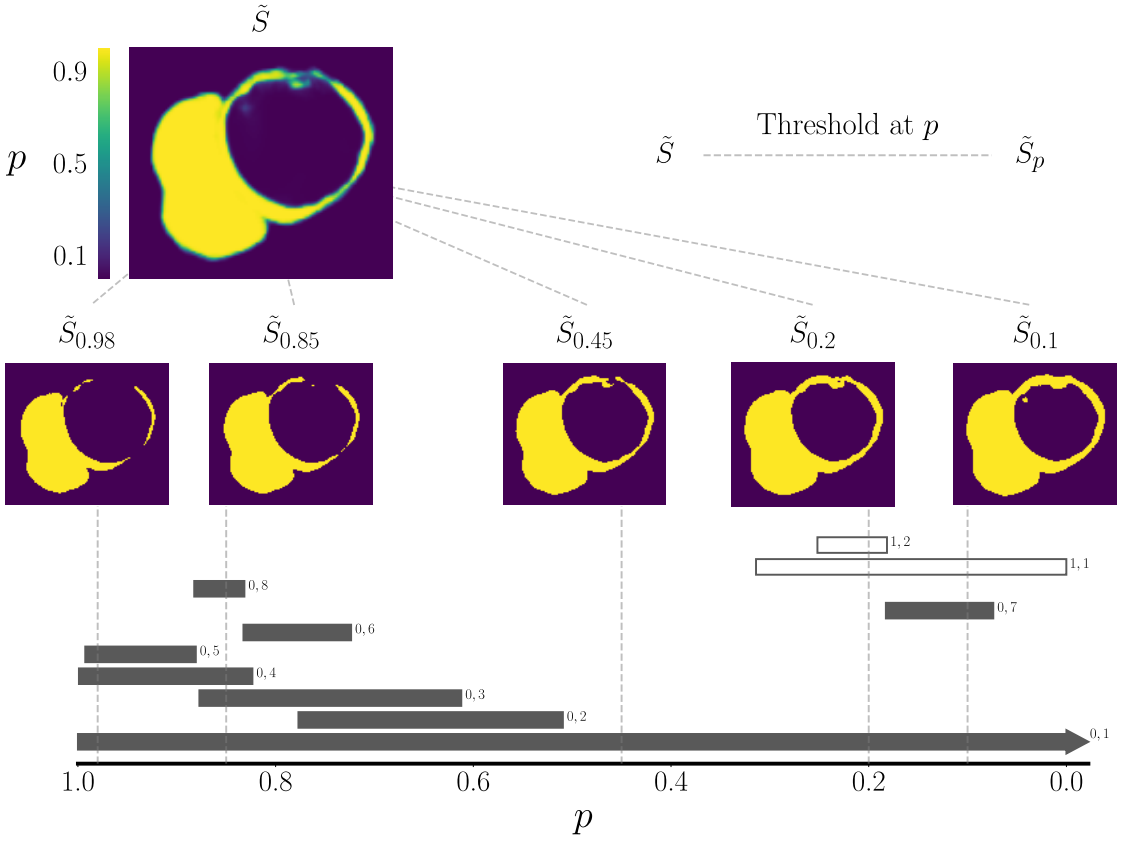


Figure 7.3: Computation of the persistent homology (PH) barcode, reflecting the topological features of the probabilistic segmentation $\tilde{S}$, when binarised at all possible probability thresholds in the interval $[0, 1]$. At a particular p: the number of vertically intersected solid bars counts connected components ($d = 0$); open bars count the number of loops ($d = 1$). Each bar is labelled with its topological dimension, and its persistence ranking in order of descending lifetime: d, l.

arguments: (1) a small change in the probabilistic segmentation $\tilde{S}$ results in a small change in the persistence of bars; and (2) respectively, the creation and destruction of each topological feature (the extremes of each bar) are associated with particular cells of the cubical complex, and therefore pixels of the predicted segmentation.

7.4 Methods

7.4.1 Notation

We address the generic multi-class segmentation task, seeking a meaningful division of the ND CMR image $\mathbf{X} : \mathbb{R}^N \rightarrow \mathbb{R}$ into meaningful anatomical segments. In the three tasks considered, relevant image segments describe the chambers of the heart, their association, and their relationship to the myocardium (see Figure 7.1). We use the following shorthand: myocardium (MY) (including left and, depending on the task, right ventricular components), left atrium (LA), left ventricle (LV), right atrium (RA) and right ventricle (RV). However, it should be noted that the formalism presented has the flexibility to accommodate the semantics of a range of multi-class segmentation tasks. We denote the ground truth image segmentation by $\mathbf{Y} : \mathbb{R}^N \rightarrow \{0, 1\}^K$, being made up by K mutually exclusive class label maps: Y_1, Y_2, ..., Y_K, including $K - 1$ foreground classes in addition to the background, Y_1.

In each task, we consider a deep learning solution, optimising the parameters, $\boldsymbol{\theta}$, of a CNN to infer the probabilistic segmentation, $\tilde{\mathbf{Y}} : \mathbb{R}^N \rightarrow [0, 1]^N$, a distribution over the per class segmentation maps: $\tilde{Y}_1$, $\tilde{Y}_2$, ..., $\tilde{Y}_K$. We write segmentation inference as $\tilde{\mathbf{Y}} = f(\mathbf{X}; \boldsymbol{\theta})$. In all cases, we achieve a discrete prediction as the segmentation which maximises pixel-wise probability: $\hat{\mathbf{Y}} : \mathbb{R}^N \rightarrow \{0, 1\}^N$, made up by K mutually exclusive classes: $\hat{Y}_1$, $\hat{Y}_2$, ..., $\hat{Y}_K$. Within our formalism, we consider the topology, not only of individual segmentation objects, but also their combination: by $Y_{i \cup j}$ we reference the Boolean union of classes i and j; by $\tilde{Y}_{i \cup j}$ we reference the pixel-wise probability of class i or j. We consider the union of a class with itself to be the segmentation of the single class: $Y_{i \cup j=i} = Y_i$ and $\tilde{Y}_{i \cup j=i} = \tilde{Y}_i$.

Given the success of CNN-based solutions, we assume that, at least with respect to spatial overlap, $\tilde{\mathbf{Y}}$ is a reasonable estimate of $\mathbf{Y}$. In this setting we describe our CNN post-processing framework for the correction of inferred segmentation topology.

7.4.2 Multi-class topological priors

As introduced in Section 7.3.1, the Betti numbers are topological invariants permitting the specification of priors for the description of foreground image segments. They can be specified on a per task (where anticipated topology is uniform across the population) or per patient basis (perhaps where anticipated topology reflects the pathology-induced and clinically relevant structural changes associated with CHD). Consider our 2D, short axis example (see Figure 7.1):

$$\begin{aligned} \mathbf{b}^{\mathrm{RV}} &= (1, 0) \\ \mathbf{b}^{\mathrm{MY}} &= (1, 1) \\ \mathbf{b}^{\mathrm{LV}} &= (1, 0) \end{aligned} \tag{7.4.1a}$$

Equation set 7.4.1a specifies that each of the RV, MY and LV should comprise a single connected component, and that the myocardium should contain a single loop. However, these equations only provide a topological specification in a segment-wise, binary fashion: they fail to capture inter-class topological relationships between cardiovascular anatomy. For instance, they make no specification that the myocardium surround the left ventricular cavity or that the right ventricle and myocardium should be adjacent. A richer topological description is returned by considering combined foreground classes:

$$\begin{aligned} \mathbf{b}^{\mathrm{RV} \cup \mathrm{MY}} &= (1, 1) \\ \mathbf{b}^{\mathrm{RV} \cup \mathrm{LV}} &= (2, 0) \\ \mathbf{b}^{\mathrm{MY} \cup \mathrm{LV}} &= (1, 0) \end{aligned} \tag{7.4.1b}$$

By the inclusion-exclusion principle, the topology of a 2D multi-class image segmentation is characterised by that of all foreground objects and all possible object pairs (Bazin et al., 2007): see Equation set 7.4.1b. For convenience, we collect all of Equation set 7.4.1 into a Betti array $\mathbf{B} : \{1, 2, 3\} \times \{1, 2, 3\} \times \{0, 1\} \to \mathbb{R}$. Each element B_d^{ij} denotes the Betti number of dimension d for the ground truth segmentation $Y_{i \cup j}$. Note that in B_d^{ij} we divide indices between sub- and superscripts to make clear the difference between class labels (i, j) and topological dimension (d), without further significance.

$$
\begin{aligned}
\mathbf{b}^{\mathrm{MY}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{LA}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{LV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{RA}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{RV}} &= (1,0,0)
\end{aligned}
\tag{7.4.2a}
$$

$$
\begin{aligned}
\mathbf{b}^{\mathrm{MY}\cup\mathrm{LA}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{LV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{RA}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{RV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{LA}\cup\mathrm{LV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{LA}\cup\mathrm{RA}} &= (2,0,0)\\
\mathbf{b}^{\mathrm{LA}\cup\mathrm{RV}} &= (2,0,0)\\
\mathbf{b}^{\mathrm{LV}\cup\mathrm{RA}} &= (2,0,0)\\
\mathbf{b}^{\mathrm{LV}\cup\mathrm{RV}} &= (2,0,0)\\
\mathbf{b}^{\mathrm{RA}\cup\mathrm{RV}} &= (1,0,0)
\end{aligned}
\tag{7.4.2b}
$$

$$
\begin{aligned}
\mathbf{b}^{\mathrm{MY}\cup\mathrm{LA}\cup\mathrm{LV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{LA}\cup\mathrm{RA}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{LA}\cup\mathrm{RV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{LV}\cup\mathrm{RA}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{LV}\cup\mathrm{RV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{MY}\cup\mathrm{RA}\cup\mathrm{RV}} &= (1,0,0)\\
\mathbf{b}^{\mathrm{LA}\cup\mathrm{LV}\cup\mathrm{RA}} &= (2,0,0)\\
\mathbf{b}^{\mathrm{LA}\cup\mathrm{LV}\cup\mathrm{RV}} &= (2,0,0)\\
\mathbf{b}^{\mathrm{LA}\cup\mathrm{RA}\cup\mathrm{RV}} &= (2,0,0)\\
\mathbf{b}^{\mathrm{LV}\cup\mathrm{RA}\cup\mathrm{RV}} &= (2,0,0)
\end{aligned}
\tag{7.4.2c}
$$

Vitally, even in the absence of the ground truth, $\mathbf{B}$ can be determined by prior knowledge of the anatomy to be segmented. Whilst this requirement may not always be fulfilled, within medical imaging, segmentation targets frequently describe anatomical structures with known topology. For example, we can extend this specification to 3D segmentation of the whole heart, in which case we consider the topological prior expressed by Equation set 7.4.2.

Firstly, note that in 3D, and according to the same inclusion-exclusion principle, multi-class topology is defined not only by that of all foreground objects (Equation set 7.4.2a) and their pair-wise combination (Equation set 7.4.2b), but extends to all object triples, as additionally described by Equation set 7.4.2c. Secondly, we highlight the low topological dimensionality of features associated with this collective 3D prior (Equation set 7.4.2). At least compared with the highly detailed segmentations presented by the ELCH dataset in Chapter 5, it expresses a simpler representation of cardiac anatomy: one which is free from endocardial trabeculation (of either the atria or ventricles). In part, such a specification allows for the more straightforward application and investigation of the topological loss functions presented. However, it also matches the semantics of the MM-WHS segmentation task (the basis of the experiment described in Section 7.5.4) whose labels include papillary muscles and trabeculation within the blood pool class of the associated chamber.

Finally, and despite the simplistic representation of individual chambers, note that Equation set 7.4.2 isolates left and right-sided structures of the normal heart in a clinically meaningful fashion. For example, $\mathbf{b}^{\mathrm{LA} \cup \mathrm{RA}} = (2, 0, 0)$ requires that the left and right atria compose two connected components, implicitly indicating their division by a complete atrial septum. In Section 7.5.5 we will manipulate this prior to specify the presence of patient-specific defects within images drawn from the ELCH dataset: $\mathbf{b}^{\mathrm{LV} \cup \mathrm{RV}} = (1, 0, 0)$ indicating the presence of a single ventricular septal defect (VSD), for example. Further details, including the interaction between defect number and high-dimensional topology, are discussed in Section 7.5.5.

7.4.3 Topological loss function

We now present our formulation for a topological loss function applicable to multi-class, CNN-based segmentation tasks. In common with previous work (Clough et al., 2019; Shin et al., 2020), we exploit the differentiable properties of the persistence barcode to construct a loss that exposes the differences between $\hat{\mathbf{Y}}$ and our prior specification $\mathbf{B}$. However, in contrast with these works, which considered single-class segmentation problems, our loss is informed by the persistence of features from multiple per class probabilistic segmentations. Accordingly, the choice of probabilistic segmentations, from which we extract topological features, is key. To align with the theory of Section 7.4.2, in 2D we consider the persistence barcode for all foreground class labels and class label pairs (see Figure 7.4). The construction of L_{T} is as follows:

$$L_{\mathrm{T}} = \sum_{d,i,j \geq i} B_d^{ij} - A_d^{ij} + Z_d^{ij} \tag{7.4.3}$$

$$A_d^{ij} = \sum_{l=1}^{B_d^{ij}} \Delta p_{d,l}(\hat{Y}_{i \cup j}) \tag{7.4.3a}$$

$$Z_d^{ij} = \sum_{l=B_d^{ij}+1}^{\infty} \Delta p_{d,l}(\hat{Y}_{i \cup j}) \tag{7.4.3b}$$

A_d^{ij} evaluates the total persistence of the B_d^{ij} longest, d-dimensional bars for the probabilistic union of segmentations for classes i and j, $\hat{Y}_{i \cup j}$. Assuming that the inferred segmentation closely approximates the ground truth, and recalling that l ranks topological features in descending order of persistence, A_d^{ij} evaluates the presence of anatomically meaningful topological features. In other words, A_d^{ij} measures the extent to which the anticipated topological features are present within $\hat{\mathbf{Y}}$. Z_d^{ij} evaluates the persistence of spurious topological features that are superfluous to B_d^{ij}. Alternatively, Z_d^{ij} is sensitive to the presence of topologically implausible CNN segmentation errors such as additional connected components or holes.

Summing over all topological dimensions d, and considering all class labels $i, j = i$ and class label pairs $i, j > i$, optimising L_{T} maximises the persistence of topological features which match the prior, and minimises those which do not.

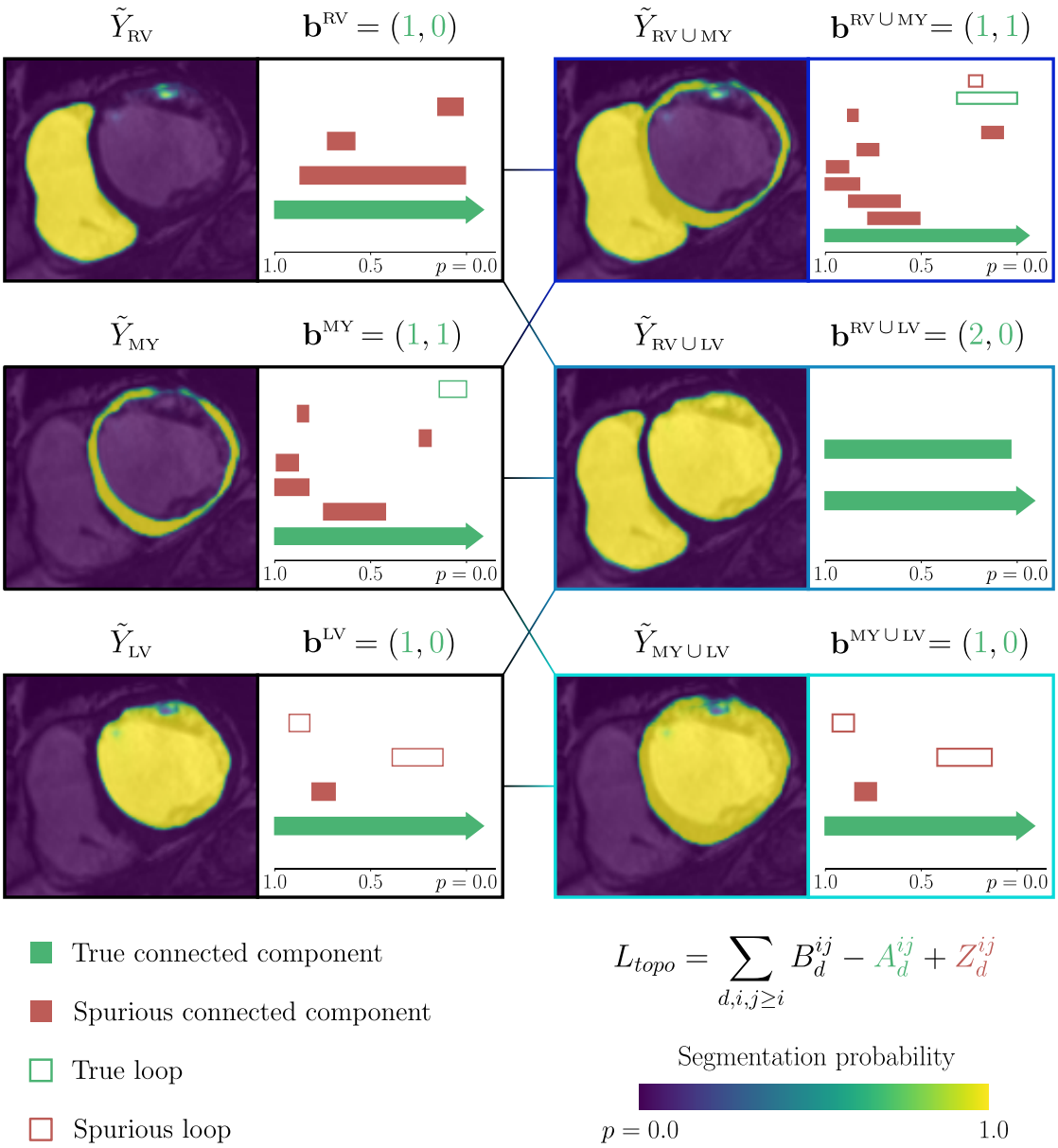


Figure 7.4: Construction of the loss L_{T}. Each probabilistic segmentation ($\tilde{Y}_i$ or $\tilde{Y}_{i \cup j}$), is accompanied by its associated persistence barcode (only bars with $\Delta p_{d,l} \geq 0.05$ are shown). L_{T} weighs the persistence of topological features which match the topological description (A_d^{ij}; depicted as green bars), against those which do not (Z_d^{ij}; depicted as red bars). To sensitise L_{T} to multi-class label map topology, these are summed over all topological dimensions (d), and individual and paired label sets ($i, j \geq i$).

In 3D, multi-class object topology is informed by that of all foreground classes, class label pairs and class label triples (Bazin et al., 2007). Correspondingly, our 3D, multi-class topological losses take the form:

$$L_{\mathrm{T}} = \sum_{d,i,j\geq i,k\geq i} B_d^{ijk} - A_d^{ijk} + Z_d^{ijk} \tag{7.4.4}$$

$$A_d^{ijk} = \sum_{l=1}^{B_d^{ijk}} \Delta p_{d,l}(\hat{Y}_{i\cup j\cup k}) \tag{7.4.4a}$$

$$Z_d^{ijk} = \sum_{l=B_d^{ijk}+1}^{\infty} \Delta p_{d,l}(\hat{Y}_{i\cup j\cup k}) \tag{7.4.4b}$$

In the following 3D experiments (see Section 7.5.4 and Section 7.5.5), we scrutinise the incremental value of label triples $(i, j \geq j, k \geq i)$ over only individual and paired labels $(i, j \geq j)$; how these gains relate to task specification; and the associated increase in computational demand.

7.4.4 CNN post-processing framework

As per previous work (Clough et al., 2019; Byrne et al., 2021), we employ our topological loss function in a CNN-based segmentation post-processing framework. In the guise of test time adaptation, this scheme seeks an improvement in inferred topology by fine tuning a pre-trained CNN, $f(\mathbf{X}; \boldsymbol{\theta})$. This achieves a new set of network parameters $\boldsymbol{\theta}_n$, optimised to correct the topology of the individual test case $\mathbf{X}_n$. It should be noted that the image features learned during this process are not assumed to be generalisable, and hence are not maintained for the segmentation of subsequent test images.

However, since topology is a global property, there are many segmentations that potentially minimise L_{T}. Hence, where V_n is the number of pixels in $\mathbf{X}_n$, a mean squared error (MSE) similarity constraint (weighted by the hyperparameter λ) limits test time adaptation to the minimal set of modifications necessary to align the segmentation and the topological prior, $\mathbf{B}$. Our topological post-processing

(TP) framework is mediated by:

$$L_{\text{TP}} = L_{\text{T}}(f(\mathbf{X}_n;\boldsymbol{\theta}_n),\mathbf{B}) + \lambda L_{\text{MSE}}(f(\mathbf{X}_n;\boldsymbol{\theta}_n), f(\mathbf{X}_n;\boldsymbol{\theta})) \tag{7.4.5}$$

$$L_{\text{MSE}} = \frac{1}{V_n}|f(\mathbf{X}_n;\boldsymbol{\theta}) - f(\mathbf{X}_n;\boldsymbol{\theta}_n)|^2 \tag{7.4.5a}$$

7.4.5 Implementation

As noted by Shin et al. (2020), computation of the persistence barcode is an intensive procedure, introducing significant inefficiencies within CNN training. This is particularly problematic in 3D and multi-task applications, in which our loss formulation demands the extraction of multiple barcodes (accounting for individual and combined label classes) per gradient update.

Previous works, including our own (Clough et al., 2019; Byrne et al., 2021), have employed the "TopologyLayer" implementation of Gabrielsson et al. (2020). This excellent resource was favoured for its close integration with the popular deep learning library PyTorch (Paszke et al., 2019) (used throughout this work) and its clear documentation. However, in the context of large, 3D image data, it has two major drawbacks: (1) without modification, TopologyLayer computes simplicial rather than cubical persistence, leading to inefficiencies and possible connectivity ambiguities in the combinatorial representation of pixel data; and (2) by its authors' own admission, TopologyLayer is best suited to computation concerning small- to medium-sized topological spaces, and makes no attempt to optimise execution in the setting of large data. These drawbacks result in lengthy computational times for the extraction of the persistence barcode, rendering such an approach impractical when applied to clinical data.

In this work, we have implemented a thin wrapper[1] around the open source "CubicalRipser" library (Kaji et al., 2020), integrating its functionality with the automatic differentiation engine of PyTorch. The CubicalRipser implementation overcomes the highlighted limitations of the TopologyLayer package, its authors suggesting that to the best of their knowledge, theirs is the fastest and most memory efficient program for computing the PH of weighted cubical complexes (Kaji et al., 2020). We build on the associated performance gains, by leveraging Python

[1] Available at https://github.com/nick-byrne/topological-losses.

multiprocessing to extract the persistence barcodes of multiple probabilistic segmentations in parallel. The performance gains and limitations of our approach are discussed in Section 7.6.2.

7.5 Experiments

7.5.1 Experimental methods and baselines

We compare the performance of our experimental methods with the following well-established baselines:

U-Net	The discrete segmentation maximising the pixel-wise probability inferred by a conventionally trained and fully supervised U-Net.
CCA	Connected component analysis (CCA): the discrete segmentation composing the per class, largest connected components of U-Net.

We test the performance of several flavours of topological post-processing. Each of these optimises the weights of the pre-trained, baseline U-Net architecture, deploying a variant of our topological loss function described in Equation 7.4.5. As for U-Net, the final result is returned by taking the discrete segmentation that maximises the topologically optimised, pixel-wise probability. We consider the following design choices:

$\text{TP}_{i,j=i}$	Post-processing according to a set of per class topological priors, specified for individual foreground labels only.
$\text{TP}_{i,j\geq i}$	Post-processing according to multi-class topological priors, specified for all individual and paired foreground labels.
$\text{TP}_{i,j\geq i,k\geq j}$	Post-processing according to multi-class topological priors, specified for individual, paired and tripled foreground labels.

We also compare performance across different pixel connectivity relations. Given the link between cubical complex construction and conventional pixel connectivity (foreground, background) (see Section 7.3.1), we use the following superscript notation: CCA^0 and TP^0 consider topology according to (4, 8) and (6, 26) pixel connectivities in 2D and 3D, respectively; CCA^N and TP^N consider topology according to (8, 4) and (26, 6) pixel connectivities. In the case of topological post-processing, the superscript also indicates the dimensionality of the cubical cells used in complex construction.

7.5.2 Metrics and statistical analysis

We use the generalised Dice similarity coefficient (Crum et al., 2006) (GDSC) as our primary, objective measure of spatial overlap performance (Crum et al., 2006), and note the change induced by post-processing as ΔGDSC:

$$\text{GDSC} = \frac{2\sum_{k=2}^{K}|Y_k \cap \hat{Y}_k|}{\sum_{k=2}^{K}|Y_k| + |\hat{Y}_k|} \tag{7.5.1}$$

We also record per class segmentation performance using the Dice similarity coefficient (Dice, 1945) (DSC), Hausdorff distance (Huttenlocher et al., 1993) (HDD) and mean surface error between predicted and ground truth segmentations.

We characterise topological accuracy using two metrics. The Betti error (BE) measures the total deviation between inferred and ground truth topology ($\mathbf{b}$ versus $\hat{\mathbf{b}}$). In 2D, this assessment is made according to the Betti numbers of all individual foreground classes and all pairs of foreground classes:

$$\text{BE} = \sum_{i,j \geq i} \|\mathbf{b}^{i \cup j} - \hat{\mathbf{b}}^{i \cup j}\|_1 \tag{7.5.2}$$

In 3D, BE also considers all class triples:

$$\text{BE} = \sum_{i,j \geq i, k \geq j} \|\mathbf{b}^{i \cup j \cup k} - \hat{\mathbf{b}}^{i \cup j \cup k}\|_1 \tag{7.5.3}$$

To understand whether improvements in segmentation topology translate into anatomically meaningful segmentations we also make a binary assessment of topo-

logical success rate (TS):

$$\mathrm{TS} = \begin{cases} 1, & \text{if BE} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{7.5.4}$$

As per the notation for experimental and baseline methods, we add superscripts to make explicit the pixel connectivity used in the calculation of these metrics: BE^0 and TS^N, for example.

Spatial overlap results are described by their median and interquartile range. To account for large differences in performance, BE is described by the i^{th} percentile (P_i) and percentile ranges indicated. Results for GDSC and BE are compared using Wilcoxon signed rank test with Bonferroni correction. TS is described by sample proportion (ρ) and its associated standard deviation (σ_ρ), and compared using exact binomial test after Bonferroni correction.

7.5.3 2D Short axis segmentation

Experimental setting

In this experiment we exemplify our approach and novel implementation within multi-class, 2D short axis segmentation, using a subset of the publicly available Automatic Cardiac Diagnosis Challenge (Bernard et al., 2018) (ACDC) training dataset (Bernard et al., 2018). Ignoring irregular anatomical appearances at apex and base, we extract the three mid ventricular slices from each short axis stack, including diastolic and systolic frames from all 100 patients. This achieves a data corpus of 600 short axis images, all sharing a common topological description summarised by the multi-class prior presented by Equation set 7.4.1. As per the winning submission to the ACDC Challenge, all image-label pairs were resampled to an isotropic pixel spacing of 1.25 mm (less than the mean and median spatial resolution of the training data) and normalised to have zero mean and unit variance (Isensee et al., 2018). Subjects were randomly divided between training, validation and test sets in the ratio 2:1:1, stratified by diagnostic group according to ACDC classification.

For its ubiquity and state of the art performance, we choose a 2D U-Net (Ronneberger et al., 2015) as the architecture for our pre-trained CNN. This was trained

using CE loss and the combined training and validation set of 450 images, for 16,000 iterations. We employed the Adam optimiser (Kingma and Ba, 2014) with a learning rate of 10^{-3}. Each minibatch contained ten patches of size 352 by 352, randomly cropped from ten different patients, zero padding where necessary. Data augmentation applied random rotations between $\pm 15°$.

Using our novel implementation based on cubical persistence, topological post-processing was performed on the inferred multi-class segmentations of the held-out test set. As described in Section 7.3.1, we investigated both 0- and N-*constructions* of the segmented images as cubical complexes. Each sought to minimise L_{TP} for the topological priors expressed by Equation set 7.4.1. In Equation 7.4.5 we used a value of $\lambda = 1000$. Test time adaptation used the Adam optimiser (Kingma and Ba, 2014) with a learning rate of 10^{-5} for 100 iterations. The reported hyperparameters for both supervised training and topological post-processing were optimised using the validation set of 150 examples.

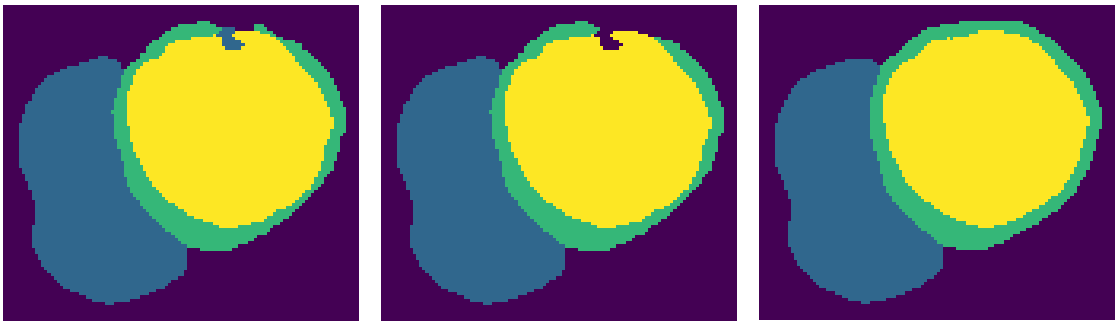
Results

Row one of Table 7.1 demonstrates that irrespective of the considered pixel connectivity, U-Net segmentations exhibit topological errors in approximately 15% of cases. This equates to around 0.35 Betti errors on average and up to 5 in the worst case. Note that these failings present despite strong spatial overlap performance that is consistent with the state of the art. This confirms our hypothesis that measures of spatial overlap do not reliably predict topological performance.

It is apparent from rows two and three of Table 7.1 that approximately two thirds of individual topological errors incurred by U-Net are resolved by rudimentary CCA, suppressing all but the largest connected component per class. Both CCA^0 ($p = 0.012$) and CCA^N ($p = 0.011$) significantly improve BE compared with U-Net prediction. These gains translate into a significant improvement in topological success rate ($p = 0.006$ and $p = 0.006$, respectively), accounting for approximately one half of the cases predicted with incorrect topology by U-Net. These results are consistent with previous works (Painchaud et al., 2020), observing the tendency of CNNs to infer the presence of spurious connected components.

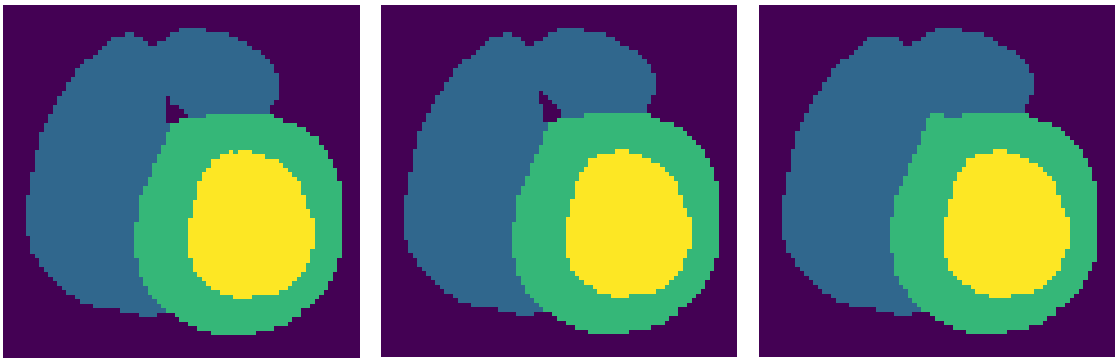
Table 7.1: Performance on 2D short axis segmentation. All results reflect performance on the held out test set, derived from the publicly available Automatic Cardiac Diagnosis Challenge (Bernard et al., 2018) (ACDC) data. Spatial overlap is measured by the generalised Dice similarity coefficient (Crum et al., 2006) (GDSC: $P_{50(P_{25},P_{75})}$) and averaged over diastolic and systolic cardiac phases. The change induced by post-processing is ΔGDSC: $P_{50(P_{25},P_{75})}$. Topological accuracy is indicated by the Betti error (BE: $P_{99(P_{98},P_{100})}$) and sample proportion gaining topological success (TS: $\rho_{(\sigma_\rho)}$). Please see Section 7.5.1 and Section 7.5.2 for definitions.

	Spatial overlap		**Topology**			
	GDSC	ΔGDSC	BE^0	BE^N	TS^0	TS^N
U-Net	0.932 (0.917,0.943)	—	5.510 (4.020,8.000)	4.530 (3.000,8.000)	0.853 (0.354)	0.873 (0.333)
$+CCA^0$	0.932 (0.917,0.943)	+0.000 (0.000,0.000)	3.020 (2.000,4.000)	—	0.927 (0.260)	—
$+CCA^N$	0.932 (0.917,0.943)	+0.000 (0.000,0.000)	—	2.000 (2.000,2.000)	—	0.947 (0.224)
$+TP^0_{i,j=i}$	0.934 (0.916,0.943)	+0.000 (0.000,0.001)	1.000 (1.000,2.000)	—	0.973 (0.162)	—
$+TP^0_{i,j\geq i}$	0.933 (0.917,0.943)	+0.000 (0.000,0.001)	0.510 (0.000,1.000)	—	0.987 (0.113)	—
$+TP^N_{i,j=i}$	0.933 (0.917,0.943)	+0.000 (0.000,0.001)	—	1.000 (0.002,1.000)	—	0.980 (0.140)
$+TP^N_{i,j\geq i}$	0.933 (0.917,0.944)	+0.000 (0.000,0.001)	—	0.000 (0.000,1.000)	—	0.993 (0.083)



(a) U-Net (b) CCA (c) $\mathrm{TP}_{i,j\geq i}$

Figure 7.5: (a) U-Net-based segmentation predicts a spurious connected component of the RV. (b) Conventionally applied connected component analysis (CCA) assumes the superiority of large connected components of the discrete segmentation, removing the offending error. However, CCA is insensitive to high-dimensional topological features of the data and cannot rectify errors associated with loops. (c) Our losses for topological post-processing (TP) are sensitive to the presence of high-dimensional topological features of the *probabilistic* segmentation. Mediated by CNN parameter optimisation, they permit expressive correction, considering the interaction between segmentation classes: suppression of the anomalous RV component is accompanied by consolidation of the LV cavity and completion of the myocardial torus.



(a) U-Net (b) $\mathrm{TP}_{i,j=i}$ (c) $\mathrm{TP}_{i,j\geq i}$

Figure 7.6: Specification of topological priors for all pairs of foreground labels captures the interaction between classes. (a) U-Net predicts a segmentation with an anomalous gap between the RV cavity and the LV myocardium. (b) Topological post-processing (TP) according to a set of per class priors ($i, j = i$) is insensitive to this error. This is because alone, both the RV and LV myocardium have correct topology. (c) Additional priors, specifying the topology of all foreground object pairs ($i, j \geq i$) allows topological post-processing to rectify the segmentation.

Sensitive only to 0D topological features, CCA cannot rectify errors related to the presence of loops. Moreover, as most frequently employed, CCA takes a discrete approach to post-processing, considering connected components of the crisp (or thresholded) segmentation. Comparatively, our topological loss functions admit improvements in *probabilistic* segmentation topology by CNN optimisation, both prediction and post-processing being conditioned on the test image. This admits topological improvement via expressive modification (see Figure 7.5).

The superior topological performance of our loss functions is quantified in Table 7.1. Row four reflects the naive extension of previous work (Clough et al., 2019, 2020) to the multi-class setting, specifying a single topological prior per class: $\mathrm{TP}^{0}_{i,j=i}$. As expected, this setup admits improved topological performance compared with CCA, resolving all topological errors in over 97% of cases. This reflects topological correction in more than half the cases for which CCA did not return an accurate result and a substantial improvement in the average BE. Despite these gains, this improvement does not return significant increases in either BE ($p = 0.069$) or TS ($p = 0.393$), when compared against CCA^{0}.

Row six outlines the performance of the full realisation of our approach in 2D, specifying topological priors per class and per pair-wise combination of classes: $\mathrm{TP}^{0}_{i,j\geq i}$. This provides topologically accurate results in almost 99% of test images, failing in just two, and returns the lowest average BE of just 0.02. Compared with the naive specification of per class topological priors only, pair-wise combinations expose a richer description of multi-class segmentation topology. This includes sensitivity to class hierarchies, adjacency and containment. The benefit of this extension is illustrated in Figure 7.6 and results in significant improvements in BE ($p = 0.026$) and TS ($p = 0.023$), when compared against CCA^{0}.

We now examine the impact of design choices within the implementation of our topological loss functions. More specifically, we address the use of either 0- or *N-constructed* cubical complexes. Comparing rows five and seven ($\mathrm{TP}^{0}_{i,j\geq i}$ versus $\mathrm{TP}^{N}_{i,j\geq i}$), we observe a small improvement in topological performance associated with the *N-constructed* cubical complex (recall that this corresponds to an 8-*connected* representation of foreground pixels). This resolves all topological errors in one further test case, achieving a topologically accurate result in more than 99% of examples. However, neither the improvement in BE ($p = 1.000$) nor

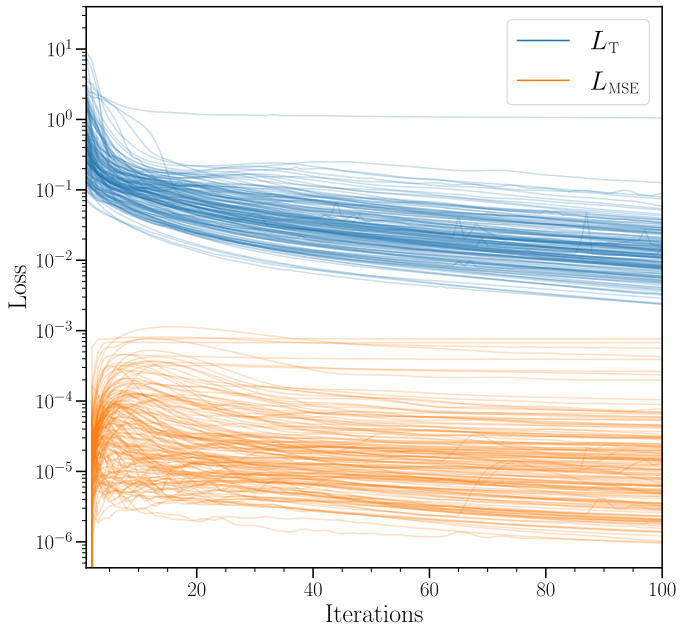


Figure 7.7: Optimisation of L_{TP} in the task of 2D short axis segmentation for the entire test set. These trajectories reflect the following design choices: specification of topological priors per class and per class pair and using 0-*constructed* cubical complexes. During topological post-processing, the topological loss, L_{T}, is smoothly and stably reduced. This is balanced by the similarity constraint, L_{MSE}. After the initial shock associated with the introduction of topological prior knowledge, these two terms balance to achieve reliable performance.

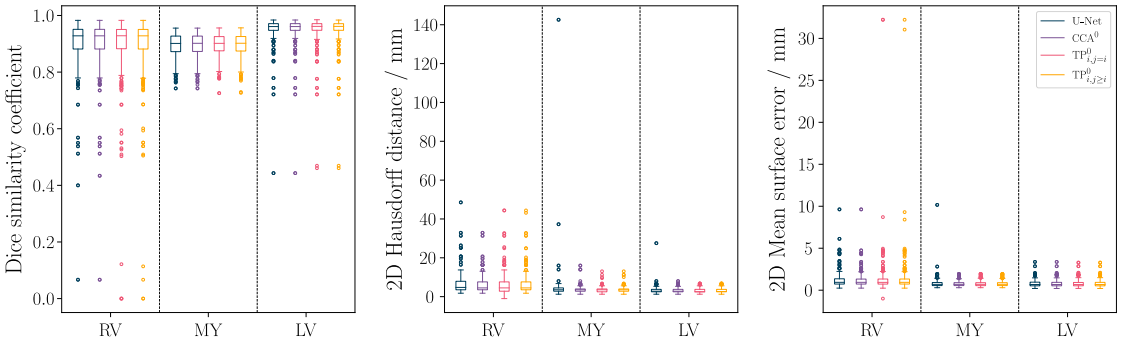


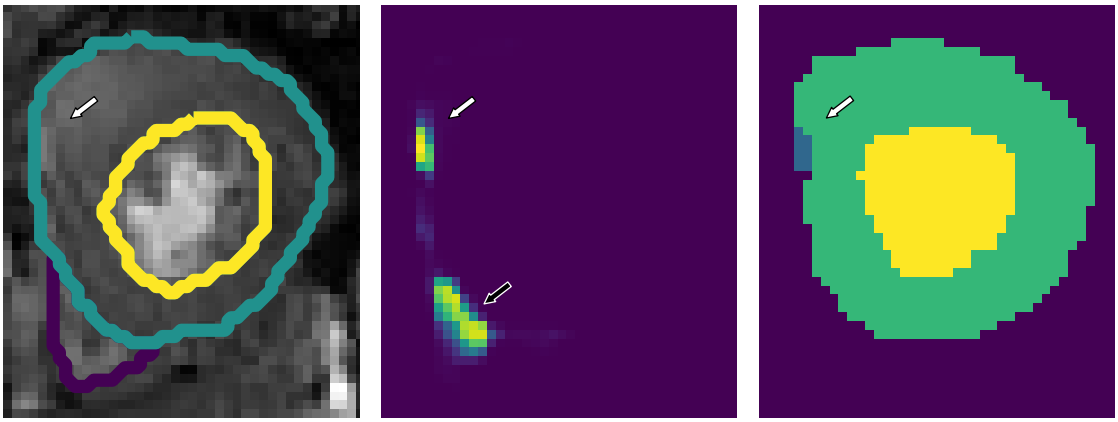
Figure 7.8: Test set performance as characterised by per class metrics for the right ventricle (RV), left ventricle (LV) and left ventricular myocardium (MY). In addition to significantly improving topological performance, our approach to TP maintains and marginally improves DSC (left), HDD (centre) and mean surface error (right) compared against U-Net segmentation. Distinctive outliers and their effect on mean surface error of the RV class are examined as failure cases in Section 7.5.3.

TS ($p = 1.000$) are significant. It should be noted that a similar trend relating topological performance to pixel connectivity is observed across U-Net and all post-processing methodologies considered. Given the apparent consistency between these options, we adopt the 0-*construction* in Section 7.5.4 and Section 7.5.5.

Taking a closer look at the performance of our topological loss functions, Figure 7.7 illustrates the evolution of the different terms of Equation 7.4.5. This demonstrates a predictable and smooth improvement in L_{T}, balanced by a small increase in the similarity constraint L_{MSE}. The latter ensures that post-processed segmentations cannot deviate dramatically from the initial U-Net prediction and maintains spatial overlap performance. This is borne out in Table 7.1: on average, all post-processing methods tested achieved marginal improvement in the GDSC compared with U-Net. Whilst this might have been statistically significant in some cases, any gain was not substantial when compared with the DSC between manual segmentations of different observers: estimated to be around 0.9 (Bai et al., 2018). Rather, we are encouraged that on this task, our topological loss functions do not degrade spatial overlap performance. Furthermore, Figure 7.8 demonstrates that our methods are at least consistent with, if not superior to, state of the art approaches across a range of per class evaluation metrics.

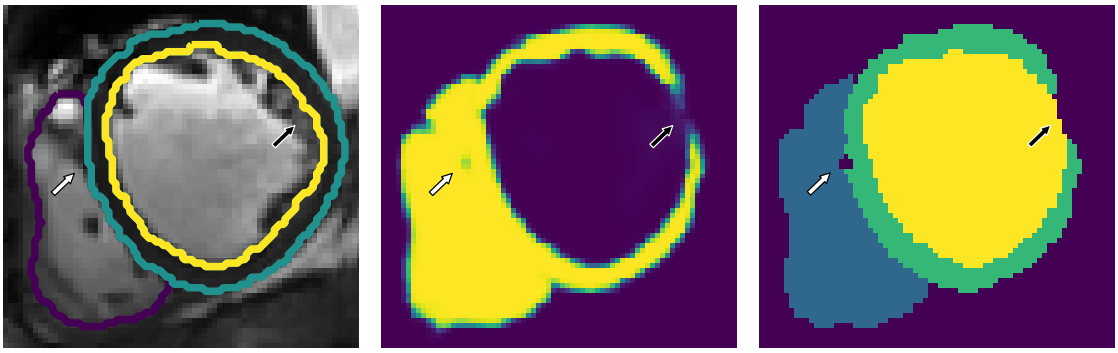
This is not to imply that the topological loss functions presented are impervious to error. Figure 7.9 demonstrates a case in which U-Net equivocally suggests the presence of more than one connected component to the right ventricular cavity. In fact, the persistence of the spurious component exceeds that most closely associated with the ground truth cavity. Compounding this challenge, both components are topologically consistent with the multi-class topological prior when considered in conjunction with the myocardium. Accordingly, optimisation of the topological loss maintains and suppresses the incorrect and correct components, respectively. As a result, topological post-processing removes any spatial overlap between predicted and ground truth RV and significantly increases the 2D mean surface error (both these results are apparent in Figure 7.8). A similar failure case is summarised by Figure 7.10.

As the former, poor performance is associated with the equivocal prediction of topological features. Here, rather than by connected components, post-processing is confounded by the prediction of more than one topologically credible loop.



(a) $\mathbf{X}$ overlay $\mathbf{Y}$ (b) $\tilde{Y}_{\mathrm{RV}}$ (c) $\mathrm{TP}_{i,j\geq i}$

Figure 7.9: Anatomically spurious components that are both persistent and consistent with the topological prior, can cause errors: (a) A mid-ventricular slice from the test set, with ground truth contours overlaid. (b) The probability of right ventricle (RV) cavity predicted by the U-Net. The white arrow indicates a false positive detection. According to the U-Net probability map, this spurious component is the most persistent 0D topological feature. (c) Furthermore, this error is topologically credible, being adjacent to the left ventricular myocardium. Consequently, the false positive component is maintained by topological post-processing (TP). Meanwhile, the probability mass correctly associated with the RV cavity is suppressed (black arrow). This error is associated with poor surface error performance and reflected as an outlier in Figure 7.8.



(a) $\mathbf{X}$ overlay $\mathbf{Y}$ (b) $\tilde{Y}_{\mathrm{RV}\cup\mathrm{MY}}$ (c) $\mathrm{TP}_{i,j\geq i}$

Figure 7.10: Anatomically spurious loops that are consistent with the topological prior can cause errors: (a) A mid-ventricular slice from the test set, with ground truth contours overlaid. (b) The probability of right ventricle (RV) cavity or myocardium (MY) predicted by the U-Net. The white arrow indicates the suggestion of a false positive loop. The black arrow points to an anomalous gap in the MY. (c) Topological post-processing (TP) with respect to the myocardial prior seeks to close the anomalous gap and maximise the persistence of a single loop. At the same time, optimising the topology of the combined RV and MY also expects a single loop. In this case, the anomalous loop wins out, is maintained in the final segmentation and actually inhibits completion of the myocardial torus.

7.5.4 3D Whole heart segmentation

Experimental setting

In this experiment, we apply our methods to the task of segmenting multi-class, whole heart anatomy from isotropic, high spatial resolution image data. We consider a semantic subset of the MM-WHS Challenge task (Zhuang et al., 2019), seeking a segmentation of the CMR volume into left and right atria (LA and RA), left and right ventricles (LV and RV) and left ventricular myocardium (MY)[2]. In the context of topological optimisation, this task is made particularly challenging by the semantics of the publicly available ground truth training data. Primarily concerned with measures of spatial overlap and surface error, these segmentations need and do not convey clinically meaningful topology. Hence, this application of our method seeks to not only refine, but also to impose meaningful segmentation topology. In the present work, we attempt this feat in the context of test time adaptation, using post-processing to introduce topological features to a CNN pre-trained in their absence.

Whilst a subset of MM-WHS patients exhibit congenital or structural heart disease, our inspection of the data suggests that the anatomy of all cases conforms to the topology of the healthy heart. Hence our topological prior (see Equation set 7.4.2) reflects: (1) continuity of the atrioventricular junctions - each atrium is trivially connected to its associated ventricle; (2) isolation of the left and right heart - no communication between left atrium and ventricle with either of the right atrium and ventricle; and (3) association of the right ventricular blood pool with medial epicardium - the RV class is trivially bound to the wall of the MY label.

Theoretically, the topological prior associated with 3D image data should reflect all foreground classes, their pair-wise combination (as per the 2D case), and finally their collection in all possible groups of three. However, for pragmatic reasons we overlook the topology of label triples. Firstly, moving from a 2D to a 3D implementation admits a significant increase in the time required to compute the PH barcode. Hence, the number of barcodes that must be computed at each topological post-processing iteration strongly influences execution time. For

[2]The full MM-WHS task also seeks segmentation of the ascending aorta and pulmonary artery. To satisfy computational resources, we disregard these segments

five foreground classes, including three-way label combinations almost doubles the number of barcodes from 15 to 25[3]. Secondly, the semantics of our topological prior do not anticipate anatomically meaningful features involving the combination of three classes (as would be the case for an atrioventricular septal defect, for example). Note that for evaluation both BE and TS assess segmentation topology using a sound theoretical basis, additionally considering foreground classes combined in groups of three.

Our experiment made use of the twenty publicly available training cases and performance was assessed against the encrypted test set of forty examples. Prior to experimentation, all training data were resampled to an isotropic spatial resolution of 1.00 mm (approximately the median spacing of the training set) and normalised to have zero mean and unit variance. We also elected to crop all volumes tightly around the foreground region of interest. This allows us to maintain high spatial resolution (granting sensitivity to topological features associated with thin tissue interfaces, such as the atrial septum) without extreme increases in computational demand. The applicability of CNN-based foreground detection to this task and dataset is well established (Zhuang et al., 2019).

For our baseline model we trained a 3D U-Net (Çiçek et al., 2016) using CE loss. We considered the entire training set and optimised over 40,000 iterations. We employed the stochastic gradient descent optimiser, using a learning rate of 0.01 and momentum of 0.99. Each minibatch contained two large 3D image patches of size 192 by 160 by 160. Given the small training set, we made use of intensive data augmentation, pre-computing 500 examples per case. Transformations included: rotation about all three spatial axes ($[-10°, 10°]$), scaling ($[0.9, 1.1]$) and non-rigid deformation. Pixel intensities were modified using additive noise.

The results of Section 7.5.3 suggested an approximate equivalence between topological post-processing using 0- and *N-constructions* in the determination of cubical persistence. We therefore limit our investigation to the former: the default option within the CubicalRipser package (Kaji et al., 2020). Accordingly, we consider $\mathrm{TP}^0_{i,j=i}$ and $\mathrm{TP}^0_{i,j\geq i}$, seeking to optimise L_{TP} for the topological priors expressed by Equation set 7.4.2a (singular anatomical classes) and their combina-

[3]Whilst our implementation allows for their parallel computation, our computational resource (at least for this experiment, see Section 7.5.5) only includes 16 CPU cores.

tion with Equation set 7.4.2b (paired anatomical classes), respectively. Topological fine tuning was mediated by the Adam optimiser (Kingma and Ba, 2014) using a learning rate of 10^{-5} for 100 iterations. Similarity with the probabilistic segmentation inferred by the pre-trained U-Net was constrained by L_{MSE} and weighted by $\lambda = 1$. The hyperparameters reported throughout this experiment (within both U-Net training and topological post-processing) were established by five fold cross-validation over the training data.

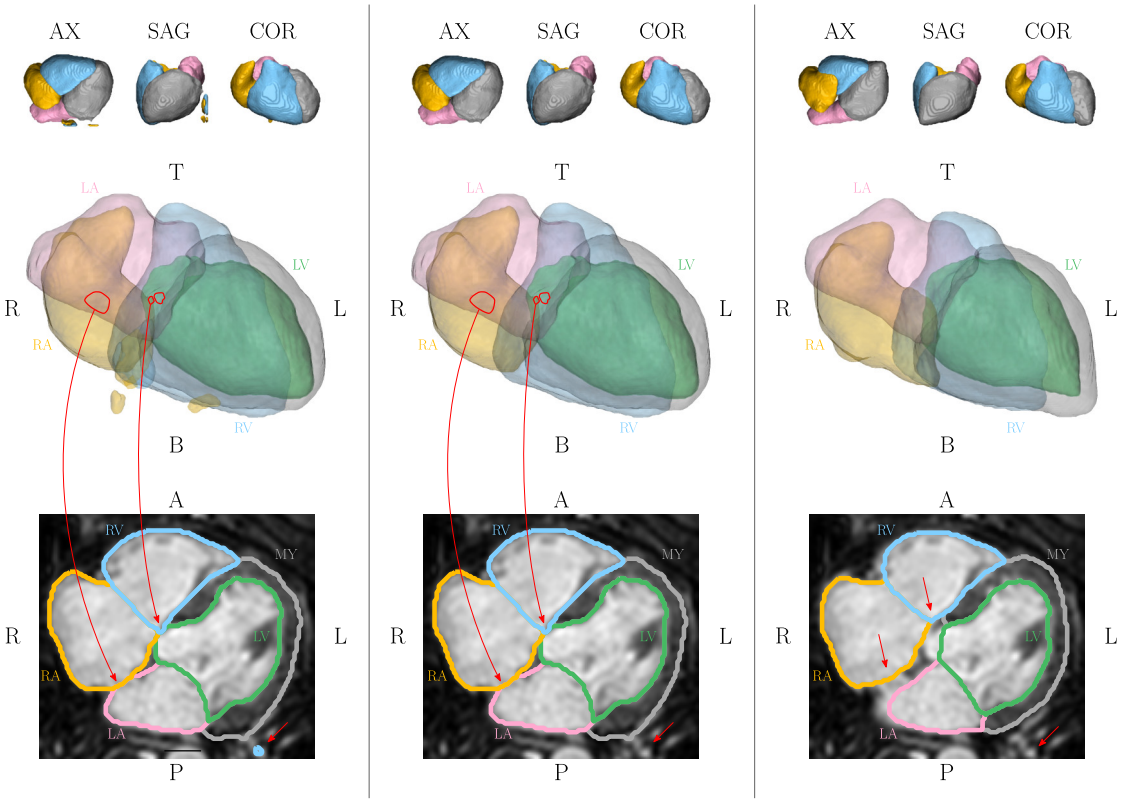
Results

Table 7.2 demonstrates that with respect to spatial overlap, our CCA^0 baseline is consistent with the leading submissions to the MM-WHS challenge (Zhuang et al., 2019). However, further assessment of topological performance requires careful consideration. Recall that the training data with which the U-Net was trained do not consistently reflect the topological prior we are seeking to impose. It could be argued, therefore, that any evaluation against this prior has limited interpretability. To inform this comparison, we examine the topological performance of the ground truth labels composing the training set: there, an average of 376.2 Betti errors excluded a single topologically accurate segmentation. Against the training labels, the topological performance of U-Net showed no statistically significant difference ($p = 0.082$).

It is clear from these results that the topology of the ground truth training labels not only diverges from our clinically relevant specification, but is actually meaningless. In this context, flavours of topological post-processing attempt to impose unseen topology at test time. However, they are not alone in this regard, the same can be said for CCA. Though widely employed as a practical approach to reduce false positives and improve spatial overlap, it must be acknowledged that in this case, CCA^0 involves the application of a limited, 0D topological prior, one that conflicts with the features of training data. Despite this, Table 7.2 endorses the former rationalisation: coupled with a significant improvement in the average GDSC ($p < 10^{-4}$), CCA^0 effectively modifies segmentation topology, reducing the number of topological errors by almost 85% ($p < 10^{-7}$). As per the results of

Table 7.2: Performance on 3D whole heart segmentation. All results reflect performance on the held out test set of the 2017 Multi-Modality Whole Heart Segmentation (Zhuang et al., 2019) (MM-WHS) data, for which images but not labels are publicly available. Spatial overlap is measured by the generalised Dice similarity coefficient (Crum et al., 2006) (GDSC: $P_{50(P_{25},\,P_{75})}$); the change induced by post-processing is ΔGDSC: $P_{50(P_{25},\,P_{75})}$. Topological accuracy is indicated by the Betti error (BE: $P_{50(P_{25},\,P_{75})}$) and sample proportion gaining Topological Success (TS: $\rho_{(\sigma_\rho)}$). Please see Section 7.5.1 and Section 7.5.2 for definitions.

	Spatial overlap		**Topology**	
	GDSC	ΔGDSC	BE^0	TS^0
U-Net	0.851 (0.812,0.877)	—	353 (277,439)	0.000 (0.000)
+CCA^0	0.865 (0.821,0.899)	+0.013 (0.005,0.025)	50 (26,72)	0.000 (0.000)
+$TP^0_{i,j=i}$	0.753 (0.684,0.786)	−0.101 (−0.148, −0.036)	11 (7,18)	0.000 (0.000)
+$TP^0_{i,j\geq i}$	0.759 (0.706,0.805)	−0.089 (−0.137, −0.047)	5 (2,8)	0.050 (0.218)



(a) U-Net (b) CCA^0 (c) $\text{TP}^0_{i,j\geq i}$

Figure 7.11: In the presence of strong spatial overlap performance, post-processing makes minimal corrections to improve segmentation topology. Notable topological features are outlined and indicated in red: (a) U-Net segmentation achieves strong spatial overlap with the ground truth but presents a number of topological errors. (b) Connected component analysis (CCA^0) successfully removes small, superfluous connected components, but the association of the left and right heart remains. (c) Topological post-processing, $\text{TP}^0_{i,j\geq i}$, successfully corrects all topological errors, including those associated with connected components and the anomalous appearance of atrial and ventricular septal defects.

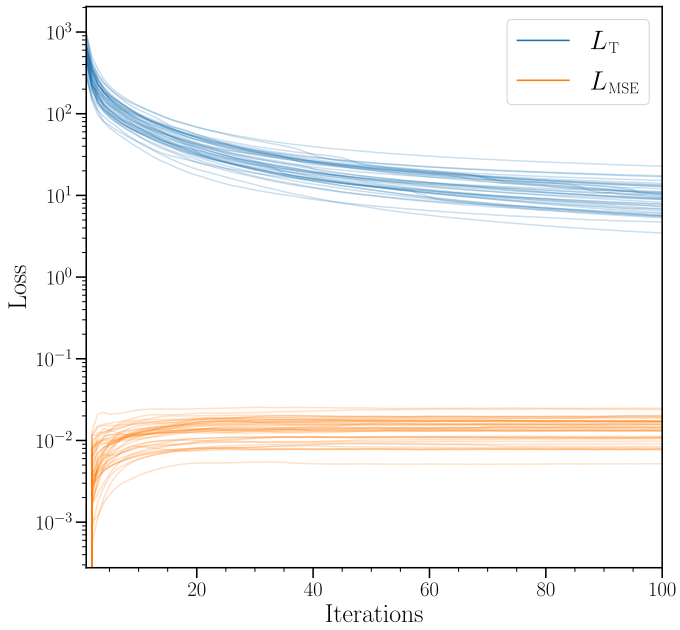


Figure 7.12: Optimisation of our topological loss in the task of 3D whole heart segmentation for the entire MM-WHS test set. Whilst predictable optimisation dynamics are achieved, compared with the 2D case, both components (L_{T} and L_{MSE}) of L_{TP} remain elevated. Here we speculate that topological optimisation is hampered by more substantial errors in U-Net prediction.

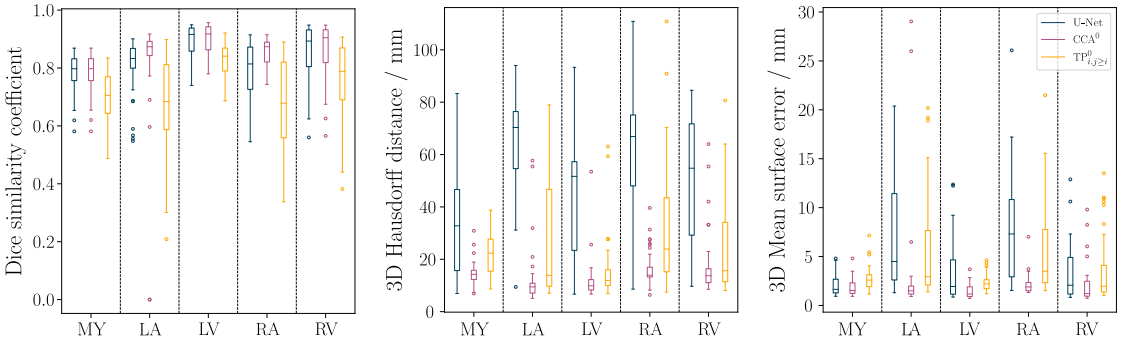


Figure 7.13: Test set performance as characterised by per class metrics for the left ventricular myocardium (MY), and blood pool cavities of the left atrium (LA), left ventricle (LV), right atrium (RA) and right ventricle (RV). Compared with U-Net prediction, topological post-processing ($\mathrm{TP}^0_{i,j\geq i}$) improves surface localisation performance, as measured by HDD (*centre*) and mean surface error (*right*). However these gains are not as large as those achieved by connected component analysis (CCA^0). Coupled with a degradation in spatial overlap performance (*left*), these results indicate the trade off between improved topological and other metrics of performance when topological post-processing is applied to this task.

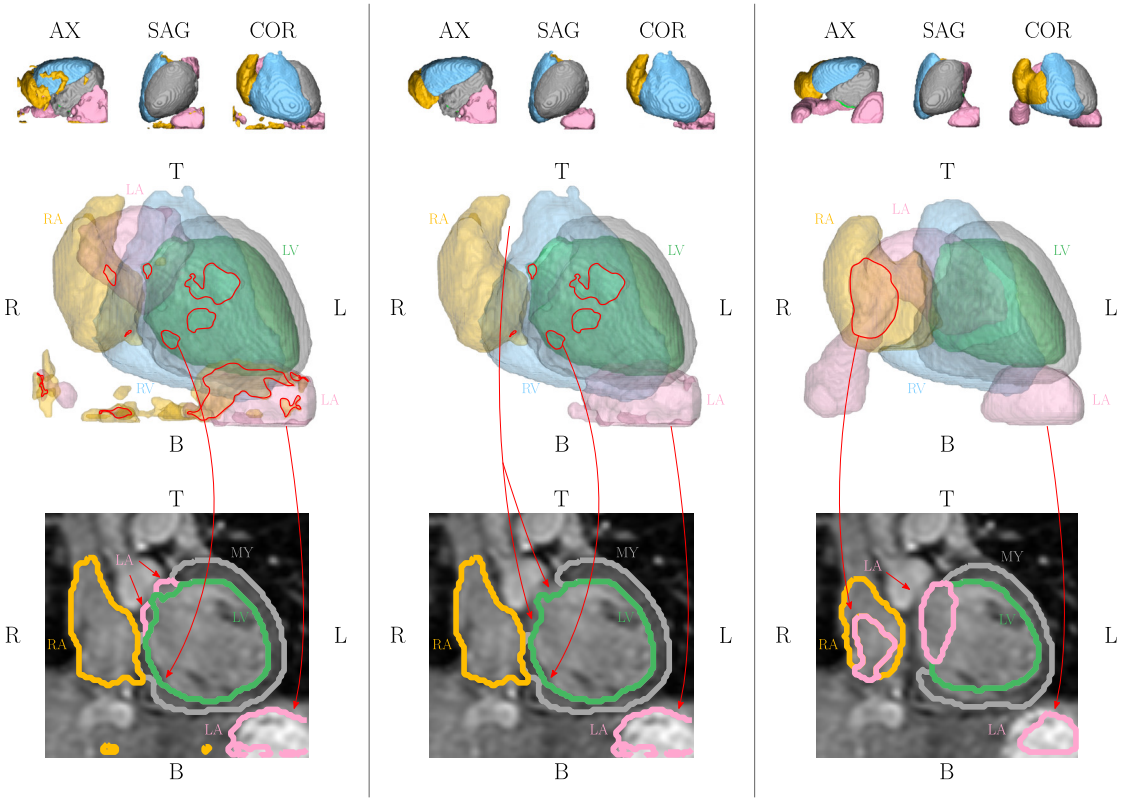
Section 7.5.3, spurious connected components make up a substantial proportion of the topological errors made by U-Net.

However, CCA^0 remains insensitive to both multi-class and high-dimensional segmentation topology. Sensitivity to such features is achieved by our topological loss functions (see Figure 7.11, Figure 7.12). Both $\text{TP}^0_{i,j=i}$ and $\text{TP}^0_{i,j\geq i}$ effectively adapt segmentation topology, significantly reducing the BE with respect to the prior specification, when compared with CCA^0 ($p < 10^{-5}$ and $p < 10^{-6}$, respectively). In the latter case, our full implementation, the average number of Betti errors is reduced by almost 90%.

It is telling that this improvement constitutes a topologically accurate segmentation in only two of forty test cases (a TS of just 5%). Coupled with the deleterious effect of topological post-processing on other metrics of segmentation performance (see Figure 7.13), this goes some way to characterise the challenge of this task and the limitations of its CNN-based solution. In this experiment, the gains in topological performance are traded off against a significant degradation in spatial overlap with the ground truth. This is true for both $\text{TP}^0_{i,j=i}$ and $\text{TP}^0_{i,j\geq i}$ ($p < 10^{-6}$ in both cases). Moreover, in addition to being statistically significant, the drop is clinically significant.

Primarily, such degradation occurs where U-Net segmentation includes multiple, probabilistically credible candidates that, to some extent, are consistent with the topological prior. Figure 7.14 demonstrates a case in which U-Net localises the LA to three distinct regions of the image (these are reflected in the persistence barcode shown in Figure 7.15). Each having a lifetime greater than 0.9, separating true from anomalous components is probabilistically equivocal. In this context, $\text{TP}^0_{i,j\geq i}$ results in a segmentation with two LA components. Remote from its true location and of significant size[4], a spurious component inferior and posterior to the apex limits spatial overlap performance. Perhaps more problematically, topological optimisation connects the LA with false positive components inferior to the base of the heart. This anatomically nonsensical communication not only limits performance, but also severely degrades spatial overlap compared with U-Net.

[4]In fact, this is the largest LA component predicted by U-Net and excludes any spatial overlap with the ground truth where CCA^0 is applied.



(a) U-Net (b) CCA^0 (c) $\mathrm{TP}^0_{i,j\geq i}$

Figure 7.14: The prediction of spurious connected components with high probability can confound topological post-processing (TP). Notable topological features are outlined and indicated in red: (a) U-Net segmentation includes significant spatial overlap and topological errors. (b) The largest connected component of the left atrium (LA) is anomalous. Naive connected component analysis (CCA^0) misidentifies the LA, resulting in zero spatial overlap for this class. (c) Probabilistic TP rectifies the majority of topological errors. However, $\mathrm{TP}^0_{i,j\geq i}$ is also confounded by anomalous LA components. The multi-class prior specifies the isolation of LA and right atrium (RA) by the expectation of two connected components. In the presence of a superfluous LA component, communication of the atria remains consistent with this specification. The result is a significant drop in spatial overlap performance.

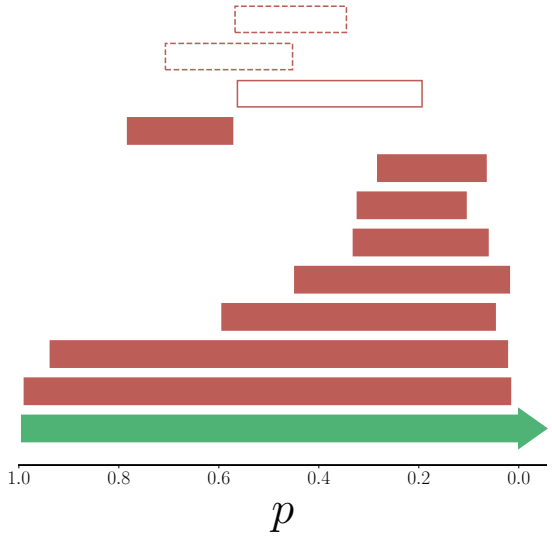
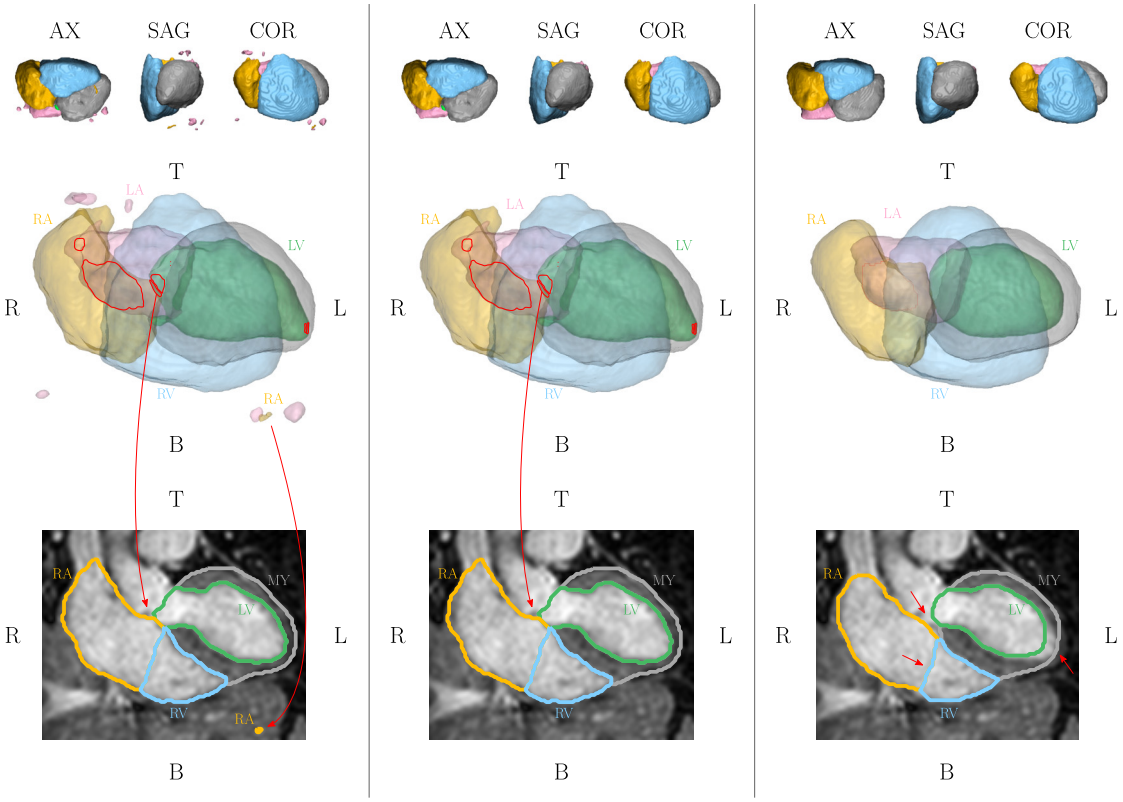


Figure 7.15: The topological credibility of candidate features is reflected by the persistence barcode. In Figure 7.14, U-Net segmentation indicates the presence of three connected components of the LA. This is captured by the barcode shown above ($\Delta p_{d,l}(\tilde{Y}_{\mathrm{LA}})$), which includes three bars with a lifetime greater than 0.9. In the context of multi-class topological optimisation, it is ambiguous which of these is the connected component correctly associated with the LA, and which are spurious features. This limits both topological and spatial overlap performance of our approach. Schematic representation follows the convention of Figure 7.3, with the addition of dashed bars to indicate 2D topological features (voids within 3D foreground objects).

Such errors are problematic in and of themselves. However, given its combinatorial nature (considering label combinations), they can also confound multi-class topology. For example, the prior seeks to isolate left and right atria by specifying that their combination comprise two connected components. Implicit within this specification, however, is an assumption that both structures demonstrate correct topology individually. In Figure 7.14, this is not the case. Provided a probabilistic segmentation that equivocally suggests multiple atrial components, CNN optimisation can reduce L_{T} by reinforcing the communication between the atria. This modification is at odds with anticipated anatomy and results from the interaction between U-Net error and the specification of multi-class topology by label combination.



(a) U-Net (b) CCA^0 (c) $\mathrm{TP}^0_{i,j\geq i}$

Figure 7.16: Insufficient CNN capacity limits the performance of topological post-processing (TP). Notable topological features are outlined and indicated in red: (a) U-Net segmentation presents several spatial overlap and topological errors, including the appearance of atrial and ventricular communications. (b) Connected component analysis (CCA^0) successfully eliminates spurious connected components, but cannot rectify the anomalous association of classes. (c) $\mathrm{TP}^0_{i,j\geq i}$ corrects errors associated with both superfluous connected components and both atrial and ventricular communications. However, the CNN features learned by optimising our topological loss, introduce associated segmentation changes remote from any topological error. This is most obvious at the apex of the heart, but also can be observed from an apparent translation of the tricuspid valve.

Secondarily, it is clear that our implementation is limited by sub-optimal learning of the image features necessary to combine strong topological and spatial overlap performance. We speculate that this might result from the limited representational capacity of the 3D U-Net in the context of constrained GPU memory. On top of the widely acknowledged computational challenges of 3D CNN implementation, our approach must contend with three further requirements: (1) relevant topological features are defined in relation to thin tissue interfaces, hence high spatial resolution is vital; (2) topological gains are introduced by test time adaptation, hence neuron activations and backpropagation are also required during inference; and (3) abstract anatomical topology is defined in relation to the entire image, so we cannot trivially exploit patch-wise prediction. In addressing these challenges, memory-saving adaptations constrain the statistical capacity of our CNN.

Consequently, we hypothesise that the U-Net architecture employed may lack the representational capacity to learn sufficiently optimal features. For the most part, $\mathrm{TP}^{0}_{i,j\geq i}$ makes sensible modifications, trading off spatial overlap performance for improved topological accuracy. However, the extent and location of these changes is neither uniformly minimal nor predictable. Figure 7.16 illustrates a case for which improved performance is accompanied by geometrical changes at the cardiac apex, a locale seemingly remote from any associated topological error. In the context of a fixed post-processing iteration budget, we speculate that such changes result from insufficient representational capacity. This phenomenon also contributes to the significant reduction in spatial overlap and surface error performance when compared with U-Net.

7.5.5 3D CHD segmentation

Patient-specific topological priors

In this experiment, we continue our investigation of applying topological loss functions to 3D, multi-class, cardiac segmentation. Rather than publicly available data, here we leverage the ELCH dataset developed in Chapter 5. Compared with the MM-WHS dataset explored in the previous Section 7.5.4, the ELCH data also comprise 3D isotropic, spatially high resolution, steady state free precession, volumetric images, segmented into left and right atrial and ventricular components (amongst

a wealth of other anatomically meaningful classes). In other respects, however, the ELCH dataset presents several semantic differences: (1) unlike the MM-WHS data, the ground truth accompanying each ELCH case describes a clinically meaningful multi-class topology, reflecting haemodynamic continuity by pixel adjacency; (2) rather than label the left ventricular myocardium, the ELCH data delineate a combined myocardium class, describing the muscular tissue surrounding both left and right ventricles; and (3) this myocardial class is manually segmented to represent the muscular annuli surrounding atrioventricular and ventriculoarterial valves. As a consequence of (2) and (3), and compared with the MM-WHS formulation (see Equation 7.4.2), the ELCH myocardium class presents notable high-dimensional features. At least in the setting of normal anatomy, we anticipate its Betti numbers to be: $\mathbf{b}^{\mathrm{MY}} = (1, 2, 0)$, the two 1D loops being associated with the passage of blood through the left and right heart respectively. For the application of our topological loss functions, however, perhaps the most important difference between ELCH and MM-WHS data concerns anatomical variety. Within the ELCH data, extant congenital defects disrupt the topology of the normal heart (presented in Equation 7.4.2) such that a uniform topological prior is not sufficient. The impact of different defects is explained in the following examples.

Consider the presence of a VSD: a communication between the normally isolated left and right ventricles. Whilst this lesion has no effect on the anticipated Betti numbers describing either ventricle alone ($\mathbf{b}^{\mathrm{LV}} = \mathbf{b}^{\mathrm{RV}} = (1, 0, 0)$), compared with the normal heart (left of arrow below) their combination now makes up a single connected component:

$$\mathbf{b}^{\mathrm{LV}\cup\mathrm{RV}} = (2, 0, 0) \quad \xrightarrow{\mathrm{VSD}} \quad \mathbf{b}^{\mathrm{LV}\cup\mathrm{RV}} = (1, 0, 0) \tag{7.5.5}$$

Each additional VSD increments the number of 1D loops by one:

$$\mathbf{b}^{\mathrm{LV}\cup\mathrm{RV}} = (2, 0, 0) \quad \xrightarrow{2\times\mathrm{VSD}} \quad \mathbf{b}^{\mathrm{LV}\cup\mathrm{RV}} = (1, 1, 0) \tag{7.5.6}$$

Such that in the general case we can write for $n > 0$:

$$\mathbf{b}^{\mathrm{LV}\cup\mathrm{RV}} = (2, 0, 0) \quad \xrightarrow{n\times\mathrm{VSD}} \quad \mathbf{b}^{\mathrm{LV}\cup\mathrm{RV}} = (1, n-1, 0) \tag{7.5.7}$$

Note that in these cases, the multiplicity of n VSDs introduces high-dimensional topological features to the prior specification.

Whilst an identical pattern holds for the presence of atrial septal defects (ASDs), the same approach can be applied for any defect or surgical modification that associates two segments isolated from one another in the normally connected heart. Though complicated somewhat by the presence of (or coincidence with) an atrioventricular septal defect, the same reasoning holds and can be applied.

In opposite fashion to septal defects, valvular atresias isolate the anatomical structures that ensure haemodynamic continuity between segments of the normal heart. For example, mitral valve atresia (MA) compromises the left atrioventricular junction, precluding communication between the left atrium and ventricle, and modifying multi-class topology. As per our VSD example, MA has no effect on the topology of the individual left atrial and ventricular labels ($\mathbf{b}^{\mathrm{LA}} = \mathbf{b}^{\mathrm{LV}} = (1, 0, 0)$), only becoming apparent upon consideration of their pairing within a multi-class prior:

$$\mathbf{b}^{\mathrm{LA}\cup\mathrm{LV}} = (1,0,0) \quad \xrightarrow{\mathrm{MA}} \quad \mathbf{b}^{\mathrm{LA}\cup\mathrm{LV}} = (2,0,0) \tag{7.5.8}$$

The atretic valve haemodynamically isolates atrium and ventricle, giving rise to two separately connected components. As per extra-anatomical communication, this principle can be applied in relation to any defect or surgical modification introducing discontinuity to great or small circulations.

The topology of more complex defects can be described by the application of both principles (association and discontinuity) simultaneously to more than one set of anatomical labels. For example, consider double inlet left ventricle (DILV), in which both left and right atria connect to the LV. In this case, the atrioventricular associations are summarised:

$$\begin{aligned}
\mathbf{b}^{\mathrm{LA}\cup\mathrm{LV}} &= (1,0,0) & & \mathbf{b}^{\mathrm{LA}\cup\mathrm{LV}} = (1,0,0) && \text{(7.5.9a)}\\
\mathbf{b}^{\mathrm{RA}\cup\mathrm{LV}} &= (2,0,0) & & \mathbf{b}^{\mathrm{RA}\cup\mathrm{LV}} = (1,0,0) && \text{(7.5.9b)}\\
\mathbf{b}^{\mathrm{LA}\cup\mathrm{RV}} &= (2,0,0) & \xrightarrow{\mathrm{DILV}} \quad & \mathbf{b}^{\mathrm{LA}\cup\mathrm{RV}} = (2,0,0) && \text{(7.5.9c)}\\
\mathbf{b}^{\mathrm{RA}\cup\mathrm{RV}} &= (1,0,0) & & \mathbf{b}^{\mathrm{RA}\cup\mathrm{RV}} = (2,0,0) && \text{(7.5.9d)}
\end{aligned}$$

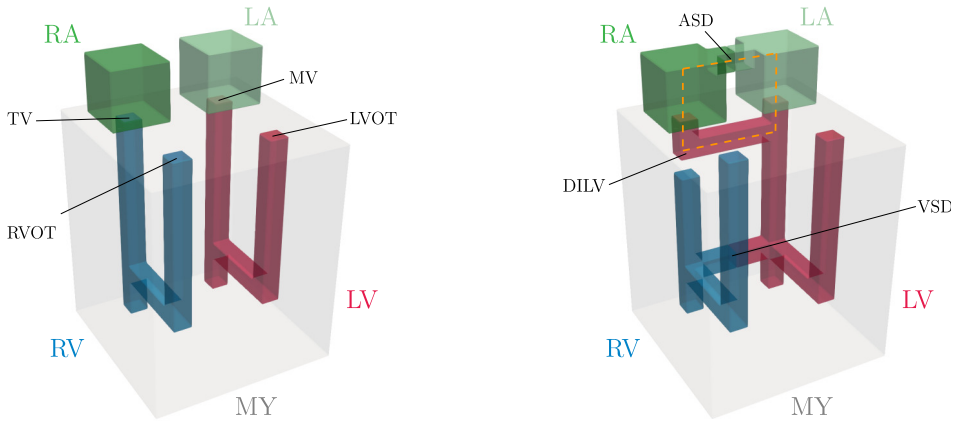
Equations 7.5.9a and 7.5.9b describe the normally related LA and LV, and the abnormal association of the RA with LV, resulting in double inlet. Equations 7.5.9c and 7.5.9d describe the normal isolation of LA from RV, and the abnormal separation of RA from RV.

This approach can be extended to the combination of defects. To the DILV example presented, consider the addition of an ASD. The association between LA and RA aside, the combination of DILV with ASD impinges on the topology of the union of LA, RA and LV classes:

$$\mathbf{b}^{\mathrm{LA}\cup\mathrm{RA}\cup\mathrm{LV}} = (2,0,0) \quad \xrightarrow[\mathrm{ASD}]{\mathrm{DILV}} \quad \mathbf{b}^{\mathrm{LA}\cup\mathrm{RA}\cup\mathrm{LV}} = (1,1,0) \tag{7.5.10}$$

This combination of defects introduces a high-dimensional topological feature (a 1D hole) to the combination of these classes. Note that this hole is not described by any pair-wise combination of LA, RA and LV.

A multi-class topological prior, considering the combination of all labels, and their paired and tripled combinations, provides a powerful, flexible and patient-specific description of CHD. Critically, this approach is also clinically relevant: the patient's diagnosis, from which these priors can be constructed, is typically known from previous echocardiography examination, ahead of the acquisition of any tomographic images (such as CMR or CT) to be segmented. Whilst the Betti numbers for such a prior can be defined through analytical consideration of anatomical association, such an approach is not trivial and particularly challenging when considering the union of three classes and high-dimensional features. Hence, we find it more efficient and convenient to build a schematic representation of the normal heart, to which the defects of a clinically known, qualitative diagnosis can be automatically added. The Betti numbers of the adjusted schematic (see Figure 7.17) can be straightforwardly computed by homology or otherwise. Note that given the topological accuracy of ELCH labels, we could have computed multi-class topological priors from the ground truth segmentations of the test set. However we prefer the presented approach, its use of a cardiac schematic illustrating a potential implementation within a clinical workflow, in which ground truth segmentations are clearly unavailable.



(a) Normal heart (b) DILV + ASD + VSD

Figure 7.17: Multi-class topological priors provide a flexible description of CHD anatomy, that can be tailored to the individual patient. For their efficient specification, we consider a schematic representation of the normal heart to which the discrete defects of a clinically known, qualitative diagnosis can be added. In the task specified in Section 7.5.5: (a) The great (left atrium (LA) and left ventricle (LV)) and small (right atrium (RA) and right ventricle (RV)) circulations of the normal heart each comprise a single connected component without high-dimensional topological features. (b) The presence of congenital defects modifies the multi-class topology of the heart: (1) a ventricular septal defect (VSD) allows communication of the left and right ventricles, the LV and RV forming a single connected component; and (2), the combined presence of double inlet left ventricle (DILV) and atrial septal defect (ASD) introduces a 1D hole (dashed orange loop) to the union of LA, RA and LV.

Experimental setting

Chapter 6 demonstrated the challenge of contending with the significant structural heterogeneity of CHD. Therefore, to most effectively expose the topological performance of our loss functions, we purposively sample data from the ELCH cohort. By selecting transposition of the great arteries (TGA) and VSD diagnostic groups (for which ventricular disproportion is reduced, see Figure 5.6 and Figure 5.7), we realise a distribution in which continuous changes in cardiac morphology (the geometry and relative proportion of anatomical segments) are minimised; but in which a broad variety in discrete label map topology is maintained.

Prior to experimentation, we manually adjust the ELCH ground truth labels to include the papillary muscles and endocardial trabeculation within the relevant chamber class: one of LA, RA, LV or RV. Were these features maintained within ground truth data, they would contribute a large number of high-dimensional topological features so as to obscure the patient-specific description of defects developed in the previous section. Whilst this change is made to facilitate methodological development, we note that a topologically accurate presentation of such defects normally outweighs that of the fine structural details composing the endocardium - at least within a 3D model for patient-specific planning. Removal of the latter simplifies intra-class topological complexity, discarding all high-dimensional features and meaning that uniformly: $\mathbf{b}^{\mathrm{MY}} = (1, 2, 0)$; $\mathbf{b}^{\mathrm{LA}} = \mathbf{b}^{\mathrm{RA}} = \mathbf{b}^{\mathrm{LV}} = \mathbf{b}^{\mathrm{RV}} = (1, 0, 0)$. At the same time, the sensitivity of patient-specific, multi-class topology to the thin tissue interfaces which divide the heart and define defects is maintained.

The importance of class triples to the topological description of patient-specific CHD presents a challenge to the parallel computation of persistence barcodes. Were our task formulation to require the segmentation of the sixteen anatomical classes considered in Chapter 6, each gradient update would demand the calculation of 696 barcodes. Whilst this experiment benefits from access to a 32-core central processing unit (CPU)[5], interrogating such an extensive set of labels with this hardware is impractical, even via our parallel implementation. Instead, we limit our experiment to the five classes shared by MM-WHS and ELCH datasets: MY, LA, RA, LV, RV. Whilst this formulation requires the computation of 30 barcodes per optimisation step, these can be entirely parallelised as per previous experiments. This allows for a practical implementation which remains both theoretically sound and clinically relevant.

In equal proportion, we split the combined TGA and VSD cases (considering steady state free precession data as the only CNN input) into training and test sets of 20 and 40, respectively (matching the separation of examples within the MM-WHS challenge). Otherwise, we prepared training data in line with the rationale presented in Section 7.5.4. Prior to experimentation, all training data were resampled to an isotropic spatial resolution of 1.00 mm, cropped tightly around the foreground classes of interest and normalised to have zero mean and unit vari-

[5] A CPU with only 24 cores was used in the completion of Section 7.5.4

ance. Our baseline U-Net optimisation (including hyperparameter settings) also mirrored the previous 3D experiment with respect to loss function, data augmentation and batch preparation. For our baseline model we trained a 3D U-Net (Çiçek et al., 2016), considering an augmented training set and applying CE loss over 40,000 iterations.

For each test case, a multi-class topological prior was automatically defined by the homology of a cardiac schematic (see Figure 7.17), adjusted for the patient-specific diagnoses established in Chapter 5. The only caveat to this approach concerns those patients exhibiting networks of many small, associated VSDs, including "Swiss cheese" septal morphology (Serraf et al., 1992). In such cases ($n = 4$), the number of inter-ventricular communications is not known *a priori*, being comprised by small septal defects demonstrated at the limit of spatial resolution of any clinical imaging modality, be it echocardiography or CMR. Therefore for these patients, we disregard the contribution of high-dimensional topological features to L_{T}, for any label pair or triple describing both LV and RV.

Apart from our choice of topological prior, post-processing (including hyperparameter settings) was performed identically to Section 7.5.4. This meant the representation of image data using the 0-*constructed* cubical complex, optimising various flavours of L_{TP} over 100 gradient updates, mediated by the Adam optimiser (Kingma and Ba, 2014) and a learning rate of 10^{-5}. Similarity with the probabilistic segmentation inferred by the pre-trained U-Net was constrained by L_{MSE} and weighted by $\lambda = 1$. Post-processing sought to align CNN prediction with priors specifying the topology of individual ($\mathrm{TP}^0_{i,j=i}$), paired ($\mathrm{TP}^0_{i,j\geq i}$) and tripled ($\mathrm{TP}^0_{i,j\geq i,k\geq i}$) labels, evaluating the incremental benefit conferred by each.

Results

At least with respect to spatial overlap (as measured by the GDSC), Table 7.3 suggests that the performance of our U-Net baseline is consistent with the ELCH state of the art established in Chapter 6. Whilst the median GDSC is slightly elevated (being 0.883 compared with 0.854, see Table 6.1) it must be remembered that the current experiment considers only a subset of both the ELCH data and task specification. Limiting segmentation to the cardiac chambers and myocardium within

Table 7.3: Performance on 3D CHD segmentation. All results reflect performance on the held out test set of the ELCH data. Spatial overlap is measured by the generalised Dice similarity coefficient (Crum et al., 2006) (GDSC: $P_{50(P_{25},P_{75})}$); the change induced by post-processing is ΔGDSC: $P_{50(P_{25},P_{75})}$. Topological accuracy is indicated by the Betti error (BE: $P_{50(P_{25},P_{75})}$) and sample proportion gaining Topological Success (TS: $\rho_{(\sigma_\rho)}$). Please see Section 7.5.1 and Section 7.5.2 for definitions.

	Spatial overlap		**Topology**	
	GDSC	ΔGDSC	BE^0	TS^0
U-Net	0.883 (0.844,0.897)	—	332 (233,620)	0.000 (0.000)
+CCA^0	0.884 (0.848,0.898)	+0.001 (0.000,0.003)	281 (191,580)	0.000 (0.000)
+$TP^0_{i,j=i}$	0.688 (0.628,0.725)	−0.176 (−0.210, −0.152)	32 (25,40)	0.000 (0.000)
+$TP^0_{i,j\geq i}$	0.693 (0.629,0.727)	−0.180 (−0.218, −0.152)	13 (10,17)	0.000 (0.000)
+$TP^0_{i,j\geq i,k\geq i}$	0.686 (0.620,0.722)	−0.186 (−0.228, −0.143)	11 (8,18)	0.000 (0.000)

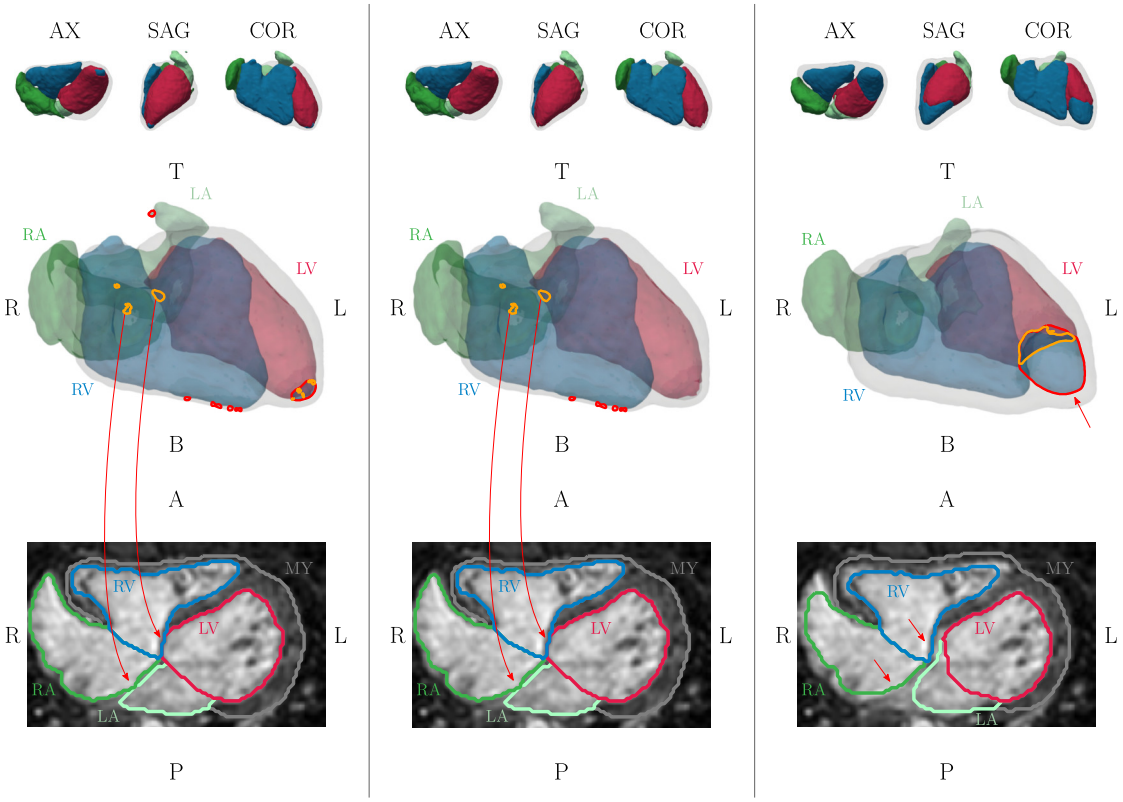
VSD and TGA diagnostic groups (for which structural heterogeneity is reduced) likely explains the gain in U-Net spatial overlap performance.

Unlike in the preceding 3D experiment, owing to the topological fidelity of the ELCH ground truth, interrogation of the topological performance of the U-Net baseline is meaningful: optimised against a topologically accurate ground truth, we may expect U-Net predictions with fewer Betti errors, culminating in an elevated rate of topological success. However, row one of Table 7.3 does not support this expectation, there being a median number of 332 topological errors per case and no entirely accurate prediction.

As in Section 7.5.4, CCA^0 improves topological performance, reducing the median number of Betti errors to 281. After Bonferroni correction, Wilcoxon signed rank test finds that this gain is statistically significant ($p < 10^{-5}$). However, compared with its application to the the MM-WHS data, CCA^0 does not as effectively reduce BE, on average accounting for only 15% of errors within U-Net segmentation of ELCH cases. In contrast, this reduction was closer to 86% in Section 7.5.4. We suggest that this difference results from the high-dimensional topological features expressed by our patient-specific topological description of CHD. Effectively a 0D topological prior, CCA is insensitive to the 1D holes presented by extra anatomical association (such as by multiple defect, see Figure 7.17) and by the ELCH specification of myocardial anatomy. Being highly relevant to the clinically meaningful representation of defects, this limits the utility of CCA to refine the predicted segmentation of CHD anatomy.

As per previous experiment, improved topological performance is achieved by the application of our topological loss functions. Whether informed by topological priors describing individual, paired or tripled labels, these make a substantial and significant reduction in the number of Betti errors compared with both U-Net and CCA^0 prediction ($p < 10^{-7}$ in all cases). Figure 7.18 illustrates a case in which segmentation topology is effectively modified, removing spurious atrial and ventricular septal defects. Whilst considerably reduced, BE remains elevated compared with the previous 3D experiment (see Table 7.2). Given the topological complexity introduced by congenital defects, this is perhaps to be expected.

Incrementally, rows three to five of Table 7.3 support the application of topological post-processing with respect to priors informed by the combination of



(a) U-Net (b) CCA^0 (c) $\mathrm{TP}^0_{i,j\geq i,k\geq i}$

Figure 7.18: Provided a high fidelity U-Net segmentation, topological post-processing (TP) effectively imparts clinical meaning to inferred segmentations. (a) U-Net prediction includes several topological errors: those involving single classes only (outlined in red and including a superfluous but persistent right ventricle (RV) connected component) in addition to spurious atrial septal defect (ASD) and ventricular septal defect (VSD) (outlined in orange). (b) Superfluous connected components (0D topological features) are effectively removed by connected component analysis (CCA^0). However, high-dimensional topological errors remain (including 1D holes at the inferior margin of the myocardium and inter-chamber communications). (c) Topological post-processing successfully corrects high-dimensional and inter-chamber errors. Importantly, after optimisation by $\mathrm{TP}^0_{i,j\geq i,k\geq i}$, both atrial and ventricular septa appear intact, as would be expected by clinical knowledge of previous imaging examination. However, note that the superfluous RV component anomalously inferred by U-Net is not suppressed, presenting a probabilistically credible candidate.

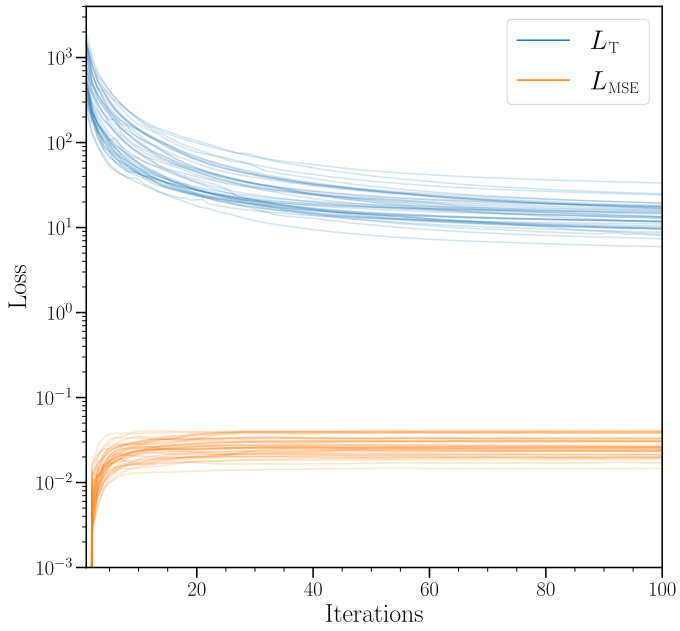


Figure 7.19: Optimisation of our topological loss in the task of 3D whole heart segmentation for the ELCH-derived test set. In common with our previous experiments, our formulation admits smooth and predictable optimisation dynamics. However, as per our previous 3D experiment (see Figure 7.12) and in contrast with the 2D case (see Figure 7.7), both the topological (L_{T}) and (dis)similarity components (L_{MSE}) remain elevated.

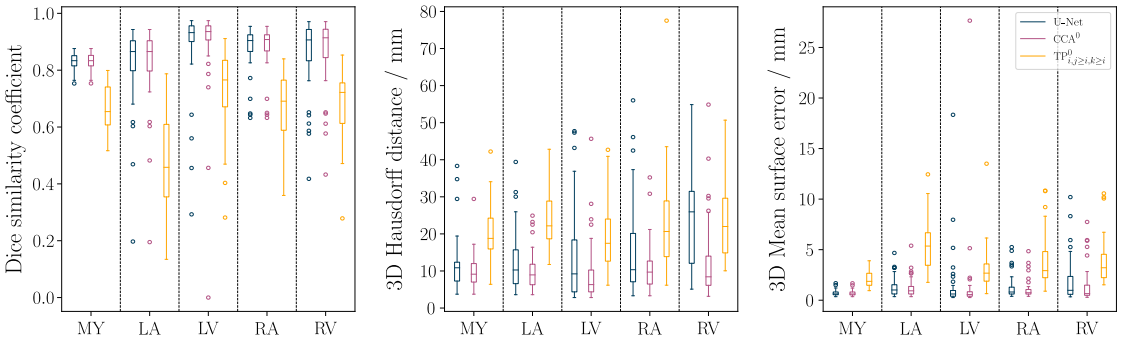
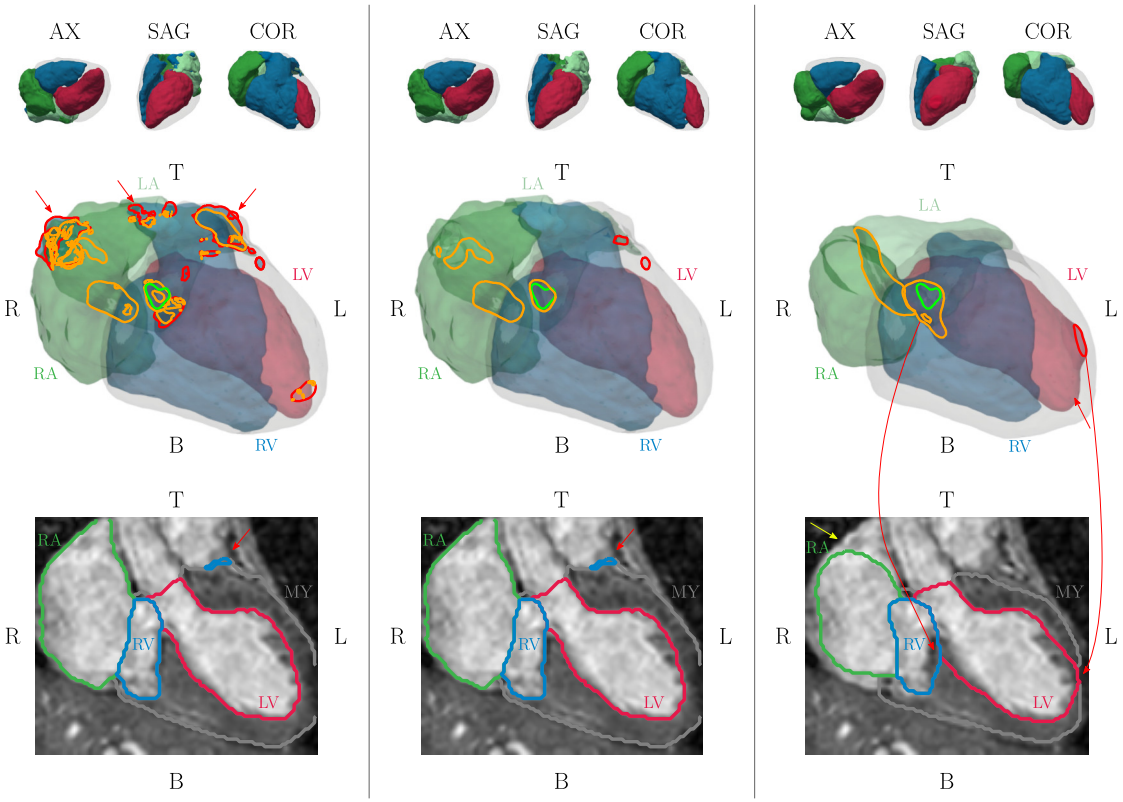


Figure 7.20: Post-processing performance on the ELCH-derived test set. Compared with U-Net prediction, topological post-processing ($\mathrm{TP}^0_{i,j\geq i,k\geq i}$) improves topological performance. However, and in contrast to our previous 3D experiment (see Figure 7.13) both surface localisation and spatial overlap performance worsen. On this challenging CHD segmentation task, these results indicate a trade off between topological and other metrics of performance. Disregarding high-dimensional topology, connected component analysis (CCA^0) provides a means of improving surface delineation whilst maintaining spatial overlap performance.

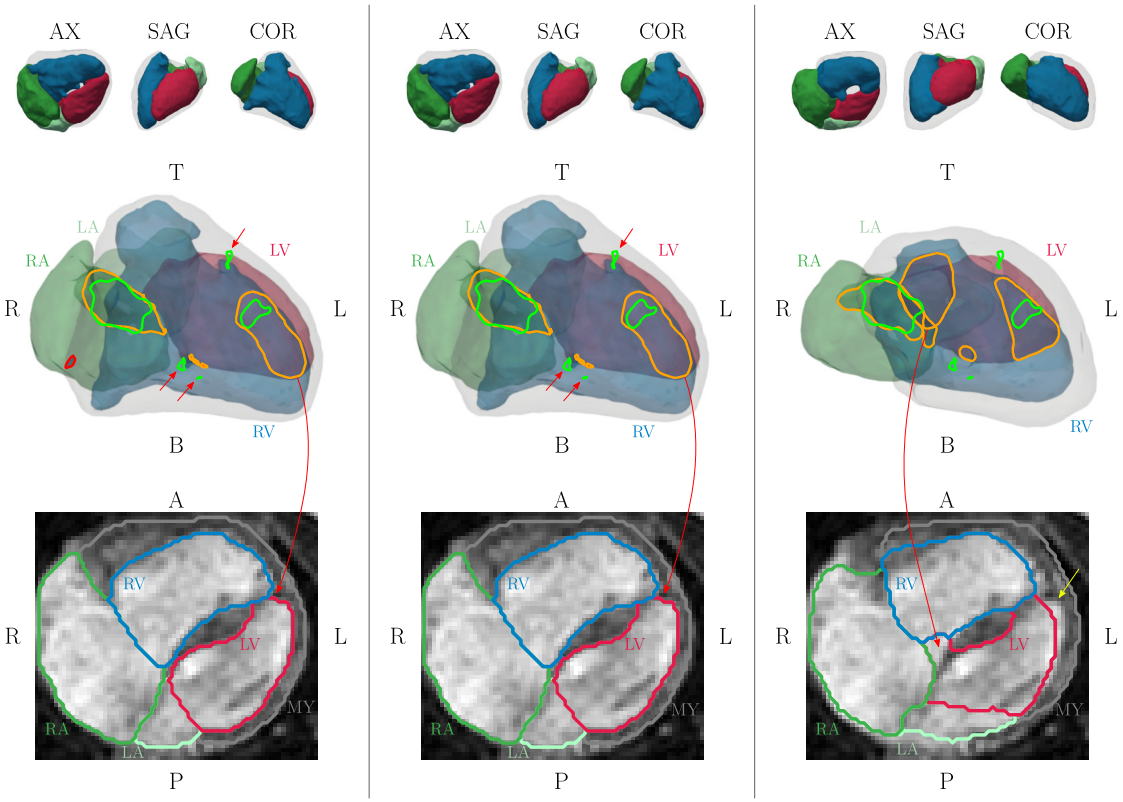
foreground classes. Considering individual label topology only, $\mathrm{TP}^0_{i,j=i}$ accounts for 90% of U-Net errors. This fraction increases to over 96% once paired labels ($\mathrm{TP}^0_{i,j\geq i}$) are added to the prior, incrementally realising a statistically significant result ($p < 10^{-6}$). Extending the prior to include the topologies of all class combinations when collected in groups of three, $\mathrm{TP}^0_{i,j\geq i,k\geq i}$ also reduces BE, the median falling to just 11 (rectifying almost 97% of U-Net errors). This result supports the theoretical description of 3D multi-class topology via the combination of all individual, paired and tripled labels (Bazin et al., 2007). Furthermore, Figure 7.19 demonstrates that such priors remain consistent with our loss formulation, demonstrating predictable behaviour on the ELCH data. However, given that the computational resources necessitated by $\mathrm{TP}^0_{i,j\geq i,k\geq i}$ deliver only a marginal (and not statistically significant) gain, $\mathrm{TP}^0_{i,j\geq i}$ might be considered the most efficient and pragmatic implementation in this case.

Despite this reduction in the number of Betti errors, and similarly to Section 7.5.4, these gains do not translate into predicted segmentations that are topologically and clinically meaningful. TS remains zero for all cases, across all baseline and post-processing methods. Substantively, this level of performance is consistent with the findings of Section 7.5.4, where we applied our approach to the normal cardiac anatomy exhibited by the MM-WHS test set. Moreover, in the current experiment we again observe a trade off between improvements in segmentation topology (as delivered by topological post-processing) and other aspects of performance. Each of $\mathrm{TP}^0_{i,j=i}$, $\mathrm{TP}^0_{i,j\geq i}$ and $\mathrm{TP}^0_{i,j\geq i,k\geq i}$ are associated with clinically and statistically significant reductions in spatial overlap performance, the median GDSC falling by almost 0.2 compared with U-Net prediction ($p < 10^{-6}$ in all cases). This compromise is also reflected by worsening surface localisation performance (see Figure 7.20). The challenge of balancing these ambitions in the context of our highly challenging, CHD segmentation task is demonstrated in Figure 7.21. As per Figure 7.14, we primarily attribute this tension to low fidelity U-Net segmentation, and in particular to the prediction of multiple probabilistically credible candidates that whilst clinically nonsensical, remain at least partially consistent with the topological prior.



(a) U-Net (b) CCA^0 (c) $\text{TP}^0_{i,j\geq i,k\geq i}$

Figure 7.21: Low fidelity U-Net prediction can severely limit the performance of topological post-processing (TP). (a) U-Net prediction includes many spurious connected components (outlined in red) and interface communications between classes (outlined in orange). Collectively, the latter far exceed the single VSD expected by clinical prior knowledge (outlined in green). (b) At least some of these topological errors are associated with spurious, extra-anatomical connected components. These are resolved by connected component analysis (CCA^0). Whereas, several anomalous ASDs remain. (c) Whilst the number of anomalous interfaces is reduced by topological post-processing, improvement is largely achieved through the agglomeration of the substantial U-Net errors, rather than their isolation and suppression. This case also reflects the challenge of resolving the high-dimensional topological features associated with the myocardium. To do this faithfully, predicted segmentations must capture the muscular annuli of the cardiac valves. In this instance, U-Net fails to capture these high resolution features, $\text{TP}^0_{i,j\geq i,k\geq i}$ instead introducing a superfluous hole towards the apex of the left ventricle (LV) (red arrow). Finally, the yellow arrow indicates the tendency for $\text{TP}^0_{i,j\geq i,k\geq i}$ to modify segmentation topology remote from relevant errors, and which we previously attributed to limited CNN capacity.



(a) U-Net (b) CCA^0 (c) $\text{TP}^0_{i,j\geq i,k\geq i}$

Figure 7.22: The insensitivity of U-Net to clinically salient high resolution features limits the performance of topological post-processing (TP), particularly where these failings are in conflict with the prior applied. (a) U-Net segmentations frequently fail to capture topologically and clinically relevant features defined at the limit of the spatial resolution of the ground truth data (outlined in green). Accordingly, there is a mismatch between the number of VSDs predicted by U-Net (outlined in orange) and that anticipated clinically: two versus four. (b) Whilst connected component analysis (CCA^0) successfully removes a small, spurious right ventricle (RV) component from the U-Net prediction (outlined in red), it makes no impact upon topological errors associated with inter-chamber communications. (c) In this context, although topological post-processing realises a segmentation that predicts the correct number of atrial and ventricular septal defects (ASDs and VSDs, outlined in orange), the absence of fine details from U-Net prediction precludes their correct spatial localisation by $\text{TP}^0_{i,j\geq i,k\geq i}$. This means that topological post-processing actually introduces a large, anomalous, inlet VSD. As per previous example, this case also illustrates sub-optimal topological modification, including changes to the left ventricular apex at a location seemingly remote from topological error (yellow arrow).

More specifically, however, the performance of our approach is constrained by the inability of U-Net to capture high resolution features of image data. Unlike in Section 7.5.4, where our loss functions imposed a topological prior to restore the gross anatomy of the normal heart, in this experiment we seek to refine topological features defined at the limit of spatial resolution. Small VSDs between LV and RV, holes within the thin tissue interface making up the atrial septum, and the myocardial annuli sourrounding the atrioventricular valves (see Figure 7.21) are all examples. As shown in Figure 7.22, topological post-processing cannot enhance or refine the representation of such features if not adequately captured by U-Net. Given that the topological specification is determined *a priori*, the absence of high resolution topological features from CNN prediction can confound our loss functions. Moreover, due to the combinatorial nature of our multi-class prior (and as observed in Section 7.5.4), the degradation in performance extends beyond the immediately compromised classes, also affecting associated label pairs and triples.

7.6 Discussion

7.6.1 Context

Frequently concerned with the delineation of structured anatomical targets, medical image segmentation often benefits from the incorporation of prior information. Our work adds to a rich body of research concerned with methods to leverage and make best use of such priors. Historically, a motivation for their use has been a desire to constrain predicted segmentations to anatomically credible morphology and, in the case of multi-class segmentation, configuration. We interpret these ambitions through the lens of topology, a property that, whilst long acknowledged as critical to anatomical plausibility, has rarely been considered explicitly. Instead, topology has more often been captured implicitly, within the examples comprising an atlas or statistical shape or appearance model, for example. In contrast, our work makes use of PH, an increasingly popular tool from topological data analysis, to expose the topological features of image data. Accordingly, topology provides both the motivation for and, unlike a significant body of previous work, the mathematical basis of our methodology and performance metrics.

Most recently, given the CNN-based state of the art, authors have considered means to inject prior information into parameter optimisation. Popularised by Oktay et al. (2017), anatomically constrained neural networks learn a compact, latent representation of anatomically plausible segmentations. A supervisory signal can subsequently be determined by the separation of ground truth and predicted segmentations in feature space. In common with Degel et al. (2018) and Yue et al. (2019), this approach assumes that the properties which characterise anatomy (morphology, topology etc.) can be implicitly encoded. In the time since, this has been shown not to be the case: CNNs optimised against a latent representation of plausible anatomy can still make predictions with anatomically implausible features. In response, Painchaud et al. (2020) engineered a highly successful solution based on nearest neighbour search within the augmented, learned latent space. At test time, their approach guaranteed a plausible result, accepting a small degradation in spatial overlap performance.

Compared with this family of methods, our topological loss functions permit optimisation against an explicit topological prior. This is beneficial to interpretability, making no assumption as to the faithful representation of anatomically relevant features within a learned representation. However, perhaps its greatest strength is drawn from the abstract quality of the prior information it employs. Methods based on a learned representation of anatomy are necessarily biased on the training data on which they depend. In contrast, topological priors are abstracted from the expert's knowledge of segmentation targets, rather than their appearance within particular examples. This enhances the generalisability of our approach, extending its application to the low data setting: in Section 7.5.4, our loss functions successfully improved CNN-based segmentation topology, training with just twenty cases. It seems unlikely that an effective latent representation of highly variable, 3D cardiac anatomy could be established from such a limited sample. Moreover, by decoupling prior information from its appearance within training data, our approach is less susceptible to performance losses in the presence of out of sample test cases, including pathology-induced structural variation. In fact, the use case presented (topological post-processing by test time adaptation) is well suited to such a scenario, affording a bespoke topological specification for the case at hand.

As demonstrated across all experiments, the use of Betti numbers to explicitly specify multi-class segmentation topology provides a powerful and interpretable description of anatomy. In the context of our ELCH cohort, this approach extended elegantly to the characterisation of patient-specific CHD. This illustrated not only the flexibility of a topological description, but also its consistency with qualitative diagnoses: our implementation automatically constructing a patient-specific prior according to a clinically predetermined medical history. The versatility of our approach extends beyond the specification of priors, but also their incorporation within bespoke PH-based loss functions and optimisation schemes. In this respect, interpretability is critical, allowing different components of the loss to be weighted or even disregarded as relevant to the clinical motivation (consider our approach to the high-dimensional topology of uncountably multiple, or "Swiss cheese" septal defects, for example).

Finally, we are keen to stress that topology is just one component of anatomical plausibility. For all the advantages presented above, our topological loss functions remain insensitive to unrealistic anatomical morphology and geometry. Therefore, the present approach should be seen as complementary to those based on an implicit, learned representation of anatomical segmentations.

7.6.2 Computational performance

Computation of the PH barcode is an intensive procedure and remains an active area of algorithmic development (Otter et al., 2017). In Section 7.4.5 we discuss the strengths and weaknesses of two such software libraries. Compared with our previous work (Byrne et al., 2021), adopting the CubicalRipser algorithm (Kaji et al., 2020) and integrating its functionality with PyTorch (Paszke et al., 2019) permitted practicable extension to the 3D setting. The performance of our implementation further benefits from the parallel computation of persistence, making trivial use of multiprocessing from the Python Standard Library.

The effect on execution time is stark. Using our previous implementation based on the TopologyLayer package (Gabrielsson et al., 2020), topological post-processing of a single, 2D short axis slice required over six minutes (Byrne et al.,

2021). Here, identical topological refinement of the same test cases had a mean execution time of only 7.12 s, an approximate fifty-fold acceleration.

In the 3D setting, such performance gains not only improve convenience, but arguably *enable* practicable scientific investigation. By our previous software implementation, computing the PH barcode for a single, random volume from the MM-WHS dataset took over an hour. When it is considered that our loss formulation necessitates the computation of at least fifteen such barcodes for all one hundred steps of iterative optimisation, such an approach is clearly impractical. In contrast, in this work, 3D topological post-processing (including all iterations) required a mean of 15.7 minutes.

Practical post-processing times for 3D volumetric data permit the specification of topological priors truly related to anatomy, rather than its appearance in 2D cross-section. This averts the technical challenge of anticipating the slice-wise topological changes associated with tomographic reconstruction and makes our formulation generalisable to many multi-class segmentation tasks.

7.6.3 Limitations and future work

We have demonstrated our approach within three CMR segmentation tasks, exhibiting varying degrees of success. Whilst the proposed topological loss functions reliably and predictably improve segmentation topology, they can also be associated with degradation in other metrics of performance when compared against U-Net prediction. Prior to topological fine tuning, CNN pre-training establishes an optimal set of parameters with respect to spatial overlap via CE loss. Hence, by introducing topological prior information, some compromise is perhaps to be expected. Our experiments demonstrate that the extent to which this trade off is felt is strongly dependent on the task considered. In 2D short axis segmentation, topological post-processing actually coincided with a minute increase in spatial overlap performance. Whereas in 3D, topological improvement was associated with a statistically and clinically significant drop in whole heart segmentation accuracy. Aside from the limitations of our approach, this result also indicates the inability of pixel-wise loss functions to promote topological feature learning.

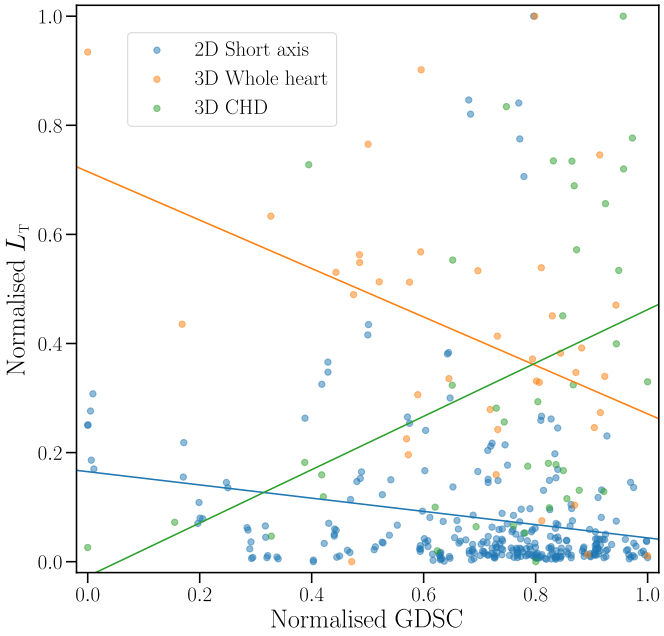


Figure 7.23: We speculate that the association between U-Net spatial overlap and topological performance informs the applicability of topological post-processing. Here we relate the value of the topological loss, prior to any post-processing, with U-Net GDSC. For ease of inspection, metrics are normalised across their respective test sets. We use linear regression to demonstrate the strength of association between variables in each case, relaxing normality requirements for the purpose of visualisation only.

As described in Section 7.5.4 and Section 7.5.5, we assert that this difference is strongly associated with pre-trained U-Net performance. A rich body of research indicates the successes of CNN-based short axis segmentation (Chen et al., 2020). Compared with the majority of 3D applications, this task benefits from relatively large amounts of training data, reduced structural variability and often a surplus of GPU memory resource in relation to adequate model capacity. This culminates in vastly improved CNN-based segmentation of the 2D short axis image compared with the 3D whole heart or CHD volumes.

Qualitatively, we suggest that the 2D task is more closely aligned with our contextual assumption for the application of topological post-processing: that pre-trained CNN segmentation closely approximates the ground truth, aside from a small number of topological errors of limited spatial extent. Quantitatively, we

speculate that the applicability of our approach is related to the degree to which pre-trained CNN topological error is explained by spatial overlap performance. Figure 7.23 illustrates this relationship per task considered.

In consideration of these results we stress a key distinction between both 2D short axis and 3D whole heart segmentation tasks, when compared with the labelling of CHD anatomy. That is: in the former tasks, anatomical topology takes a structural appearance that is consistent across both training and test sets, reflecting the normal heart. Consequently, and allowing for the limitations of pixel-wise optimisation, we might anticipate a crude association between improved spatial overlap and topological performance. This expectation is borne out in Figure 7.23. Comparing the strength of association between these two, Spearman's Rho suggests that in the 3D task, there is greater association between U-Net spatial overlap and topological performance (ρ = -0.364 versus -0.228, $p < 0.05$ in both cases). This indicates that the topological errors presented by U-Net are more likely to be associated with significant deficits in spatial overlap. Equivocation between true and spurious topological features results.

In contrast, in the task of CHD segmentation, we apply patient-specific topological priors per case. Not only does anatomical topology vary in a clinically meaningful sense, but so does the structural appearance of these topological features. For example, a VSD may be located at the atrioventricular inlet, at the apex, or at the ventriculoarterial outlet; can take different size; and occur in different number. The inability of pixel-wise optimisation to capture an abstract representation of this structurally and topologically heterogeneous distribution is demonstrated by Figure 7.23. This shows that for the CHD segmentation task there is a *negative* correlation between U-Net spatial overlap and topological performance. Presumably, this indicates that parameter optimisation according to CE loss culminates in a set of learned features that are insensitive to clinically meaningful anatomical topology. As a result, we assert that the application of our topological post-processing scheme is hampered by a highly disadvantageous parameter initialisation: one that is entirely ignorant of the topological features we aim to promote. Hence, and as observed in Section 7.5.5, any improvement in topological performance is associated with substantial degradation in other aspects of segmentation performance. It is possible that this obstacle might be alleviated

by introducing topological supervision within pre-training, perhaps in combination with those losses based on a latent representation of plausible anatomy. Validation of these assertions across a range of tasks will be necessary to fully understand the generalisability of our approach.

From a theoretical perspective, our formulation presents a framework that is applicable to any multi-class image segmentation task, admitting a variety future applications. More generally, we speculate that our topological loss functions may enhance a range of CNN training paradigms. Given our efficient implementation, our approach could easily be incorporated into conventional CNN optimisation as part of a multi-component loss. Furthermore, the abstract and explicit topological priors on which such losses are based could plausibly provide a supervisory signal for weakly or semi-supervised learning.

Finally, whilst we remain concerned by degradation in GDSC, we stress the importance of anatomically meaningful segmentation to an array of downstream tasks, including surgical planning (Valverde et al., 2017a). For these purposes, we think it important that diverse aspects of performance, including topology, are represented and considered alongside spatial overlap in future.

7.7 Conclusion

Building on previous works, we have extended PH-based loss functions to multi-class image segmentation. In the context of state of the art spatial overlap performance, our novel approach made statistically significant and predictable improvements in label map topology within 2D and 3D tasks (including within our bespoke ELCH dataset representative of CHD anatomy).

Our approach is theoretically founded, building a multi-class prior as the collection of both individual and combined label map topologies. Compared with the naive consideration of singular labels, the superiority of this scheme is borne out experimentally. Moreover, we considered different approaches to the construction of cubical complexes and demonstrated their consistency in relation to the proposed loss functions. Crucially, we adopted a highly efficient algorithmic backbone for the computation of PH, achieving dramatic improvements in execution time and permitting practicable extension to the 3D setting.

A careful analysis of both quantitative and qualitative performance allowed us to faithfully reflect the limitations of our approach and enables consideration of its wider application. Whilst impressive topologically, degradation in other aspects of performance currently limit the clinical application of such losses to the segmentation of CHD anatomy from 3D CMR. Despite these obstacles, our consideration of CHD anatomy through the lens of topology and PH shows great promise for future work, in particular the description of patient-specific disease by multi-class Betti number specification.

Whilst demonstrated in the field of CMR image analysis, our formulation is generalisable to any multi-class segmentation task: we envisage many applications across a diverse range of anatomical and pathological targets.

Chapter 8

Conclusions

8.1 Summary

This thesis set out to investigate the application of convolutional neural network (CNN)-based methodologies to the segmentation of 3D congenital heart disease (CHD) anatomy from cardiac magnetic resonance (CMR) data. In our consideration of this (and related) tasks, we firstly observed the paucity of relevant training data. The most closely aligned, pre-existing and publicly available dataset, that provided by the Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease (Pace et al., 2015) (HVSMR) Challenge, included just twenty examples. Whilst a larger collection of 110 congenital scans had been presented by Xu et al. (2019a), the ImageCHD dataset caters only to computed tomography (CT) acquisition. Hence, our first ambition set out to curate our own training dataset, one that might admit experimentation in support of our broader investigation.

In so doing, we achieved the Evelina London Children's Hospital (ELCH) dataset, a resource of 150 volumetric CMR acquisitions, each: acquired at isotropically high resolution; accompanied by associated 4D time-resolved magnetic resonance angiography (TR-MRA); and labelled according to a multi-class manual segmentation protocol including eighteen different anatomical structures. Supplementing our account of data curation, Chapter 5 advanced a comprehensive characterisation of our dataset, the patients contained (including clinically rele-

vant diagnoses and medical histories) and the imaging protocols used. Moreover, our analysis made stark the significant structural variation exhibited by our patient population, even outstripping that captured by the comparable ImageCHD dataset. Taken together, we believe that the ELCH dataset, its characterisation and its diverse distribution of CHD, represents much more than the first credible resource for CNN-based segmentation of volumetric CMR. More generally, we hope that it will prove an invaluable starting point for those seeking to understand, and improve the care of patients with, CHD in the future.

Critical to its value, our labelling scheme encoded haemodynamic continuity via pixel adjacency, allowing each segmentation to capture the presence of congenital heart defects in a clinically meaningful fashion. This is just one way that our unrelenting focus on the clinical requirements of our segmentation task (a consideration that we sometimes perceive to be lacking from, or only superficially addressed in the technical literature) has shaped the progression of our work. Concentrating on these clinical requirements has conferred great value to this project.

Our focus was brought to bear in Chapter 6, in which we leveraged the ELCH dataset to investigate CNN-based segmentation of patient-specific anatomy from 3D CMR images. Firstly, it prompted us to make use of all $\geq$ 3D data made available by clinically routine CMR protocols. This meant the inclusion of 4D TR-MRA and the exploration of different approaches to incorporate temporally resolved data, concatenated as additional channels of our input. Though these design choices failed to deliver on their hypothesised advantages (that cardiac segments in close spatial proximity, and perhaps isolated by an indistinct boundary, could be separated by their differential dynamic contrast enhancement), they did make marginal improvements in the segmentation of the extracardiac vasculature. Perhaps more importantly, our clinical focus motivated the design of novel performance metrics, sensitive to the accurate delineation of defects. Moving beyond the technical metrics considered in the majority of the literature, we were able to identify limitations in conventional CNN-based methods, as applied to our task. Namely, these centre on the anatomically spurious predictions that can result when CNNs are trained with pixel-wise loss functions, ignorant of extended spatial coherence. Problematically, errors of this sort also limit the clinically meaningful representation of heart defects.

Considering our previous metrics through the lens of topology, in Chapter 7 we extended existing persistent homology (PH)-based loss functions for binary segmentation, to the multi-class setting. These expose the differences between a predicted segmentation and its anticipated topology, as specified by a Betti number prior. Key to the success of our approach, coupled with our combinatorial multi-class framework, we presented an efficient formulation based on cubical complexes and parallel execution. This allowed the application of PH-based topological post-processing (in the guise of CNN parameter adaptation at test time) in 3D, for the first time. The fact that such priors can be expressed and optimised in 3D allows the specification of anatomical topology, rather than its appearance in 2D cross-section. This extension is key to the generalisability of our approach and allowed us to apply multi-class topological losses to the ELCH dataset. Demonstrating that topology can provide a compact description of CHD, we were able to construct and apply *patient-specific* topological priors. In all experiments, our multi-class PH-based losses made statistically significant improvements in topology. We anticipate many applications of this approach across a range of different medical image segmentation tasks.

8.2 Clinical impact

Having maintained a strong clinical focus throughout, including within its design, execution and consideration, we believe our work has the potential for significant impact in future. Given our motives, we hope this will be felt through the eventual consolidation of 3D, patient-specific anatomical modelling as a gold standard tool, routinely deployed in the management of all patients with CHD. Whilst we make no claims as to the clinical suitability of CNN-based solutions to this task at present, we hope that our work has fostered, and can inspire, progress in pursuit of this ambition. Were this objective to be fulfilled, many clinical applications - some established, others only presently idealised - might become possible.

Firstly, our motivations concern the extraction of such models for the purposes of multi-disciplinary communication. Where readily available, patient-specific models of disease promote a shared understanding of anatomy that is accessible to all team members. Within the CHD population, this level of 3D structural

appreciation serves to improve consensus decision-making, and, where appropriate, planning of surgical and cardiac catheter-based intervention. We anticipate that improved image segmentation methods, possibly building on our own work, will foster the growth of such techniques. In turn, this might extend the promise of personalised care to more patients treated in more centres, improving outcomes.

As argued in Chapter 2, greater patient throughput raises the prospect of effectively powered Health Technology Assessment. This might rigorously establish the clinical effectiveness of patient-specific anatomical modelling. Moreover, in the same chapter we reviewed a range of other applications of such models, including within: device development; patient consenting, communication and education; medical training; and academic research. Each of these applications stands to gain from further developments in CNN-based segmentation, particularly when conducted with the strength of clinical focus advanced by our work.

Secondly, and perhaps more speculatively, if the burden of segmentation is entirely eliminated, realising the automated preparation of patient-specific models, population-based analyses may become possible. The accepted wisdom suggests that no two cases of CHD are alike. However, were a representative and computationally tractable distribution of anatomical geometry made available, subtle trends and associations between currently disparate diagnostic groups may emerge. Other clinically relevant structural insights may be revealed. The academic investigation of such topics would further our collective 3D understanding of cardiac anatomy, including its modes of defective development. In this sense, automated segmentation may present the opportunity for a new field of study, one investigating cardiac morphology *in vivo*.

8.3 Directions for future work

Seeking to expedite the segmentation of patient-specific CHD anatomy from volumetric imaging data, we suggest several directions for future work. These are closely linked to the limitations of the three main contributions made by our work.

Firstly, and perhaps most importantly, we are sufficiently encouraged by our assessments of CNN-based segmentation to endorse continued investigation. However, we think it important to acknowledge the limited representational capacity

of the ELCH dataset (despite it being the largest of its kind). There are clear ways that this might be improved, including scan data from another centre or scanner, for example. Fundamentally, however, the CHD population remains so structurally diverse as to likely preclude its effective representation within a finite dataset. Hence, instead of promoting the accumulation of more labels from retrospective scan data, we argue for the incorporation of dedicated manual annotation (meeting the requirements of a protocol similar to that outlined in Chapter 5) within prospectively administered care. In such a scenario, the various motivations and ambitions of clinical practice, and its underpinning scientific methods act cooperatively. More specifically, prospective labelling provides a basis for both: (1) the patient-specific anatomical models that enhance the delivery of care; and for (2) the accumulation of training examples to support the development of data-driven methodologies (possibly including CNN-based segmentation of the same, prospectively acquired image data).

The limitations of the experimental work contained in Chapter 6 prompt an obvious technical extension for the treatment of 4D TR-MRA data within CNN-based segmentation. Given their sequential nature, salient features might be better extracted using a convolutional recurrent neural network. However, the broader thrust of this experiment suggests a more general question: why only incorporate the 3D structural and 4D dynamic acquisitions? A routine CMR protocol includes an array of different series, including black blood and cine stacks. These are acquired for their complementary semantics, which could, or perhaps should, also be leveraged within CNN-based segmentation. Going one step further, future work might look to incorporate scan data acquired by other modalities. In particular, neither CMR nor CT visualise the heart's valves with great acuity. In contrast, 3D echocardiography demonstrates these anatomical features faithfully. Key to certain congenital diagnoses (for example where the valve apparatus might straddle the ventricular septum), such data may provide another source of complementary, discriminative features in the future.

Lastly, the limitations of our multi-class topological loss functions suggest a number of technically focused avenues for future work. In Chapter 7, our experiments observed that while such losses reliably confer statistically significant improvements in segmentation topology, they can also be associated with degra-

dation in spatial overlap performance. We theorised that such deterioration might be associated with the difficulty of the task considered, speculating as to a dependence on the extent to which topological errors in U-Net prediction coincide with inaccuracies in spatial overlap. Establishing the veracity of this hypothesis is fertile ground for future work. More generally it seems at least plausible that the conflict between topological and spatial overlap performance might be resolved by incorporating PH-based losses within conventional pre-training, or fine tuning. Optimisation against an abstract topological prior also allows for the investigation of semi-supervised learning.

8.4 Closing statement

The marriage of medical imaging and 3D modelling technologies has shown great promise. Through their support of personalised care, the patient-specific models of anatomy which result have changed the lives of individual patients with CHD. Problematically , however, the burden of (manual) image segmentation, an unavoidable step in their construction, precludes the application of these exciting technologies to all but the most complex cases, treated at the most specialist hospitals.

In this thesis, we have investigated the possibility of automating this segmentation task, depending on CNN-based methods from the field of deep learning. By curating a unique training dataset of 150 patients and applying a clinically focused analysis, we have highlighted the limitations of existing approaches. In response, we have developed and leveraged a multi-class topological description of patient-specific anatomy, one that we have incorporated into a loss function, and that is sensitive to the presence of congenital heart defects. Whether through our unique training dataset, our keen clinical assessments or our highly generalisable topological loss functions, we anticipate many applications and extensions of our work. I am incredibly grateful to have had the opportunity to make these contributions, and hope that they drive developments in the personalised care of all members of the CHD population in the future.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M. et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Preprint* (arXiv:1603.04467).

Abdel-Sayed, P., Kalejs, M. and Von Segesser, L. K. (2009). A new training set-up for trans-apical aortic valve replacement. *Interactive cardiovascular and thoracic surgery* **8**(6): 599–601.

Abdelkarim, A., Hageman, A., Levi, D. S. and Aboulhosn, J. (2018). Operationalizing low-cost three-dimensional printing in planning for complex congenital cardiac interventions. *Materials Today Communications* **15**: 171–174.

Adams, R. and Bischof, L. (1994). Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence* **16**(6): 641–647.

Al Jabbari, O., Abu Saleh, W. K., Patel, A. P., Igo, S. R. and Reardon, M. J. (2016). Use of three-dimensional models to assist in the resection of malignant cardiac tumors. *Journal of cardiac surgery* **31**(9): 581–583.

Al-Sarraf, N., Thalib, L., Hughes, A., Houlihan, M., Tolan, M., Young, V. and McGovern, E. (2011). Cross-clamp time is an independent predictor of mortality and morbidity in low-and high-risk cardiac patients. *International Journal of Surgery* **9**(1): 104–109.

Albà, X., Lekadir, K., Hoogendoorn, C., Pereanez, M., Swift, A. J., Wild, J. M. and Frangi, A. F., Reusability of statistical shape models for the segmentation of severely abnormal hearts. In *Statistical Atlases and Computational Models of the Heart-Imaging and Modelling Challenges: 5th International Workshop, STACOM 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 18, 2014, Revised Selected Papers 5*. Springer, 2015, 257–264.

Andrushchuk, U., Adzintsou, V., Nevyglas, A. and Model, H. (2018). Virtual and real septal myectomy using 3-dimensional printed models. *Interactive CardioVascular and Thoracic Surgery* **26**(5): 881–882.

Anwar, S., Rockefeller, T., Raptis, D. A., Woodard, P. K. and Eghtesady, P. (2018a). 3D printing provides a precise approach in the treatment of tetralogy of Fallot, pulmonary atresia with major aortopulmonary collateral arteries. *Current treatment options in cardiovascular medicine* **20**: 1–9.

Anwar, S., Singh, G. K., Miller, J., Sharma, M., Manning, P., Billadello, J. J., Eghtesady, P. and Woodard, P. K. (2018b). 3D printing is a transformative technology in congenital heart disease. *JACC: Basic to Translational Science* **3**(2): 294–312.

Apitz, C., Webb, G. D. and Redington, A. N. (2009). Tetralogy of fallot. *The Lancet* **374**(9699): 1462–1471.

Armillotta, A., Bonhoeffer, P., Dubini, G., Ferragina, S., Migliavacca, F., Sala, G. and Schievano, S. (2007). Use of rapid prototyping models in the planning of percutaneous pulmonary valved stent implantation. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* **221**(4): 407–416.

Aroney, N., Markham, R., Putrino, A., Crowhurst, J., Wall, D., Scalia, G. and Walters, D. (2019). Three-dimensional printed cardiac fistulae: a case series. *European Heart Journal-Case Reports* **3**(2): ytz060.

Assaf, R., Goupil, A., Vrabie, V. and Kacim, M., Homology functionality for grayscale image segmentation. In *CRIMSTIC*, volume 8. 2016, 281–286.

Avendi, M. R., Kheradvar, A. and Jafarkhani, H. (2017). Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. *Magnetic resonance in medicine* **78**(6): 2439–2448.

Averkin, I. I., Grehov, E. V., Pervunina, T. M., Komlichenko, E. V., Vasichkina, E. S., Zaverza, V. M., Nikiforov, V. G., Latipova, M. L., Govorov, I. E., Kozyrev, I. A. et al. (2022). 3D-printing in preoperative planning in neonates with complex congenital heart defects. *The Journal of Maternal-Fetal & Neonatal Medicine* **35**(10): 2020–2024.

Aylward, S. R. and Bullitt, E. (2002). Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE transactions on medical imaging* **21**(2): 61–75.

Backhaus, S. J., Staab, W., Steinmetz, M., Ritter, C. O., Lotz, J., Hasenfuß, G., Schuster, A. and Kowallick, J. T. (2019). Fully automated quantification of biventricular volumes and function in cardiovascular magnetic resonance: applicability to clinical routine settings. *Journal of Cardiovascular Magnetic Resonance* **21**: 1–13.

Bagur, R., Cheung, A., Chu, M. W. and Kiaii, B. (2018). 3-dimensional–printed model for planning transcatheter mitral valve replacement. *JACC: Cardiovascular Interventions* **11**(8): 812–813.

Bai, W., Chen, C., Tarroni, G., Duan, J., Guitton, F., Petersen, S. E., Guo, Y., Matthews, P. M. and Rueckert, D., Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019, 541–549.

Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P. M. and Rueckert, D., Semi-supervised learning for network-based cardiac MR image segmentation. In *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20*. Springer International Publishing, 2017, 253–260.

Bai, W., Shi, W., de Marvao, A., Dawes, T. J., O'Regan, D. P., Cook, S. A. and Rueckert, D. (2015a). A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. *Medical image analysis* **26**(1): 133–145.

Bai, W., Shi, W., Ledig, C. and Rueckert, D. (2015b). Multi-atlas segmentation with augmented features for cardiac MR images. *Medical image analysis* **19**(1): 98–109.

Bai, W., Shi, W., Wang, H., Peters, N. S. and Rueckert, D. (2012). Multiatlas based segmentation with local label fusion for right ventricle MR images. *image* **6**: 9.

Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A. M., Aung, N., Lukaschuk, E., Sanghvi, M. M. et al. (2018). Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance* **20**(1): 1–12.

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J. and Dalca, A. V. (2019). Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging* **38**(8): 1788–1800.

Bateman, M. G., Durfee, W. K., Iles, T. L., Martin, C. M., Liao, K., Erdman, A. G. and Iaizzo, P. A. (2020). Cardiac patient–specific three-dimensional models as surgical planning tools. *Surgery* **167**(2): 259–263.

Batteux, C., Haidar, M. A. and Bonnet, D. (2019). 3D-printed models for surgical planning in complex congenital heart diseases: a systematic review. *Frontiers in pediatrics* **7**: 23.

Bazin, P.-L., Ellingsen, L. M. and Pham, D. L., Digital homeomorphisms in deformable registration. In *IPMI*, volume 4584. 2007, 211–222.

Bengio, Y., Courville, A. and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* **35**(8): 1798–1828.

Bengio, Y., Louradour, J., Collobert, R. and Weston, J., Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*. 2009, 41–48.

Bengs, M., Gessert, N. and Schlaefer, A. (2020). 4d spatio-temporal deep learning with 4d fmri data for autism spectrum disorder classification. *arXiv Preprint* (arXiv:2004.10165).

BenTaieb, A. and Hamarneh, G., Topology aware fully convolutional networks for histology gland segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer International Publishing, 2016, 460–468.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Ballester, M. A. G. et al. (2018). Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging* **37**(11): 2514–2525.

Bertolini, M., Rossoni, M. and Colombo, G. (2021). Operative Workflow from CT to 3D Printing of the Heart: Opportunities and Challenges. *Bioengineering* **8**(10): 130.

Bezdek, J. C., Ehrlich, R. and Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & geosciences* **10**(2-3): 191–203.

Bhatla, P., Mosca, R. S. and Tretter, J. T. (2017a). Altering management decisions with gained anatomical insight from a 3D printed model of a complex ventricular septal defect. *Cardiology in the Young* **27**(2): 377–380.

Bhatla, P., Tretter, J. T., Ludomirsky, A., Argilla, M., Latson, L. A., Chakravarti, S., Barker, P. C., Yoo, S.-J., McElhinney, D. B., Wake, N. et al. (2017b). Utility and scope of rapid prototyping in patients with complex muscular ventricular septal defects or double-outlet right ventricle: does it alter management decisions? *Pediatric cardiology* **38**: 103–114.

Biglino, G., Capelli, C., Binazzi, A., Reggiani, R., Cosentino, D., Migliavacca, F., Bonhoeffer, P., Taylor, A. M. and Schievano, S. (2012a). Virtual and real bench testing of a new percutaneous valve device: a case study. *Eurointervention: Journal of Europcr in Collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology* **8**(1): 120–128.

Biglino, G., Capelli, C., Bruse, J., Bosi, G. M., Taylor, A. M. and Schievano, S. (2017a). Computational modelling for congenital heart disease: how far are we from clinical translation? *Heart* **103**(2): 98–103.

Biglino, G., Capelli, C., Koniordou, D., Robertshaw, D., Leaver, L.-K., Schievano, S., Taylor, A. M. and Wray, J. (2017b). Use of 3D models of congenital heart disease as an education tool for cardiac nurses. *Congenital heart disease* **12**(1): 113–118.

Biglino, G., Capelli, C., Wray, J., Schievano, S., Leaver, L.-K., Khambadkone, S., Giardini, A., Derrick, G., Jones, A. and Taylor, A. M. (2015). 3D-manufactured patient-specific models of congenital heart defects for communication in clinical practice: feasibility and acceptability. *BMJ open* **5**(4): e007165.

Biglino, G., Corsini, C., Schievano, S., Dubini, G., Giardini, A., Hsia, T.-Y., Pennati, G., Taylor, A. M. and Group, M. C. (2014). Computational models of aortic coarctation in hypoplastic left heart syndrome: considerations on validation of a detailed 3D model. *The International journal of artificial organs* **37**(5): 371–381.

Biglino, G., Giardini, A., Baker, C., Figliola, R. S., Hsia, T.-Y., Taylor, A. M., Schievano, S., Group, M. C. et al. (2012b). In vitro study of the Norwood palliation: a patient-specific mock circulatory system. *ASAIO journal* **58**(1): 25–31.

Biglino, G., Giardini, A., Baker, C., Figliola, R. S., Hsia, T.-Y., Taylor, A. M., Schievano, S., Group, M. C. et al. (2013). Implementing the Sano modification in an experimental model of first-stage palliation of hypoplastic left heart syndrome. *ASAIO Journal* **59**(1): 86–89.

Biglino, G., Koniordou, D., Gasparini, M., Capelli, C., Leaver, L.-K., Khambadkone, S., Schievano, S., Taylor, A. M. and Wray, J. (2017c). Piloting the use of patient-specific cardiac models as a novel tool to facilitate communication during cinical consultations. *Pediatric Cardiology* **38**: 813–818.

Biglino, G., Layton, S., Lee, M., Sophocleous, F., Hall, S. and Wray, J. (2019). 'Making the Invisible Visible': an audience response to an art installation representing the complexity of congenital heart disease and heart transplantation. *Medical Humanities* **45**(4): 399–405.

Biglino, G. and Milano, E. G. (2018). Applications of 3D printing in paediatric cardiology: its potential and the need for gathering evidence. *Translational Pediatrics* **7**(3): 219.

Birbara, N. S., Otton, J. M. and Pather, N. (2019). 3D modelling and printing technology to produce patient-specific 3D models. *Heart, Lung and Circulation* **28**(2): 302–313.

Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural computation* **7**(1): 108–116.

Blacksun Software (2021). Mousotron : Mouse and keyboard activity monitor (Version 12.1) [Computer Program]. Available from: http://www.blacksunsoftware.com/mousotron.html. Downloaded on 28 April 2021.

Borra, D., Masci, A., Esposito, L., Andalò, A., Fabbri, C. and Corsi, C., A semantic-wise convolutional neural network approach for 3-D left atrium segmentation from late gadolinium enhanced magnetic resonance imaging. In *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*. Springer International Publishing, 2019, 329–338.

Borracci, R. A., Ferreira, L. M., Alvarez Gallesio, J. M., Tenorio Nunez, O. M., David, M. and Eyheremendy, E. P. (2021). Three-dimensional virtual and printed models for planning adult cardiovascular surgery. *Acta Cardiologica* **76**(5): 534–543.

Bouthillier, X., Konda, K., Vincent, P. and Memisevic, R. (2015). Dropout as data augmentation. *arXiv Preprint* (arXiv:1506.08700).

Boykov, Y. and Funka-Lea, G. (2006). Graph cuts and efficient nd image segmentation. *International journal of computer vision* **70**(2).

Boykov, Y. Y. and Jolly, M.-P., Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, volume 1. IEEE, 2001, 105–112.

Bramlet, M., Olivieri, L., Farooqi, K., Ripley, B. and Coakley, M. (2017). Impact of three-dimensional printing on the study and treatment of congenital heart disease. *Circulation research* **120**(6): 904–907.

Breiman, L. (2001). Random forests. *Machine learning* **45**: 5–32.

Briggman, K., Denk, W., Seung, S., Helmstaedter, M. and Turaga, S. C., Maximin affinity learning of image segmentation. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C. and Culotta, A. (Editors), *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc., 2009.

Brun, H., Bugge, R. A. B., Suther, L., Birkeland, S., Kumar, R., Pelanis, E. and Elle, O. J. (2019). Mixed reality holograms for heart surgery planning: first user experience in congenital heart disease. *European Heart Journal-Cardiovascular Imaging* **20**(8): 883–888.

Bruns, S., Wolterink, J. M., Takx, R. A., van Hamersvelt, R. W., Suchá, D., Viergever, M. A., Leiner, T. and Išgum, I. (2020). Deep learning from dual-energy information for whole-heart segmentation in dual-energy and single-energy non-contrast-enhanced cardiac CT. *Medical physics* **47**(10): 5048–5060.

Budd, S., Robinson, E. C. and Kainz, B. (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis* **71**: 102062.

Bui, V., Hsu, L.-Y., Chang, L.-C. and Chen, M. Y., An automatic random walk based method for 3D segmentation of the heart in cardiac computed tomography images. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, 1352–1355.

Bui, V., Hsu, L.-Y., Shanbhag, S. M., Tran, L., Bandettini, W. P., Chang, L.-C. and Chen, M. Y. (2020a). Improving multi-atlas cardiac structure segmentation of computed tomography angiography: A performance evaluation based on a heterogeneous dataset. *Computers in biology and medicine* **125**: 104019.

Bui, V., Shanbhag, S. M., Levine, O., Jacobs, M., Bandettini, W. P., Chang, L.-C., Chen, M. Y. and Hsu, L.-Y. (2020b). Simultaneous multi-structure segmentation of the heart and peripheral tissues in contrast enhanced cardiac computed tomography angiography. *IEEE Access* **8**: 16187–16202.

Byl, J. L., Sholler, R., Gosnell, J. M., Samuel, B. P. and Vettukattil, J. J. (2020). Moving beyond two-dimensional screens to interactive three-dimensional visualization in congenital heart disease. *The international journal of cardiovascular imaging* **36**(8): 1567–1573.

Byrne, N., Clough, J. R., Montana, G. and King, A. P., A persistent homology-based topological loss function for multi-class CNN segmentation of cardiac MRI. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges: 11th International Workshop, STACOM 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers 11*. Springer, 2021, 3–13.

Byrne, N., Clough, J. R., Valverde, I., Montana, G. and King, A. P., Topology-preserving augmentation for CNN-based segmentation of congenital heart defects from 3D paediatric CMR. In *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis: First International Workshop, SUSI 2019, and 4th International Workshop, PIPPI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings 4*. Springer International Publishing, 2019, 181–188.

Byrne, N., Velasco Forte, M., Tandon, A., Valverde, I. and Hussain, T. (2016). A systematic review of image segmentation methodology, used in the additive manufacture of patient-specific 3D printed models of the cardiovascular system. *JRSM cardiovascular disease* **5**: 2048004016645467.

Cai, K., Yang, R., Chen, H., Li, L., Zhou, J., Ou, S. and Liu, F. (2017a). A framework combining window width-level adjustment and Gaussian filter-based multi-resolution for automatic whole heart segmentation. *Neurocomputing* **220**: 138–150.

Cai, K., Yang, R., Yue, H., Li, L., Ou, S. and Liu, F. (2017b). Dynamic updating atlas for heart segmentation with a nonlinear field-based model. *The International Journal of Medical Robotics and Computer Assisted Surgery* **13**(3): e1785.

Campello, V. M., Gkontra, P., Izquierdo, C., Martin-Isla, C., Sojoudi, A., Full, P. M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J. et al. (2021). Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. *IEEE Transactions on Medical Imaging* **40**(12): 3543–3554.

Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**(6): 679–698.

Canstein, C., Cachot, P., Faust, A., Stalder, A., Bock, J., Frydrychowicz, A., Küffer, J., Hennig, J. and Markl, M. (2008). 3D MR flow analysis in realistic rapid-prototyping model systems of the thoracic aorta: comparison with in vivo data and computational fluid dynamics in identical vessel geometries. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **59**(3): 535–546.

Capelli, C., Bosi, G. M., Cerri, E., Nordmeyer, J., Odenwald, T., Bonhoeffer, P., Migliavacca, F., Taylor, A. M. and Schievano, S. (2012). Patient-specific simulations of transcatheter aortic valve stent implantation. *Medical & biological engineering & computing* **50**: 183–192.

Capelli, C., Taylor, A. M., Migliavacca, F., Bonhoeffer, P. and Schievano, S. (2010). Patient-specific reconstructed anatomies and computer simulations are fundamental for selecting medical device treatment: application to a new percutaneous pulmonary valve. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **368**(1921): 3027–3038.

Carberry, T., Murthy, R., Hsiao, A., Petko, C., Moore, J., Lamberti, J. and Hegde, S. (2019). Fontan revision: Presurgical planning using four-dimensional (4D) flow and three-dimensional (3D) printing. *World Journal for Pediatric and Congenital Heart Surgery* **10**(2): 245–249.

Cardoso, M. J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N. C., Ourselin, S., Initiative, A. D. N. et al. (2013). STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. *Medical image analysis* **17**(6): 671–684.

Caselles, V., Kimmel, R. and Sapiro, G. (1997). Geodesic active contours. *International journal of computer vision* **22**(1): 61.

Chaitanya, K., Karani, N., Baumgartner, C. F., Becker, A., Donati, O. and Konukoglu, E., Semi-supervised and task-driven data augmentation. In *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26*. Springer, 2019, 29–41.

Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on image processing* **10**(2): 266–277.

Chaowu, Y., Hua, L. and Xin, S. (2016). Three-dimensional printing as an aid in transcatheter closure of secundum atrial septal defect with rim deficiency: in vitro trial occlusion based on a personalized heart model. *Circulation* **133**(17): e608–e610.

Chen, C., Biffi, C., Tarroni, G., Petersen, S., Bai, W. and Rueckert, D., Learning shape priors for robust cardiac MR segmentation from multi-view images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer International Publishing, 2019, 523–531.

Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W. and Rueckert, D. (2020). Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine* **7**: 25.

Chen, J., Yang, L., Zhang, Y., Alber, M. and Chen, D. Z., Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I. and Garnett, R. (Editors), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Chen, L., Wu, Y., DSouza, A. M., Abidin, A. Z., Wismüller, A. and Xu, C., MRI tumor segmentation with densely connected 3D CNN. In *Medical Imaging 2018: Image Processing*, volume 10574. SPIE, 2018a, 357–364.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* **40**(4): 834–848.

Chen, S. A., Ong, C. S., Malguria, N., Vricella, L. A., Garcia, J. R. and Hibino, N. (2018b). Digital design and 3D printing of aortic arch reconstruction in HLHS for surgical simulation and training. *World Journal for Pediatric and Congenital Heart Surgery* **9**(4): 454–458.

Choy, C., Gwak, J. and Savarese, S., 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, 3075–3084.

Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D'Anastasi, M. et al., Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, 415–423.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. and Ronneberger, O., 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer International Publishing, 2016, 424–432.

Ciresan, D., Giusti, A., Gambardella, L. and Schmidhuber, J., Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. In Pereira, F., Burges, C., Bottou, L. and Weinberger, K. (Editors), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

Clark, D. and Badea, C., Convolutional regularization methods for 4D, x-ray CT reconstruction. In *Medical imaging 2019: physics of medical imaging*, volume 10948. SPIE, 2019, 574–585.

Clough, J. R., Byrne, N., Oksuz, I., Zimmer, V. A., Schnabel, J. A. and King, A. P. (2020). A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(12): 8766–8778.

Clough, J. R., Oksuz, I., Byrne, N., Schnabel, J. A. and King, A. P., Explicit topological priors for deep-learning based image segmentation using persistent homology. In *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26*. Springer International Publishing, 2019, 16–28.

Cocosco, C. A., Niessen, W. J., Netsch, T., Vonken, E.-j. P., Lund, G., Stork, A. and Viergever, M. A. (2008). Automatic image-driven segmentation of the ventricles in cardiac cine MRI. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **28**(2): 366–374.

Collins, D. L. and Evans, A. C. (1997). Animal: validation and applications of nonlinear registration-based segmentation. *International journal of pattern recognition and artificial intelligence* **11**(08): 1271–1294.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J. and Evans, A. C. (1998). Design and construction of a realistic digital brain phantom. *IEEE transactions on medical imaging* **17**(3): 463–468.

Conti, M., Marconi, S., Muscogiuri, G., Guglielmo, M., Baggiano, A., Italiano, G., Mancini, M. E., Auricchio, F., Andreini, D., Rabbat, M. G. et al. (2019). Left atrial appendage closure guided by 3D computed tomography printing technology: A case control study. *Journal of Cardiovascular Computed Tomography* **13**(6): 336–339.

Cootes, T. F., Edwards, G. J. and Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence* **23**(6): 681–685.

Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding* **61**(1): 38–59.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning* **20**: 273–297.

Costello, J. P., Olivieri, L. J., Krieger, A., Thabit, O., Marshall, M. B., Yoo, S.-J., Kim, P. C., Jonas, R. A. and Nath, D. S. (2014). Utilizing three-dimensional printing technology to assess the feasibility of high-fidelity synthetic ventricular septal defect models for simulation in medical education. *World Journal for Pediatric and Congenital Heart Surgery* **5**(3): 421–426.

Costello, J. P., Olivieri, L. J., Su, L., Krieger, A., Alfares, F., Thabit, O., Marshall, M. B., Yoo, S.-J., Kim, P. C., Jonas, R. A. et al. (2015). Incorporating three-dimensional printing into a simulation-based congenital heart disease and critical care training curriculum for resident physicians. *Congenital heart disease* **10**(2): 185–190.

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory* **13**(1): 21–27.

Crum, W. R., Camara, O. and Hill, D. L. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE transactions on medical imaging* **25**(11): 1451–1461.

Daemen, J. H., Heuts, S., Olsthoorn, J. R., Maessen, J. G. and Sardari Nia, P. (2019). Mitral valve modelling and three-dimensional printing for planning and simulation of mitral valve repair. *European Journal of Cardio-Thoracic Surgery* **55**(3): 543–551.

Dankowski, R., Baszko, A., Sutherland, M., Firek, L., Kałmucki, P., Wróblewska, K., Szyszka, A., Groothuis, A. and Siminiak, T. (2014). 3D heart model printing for preparation of percutaneous structural interventions: description of the technology and case report. *Kardiologia Polska (Polish Heart Journal)* **72**(6): 546–551.

Deakyne, A. J., Iles, T. L., Mattson, A. R. and Iaizzo, P. A. (2019). Virtual Prototyping: Computational Device Placements within Detailed Human Heart Models. *Applied Sciences* **10**(1): 175.

Degel, M. A., Navab, N. and Albarqouni, S., Domain and geometry agnostic CNNs for left atrium segmentation in 3D ultrasound. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11*. Springer International Publishing, 2018, 630–637.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)* **39**(1): 1–22.

Deng, S., Wheeler, G., Toussaint, N., Munroe, L., Bhattacharya, S., Sajith, G., Lin, E., Singh, E., Chu, K. Y. K., Kabir, S. et al. (2021). A virtual reality system for improved image-based planning of complex cardiac procedures. *Journal of imaging* **7**(8): 151.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* **26**(3): 297–302.

Ding, W., Li, L., Zhuang, X. and Huang, L., Cross-modality multi-atlas segmentation using deep neural networks. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III*. Springer International Publishing Cham, 2020, 233–242.

Ding, X., Guo, Y., Ding, G. and Han, J., Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, 1911–1920.

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K. and Darrell, T., Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, 2625–2634.

Dong, S., Luo, G., Wang, K., Cao, S., Li, Q. and Zhang, H. (2018). A combined fully convolutional networks and deformable model for automatic left ventricle segmentation based on 3D echocardiography. *BioMed research international* **2018**.

Dong, Z., Men, J., Yang, Z., Jerwick, J., Li, A., Tanzi, R. E. and Zhou, C. (2020). FlyNet 2.0: drosophila heart 3D (2D+ time) segmentation in optical coherence microscopy images using a convolutional long short-term memory neural network. *Biomedical Optics Express* **11**(3): 1568–1579.

Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J. and Heng, P.-A. (2017). 3D deeply supervised network for automated segmentation of volumetric medical images. *Medical image analysis* **41**: 40–54.

Du, X., Song, Y., Liu, Y., Zhang, Y., Liu, H., Chen, B. and Li, S. (2020). An integrated deep learning framework for joint segmentation of blood pool and myocardium. *Medical image analysis* **62**: 101685.

Duan, J., Bello, G., Schlemper, J., Bai, W., Dawes, T. J., Biffi, C., de Marvao, A., Doumoud, G., O'Regan, D. P. and Rueckert, D. (2019). Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. *IEEE transactions on medical imaging* **38**(9): 2151–2164.

Duda, R. O. and Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM* **15**(1): 11–15.

Dumoulin, V. and Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv Preprint* (arXiv:1603.07285).

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics* **3**(3): 32–57.

Dydynski, P. B., Kiper, C., Kozik, D., Keller, B. B., Austin, E. and Holland, B. (2016). Three-dimensional reconstruction of intracardiac anatomy using CTA and surgical planning for double outlet right ventricle: early experience at a tertiary care congenital heart center. *World Journal for Pediatric and Congenital Heart Surgery* **7**(4): 467–474.

Ecabert, O., Peters, J., Lorenz, C., von Berg, J., Vembar, M., Subramanyan, K., Lavi, G. and Weese, J., Towards automatic full heart segmentation in computed-tomography images. In *Computers in Cardiology, 2005*. IEEE, 2005, 223–226.

Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M. J., Vembar, M., Olszewski, M. E., Subramanyan, K., Lavi, G. et al. (2008). Automatic model-based segmentation of the heart in CT images. *IEEE transactions on medical imaging* **27**(9): 1189–1201.

Ecabert, O., Peters, J., Walker, M. J., Ivanc, T., Lorenz, C., von Berg, J., Lessick, J., Vembar, M. and Weese, J. (2011). Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. *Medical image analysis* **15**(6): 863–876.

Edelsbrunner, H., Harer, J. et al. (2008). Persistent homology-a survey. *Contemporary mathematics* **453**(26): 257–282.

Ehret, N., Alkassar, M., Dittrich, S., Cesnjevar, R., Rüffer, A., Uder, M., Rompel, O., Hammon, M. and Glöckler, M. (2018). A new approach of three-dimensional guidance in paediatric cath lab: segmented and tessellated heart models for cardiovascular interventions in CHD. *Cardiology in the Young* **28**(5): 661–667.

El Jurdi, R., Petitjean, C., Honeine, P., Cheplygina, V. and Abdallah, F. (2021). High-level prior-based loss functions for medical image segmentation: A survey. *Computer Vision and Image Understanding* **210**: 103248.

El Sabbagh, A., Eleid, M. F., Al-Hijji, M., Anavekar, N. S., Holmes, D. R., Nkomo, V. T., Oderich, G. S., Cassivi, S. D., Said, S. M., Rihal, C. S. et al. (2018). The various applications of 3D printing in cardiovascular diseases. *Current cardiology reports* **20**: 1–9.

Faletti, R., Gatti, M., Cosentino, A., Bergamasco, L., Stura, E. C., Garabello, D., Pennisi, G., Salizzoni, S., Veglia, S., Ottavio, D. et al. (2018). 3D printing of the aortic annulus based on cardiovascular computed tomography: Preliminary experience in pre-procedural planning for aortic valve sizing. *Journal of cardiovascular computed tomography* **12**(5): 391–397.

Falk, K. L., Zhou, H., Trampe, B., Heiser, T., Srinivasan, S., Iruretagoyena, J. I. and Roldán-Alzate, A. (2018). Modeling fetal cardiac anomalies from prenatal echocardiography with 3-dimensional printing and 4-dimensional flow magnetic resonance imaging. *Circulation: Cardiovascular Imaging* **11**(9): e007705.

Fan, Y., Kwok, K.-W., Zhang, Y., Cheung, G. S.-H., Chan, A. K.-Y. and Lee, A. P.-W. (2016). Three-dimensional printing for planning occlusion procedure for a double-lobed left atrial appendage. *Circulation: Cardiovascular Interventions* **9**(3): e003561.

Fan, Y., Wong, R. H. and Lee, A. P.-W. (2019). Three-dimensional printing in structural heart disease and intervention. *Annals of translational medicine* **7**(20).

Farooqi, K. M., Lengua, C. G., Weinberg, A. D., Nielsen, J. C. and Sanz, J. (2016). Blood pool segmentation results in superior virtual cardiac models than myocardial segmentation for 3D printing. *Pediatric cardiology* **37**: 1028–1036.

Farooqi, K. M. and Mahmood, F. (2018). Innovations in preoperative planning: insights into another dimension using 3D printing for cardiac disease. *Journal of Cardiothoracic and Vascular Anesthesia* **32**(4): 1937–1945.

Farotto, D. and Maes, J. (2019). Mimics ct heart tool for chamber segmentation: quantitative validation. *Technical report*, Materialise NV.

Fehling, M. K., Grosch, F., Schuster, M. E., Schick, B. and Lohscheller, J. (2020). Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. *Plos one* **15**(2): e0227791.

Fernandez-Alvarez, J.-A., Infante-Cossio, P., Barrera-Pulido, F., Gacto-Sanchez, P., Suarez-Mejias, C., Gomez-Ciriza, G., Sicilia-Castro, D. and Gomez-Cia, T. (2014). Virtual reality AYRA software for preoperative planning in facial allotransplantation. *Journal of Craniofacial Surgery* **25**(5): 1805–1809.

Ferrari, E., Gallo, M., Wang, C., Zhang, L., Taramasso, M., Maisano, F., Pirelli, L., Berdajs, D. and von Segesser, L. K. (2020). Three-dimensional printing in adult cardiovascular medicine for surgical and transcatheter procedural planning, teaching and technological innovation. *Interactive cardiovascular and thoracic surgery* **30**(2): 203–214.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S. et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**(3): 341–355.

Fix, E. (1985). *Discriminatory analysis: nonparametric discrimination, consistency properties*, volume 1. USAF school of Aviation Medicine.

Foley, T. A., El Sabbagh, A., Anavekar, N. S., Williamson, E. E. and Matsumoto, J. M. (2017). 3D-printing: applications in cardiovascular imaging. *Current Radiology Reports* **5**: 1–13.

Forte, M. N. V., Hussain, T., Roest, A., Gomez, G., Jongbloed, M., Simpson, J., Pushparajah, K., Byrne, N. and Valverde, I. (2019). Living the heart in three dimensions: applications of 3D printing in CHD. *Cardiology in the Young* **29**(6): 733–743.

Forte, V., Byrne, N., Perez, V., Bell, A., Gómez-Ciriza, G., Krasemann, T., Sievert, H., Simpson, J., Pushparajah, K., Razavi, R. et al. (2017). 3D printed models in patients with coronary artery fistulae: anatomical assessment and interventional planning. *Eurointervention: Journal of Europcr in Collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology* **13**(9): e1080–e1083.

Frangi, A. F., Niessen, W. J., Hoogeveen, R. M., Van Walsum, T. and Viergever, M. A. (1999). Model-based quantitation of 3-D magnetic resonance angiographic images. *IEEE Transactions on medical imaging* **18**(10): 946–956.

Frangi, A. F., Rueckert, D., Schnabel, J. A. and Niessen, W. J. (2002). Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE transactions on medical imaging* **21**(9): 1151–1166.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R. and Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American journal of epidemiology* **186**(9): 1026–1034.

Fujita, T., Saito, N., Minakata, K., Imai, M., Yamazaki, K. and Kimura, T. (2017). Transfemoral transcatheter aortic valve implantation in the presence of a mechanical mitral valve prosthesis using a dedicated TAVI guidewire: utility of a patient-specific three-dimensional heart model. *Cardiovascular intervention and therapeutics* **32**: 308–311.

Fukushima, K. and Miyake, S., Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets: Proceedings of the US-Japan Joint Seminar held at Kyoto, Japan February 15–19, 1982*. Springer, 1982, 267–285.

Funka-Lea, G., Boykov, Y., Florin, C., Jolly, M.-P., Moreau-Gobard, R., Ramaraj, R. and Rinck, D., Automatic heart isolation for CT coronary visualization using graph-cuts. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.* IEEE, 2006, 614–617.

Gabrielsson, R. B., Nelson, B. J., Dwaraknath, A. and Skraba, P., A Topology Layer for Machine Learning. In Chiappa, S. and Calandra, R. (Editors), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*. PMLR, 2020, 1553–1563.

Galisot, G., Brouard, T. and Ramel, J.-Y., Local probabilistic atlases and a posteriori correction for the segmentation of heart images. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, 207–214.

Ganaye, P.-A., Sdika, M. and Benoit-Cattin, H., Semi-supervised learning for segmentation under semantic constraint. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11*. Springer International Publishing, 2018, 595–602.

Gao, M., Chen, C., Zhang, S., Qian, Z., Metaxas, D. and Axel, L., Segmenting the papillary muscles and the trabeculae from high resolution cardiac CT through restoration of topological handles. In *Information Processing in Medical Imaging: 23rd International Conference, IPMI 2013, Asilomar, CA, USA, June 28–July 3, 2013. Proceedings 23*. Springer Berlin Heidelberg, 2013, 184–195.

Gao, M., Huang, J., Zhang, S., Qian, Z., Voros, S., Metaxas, D. and Axel, L., 4D cardiac reconstruction using high resolution CT images. In *Functional Imaging and Modeling of the Heart: 6th International Conference, FIMH 2011, New York City, NY, USA, May 25-27, 2011. Proceedings 6*. Springer, 2011, 153–160.

Gao, Y., Phillips, J. M., Zheng, Y., Min, R., Fletcher, P. T. and Gerig, G., Fully convolutional structured LSTM networks for joint 4D medical image segmentation. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, 1104–1108.

Garcia, J., Yang, Z., Mongrain, R., Leask, R. L. and Lachapelle, K. (2018). 3D printing materials and their use in medical education: a review of current technology and trends for the future. *BMJ simulation & technology enhanced learning* **4**(1): 27.

Gardin, C., Ferroni, L., Latremouille, C., Chachques, J. C., Mitrečić, D. and Zavan, B. (2020). Recent applications of three dimensional printing in cardiovascular medicine. *Cells* **9**(3): 742.

Garekar, S., Bharati, A., Chokhandre, M., Mali, S., Trivedi, B., Changela, V. P., Solanki, N., Gaikwad, S. and Agarwal, V. (2016). Clinical application and multidisciplinary assessment of three dimensional printing in double outlet right ventricle with remote ventricular septal defect. *World Journal for Pediatric and Congenital Heart Surgery* **7**(3): 344–350.

Garg, R. and Zahn, E. M. (2020). Utility of three-dimensional (3D) modeling for planning structural heart interventions (with an emphasis on valvular heart disease). *Current Cardiology Reports* **22**: 1–11.

Garin, A., Heiss, T., Maggs, K., Bleile, B. and Robins, V. (2020). Duality in persistent homology of images. *arXiv Preprint* (arXiv:2005.04597).

Gessert, N., Bengs, M., Schlüter, M. and Schlaefer, A. (2020). Deep learning with 4D spatio-temporal data representations for OCT-based force estimation. *Medical image analysis* **64**: 101730.

Gharleghi, R., Dessalles, C. A., Lal, R., McCraith, S., Sarathy, K., Jepson, N., Otton, J., Barakat, A. I. and Beier, S. (2021). 3D printing for cardiovascular applications: from end-to-end processes to emerging developments. *Annals of Biomedical Engineering* **49**(7): 1598–1618.

Ghisiawan, N., Herbert, C. E., Zussman, M., Verigan, A. and Stapleton, G. E. (2016). The use of a three-dimensional print model of an aortic arch to plan a complex percutaneous intervention in a patient with coarctation of the aorta. *Cardiology in the Young* **26**(8): 1568–1572.

Ghosh, T. K., Hasan, M. K., Roy, S., Alam, M. A., Hossain, E. and Ahmad, M. (2021). Multi-class probabilistic atlas-based whole heart segmentation method in cardiac CT and MRI. *IEEE Access* **9**: 66948–66964.

Giannakidis, A., Kamnitsas, K., Spadotto, V., Keegan, J., Smith, G., Glocker, B., Rueckert, D., Ernst, S., Gatzoulis, M. A., Pennell, D. J. et al., Fast fully automatic segmentation of the severely abnormal human right ventricle from cardiovascular magnetic resonance images using a multi-scale 3D convolutional neural network. In *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2016, 42–46.

Giannopoulos, A. A., Mitsouras, D., Yoo, S.-J., Liu, P. P., Chatzizisis, Y. S. and Rybicki, F. J. (2016). Applications of 3D printing in cardiovascular diseases. *Nature Reviews Cardiology* **13**(12): 701–718.

Gilliland, P. R., Uus, A., van Poppel, M. P., Grigorescu, I., Steinweg, J. K., Lloyd, D. F., Pushparajah, K., King, A. P. and Deprez, M. (2022). Automated atlas-based multi-label fetal cardiac vessel segmentation in Congenital Heart Disease. *bioRxiv* : 2022–01.

Girshick, R., Donahue, J., Darrell, T. and Malik, J., Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, 580–587.

Gomez, A., Gomez, G., Simpson, J. and Valverde, I. (2020). 3D hybrid printed models in complex congenital heart disease: 3D echocardiography and cardiovascular magnetic resonance imaging fusion. *European Heart Journal* **41**(43): 4214–4214.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep learning*. MIT press.

Gosnell, J., Pietila, T., Samuel, B. P., Kurup, H. K., Haw, M. P. and Vettukattil, J. J. (2016). Integration of computed tomography and three-dimensional echocardiography for hybrid three-dimensional printing in congenital heart disease. *Journal of digital imaging* **29**: 665–669.

Grady, L. (2006). Random walks for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **28**(11): 1768–1783.

Grant, E. K. and Olivieri, L. J. (2017). The role of 3-D heart models in planning and executing interventional procedures. *Canadian Journal of Cardiology* **33**(9): 1074–1081.

Greil, G. F., Kuettner, A., Flohr, T., Grasruck, M., Sieverding, L., Meinzer, H.-P. and Wolf, I. (2007). High-resolution reconstruction of a waxed heart specimen with flat panel volume computed tomography and rapid prototyping. *Journal of computer assisted tomography* **31**(3): 444–448.

Grosgeorge, D., Petitjean, C., Caudron, J., Fares, J. and Dacher, J.-N. (2011). Automatic cardiac ventricle segmentation in MR images: a validation study. *International journal of computer assisted radiology and surgery* **6**: 573–581.

Grosgeorge, D., Petitjean, C., Dacher, J.-N. and Ruan, S. (2013). Graph cut segmentation with a statistical shape model in cardiac MRI. *Computer Vision and Image Understanding* **117**(9): 1027–1035.

Ha, H., Kim, G. B., Kweon, J., Lee, S. J., Kim, Y.-H., Kim, N. and Yang, D. H. (2016). The influence of the aortic valve angle on the hemodynamic features of the thoracic aorta. *Scientific reports* **6**(1): 1–14.

Haak, A., Vegas-Sanchez-Ferrero, G., Mulder, H. H., Kirisli, H. A., Baka, N., Metz, C., Klein, S., Ren, B., van Burken, G., Pluim, J. P. et al., Simultaneous segmentation of multiple heart cavities in 3D transesophageal echocardiograms. In *2013 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2013, 659–662.

Habijan, M., Babin, D., Galić, I., Leventić, H., Romić, K., Velicki, L. and Pižurica, A. (2020). Overview of the whole heart and heart chamber segmentation methods. *Cardiovascular Engineering and Technology* **11**: 725–747.

Habijan, M., Galić, I., Leventić, H. and Romić, K. (2021). Whole heart segmentation using 3d fm-pre-resnet encoder–decoder based architecture with variational autoencoder regularization. *Applied Sciences* **11**(9): 3912.

Habijan, M., Leventić, H., Galić, I. and Babin, D., Whole heart segmentation from CT images using 3D U-net architecture. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2019, 121–126.

Hachulla, A.-L., Noble, S., Guglielmi, G., Agulleiro, D., Müller, H. and Vallée, J.-P. (2019). 3D-printed heart model to guide LAA closure: useful in clinical practice? *European radiology* **29**: 251–258.

Haft-Javaherian, M., Villiger, M., Schaffer, C. B., Nishimura, N., Golland, P. and Bouma, B. E., A topological encoding convolutional neural network for segmentation of 3D multiphoton images of brain vasculature using persistent homology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, 990–991.

Hajij, M., Zamzmi, G. and Batayneh, F., TDA-Net: fusion of persistent homology and deep learning features for COVID-19 detection from chest X-Ray images. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, 4115–4119.

Han, F., Co-Vu, J., Lopez-Colon, D., Forder, J., Bleiweis, M., Reyes, K., DeGroff, C. and Chandran, A. (2019). Impact of 3D printouts in optimizing surgical results for complex congenital heart disease. *World Journal for Pediatric and Congenital Heart Surgery* **10**(5): 533–538.

Han, T., Ivo, R. F., Rodrigues, D. d. A., Peixoto, S. A., de Albuquerque, V. H. C. and Reboucas Filho, P. P. (2020). Cascaded volumetric fully convolutional networks for whole-heart and great vessel 3D segmentation. *Future Generation Computer Systems* **108**: 198–209.

Hangge, P., Pershad, Y., Witting, A. A., Albadawi, H. and Oklu, R. (2018). Three-dimensional (3D) printing and its applications for aortic diseases. *Cardiovascular diagnosis and therapy* **8**(Suppl 1): S19.

Hansen, J. H., Duong, P., Jivanji, S. G., Jones, M., Kabir, S., Butera, G., Qureshi, S. A. and Rosenthal, E. (2020). Transcatheter correction of superior sinus venosus atrial septal defects as an alternative to surgical treatment. *Journal of the American College of Cardiology* **75**(11): 1266–1278.

Harb, S. C., Rodriguez, L. L., Vukicevic, M., Kapadia, S. R. and Little, S. H. (2019). Three-dimensional printing applications in percutaneous structural heart interventions. *Circulation: Cardiovascular Imaging* **12**(10): e009014.

Hardoon, D. R., Szedmak, S. and Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation* **16**(12): 2639–2664.

Harms, J., Lei, Y., Tian, S., McCall, N. S., Higgins, K. A., Bradley, J. D., Curran, W. J., Liu, T. and Yang, X. (2021). Automatic delineation of cardiac substructures using a region-based fully convolutional network. *Medical Physics* **48**(6): 2867–2876.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)* **28**(1): 100–108.

He, K., Gkioxari, G., Dollár, P. and Girshick, R., Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 2017a, 2961–2969.

He, K., Zhang, X., Ren, S. and Sun, J., Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 2015, 1026–1034.

He, K., Zhang, X., Ren, S. and Sun, J., Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 770–778.

He, L., Ren, X., Gao, Q., Zhao, X., Yao, B. and Chao, Y. (2017b). The connected-component labeling problem: A review of state-of-the-art algorithms. *Pattern Recognition* **70**: 25–43.

He, Y., Carass, A., Liu, Y., Jedynak, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A. and Prince, J. L., Fully convolutional boundary regression for retina OCT segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*. Springer International Publishing, 2019, 120–128.

Heathfield, E., Hussain, T., Qureshi, S., Valverde, I., Witter, T., Douiri, A., Bell, A., Beerbaum, P., Razavi, R. and Greil, G. F. (2013). Cardiovascular magnetic resonance imaging in congenital heart disease as an alternative to diagnostic invasive cardiac catheterization: a single center experience. *Congenital Heart Disease* **8**(4): 322–327.

Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D. and Hammers, A. (2006). Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* **33**(1): 115–126.

Heinrich, M. P. and Oster, J., MRI whole heart segmentation using discrete nonlinear registration and fast non-local fusion. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, 233–241.

Hermsen, J. L., Burke, T. M., Seslar, S. P., Owens, D. S., Ripley, B. A., Mokadam, N. A. and Verrier, E. D. (2017). Scan, plan, print, practice, perform: development and use of a patient-specific 3-dimensional printed model in adult cardiac surgery. *The Journal of thoracic and cardiovascular surgery* **153**(1): 132–140.

Hermsen, J. L., Roldan-Alzate, A. and Anagnostopoulos, P. V. (2020). Three-dimensional printing in congenital heart disease. *Journal of thoracic disease* **12**(3): 1194.

Hernández-García, A. and König, P. (2018). Data augmentation instead of explicit regularization. *arXiv Preprint* (arXiv:1806.03852).

Hernández-Hoyos, M., Orkisz, M., Puech, P., Mansard-Desbleds, C., Douek, P. and Magnin, I. E. (2002). Computer-assisted analysis of three-dimensional MR angiograms. *RadioGraphics* **22**(2): 421–436.

Hesamian, M. H., Jia, W., He, X. and Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging* **32**: 582–596.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv Preprint* (arXiv:1207.0580).

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence* **20**(8): 832–844.

Hoashi, T., Ichikawa, H., Nakata, T., Shimada, M., Ozawa, H., Higashida, A., Kurosaki, K., Kanzaki, S. and Shiraishi, I. (2018). Utility of a super-flexible three-dimensional printed heart model in congenital heart surgery. *Interactive CardioVascular and Thoracic Surgery* **27**(5): 749–755.

Hofer, C., Kwitt, R., Niethammer, M. and Dixit, M., Connectivity-Optimized Representation Learning via Persistent Homology. In Chaudhuri, K. and Salakhutdinov, R. (Editors), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*. PMLR, 2019, 2751–2760.

Hofer, C., Kwitt, R., Niethammer, M. and Uhl, A., Deep Learning with Topological Signatures. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R. (Editors), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Hoffman, J. I. and Kaplan, S. (2002). The incidence of congenital heart disease. *Journal of the American college of cardiology* **39**(12): 1890–1900.

Hu, X., Li, F., Samaras, D. and Chen, C., Topology-Preserving Deep Image Segmentation. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. and Garnett, R. (Editors), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q., Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, 4700–4708.

Huang, S., Aregullin, E. O., Gosnell, J. M., Samuel, B. P., Kaley, V. R., Castiaux, A., Pinger, C., Apkinar, M. H., Chinnadurai, P., Spence, D. M. et al. (2019). Rapid prototyping and image fusion guidance for transcatheter closure of superior sinus venosus atrial septal defect. *SN Comprehensive Clinical Medicine* **1**: 996–1000.

Hudec, M., Chovancik, J., Jiravsky, O., Hecko, J., Miklík, R. and Sknouril, L. (2021). Prediction of Left Atrial Appendage occluder size based on 3D printed models. *EP Europace* **23**(Supplement_3): euab116–285.

Hunter, A. L. and Swan, L. (2016). Quality of life in adults living with congenital heart disease: beyond morbidity and mortality. *Journal of thoracic disease* **8**(12): E1632.

Hussein, N., Honjo, O., Barron, D. J., Haller, C., Coles, J. G., Van Arsdell, G., Lim, A. and Yoo, S.-J. (2021). Assessment tool validation and technical skill improvement in the simulation of the Norwood operation using three-dimensional-printed heart models. *European Journal of Cardio-Thoracic Surgery* **59**(2): 316–324.

Hussein, N., Honjo, O., Haller, C., Hickey, E., Coles, J. G., Williams, W. G. and Yoo, S.-J. (2020). Hands-On Surgical Simulation in Congenital Heart Surgery: Literature Review and Future Perspective. *Seminars in Thoracic and Cardiovascular Surgery* **32**(1): 98–105.

Huttenlocher, D. P., Klanderman, G. A. and Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence* **15**(9): 850–863.

Imai, M., Yoshida, M., Toyota, T., Shiomi, H., Shizuta, S., Saito, N. and Kimura, T. (2018). Successful Catheter Treatment Using Pre-Operative 3D Organ Model Simulation for Atrial Septal Defect With Dextrocardia and Interrupted Inferior Vena Cava to the Superior Vena Cava. *JACC: Cardiovascular Interventions* **11**(8): e63–e64.

Ioffe, S. and Szegedy, C., Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Bach, F. and Blei, D. (Editors), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*. Lille, France: PMLR, 2015, 448–456.

Iriart, X., Ciobotaru, V., Martin, C., Cochet, H., Jalal, Z., Thambo, J.-B. and Quessard, A. (2018). Role of cardiac imaging and three-dimensional printing in percutaneous appendage closure. *Archives of Cardiovascular Diseases* **111**(6-7): 411–420.

Isensee, F., Jaeger, P. F., Full, P. M., Wolf, I., Engelhardt, S. and Maier-Hein, K. H., Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, 120–129.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2): 203–211.

Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M. A. and Van Ginneken, B. (2009). Multi-atlas-based segmentation with local decision fusion–application to cardiac and aortic segmentation in CT scans. *IEEE transactions on medical imaging* **28**(7): 1000–1010.

Jacobs, S., Grunert, R., Mohr, F. W. and Falk, V. (2008). 3D-Imaging of cardiac structures using 3D heart models for planning in heart surgery: a preliminary study. *Interactive cardiovascular and thoracic surgery* **7**(1): 6–9.

Jaderberg, M., Simonyan, K., Zisserman, A. and kavukcuoglu, k., Spatial Transformer Networks. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R. (Editors), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Jain, A. K. and Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern recognition* **24**(12): 1167–1186.

Jain, V., Bollmann, B., Richardson, M., Berger, D. R., Helmstaedter, M. N., Briggman, K. L., Denk, W., Bowden, J. B., Mendenhall, J. M., Abraham, W. C. et al., Boundary learning by optimization with topological constraints. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, 2488–2495.

Jang, S.-J., Torabinia, M., Dhrif, H., Caprio, A., Liu, J., Wong, S. C. and Mosadegh, B. (2020). Development of a hybrid training simulator for structural heart disease interventions. *Advanced Intelligent Systems* **2**(12): 2000109.

Ji, S., Xu, W., Yang, M. and Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence* **35**(1): 221–231.

Jivanji, S. G., Qureshi, S. A. and Rosenthal, E. (2019). Novel use of a 3D printed heart model to guide simultaneous percutaneous repair of severe pulmonary regurgitation and right ventricular outflow tract aneurysm. *Cardiology in the Young* **29**(4): 534–537.

Ju, T., Losasso, F., Schaefer, S. and Warren, J., Dual contouring of hermite data. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 2002, 339–346.

Kaczynski, T., Mischaikow, K. and Mrozek, M. (2006). *Computational Homology*, volume 157. Springer Science & Business Media.

Kaggle (2016). Second Annual Data Science Bowl: Transforming How We Diagnose Heart Disease. Retrieved 6 June 2021 from https://www.kaggle.com/c/second-annual-data-science-bowl.

Kaji, S., Sudo, T. and Ahara, K. (2020). Cubical ripser: Software for computing persistent homology of image and volume data. *arXiv Preprint* (arXiv:2005.12692).

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D. and Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis* **36**: 61–78.

Kanakatte, A., Bhatia, D. and Ghose, A., Heart Region Segmentation using Dense VNet from Multimodality Images. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, 3255–3258.

Kang, D., Woo, J., Slomka, P. J., Dey, D., Germano, G. and Jay Kuo, C.-C. (2012). Heart chambers and whole heart segmentation techniques. *Journal of Electronic Imaging* **21**(1): 010901–010901.

Kang, J., Samarasinghe, G., Senanayake, U., Conjeti, S. and Sowmya, A., Deep learning for volumetric segmentation in spatio-temporal data: application to segmentation of prostate in DCE-MRI. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, 61–65.

Kanopoulos, N., Vasanthavada, N. and Baker, R. L. (1988). Design of an image edge detection filter using the Sobel operator. *IEEE Journal of solid-state circuits* **23**(2): 358–367.

Kappanayil, M., Koneti, N. R., Kannan, R. R., Kottayil, B. P. and Kumar, K. (2017). Three-dimensional-printed cardiac prototypes aid surgical decision-making and preoperative planning in selected cases of complex congenital heart diseases: Early experience and proof of concept in a resource-limited environment. *Annals of pediatric cardiology* **10**(2): 117.

Karimi-Bidhendi, S., Arafati, A., Cheng, A. L., Wu, Y., Kheradvar, A. and Jafarkhani, H. (2020). Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *Journal of cardiovascular magnetic resonance* **22**(1): 80.

Karsenty, C., Guitarte, A., Dulac, Y., Briot, J., Hascoet, S., Vincent, R., Delepaul, B., Vignaud, P., Djeddai, C., Hadeed, K. et al. (2021). The usefulness of 3D printed heart models for medical student education in congenital heart disease. *BMC medical education* **21**(1): 1–8.

Kass, M., Witkin, A. and Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision* **1**(4): 321–331.

Kausar, A., Razzak, I., Shapiai, M. I. and Beheshti, A. (2021). 3D shallow deep neural network for fast and precise segmentation of left atrium. *Multimedia Systems* : 1–11.

Khalili, H., Gentry, R. E., Stevens, M. A., Almany, S. L., Banerjee, S., Haines, D. E. and Hanzel, G. S. (2017). Rapid and affordable 3-dimensional prototyping for left atrial appendage closure planning. *Circulation: Cardiovascular Interventions* **10**(2): e004710.

Kikinis, R., Pieper, S. D. and Vosburgh, K. G. (2014). 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. *Intraoperative imaging and image-guided therapy* : 277–289.

Kim, M. S., Hansgen, A. R. and Carroll, J. D. (2008a). Use of rapid prototyping in the care of patients with structural heart disease. *Trends in cardiovascular medicine* **18**(6): 210–216.

Kim, M. S., Hansgen, A. R., Wink, O., Quaife, R. A. and Carroll, J. D. (2008b). Rapid prototyping: a new tool in understanding and treating structural heart disease. *Circulation* **117**(18): 2388–2394.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv Preprint* (arXiv:1412.6980).

Kiraly, L. (2018). Three-dimensional modelling and three-dimensional printing in pediatric and congenital cardiac surgery. *Translational pediatrics* **7**(2): 129.

Kiraly, L., Kiraly, B., Szigeti, K., Tamas, C. Z. and Daranyi, S. (2019). Virtual museum of congenital heart defects: digitization and establishment of a database for cardiac specimens. *Quantitative imaging in medicine and surgery* **9**(1): 115.

Kiraly, L., Shah, N. C., Abdullah, O., Al-Ketan, O. and Rowshan, R. (2021). Three-dimensional virtual and printed prototypes in complex congenital and pediatric cardiac surgery–a multidisciplinary team-learning experience. *Biomolecules* **11**(11): 1703.

Kirişli, H., Schaap, M., Klein, S., Papadopoulou, S.-L., Bonardi, M., Chen, C.-H., Weustink, A. C., Mollet, N. R., Vonken, E.-J., Van Der Geest, R. J. et al. (2010). Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: a large-scale, multicenter, and multivendor study. *Medical physics* **37**(12): 6279–6291.

Koehler, S., Tandon, A., Hussain, T., Latus, H., Pickardt, T., Sarikouch, S., Beerbaum, P., Greil, G., Engelhardt, S. and Wolf, I., How well do U-Net-based segmentation trained on adult cardiac magnetic resonance imaging data generalize to rare congenital heart diseases for surgical planning? In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11315. SPIE, 2020, 409–421.

Koikkalainen, J., Tolli, T., Lauerma, K., Antila, K., Mattila, E., Lilja, M. and Lotjonen, J. (2008). Methods of artificial enlargement of the training set for statistical shape models. *IEEE Transactions on Medical Imaging* **27**(11): 1643–1654.

Kong, T. Y. and Rosenfeld, A. (1989). Digital topology: Introduction and survey. *Computer Vision, Graphics, and Image Processing* **48**(3): 357–393.

Krähenbühl, P. and Koltun, V., Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F. and Weinberger, K. (Editors), *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **60**(6): 84–90.

Kuk, M., Mitsouras, D., Dill, K. E., Rybicki, F. J. and Dwivedi, G. (2017). 3D printing from cardiac computed tomography for procedural planning. *Current Cardiovascular Imaging Reports* **10**: 1–9.

Kurup, H. K., Samuel, B. P. and Vettukattil, J. J. (2015). Hybrid 3D printing: a game-changer in personalized cardiac medicine? .

Lapp, R. M., Lorenzo-Valdés, M. and Rueckert, D., 3D/4D cardiac segmentation using active appearance models, non-rigid registration, and the insight toolkit. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004: 7th International Conference, Saint-Malo, France, September 26-29, 2004. Proceedings, Part I 7*. Springer, 2004, 419–426.

Larrey-Ruiz, J., Morales-Sánchez, J., Bastida-Jumilla, M. C., Menchón-Lara, R. M., Verdú-Monedero, R. and Sancho-Gómez, J. L. (2014). Automatic image-based segmentation of the heart from CT scans. *EURASIP Journal on Image and Video Processing* **2014**(1): 1–13.

Larsson, G., Maire, M. and Shakhnarovich, G. (2016). Fractalnet: Ultra-deep neural networks without residuals. *arXiv Preprint* (arXiv:1605.07648).

Lartaud, P.-J., Hallé, D., Schleef, A., Dessouky, R., Vlachomitrou, A. S., Douek, P., Rouet, J.-M., Nempont, O. and Boussel, L. (2021). Spectral augmentation for heart chambers segmentation on conventional contrasted and unenhanced CT scans: an in-depth study. *International Journal of Computer Assisted Radiology and Surgery* **16**(10): 1699–1709.

Lavie-Badie, Y., Cazalbou, S., Itier, R., Briot, J. and Lhermusier, T. (2021). Planning a transcatheter intervention for a surgical mitral valve repair failure: insights from 3D printing. *European Heart Journal-Cardiovascular Imaging* **22**(5): e18–e18.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *nature* **521**(7553): 436–444.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation* **1**(4): 541–551.

Lee, C. and Lee, J. Y. (2020). Utility of three-dimensional printed heart models for education on complex congenital heart diseases. *Cardiology in the Young* **30**(11): 1637–1642.

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z. and Tu, Z., Deeply-Supervised Nets. In Lebanon, G. and Vishwanathan, S. V. N. (Editors), *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*. San Diego, California, USA: PMLR, 2015, 562–570.

Lee, M. C. H., Petersen, K., Pawlowski, N., Glocker, B. and Schaap, M. (2019). TETRIS: Template transformer networks for image segmentation with shape priors. *IEEE transactions on medical imaging* **38**(11): 2596–2606.

Levin, D., Mackensen, G. B., Reisman, M., McCabe, J. M., Dvir, D. and Ripley, B. (2020). 3D printing applications for transcatheter aortic valve replacement. *Current cardiology reports* **22**: 1–9.

Li, H., Shu, M., Wang, X., Song, Z. et al. (2017a). Application of 3D printing technology to left atrial appendage occlusion. *International Journal of Cardiology* **231**: 258–263.

Li, J., Zhang, R., Shi, L. and Wang, D., Automatic whole-heart segmentation in congenital heart disease using deeply-supervised 3D FCN. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer International Publishing, 2017b, 111–118.

Li, P., Fang, F., Qiu, X., Xu, N., Wang, Y., Ouyang, W.-B., Zhang, F.-W., Hu, H.-B. and Pan, X.-B. (2020). Personalized three-dimensional printing and echoguided procedure facilitate single device closure for multiple atrial septal defects. *Journal of Interventional Cardiology* **2020**.

Li, S. Z., Markov random field models in computer vision. In *Computer Vision–ECCV'94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6 1994 Proceedings, Volume II 3*. Springer, 1994, 361–370.

Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M. J. and Vercauteren, T., On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. In *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25*. Springer, 2017c, 348–360.

Liang, J., Zhao, X., Pan, G., Zhang, G., Zhao, D., Xu, J., Li, D. and Lu, B. (2021). Blood-Pool or Myocardial 3D Printing, Which One is Better for the Diagnosis of Types of Congenital Heart Disease? *Research Square Preprint* (rs.3.rs-722298/v1).

Liang, P., Chen, J., Zheng, H., Yang, L., Zhang, Y. and Chen, D. Z., Cascade decoder: A universal decoding method for biomedical image segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, 339–342.

Lillehei, C. (1957). Surgical treatment of congenital and acquired heart disease by use of total cardiopulmonary bypass; analysis of result in 350 patients. *Acta chirurgica Scandinavica* **113**(6): 496–501.

Lin, D., Dai, J., Jia, J., He, K. and Sun, J., Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 3159–3167.

Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollár, P., Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*. 2017, 2980–2988.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis* **42**: 60–88.

Little, S. H., Vukicevic, M., Avenatti, E., Ramchandani, M. and Barker, C. M. (2016). 3D printed modeling for patient-specific mitral valve intervention: repair with a clip and a plug. *JACC: Cardiovascular Interventions* **9**(9): 973–975.

Liu, P., Liu, R., Zhang, Y., Liu, Y., Tang, X. and Cheng, Y. (2016). The value of 3D printing models of left atrial appendage using real-time 3D transesophageal echocardiographic data in left atrial appendage occlusion: applications toward an era of truly personalized medicine. *Cardiology* **135**(4): 255–261.

Liu, T., Tian, Y., Zhao, S. and Huang, X., Graph Reasoning and Shape Constraints for Cardiac Segmentation in Congenital Heart Defect. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*. Springer International Publishing, 2020a, 607–616.

Liu, T., Tian, Y., Zhao, S., Huang, X. and Wang, Q. (2019). Automatic whole heart segmentation using a two-stage u-net framework and an adaptive threshold window. *IEEE Access* **7**: 83628–83636.

Liu, Z., Li, S., Chen, Y.-k., Liu, T., Liu, Q., Xu, X., Shi, Y. and Wen, W., Orchestrating medical image compression and remote segmentation networks. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*. Springer, 2020b, 406–416.

Lobregt, S., Verbeek, P. W. and Groen, F. C. (1980). Three-dimensional skeletonization: principle and algorithm. *IEEE Transactions on pattern analysis and machine intelligence* **PAMI-2**(1): 75–77.

Loke, Y.-H., Harahsheh, A. S., Krieger, A. and Olivieri, L. J. (2017). Usage of 3D models of tetralogy of Fallot for medical education: impact on learning congenital heart disease. *BMC medical education* **17**(1): 1–8.

Lombaert, H., Sun, Y., Grady, L. and Xu, C., A multilevel banded graph cuts method for fast image segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1. IEEE, 2005, 259–265.

Long, J., Shelhamer, E. and Darrell, T., Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, 3431–3440.

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM siggraph computer graphics* **21**(4): 163–169.

Lorenz, C. and von Berg, J. (2006). A comprehensive shape model of the heart. *Medical image analysis* **10**(4): 657–670.

Lorenz, C. H., Walker, E. S., Graham Jr, T. P. and Powers, T. A. (1995). Right ventricular performance and mass by use of cine MRI late after atrial repair of transposition of the great arteries. *Circulation* **92**(9): 233–239.

Lorenzo-Valdés, M., Sanchez-Ortiz, G. I., Elkington, A. G., Mohiaddin, R. H. and Rueckert, D. (2004). Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis* **8**(3): 255–265.

Lösel, P. and Heuveline, V., A GPU based diffusion method for whole-heart and great vessel segmentation. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer International Publishing, 2017, 121–128.

Luo, X. and Zhuang, X., MvMM-RegNet: A new image registration framework based on multivariate mixture model and neural network estimation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23*. Springer International Publishing, 2020, 149–159.

Ma, X., Tao, L., Chen, X., Li, W., Peng, Z., Chen, Y., Jin, J., Zhang, X., Xiong, Q., Zhong, Z. et al. (2015). Clinical application of three-dimensional reconstruction and rapid prototyping technology of multislice spiral computed tomography angiography for the repair of ventricular septal defect of tetralogy of Fallot. *Genet Mol Res* **14**(1): 1301–1309.

Ma, Y., Ding, P., Li, L., Liu, Y., Jin, P., Tang, J. and Yang, J. (2021). Three-dimensional printing for heart diseases: clinical application review. *Bio-design and Manufacturing* **4**(3): 675–687.

Mahapatra, D. (2014). Analyzing training information from random forests for improved image segmentation. *IEEE Transactions on Image Processing* **23**(4): 1504–1512.

Makowski, P., Sørensen, T. S., Therkildsen, S. V., Materka, A., Stødkilde-Jørgensen, H. and Pedersen, E. M. (2002). Two-phase active contour method for semiautomatic segmentation of the heart and blood vessels from MRI images for 3D visualization. *Computerized Medical Imaging and Graphics* **26**(1): 9–17.

Mankovich, N. J., Cheeseman, A. M. and Stoker, N. G. (1990). The display of three-dimensional anatomy with stereolithographic models. *Journal of digital imaging* **3**: 200–203.

Mansi, T., Voigt, I., Leonardi, B., Pennec, X., Durrleman, S., Sermesant, M., Delingette, H., Taylor, A. M., Boudjemline, Y., Pongiglione, G. et al. (2011). A statistical model for quantification and prediction of cardiac remodelling: Application to tetralogy of fallot. *IEEE transactions on medical imaging* **30**(9): 1605–1616.

Marelli, A. J., Ionescu-Ittu, R., Mackie, A. S., Guo, L., Dendukuri, N. and Kaouache, M. (2014). Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. *Circulation* **130**(9): 749–756.

Markl, M., Frydrychowicz, A., Kozerke, S., Hope, M. and Wieben, O. (2012). 4D flow MRI. *Journal of Magnetic Resonance Imaging* **36**(5): 1015–1036.

Medero, R., García-Rodríguez, S., François, C. J. and Roldán-Alzate, A. (2017). Patient-specific in vitro models for hemodynamic analysis of congenital heart disease–Additive manufacturing approach. *Journal of Biomechanics* **54**: 111–116.

Meier, L., Meineri, M., Qua Hiansen, J. and Horlick, E. (2017). Structural and congenital heart disease interventions: the role of three-dimensional printing. *Netherlands Heart Journal* **25**: 65–75.

Meyer-Szary, J., Woźniak-Mielczarek, L., Sabiniewicz, D. and Sabiniewicz, R. (2019). Feasibility of in-house rapid prototyping of cardiovascular three-dimensional models for planning and training non-standard interventional procedures. *Cardiology Journal* **26**(6): 790–792.

Milletari, F., Navab, N. and Ahmadi, S.-A., V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, 565–571.

Milletari, F., Rothberg, A., Jia, J. and Sofka, M., Integrating statistical prior knowledge into convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*. Springer International Publishing, 2017, 161–168.

Min, S., Chen, X., Xie, H., Zha, Z.-J. and Zhang, Y. (2020). A mutually attentive co-training framework for semi-supervised recognition. *IEEE Transactions on Multimedia* **23**: 899–910.

Min, S., Chen, X., Zha, Z.-J., Wu, F. and Zhang, Y., A two-stream mutual attention network for semi-supervised biomedical segmentation with noisy labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33(01). 2019, 4578–4585.

Minnema, J., Wolff, J., Koivisto, J., Lucka, F., Batenburg, K. J., Forouzanfar, T. and van Eijnatten, M. (2021). Comparison of convolutional neural network training strategies for cone-beam CT image segmentation. *Computer Methods and Programs in Biomedicine* **207**: 106192.

Mitchell, S., Korones, S. and Berendes, H. (1971). Congenital heart disease in 56,109 births incidence and natural history. *Circulation* **43**(3): 323–332.

Mitchell, S. C., Bosch, J. G., Lelieveldt, B. P., Van der Geest, R. J., Reiber, J. H. and Sonka, M. (2002). 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *IEEE transactions on medical imaging* **21**(9): 1167–1178.

Mitchell, S. C., Lelieveldt, B. P., Van Der Geest, R. J., Bosch, H. G., Reiver, J. and Sonka, M. (2001). Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. *IEEE Transactions on medical imaging* **20**(5): 415–423.

Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C. and Ourselin, S. (2010). Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine* **98**(3): 278–284.

Mokkarala, M., Ballard, D. H., Wesley, R. A., Gutierrez, F. R., Javidan-Nejad, C., Singh, G. K., Woodard, P. K. and Lindley, K. J. (2020). Coronary-cameral fistula with double-chambered right ventricle: appearance on cardiac magnetic resonance imaging and 3D printed anatomic modeling. *Clinical imaging* **59**(1): 84–87.

Moons, P., Sluysmans, T., De Wolf, D., Massin, M., Suys, B., Benatar, A. and Gewillig, M. (2009). Congenital heart disease in 111 225 births in Belgium: birth prevalence, treatment and survival in the 21st century. *Acta paediatrica* **98**(3): 472–477.

Moore, R. A., Riggs, K. W., Kourtidou, S., Schneider, K., Szugye, N., Troja, W., D'Souza, G., Rattan, M., Bryant III, R., Taylor, M. D. et al. (2018). Three-dimensional printing and virtual surgery for congenital heart procedural planning. *Birth defects research* **110**(13): 1082–1090.

Mortazi, A., Burt, J. and Bagci, U., Multi-planar deep segmentation networks for cardiac substructures from MRI and CT. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, 199–206.

Mortazi, A., Karim, R., Rhode, K., Burt, J. and Bagci, U., CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20*. Springer, 2017, 377–385.

Mottl-Link, S., Hübler, M., Kühne, T., Rietdorf, U., Krueger, J. J., Schnackenburg, B., De Simone, R., Berger, F., Juraszek, A., Meinzer, H.-P. et al. (2008). Physical models aiding in complex congenital heart surgery. *The Annals of thoracic surgery* **86**(1): 273–277.

Mowers, K., Fullerton, J., Hicks, D., Singh, G., Johnson, M. and Anwar, S. (2021). 3D echocardiography provides highly accurate 3D printed models in congenital heart disease. *Pediatric Cardiology* **42**: 131–141.

Mukhopadhyay, A., Total variation random forest: Fully automatic mri segmentation in congenital heart diseases. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer International Publishing, 2017, 165–171.

Muraru, D., Veronesi, F., Maddalozzo, A., Dequal, D., Frajhof, L., Rabischoffsky, A., Iliceto, S. and Badano, L. P. (2017). 3D printing of normal and pathologic tricuspid valves from transthoracic 3D echocardiography data sets. *European Heart Journal-Cardiovascular Imaging* **18**(7): 802–808.

Myronenko, A., Yang, D., Buch, V., Xu, D., Ihsani, A., Doyle, S., Michalski, M., Tenenholtz, N. and Roth, H., 4D CNN for semantic segmentation of cardiac volumetric sequences. In *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers 10*. Springer, 2020, 72–80.

Nainamalai, V., Lippert, M., Brun, H., Elle, O. J. and Kumar, R. P. (2022). Local integration of deep learning for advanced visualization in congenital heart disease surgical planning. *Intelligence-Based Medicine* **6**: 100055.

Nair, V. and Hinton, G. E., Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010, 807–814.

Negi, S., Dhiman, S. and Sharma, R. K. (2014). Basics and applications of rapid prototyping medical models. *Rapid Prototyping Journal* **20**(3): 256–267.

Nesterov, Y. (2003). *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media.

New Media Centre - University of Basel (2021). Congenital heart disease. Retrieved 22 November 2021 from https://www.congenital-heart-disease.ch/heart-disease.

Ngan, E. M., Rebeyka, I. M., Ross, D. B., Hirji, M., Wolfaardt, J. F., Seelaus, R., Grosvenor, A. and Noga, M. L. (2006). The rapid prototyping of anatomic models in pulmonary atresia. *The Journal of thoracic and cardiovascular surgery* **132**(2): 264–269.

Ngo, T. A., Lu, Z. and Carneiro, G. (2017). Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Medical image analysis* **35**: 159–171.

Niederer, S. A., Lumens, J. and Trayanova, N. A. (2019). Computational models in cardiology. *Nature Reviews Cardiology* **16**(2): 100–111.

Noecker, A. M., Chen, J.-F., Zhou, Q., White, R. D., Kopcak, M. W., Arruda, M. J. and Duncan, B. W. (2006). Development of patient-specific three-dimensional pediatric cardiac models. *ASAIO journal* **52**(3): 349–353.

Noh, H., Hong, S. and Han, B., Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*. 2015, 1520–1528.

Nova, R., Nurmaini, S., Partan, R. U. and Putra, S. T. (2021). Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study. *Informatics in Medicine Unlocked* **24**: 100601.

Ntsinjana, H. N., Hughes, M. L. and Taylor, A. M. (2011). The role of cardiovascular magnetic resonance in pediatric congenital heart disease. *Journal of Cardiovascular Magnetic Resonance* **13**: 1–20.

Nurmaini, S., Rachmatullah, M. N., Sapitri, A. I., Darmawahyuni, A., Jovandy, A., Firdaus, F., Tutuko, B. and Passarella, R. (2020). Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation. *IEEE Access* **8**: 196160–196174.

Obasare, E., Mainigi, S. K., Morris, D. L., Slipczuk, L., Goykhman, I., Friend, E., Ziccardi, M. R. and Pressman, G. S. (2018). CT based 3D printing is superior to transesophageal echocardiography for pre-procedure planning in left atrial appendage device closure. *The international journal of cardiovascular imaging* **34**: 821–831.

Ochoa, S., Segal, J., Garcia, N. and Fischer, E. A. (2019). Three-dimensional printed cardiac models for focused cardiac ultrasound instruction. *Journal of Ultrasound in Medicine* **38**(6): 1405–1409.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S. A., De Marvao, A., Dawes, T., O'Regan, D. P. et al. (2017). Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE transactions on medical imaging* **37**(2): 384–395.

Olejnik, P., Juskanic, D., Patrovic, L. and Halaj, M. (2018). First printed 3D heart model based on cardiac magnetic resonance imaging data in Slovakia. *Bratislavske Lekarske Listy* **119**(12): 781–784.

Oliveira-Santos, M., Santos, E. O., Marinho, A. V., Leite, L., Guardado, J., Matos, V., Pego, G. M. and Marques, J. S. (2018). Patient-specific 3D printing simulation to guide complex coronary intervention. *Revista Portuguesa de Cardiologia* **37**(6): 541–e1.

Ong, C. S., Krishnan, A., Huang, C. Y., Spevak, P., Vricella, L., Hibino, N., Garcia, J. R. and Gaur, L. (2018). Role of virtual reality in congenital heart disease. *Congenital heart disease* **13**(3): 357–361.

Oster, M. E., Lee, K. A., Honein, M. A., Riehle-Colarusso, T., Shin, M. and Correa, A. (2013). Temporal trends in survival among infants with critical congenital heart defects. *Pediatrics* **131**(5): e1502–e1508.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* **9**(1): 62–66.

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P. and Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science* **6**: 1–38.

Otton, J. M., Birbara, N. S., Hussain, T., Greil, G., Foley, T. A. and Pather, N. (2017). 3D printing from cardiovascular CT: a practical guide and review. *Cardiovascular diagnosis and therapy* **7**(5): 507.

Pace, D. F., Dalca, A. V., Brosch, T., Geva, T., Powell, A. J., Weese, J., Moghari, M. H. and Golland, P., Iterative segmentation from limited training data: applications to congenital heart disease. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer International Publishing, 2018, 334–342.

Pace, D. F., Dalca, A. V., Geva, T., Powell, A. J., Moghari, M. H. and Golland, P., Interactive whole-heart segmentation in congenital heart disease. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer International Publishing, 2015, 80–88.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery* **88**: 105906.

Painchaud, N., Skandarani, Y., Judge, T., Bernard, O., Lalande, A. and Jodoin, P.-M. (2020). Cardiac segmentation with strong anatomical guarantees. *IEEE transactions on medical imaging* **39**(11): 3703–3713.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S., PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. and Garnett, R. (Editors), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Payer, C., Štern, D., Bischof, H. and Urschler, M., Multi-label whole heart segmentation using CNNs and anatomical label configurations. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers*. Springer, 2018, 190–198.

Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S. E. and Frangi, A. F. (2016). A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine* **29**: 155–195.

Perens, G., Chyu, J., McHenry, K., Yoshida, T. and Finn, J. P. (2020). Three-dimensional congenital heart models created with free software and a desktop printer: assessment of accuracy, technical aspects, and clinical use. *World Journal for Pediatric and Congenital Heart Surgery* **11**(6): 797–801.

Pérez-García, F., Sparks, R. and Ourselin, S. (2021). TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine* **208**: 106236.

Peters, J., Ecabert, O., Meyer, C., Kneser, R. and Weese, J. (2010). Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation. *Medical image analysis* **14**(1): 70–84.

Peters, J., Ecabert, O., Meyer, C., Schramm, H., Kneser, R., Groth, A. and Weese, J., Automatic whole heart segmentation in static magnetic resonance image volumes. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29-November 2, 2007, Proceedings, Part II 10*. Springer Berlin Heidelberg, 2007, 402–410.

Petitjean, C. and Dacher, J.-N. (2011). A review of segmentation methods in short axis cardiac MR images. *Medical image analysis* **15**(2): 169–184.

Petitjean, C., Zuluaga, M. A., Bai, W., Dacher, J.-N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I. B., Cardoso, M. J., Chen, H.-C. et al. (2015). Right ventricle segmentation from cardiac MRI: a collation study. *Medical image analysis* **19**(1): 187–202.

Peyrat, J.-M., Delingette, H., Sermesant, M., Xu, C. and Ayache, N. (2010). Registration of 4D cardiac CT sequences under trajectory constraints with multichannel diffeomorphic demons. *IEEE transactions on medical imaging* **29**(7): 1351–1368.

Pracon, R., Grygoruk, R., Konka, M., Kepka, C. and Demkow, M. (2018). Percutaneous Closure of Ventricular Septal Defect Resulting From Chest Stab Wound in an 18-Year-Old Boy: A 3-Dimensional Heart Model–Guided Procedure. *Circulation: Cardiovascular Imaging* **11**(11): e008326.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E. and Nielsen, M., Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16*. Springer, 2013, 246–253.

Qaiser, T., Sirinukunwattana, K., Nakane, K., Tsang, Y.-W., Epstein, D. and Rajpoot, N. (2016). Persistent homology for fast tumor segmentation in whole slide histology images. *Procedia Computer Science* **90**: 119–124.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural networks* **12**(1): 145–151.

Qian, Z., Wang, K., Liu, S., Zhou, X., Rajagopal, V., Meduri, C., Kauten, J. R., Chang, Y.-H., Wu, C., Zhang, C. et al. (2017). Quantitative prediction of paravalvular leak in transcatheter aortic valve replacement based on tissue-mimicking 3D printing. *JACC: Cardiovascular Imaging* **10**(7): 719–731.

Queirós, S., Barbosa, D., Heyde, B., Morais, P., Vilaça, J. L., Friboulet, D., Bernard, O. and D'hooge, J. (2014). Fast automatic myocardial segmentation in 4D cine CMR datasets. *Medical image analysis* **18**(7): 1115–1131.

Quimby Jr, D. L., Ford, J., Tanner, G. J., Mencer, N., Decker, S. and Matar, F. (2022). Three-dimensional cardiac print assisted percutaneous closure of left ventricular pseudoaneurysm in patient with Behçet's disease. *Catheterization and Cardiovascular Interventions* **99**(2): 512–517.

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A. and Wright, G. (2009). Evaluation framework for algorithms segmenting short axis cardiac MRI. *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge* **49**: 4.

Ran, C., Liu, P., Qian, Y., He, Y. and Wang, Q., U-shaped densely connected convolutional networks for automatic 3D cardiovascular MR segmentation. In *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2018, 1010–1015.

Razavi, R. S., Hill, D. L., Muthurangu, V., Miquel, M. E., Taylor, A. M., Kozerke, S. and Baker, E. J. (2003). Three-dimensional magnetic resonance imaging of congenital cardiac anomalies. *Cardiology in the Young* **13**(5): 461–465.

Ren, S., He, K., Girshick, R. and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **28**.

Rengier, F., Mehndiratta, A., Von Tengg-Kobligk, H., Zechmann, C. M., Unterhinninghofen, R., Kauczor, H.-U. and Giesel, F. L. (2010). 3D printing based on imaging data: review of medical applications. *International journal of computer assisted radiology and surgery* **5**: 335–341.

Rezaei, M., Yang, H. and Meinel, C. (2020). Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation. *Multimedia Tools and Applications* **79**(21-22): 15329–15348.

Riahi, M., Velasco Forte, M., Byrne, N., Hermuzi, A., Jones, M., Baruteau, A.-E., Valverde, I., Qureshi, S. A. and Rosenthal, E. (2018). Early experience of transcatheter correction of superior sinus venosus atrial septal defect with partial anomalous pulmonary venous drainage. *EuroIntervention* **14**(8): 868–876.

Riesenkampff, E., Rietdorf, U., Wolf, I., Schnackenburg, B., Ewert, P., Huebler, M., Alexi-Meskishvili, V., Anderson, R. H., Engel, N., Meinzer, H.-P. et al. (2009). The practical clinical value of three-dimensional models of complex congenitally malformed hearts. *The Journal of thoracic and cardiovascular surgery* **138**(3): 571–580.

Ringenberg, J., Deo, M., Devabhaktuni, V., Berenfeld, O., Boyers, P. and Gold, J. (2014). Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI. *Computerized Medical Imaging and Graphics* **38**(3): 190–201.

Rohlfing, T., Brandt, R., Menzel, R. and Maurer Jr, C. R. (2004). Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage* **21**(4): 1428–1442.

Ronneberger, O., Fischer, P. and Brox, T., U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer International Publishing, 2015, 234–241.

Rosset, A., Spadola, L. and Ratib, O. (2004). OsiriX: an open-source software for navigating in multidimensional DICOM images. *Journal of digital imaging* **17**: 205–216.

Roth, H. R., Shen, C., Oda, H., Sugino, T., Oda, M., Hayashi, Y., Misawa, K. and Mori, K., A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11*. Springer International Publishing, 2018, 417–425.

Ruijsink, B., Puyol-Antón, E., Oksuz, I., Sinclair, M., Bai, W., Schnabel, J. A., Razavi, R. and King, A. P. (2020). Fully automated, quality-controlled cardiac analysis from CMR: validation and large-scale application to characterize cardiac function. *Cardiovascular Imaging* **13**(3): 684–695.

Rumelhart, D. E., Durbin, R., Golden, R. and Chauvin, Y. (1995). Backpropagation: The basic theory. *Backpropagation: Theory, architectures and applications* : 1–34.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature* **323**(6088): 533–536.

Rupprecht, C., Huaroc, E., Baust, M. and Navab, N. (2016). Deep active contours. *arXiv Preprint* (arXiv:1607.05074).

Ryan, J., Plasencia, J., Richardson, R., Velez, D., Nigro, J. J., Pophal, S. and Frakes, D. (2018). 3D printing for congenital heart disease: a single site's initial three-yearexperience. *3D printing in medicine* **4**(1): 1–9.

Saeed, D., Ootaki, Y., Noecker, A., Weber, S., Smith, W. A., Duncan, B. W. and Fukamachi, K. (2008). The Cleveland Clinic PediPump: Virtual fitting studies in children using three-dimensional reconstructions of cardiac computed tomography scans. *Asaio Journal* **54**(1): 133–137.

Schaffer, M. (2013). Spectrum of congenital cardiac defects. *Pediatric and congenital cardiology, cardiac surgery and intensive care, London: Springer* : 1419–1123.

Schievano, S., Migliavacca, F., Coats, L., Khambadkone, S., Carminati, M., Wilson, N., Deanfield, J. E., Bonhoeffer, P. and Taylor, A. M. (2007). Percutaneous pulmonary valve implantation based on rapid prototyping of right ventricular outflow tract and pulmonary trunk from MR data. *Radiology* **242**(2): 490–497.

Schievano, S., Taylor, A. M., Capelli, C., Coats, L., Walker, F., Lurz, P., Nordmeyer, J., Wright, S., Khambadkone, S., Tsang, V. et al. (2010). First-in-man implantation of a novel percutaneous valve: a new approach to medical device development. *EuroIntervention: journal of EuroPCR in collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology* **5**(6): 745–750.

Schmauss, D., Haeberle, S., Hagl, C. and Sodian, R. (2015). Three-dimensional printing in cardiac surgery and interventional cardiology: a single-centre experience. *European Journal of Cardio-Thoracic Surgery* **47**(6): 1044–1052.

Schneider, K., Ghaleb, S., Morales, D. L. and Tretter, J. T. (2019). Virtual dissection and endocast three-dimensional reconstructions: maximizing computed tomographic data for procedural planning of an obstructed pulmonary venous baffle. *Cardiology in the Young* **29**(8): 1104–1106.

Serraf, A., Lacour-Gayet, F., Bruniaux, J., Ouaknine, R., Losay, J., Petit, J., Binet, J.-P., Planché, C. and Kirklin, J. W. (1992). Surgical management of isolated multiple ventricular septal defects: logical approach in 130 cases. *The Journal of Thoracic and Cardiovascular Surgery* **103**(3): 437–443.

Sezgin, M. and Sankur, B. l. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging* **13**(1): 146–168.

Shahzad, R., Bos, D., Budde, R. P., Pellikaan, K., Niessen, W. J., Van Der Lugt, A. and Van Walsum, T. (2017a). Automatic segmentation and quantification of the cardiac structures from non-contrast-enhanced cardiac CT scans. *Physics in Medicine & Biology* **62**(9): 3798.

Shahzad, R., Gao, S., Tao, Q., Dzyubachyk, O. and Van Der Geest, R., Automated cardiovascular segmentation in patients with congenital heart disease from 3D CMR scans: combining multi-atlases and level-sets. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer, 2017b, 147–155.

Sharobeem, S., Le Breton, H., Lalys, F., Lederlin, M., Lagorce, C., Bedossa, M., Boulmier, D., Leurent, G., Haigron, P. and Auffret, V. (2022). Validation of a whole heart segmentation from computed tomography imaging using a deep-learning approach. *Journal of Cardiovascular Translational Research* **15**(2): 427–437.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k. and Woo, W.-c., Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R. (Editors), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Shin, H.-C., Orton, M. R., Collins, D. J., Doran, S. J. and Leach, M. O. (2012). Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. *IEEE transactions on pattern analysis and machine intelligence* **35**(8): 1930–1943.

Shin, J. and Truong, Q. A. (2018). Manufacturing better outcomes in cardiovascular intervention: 3D printing in clinical practice today. *Current Treatment Options in Cardiovascular Medicine* **20**: 1–13.

Shin, S. Y., Lee, S., Elton, D., Gulley, J. L. and Summers, R. M., Deep small bowel segmentation with cylindrical topological constraints. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*. Springer International Publishing, 2020, 207–215.

Shiraishi, I., Yamagishi, M., Hamaoka, K., Fukuzawa, M. and Yagihara, T. (2010). Simulative operation on congenital heart disease using rubber-like urethane stereolithographic biomodels based on 3D datasets of multislice computed tomography. *European journal of cardio-thoracic surgery* **37**(2): 302–306.

Shrivastava, A., Gupta, A. and Girshick, R., Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 761–769.

Simonyan, K. and Zisserman, A., Two-Stream Convolutional Networks for Action Recognition in Videos. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. and Weinberger, K. (Editors), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

Sinclair, M., Schuh, A., Hahn, K., Petersen, K., Bai, Y., Batten, J., Schaap, M. and Glocker, B. (2022). Atlas-ISTN: joint segmentation, registration and atlas construction with image-and-spatial transformer networks. *Medical Image Analysis* **78**: 102383.

Smerling, J., Marboe, C. C., Lefkowitch, J. H., Pavlicova, M., Bacha, E., Einstein, A. J., Naka, Y., Glickstein, J. and Farooqi, K. M. (2019). Utility of 3D printed cardiac models for medical student education in congenital heart disease: across a spectrum of disease severity. *Pediatric Cardiology* **40**: 1258–1265.

Smith, M., McGuinness, J., O'Reilly, M., Nolke, L., Murray, J. and Jones, J. (2017). The role of 3D printing in preoperative planning for heart transplantation in complex congenital heart disease. *Irish Journal of Medical Science (1971-)* **186**: 753–756.

So, K. C.-Y., Fan, Y., Sze, L., Kwok, K.-w., Chan, A. K.-y., Cheung, G. S.-H. and Lee, A. P.-W. (2017). Using multimaterial 3-dimensional printing for personalized planning of complex structural heart disease intervention. *JACC: Cardiovascular Interventions* **10**(11): e97–e98.

Sodian, R., Kruttschnitt, M., Hitschrich, N., Mumm, B., Schnell, C., Hagl, C., Thierfelder, N. and König, F. (2021). 3-dimensional printing for the diagnosis of left ventricular outflow tract obstruction after mitral valve replacement. *Interactive CardioVascular and Thoracic Surgery* **32**(5): 724–726.

Sodian, R., Schmauss, D., Markert, M., Weber, S., Nikolaou, K., Haeberle, S., Vogt, F., Vicol, C., Lueth, T., Reichart, B. et al. (2008a). Three-dimensional printing creates models for surgical planning of aortic valve replacement after previous coronary bypass grafting. *The Annals of thoracic surgery* **85**(6): 2105–2108.

Sodian, R., Schmauss, D., Schmitz, C., Bigdeli, A., Haeberle, S., Schmoeckel, M., Markert, M., Lueth, T., Freudenthal, F., Reichart, B. et al. (2009). 3-dimensional printing of models to create custom-made devices for coil embolization of an anastomotic leak after aortic arch replacement. *The Annals of thoracic surgery* **88**(3): 974–978.

Sodian, R., Weber, S., Markert, M., Loeff, M., Lueth, T., Weis, F. C., Daebritz, S., Malec, E., Schmitz, C. and Reichart, B. (2008b). Pediatric cardiac transplantation: three-dimensional printing of anatomic models for surgical planning of heart transplantation in patients with univentricular heart. *The Journal of Thoracic and Cardiovascular Surgery* **136**(4): 1098–1099.

Sodian, R., Weber, S., Markert, M., Rassoulian, D., Kaczmarek, I., Lueth, T. C., Reichart, B. and Daebritz, S. (2007). Stereolithographic models for surgical planning in congenital heart surgery. *The Annals of thoracic surgery* **83**(5): 1854–1857.

Sommer, K. N., Shepard, L. M., Mitsouras, D., Iyer, V., Angel, E., Wilson, M. F., Rybicki, F. J., Kumamaru, K. K., Sharma, U. C., Reddy, A. et al. (2020). Patient-specific 3D-printed coronary models based on coronary computed tomography angiography volumes to investigate flow conditions in coronary artery disease. *Biomedical Physics & Engineering Express* **6**(4): 045007.

Song, H., Zhou, Q., Zhang, L., Deng, Q., Wang, Y., Hu, B., Tan, T., Chen, J., Pan, Y. and He, F. (2017). Evaluating the morphology of the left atrial appendage by a transesophageal echocardiographic 3-dimensional printed model. *Medicine* **96**(38).

Sørensen, T. S., Beerbaum, P., Mosegaard, J., Rasmusson, A., Schaeffter, T., Austin, C., Razavi, R. and Greil, G. F. (2008). Virtual cardiotomy based on 3-D MRI for preoperative planning in congenital heart disease. *Pediatric radiology* **38**: 1314–1322.

Sorensen, T. S., Pedersen, E. M., Hansen, O. K. and Sorensen, K. (2003). Visualization of morphological details in congenitally malformed hearts: virtual three-dimensional reconstruction from magnetic resonance imaging. *Cardiology in the Young* **13**(5): 451–460.

Springenberg, J. T., Dosovitskiy, A., Brox, T. and Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *arXiv Preprint* (arXiv:1412.6806).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**(1): 1929–1958.

St. Louis, J. D., Introduction to Congenital Heart Disease: Nomenclature and Classification. In *Cardiovascular Pediatric Critical Illness and Injury*. Springer London London, 2008, 1–3.

Su, W., Xiao, Y., He, S., Huang, P. and Deng, X. (2018). Three-dimensional printing models in congenital heart disease education for medical students: a controlled comparative study. *BMC medical education* **18**(1): 1–6.

Subat, A., Goldberg, A., Demaria, S. and Katz, D. (2018). The utility of simulation in the management of patients with congenital heart disease: past, present, and future. *Seminars in Cardiothoracic and Vascular Anesthesia* **22**(1): 81–90.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. and Jorge Cardoso, M., Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*. Springer, 2017, 240–248.

Suinesiaputra, A., Frangi, A. F., Kaandorp, T. A., Lamb, H. J., Bax, J. J., Reiber, J. H. and Lelieveldt, B. P. (2009). Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. *IEEE Transactions on Medical Imaging* **28**(4): 595–607.

Sulaiman, A., Boussel, L., Taconnet, F., Serfaty, J. M., Alsaid, H., Attia, C., Huet, L. and Douek, P. (2008). In vitro non-rigid life-size model of aortic arch aneurysm for endovascular prosthesis assessment. *European journal of cardio-thoracic surgery* **33**(1): 53–57.

Sun, H., Frangi, A. F., Wang, H., Sukno, F. M., Tobon-Gomez, C. and Yushkevich, P. A., Automatic cardiac MRI segmentation using a biventricular deformable medial model. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part I 13*. Springer Berlin Heidelberg, 2010, 468–475.

Sun, L., Jia, K., Yeung, D.-Y. and Shi, B. E., Human action recognition using factorized spatio-temporal convolutional networks. In *Proceedings of the IEEE international conference on computer vision*. 2015, 4597–4605.

Sun, Z. (2020). Clinical applications of patient-specific 3D printed models in cardiovascular disease: current status and future directions. *Biomolecules* **10**(11): 1577.

Sun, Z., Lau, I., Wong, Y. H. and Yeong, C. H. (2019). Personalized three-dimensional printed models in congenital heart disease. *Journal of Clinical Medicine* **8**(4): 522.

Sundgaard, J. V., Juhl, K. A., Kofoed, K. F. and Paulsen, R. R., Multi-planar whole heart segmentation of 3D CT images using 2D spatial propagation CNN. In *Medical Imaging 2020: Image Processing*, volume 11313. SPIE, 2020, 477–484.

Sutskever, I., Martens, J., Dahl, G. and Hinton, G., On the importance of initialization and momentum in deep learning. In Dasgupta, S. and McAllester, D. (Editors), *Proceedings of the 30th International Conference on Machine Learning*, volume 28(3) of *Proceedings of Machine Learning Research*. Atlanta, Georgia, USA: PMLR, 2013, 1139–1147.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, 1–9.

Tavakoli, V. and Amini, A. A. (2013). A survey of shaped-based registration and segmentation techniques for cardiac images. *Computer Vision and Image Understanding* **117**(9): 966–989.

Thakkar, A. N., Chinnadurai, P., Breinholt, J. P. and Lin, C. H. (2018). Transcatheter closure of a sinus venosus atrial septal defect using 3D printing and image fusion guidance. *Catheterization and Cardiovascular Interventions* **92**(2): 353–357.

The SCOT-Heart Investigators (2015). CT coronary angiography in patients with suspected angina due to coronary heart disease (SCOT-HEART): an open-label, parallel-group, multicentre trial. *The Lancet* **385**(9985): 2383–2391.

Thiene, G. and Frescura, C. (2010). Anatomical and pathophysiological classification of congenital heart disease. *Cardiovascular Pathology* **19**(5): 259–274.

Tiwari, N., Ramamurthy, H. R., Kumar, V., Kumar, A., Dhanalakshmi, B. and Kumar, G. (2021). The role of three-dimensional printed cardiac models in the management of complex congenital heart diseases. *medical journal armed forces india* **77**(3): 322–330.

Tobon-Gomez, C., Geers, A. J., Peters, J., Weese, J., Pinto, K., Karim, R., Ammar, M., Daoudi, A., Margeta, J., Sandoval, Z. et al. (2015). Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *IEEE transactions on medical imaging* **34**(7): 1460–1473.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y. and Bregler, C., Efficient object localization using convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, 648–656.

Tóthová, K., Parisot, S., Lee, M. C., Puyol-Antón, E., Koch, L. M., King, A. P., Konukoglu, E. and Pollefeys, M., Uncertainty quantification in CNN-based surface prediction using shape priors. In *Shape in Medical Imaging: International Workshop, ShapeMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*. Springer, 2018, 300–310.

Tran, D., Bourdev, L., Fergus, R., Torresani, L. and Paluri, M., Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*. 2015, 4489–4497.

Triedman, J. K. and Newburger, J. W. (2016). Trends in congenital heart disease: the next decade. *Circulation* **133**(25): 2716–2733.

Tu, Z., Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2. IEEE, 2005, 1589–1596.

Tuncay, V. and van Ooijen, P. (2019). 3D printing for heart valve disease: a systematic review. *European radiology experimental* **3**(1): 1–10.

Tziritas, G., Fully-automatic segmentation of cardiac images using 3-D MRF model optimization and substructures tracking. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer International Publishing, 2017, 129–136.

Uccheddu, F., Carfagni, M., Governi, L., Furferi, R., Volpe, Y. and Nocerino, E. (2018). 3D printing of cardiac structures from medical images: an overview of methods and interactive tools. *International Journal on Interactive Design and Manufacturing (IJIDeM)* **12**: 597–609.

Uijlings, J. R., Van De Sande, K. E., Gevers, T. and Smeulders, A. W. (2013). Selective search for object recognition. *International journal of computer vision* **104**: 154–171.

Valverde, I. (2017). Three-dimensional printed cardiac models: applications in the field of medical education, cardiovascular surgery, and structural heart interventions. *Revista Española de Cardiología (English Edition)* **70**(4): 282–291.

Valverde, I., Gomez, G., Byrne, N., Anwar, S., Silva Cerpa, M. A., Martin Talavera, M., Pushparajah, K. and Velasco Forte, M. N. (2022). Criss-cross heart three-dimensional printed models in medical education: A multicenter study on their value as a supporting tool to conventional imaging. *Anatomical Sciences Education* **15**(4): 719–730.

Valverde, I., Gomez, G., Coserria, J. F., Suarez-Mejias, C., Uribe, S., Sotelo, J., Velasco, M. N., Santos De Soto, J., Hosseinpour, A.-R. and Gomez-Cia, T. (2015a). 3 D printed models for planning endovascular stenting in transverse aortic arch hypoplasia. *Catheterization and Cardiovascular Interventions* **85**(6): 1006–1012.

Valverde, I., Gomez, G., Gonzalez, A., Suarez-Mejias, C., Adsuar, A., Coserria, J. F., Uribe, S., Gomez-Cia, T. and Hosseinpour, A. R. (2015b). Three-dimensional patient-specific cardiac model for surgical planning in Nikaidoh procedure. *Cardiology in the Young* **25**(4): 698–704.

Valverde, I., Gomez-Ciriza, G., Hussain, T., Suarez-Mejias, C., Velasco-Forte, M. N., Byrne, N., Ordoñez, A., Gonzalez-Calle, A., Anderson, D., Hazekamp, M. G. et al. (2017a). Three-dimensional printed models for surgical planning of complex congenital heart defects: an international multicentre study. *European Journal of Cardio-Thoracic Surgery* **52**(6): 1139–1148.

Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Oliver, A. and Lladó, X. (2017b). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage* **155**: 159–168.

Van De Leemput, S. C., Prokop, M., van Ginneken, B. and Manniesing, R. (2019). Stacked bidirectional convolutional LSTMs for deriving 3D non-contrast CT from spatiotemporal 4D CT. *IEEE transactions on medical imaging* **39**(4): 985–996.

Van Der Bom, T., Zomer, A. C., Zwinderman, A. H., Meijboom, F. J., Bouma, B. J. and Mulder, B. J. (2011). The changing epidemiology of congenital heart disease. *Nature Reviews Cardiology* **8**(1): 50–60.

Van Der Linde, D., Konings, E. E., Slager, M. A., Witsenburg, M., Helbing, W. A., Takkenberg, J. J. and Roos-Hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *Journal of the American College of Cardiology* **58**(21): 2241–2247.

Vandaele, R., Nervo, G. A. and Gevaert, O. (2020). Topological image modification for object detection and topological image processing of skin lesions. *Scientific Reports* **10**(1): 21061.

Vargas, J., Spiotta, A. and Chatterjee, A. R. (2019). Initial experiences with artificial neural networks in the detection of computed tomography perfusion deficits. *World neurosurgery* **124**: e10–e16.

Velasco Forte, M. N., Byrne, N., Valverde, I., Gomez Ciriza, G., Hermuzi, A., Prachasilchai, P., Mainzer, G., Pushparajah, K., Henningsson, M., Hussain, T. et al. (2018). Interventional correction of sinus venosus atrial septal defect and partial anomalous pulmonary venous drainage: procedural planning using 3D printed models. *JACC: Cardiovascular Imaging* **11**(2 Part 1): 275–278.

Vercauteren, T., Pennec, X., Perchant, A. and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* **45**(1): S61–S72.

Viola, P. and Jones, M., Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1. Ieee, 2001, I–I.

Vosshenrich, J. and Breit, H.-C. (2021). Radiologist mouse movements at a PACS workstation. *Radiology* **299**(1): 52–52.

Vranicar, M., Gregory, W., Douglas, W. I., Di Sessa, P. and Di Sessa, T. G. (2008). The use of stereolithographic hand held models for evaluation of congenital anomalies of the great arteries. *Studies in health technology and informatics* **132**: 538.

Vukicevic, M., Mosadegh, B., Min, J. K. and Little, S. H. (2017a). Cardiac 3D printing and its future directions. *JACC: Cardiovascular Imaging* **10**(2): 171–184.

Vukicevic, M., Puperi, D. S., Jane Grande-Allen, K. and Little, S. H. (2017b). 3D printed modeling of the mitral valve for catheter-based structural interventions. *Annals of biomedical engineering* **45**: 508–519.

Wang, C. and Smedby, Ö., Automatic whole heart segmentation using deep learning and shape context. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018, 242–249.

Wang, C., Wang, Q. and Smedby, Ö., Automatic heart and vessel segmentation using random forests and a local phase guided level set method. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer, 2017a, 159–164.

Wang, C., Zhang, L., Qin, T., Xi, Z., Sun, L., Wu, H. and Li, D. (2020a). 3D printing in adult cardiovascular surgery and interventions: a systematic review. *Journal of Thoracic Disease* **12**(6): 3227.

Wang, D. D., Qian, Z., Vukicevic, M., Engelhardt, S., Kheradvar, A., Zhang, C., Little, S. H., Verjans, J., Comaniciu, D., O'Neill, W. W. et al. (2021a). 3D printing, computational modeling, and artificial intelligence for structural heart disease. *Cardiovascular Imaging* **14**(1): 41–60.

Wang, F., Liu, H., Samaras, D. and Chen, C., Topogan: A topology-aware generative adversarial network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 2020b, 118–136.

Wang, H., Song, H., Yang, Y., Cao, Q., Hu, Y., Chen, J., Guo, J., Wang, Y., Jia, D., Cao, S. et al. (2020c). Three-dimensional printing for cardiovascular diseases: From anatomical modeling to dynamic functionality. *BioMedical Engineering OnLine* **19**(1): 1–16.

Wang, J., Coles-Black, J., Matalanis, G. and Chuen, J. (2018). Innovations in cardiac surgery: techniques and applications of 3D printing. *Journal of 3D printing in medicine* **2**(4): 179–186.

Wang, W., Xia, Q., Hu, Z., Yan, Z., Li, Z., Wu, Y., Huang, N., Gao, Y., Metaxas, D. and Zhang, S. (2021b). Few-shot learning by a cascaded framework with shape-constrained pseudo label assessment for whole heart segmentation. *IEEE Transactions on Medical Imaging* **40**(10): 2629–2641.

Wang, Z., Liu, Y., Luo, H., Gao, C., Zhang, J. and Dai, Y. (2017b). Is a three-dimensional printing model better than a traditional cardiac model for medical education? A pilot randomized controlled study. *Acta Cardiologica Sinica* **33**(6): 664.

Warfield, S. K., Zou, K. H. and Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging* **23**(7): 903–921.

Weese, J., Kaus, M., Lorenz, C., Lobregt, S., Truyen, R. and Pekar, V., Shape constrained deformable models for 3D medical image segmentation. In *Information Processing in Medical Imaging: 17th International Conference, IPMI 2001 Davis, CA, USA, June 18–22, 2001 Proceedings 17*. Springer, 2001, 380–387.

Weiss, K., Khoshgoftaar, T. M. and Wang, D. (2016). A survey of transfer learning. *Journal of Big data* **3**(1): 1–40.

Wen, C.-Y., Yang, A.-S., Tseng, L.-Y. and Chai, J.-W. (2010). Investigation of pulsatile flowfield in healthy thoracic aorta models. *Annals of biomedical engineering* **38**: 391–402.

Wierzbicki, M., Moore, J., Drangova, M. and Peters, T. (2008). Subject-specific models for image-guided cardiac surgery. *Physics in Medicine & Biology* **53**(19): 5295.

Witschey, W. R., Pouch, A. M., McGarvey, J. R., Ikeuchi, K., Contijoch, F., Levack, M. M., Yushkevick, P. A., Sehgal, C. M., Jackson, B. M., Gorman, R. C. et al. (2014). Three-dimensional ultrasound-derived physical mitral valve modeling. *The Annals of thoracic surgery* **98**(2): 691–694.

Wolf, I., Vetter, M., Wegner, I., Böttger, T., Nolden, M., Schöbinger, M., Hastenteufel, M., Kunert, T. and Meinzer, H.-P. (2005). The medical imaging interaction toolkit. *Medical image analysis* **9**(6): 594–604.

Wolterink, J. M., Leiner, T., Viergever, M. A. and Išgum, I., Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers*. Springer International Publishing Cham, 2017, 95–102.

Wu, Y., Wang, Y. and Jia, Y. (2013). Segmentation of the left ventricle in cardiac cine MRI using a shape-constrained snake model. *Computer Vision and Image Understanding* **117**(9): 990–1003.

Xia, Q., Yao, Y., Hu, Z. and Hao, A., Automatic 3D atrial segmentation from GE-MRIs using volumetric fully convolutional networks. In *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*. Springer International Publishing, 2019, 211–220.

Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X. et al. (2021). A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical image analysis* **67**: 101832.

Xu, H., Niederer, S. A., Williams, S. E., Newby, D. E., Williams, M. C. and Young, A. A., Whole heart anatomical refinement from ccta using extrapolation and parcellation. In *Functional Imaging and Modeling of the Heart: 11th International Conference, FIMH 2021, Stanford, CA, USA, June 21-25, 2021, Proceedings*. Springer International Publishing Cham, 2021, 63–70.

Xu, X., Wang, T., Shi, Y., Yuan, H., Jia, Q., Huang, M. and Zhuang, J., Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019a, 477–485.

Xu, X., Wang, T., Zeng, D., Shi, Y., Jia, Q., Yuan, H., Huang, M. and Zhuang, J., Accurate congenital heart disease model generation for 3d printing. In *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*. IEEE, 2019b, 127–130.

Xu, X., Wang, T., Zhuang, J., Yuan, H., Huang, M., Cen, J., Jia, Q., Dong, Y. and Shi, Y., Imagechd: A 3d computed tomography image dataset for classification of congenital heart disease. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*. Springer, 2020, 77–87.

Xu, Z., Wu, Z. and Feng, J. (2018). CFUN: Combining faster R-CNN and U-net network for efficient whole heart segmentation. *arXiv Preprint* (arXiv:1812.04914).

Yang, G., Sun, C., Chen, Y., Tang, L., Shu, H. and Dillenseger, J.-l., Automatic whole heart segmentation in CT images based on multi-atlas image registration. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018a, 250–257.

Yang, X., Bian, C., Yu, L., Ni, D. and Heng, P.-A., Hybrid loss guided convolutional networks for whole heart parsing. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*. Springer, 2018b, 215–223.

Ye, C., Wang, W., Zhang, S. and Wang, K. (2019). Multi-depth fusion network for whole-heart CT image segmentation. *IEEE Access* **7**: 23421–23429.

Yıldız, O., Köse, B., Tanıdır, İ. C., Pekkan, K., Güzeltaş, A. and Haydin, S. (2021). Single-center experience with routine clinical use of 3D technologies in surgical planning for pediatric patients with complex congenital heart disease. *Diagnostic and Interventional Radiology* **27**(4): 488.

Yoo, J. S., Reddy, Y. N. and Kim, K.-H. (2020). Heart transplantation for dextrocardia: preoperative planning using 3D printing. *European Heart Journal-Cardiovascular Imaging* **21**(3): 346–346.

Yoo, S.-J., Spray, T., Austin III, E. H., Yun, T.-J. and van Arsdell, G. S. (2017). Hands-on surgical training of congenital heart surgery using 3-dimensional print models. *The Journal of thoracic and cardiovascular surgery* **153**(6): 1530–1540.

Yoo, S.-J., Thabit, O., Kim, E. K., Ide, H., Yim, D., Dragulescu, A., Seed, M., Grosse-Wortmann, L. and van Arsdell, G. (2016). 3D printing in medicine of congenital heart diseases. *3D Printing in Medicine* **2**: 1–12.

Yoo, S.-J. and Van Arsdell, G. S. (2018). 3D printing in surgical management of double outlet right ventricle. *Frontiers in pediatrics* **5**: 289.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv Preprint* (arXiv:1506.06579).

Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv Preprint* (arXiv:1511.07122).

Yu, L., Cheng, J.-Z., Dou, Q., Yang, X., Chen, H., Qin, J. and Heng, P.-A., Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20*. Springer International Publishing, 2017a, 287–295.

Yu, L., Yang, X., Qin, J. and Heng, P.-A., 3D FractalNet: dense volumetric segmentation for cardiovascular MRI volumes. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer International Publishing, 2017b, 103–110.

Yue, Q., Luo, X., Ye, Q., Xu, L. and Zhuang, X., Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019, 559–567.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C. and Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**(3): 1116–1128.

Zablah, J. E., Rodriguez, S. A., Lorenz, A. and Morgan, G. J. (2021). Cardiac catheterization laboratory and the role in effective patient education: A model approach. *Progress in Pediatric Cardiology* **63**: 101396.

Zeiler, M. D., Taylor, G. W. and Fergus, R., Adaptive deconvolutional networks for mid and high level feature learning. In *2011 international conference on computer vision*. IEEE, 2011, 2018–2025.

Zeng, D., Wu, Y., Hu, X., Xu, X., Yuan, H., Huang, M., Zhuang, J., Hu, J. and Shi, Y., Positional contrastive learning for volumetric medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*. Springer, 2021, 221–230.

Zhang, H., Wahle, A., Johnson, R. K., Scholz, T. D. and Sonka, M. (2009). 4-D cardiac MR image analysis: left and right ventricular morphology and function. *IEEE transactions on medical imaging* **29**(2): 350–364.

Zhang, J., Ma, W., Zhang, W. and Kong, Y. (2019a). Three-dimensional printed models–guided surgical repair for recurrent coronary artery fistula. *The Annals of Thoracic Surgery* **107**(3): e161–e163.

Zhang, Y., Ying, M. T. and Chen, D. Z., Decompose-and-integrate learning for multi-class segmentation in medical images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019b, 641–650.

Zhao, Y., Li, X., Zhang, W., Zhao, S., Makkie, M., Zhang, M., Li, Q. and Liu, T., Modeling 4d fmri data via spatio-temporal convolutional neural networks (st-cnn). In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11*. Springer, 2018, 181–189.

Zheng, H., Yang, L., Han, J., Zhang, Y., Liang, P., Zhao, Z., Wang, C. and Chen, D. Z., HFA-Net: 3D cardiovascular image segmentation with asymmetrical pooling and content-aware fusion. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer International Publishing, 2019a, 759–767.

Zheng, H., Zhang, Y., Yang, L., Liang, P., Zhao, Z., Wang, C. and Chen, D. Z., A new ensemble learning framework for 3D biomedical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33(01). 2019b, 5909–5916.

Zheng, H., Zhang, Y., Yang, L., Wang, C. and Chen, D. Z., An annotation sparsification strategy for 3D medical image segmentation via representative selection and self-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34(04). 2020, 6925–6932.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M. and Comaniciu, D., Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features. In *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, 1–8.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M. and Comaniciu, D. (2008). Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE transactions on medical imaging* **27**(11): 1668–1681.

Zhou, R., Liao, Z., Pan, T., Milgrom, S. A., Pinnix, C. C., Shi, A., Tang, L., Yang, J., Liu, Y., Gomez, D. et al. (2017). Cardiac atlas development and validation for automatic segmentation of cardiac substructures. *Radiotherapy and Oncology* **122**(1): 66–71.

Zhuang, X. (2013). Challenges and methodologies of fully automatic whole heart segmentation: a review. *Journal of healthcare engineering* **4**(3): 371–407.

Zhuang, X., Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer, 2016, 581–588.

Zhuang, X., Bai, W., Song, J., Zhan, S., Qian, X., Shi, W., Lian, Y. and Rueckert, D. (2015). Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection. *Medical physics* **42**(7): 3822–3833.

Zhuang, X., Leung, K., Rhode, K., Razavi, R., Hawkes, D. and Ourselin, S., Whole heart segmentation of cardiac MRI using multiple path propagation strategy. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part I 13*. Springer, 2010a, 435–443.

Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M. P., Oster, J., Wang, C., Smedby, Ö., Bian, C. et al. (2019). Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Medical image analysis* **58**: 101537.

Zhuang, X., Rhode, K. S., Razavi, R. S., Hawkes, D. J. and Ourselin, S. (2010b). A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE transactions on medical imaging* **29**(9): 1612–1625.

Zhuang, X. and Shen, J. (2016). Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Medical image analysis* **31**: 77–87.

Zhuang, X., Yao, C., Ma, Y., Hawkes, D., Penney, G. and Ourselin, S., Registration-based propagation for whole heart segmentation from compounded 3D echocardiography. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2010c, 1093–1096.

Zotti, C., Luo, Z., Lalande, A. and Jodoin, P.-M. (2018). Convolutional neural network with shape prior applied to cardiac MRI segmentation. *IEEE journal of biomedical and health informatics* **23**(3): 1119–1128.

Zuluaga, M. A., Biffi, B., Taylor, A. M., Schievano, S., Vercauteren, T. and Ourselin, S., Strengths and pitfalls of whole-heart atlas-based segmentation in congenital heart disease patients. In *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*. Springer International Publishing, 2017, 139–146.

Zuluaga, M. A., Burgos, N., Mendelson, A. F., Taylor, A. M. and Ourselin, S. (2015). Voxelwise atlas rating for computer assisted diagnosis: Application to congenital heart diseases of the great arteries. *Medical image analysis* **26**(1): 185–194.

Zuluaga, M. A., Cardoso, M. J., Modat, M. and Ourselin, S., Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion. In *Functional Imaging and Modeling of the Heart: 7th International Conference, FIMH 2013, London, UK, June 20-22, 2013. Proceedings 7*. Springer, 2013, 174–181.