Simulation and Synthesis for Cardiac Magnetic Resonance Image Analysis

Simulation and Synthesis for Cardiac Magnetic Resonance Image Analysis

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op donderdag 20 april 2023 om 13:30 uur

door

Sina Amirrajab

geboren te Dezful, Iran

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. Keita Ito
1^e promotor:	prof.dr.ir. Marcel Breeuwer
2^e promotor:	prof.dr. Josien Pluim
leden:	prof.dr. Amedeo Chiribiri
	(KCL St Thomas' Hospital, London, UK)
	prof.dr. Ivana Isgum
	(Amsterdam UMC – UvA, Amsterdam, NL)
	prof.dr.ir. Maarten Paulides
	prof.dr. Tony Stöcker
	(DZNE, Bonn, DE)
adviseur:	dr. Jürgen Weese
	(Philips Research Hamburg, Hamburg, DE)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Simulation and Synthesis for Cardiac Magnetic Resonance Image Analysis

Sina Amirrajab

This research is a part of the openGTN project, supported by the European Union in the Marie Curie Innovative Training Networks (ITN) fellowship program under project No. 764465.

Cover design by DALL · E2 using the prompt: “An impressive black pencil painting of many generated cardiac magnetic resonance images with dark background”.

Printed by ADC DEREUMAUX.

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-5706-6

To the person who has everything I don't have
and has nothing I don't need!

Contents

CHAPTER 1

Introduction

1.1 Simulation and Synthesis for Cardiac MRI

Cardiovascular magnetic resonance (CMR) imaging has become an essential tool in cardiology as a non-invasive modality and the gold standard technique to assess heart anatomy, regional and global function, as well as myocardial viability. Functional analysis of the heart through calculation of the left and right ventricular volumes, ejection fraction, and stroke volumes requires accurate delineation of the heart structures at the end-diastolic and end-systolic phases. Due to the limited accuracy of fully-automated cardiac image segmentation methods, semi-automated approaches are still preferred in clinical practice. Applying state-of-the-art deep learning (DL) methods to the field of medical image analysis faces several challenges and issues regarding the availability of high-quality medical data. A limited number of CMR images with ground truth labels hampers the development of data-hungry supervised deep convolutional neural networks (CNNs) for automated image segmentation.

In this thesis, we develop and investigate different approaches to generate a substantial number of realistic CMR images with corresponding ground truth labels that can be utilized for training supervised CNNs for the task of cardiac tissue segmentation. In particular, we present solutions for two main categories of image generation, namely physics-driven image simulation and data-driven image synthesis. The former is grounded on the underlying physics of MR image formation based on the so-called Bloch equations to simulate data, whereas the latter applies recent methods based on conditional deep generative models to synthesize data.

We refer to the terms image **Simulation** and image **Synthesis** many times throughout this thesis. In simulation, which refers to mechanistic (hypothesis-driven) models, we commonly employ first principle formulations for image generation while in synthesis, which refers to phenomenological (data-driven) models we begin with available data [1].

Image simulation is considered the modeling of the process of magnetic resonance image formation using the underlying physics of imaging governed by so called Bloch equations. Three main ingredients are involved in the process of MR image simulation; 1) computerized anatomical model to provide human anatomy with sufficient detail of organs and structures of interest, i.e. the heart and surrounding organs visible in the cardiac MR image field-of-view (FOV), 2) MR related tissue properties such as longitudinal relaxation time (T1), transversal relaxation time (T2) and proton density (PD) for each tissue and organ in the FOV of simulation, and 3) mathematical equations and operations for the specific imaging sequence including acquisition parameters such as repetition (TR), echo time (TE), radio frequency (RF) flip angle (FA), matrix size and the spatial resolution of the imaging.

Image synthesis involves learning imaging features and anatomical characteristics from abundant data and in turn generating new images. Deep generative models such as generative adversarial networks (GANs) [2] and variational auto-encoder (VAEs) [3] are two well-known deep learning models for learning probability density function of the data. We utilize such algorithms to learn the generating factors of the training data and synthesize new examples with corresponding ground truth labels.

1.2 Summary of this Thesis

The scope of the research mainly concerns with generating training data for deep learning application and investigating the benefit of that for the clinical application of cardiac MR image segmentation, among other applications which remained unexplored. Two main methods of physics-based image simulation and data-driven image synthesis are proposed and evaluated for cardiac image analysis throughout this thesis. Each chapter includes both methodological developments and specific usability evaluation of the results.

In Chapter 2, we develop a flexible framework for simulating CMR images with variable anatomical and imaging characteristics for the purpose of creating a diversified virtual population. We advance previous works on both cardiac MR image simulation and anatomical modeling to increase the realism in terms of both image appearance and underlying anatomy. A database of virtual subjects is simulated and its usefulness for aiding a DL segmentation method is evaluated. Our experiments show that training completely with simulated images achieves comparable performance to a model trained with real images for heart cavity segmentation in mid-ventricular slices. Moreover, we demonstrate that such simulated data can be used in addition to classical augmentation for boosting the performance of the augmented model, and in a scenario that only 45% of the real data is available, the baseline performance (trained with all real images) is retained when the simulated data is added during training.

Next, in Chapter 3 we propose a novel framework consisting of image segmentation and synthesis based on mask-conditional generative adversarial networks (GANs) for generating high-fidelity and diverse CMR images with variable anatomical representation. Extensive experiments are conducted to analyze the importance of different modules in the framework for synthesizing highly-realistic images. Leveraging label-conditioned normalization layers throughout the generator architecture allows for the preservation of content information, while at the same time accurately transferring the image characteristics of real data. Furthermore, one of the main findings of this work is the importance of introducing detailed labels in the form of multi-tissue maps for generation of highly realistic images with accurate anatomies even with a significantly smaller number of training data, compared to utilizing only cavity-tissue labels during training. Furthermore, we evaluate the effectiveness and usability of the synthetic data in the scenarios of data replacement and augmentation for training a segmentation network. The results of the replacement study indicate that segmentation models trained with only synthetic data can achieve comparable performance to the baseline model trained with real data, indicating that the synthetic data captures the essential characteristics of its real counterpart. Furthermore, augmenting real with synthetic data during training can significantly improve both the Dice score (maximum increase of 4%) and Hausdorff Distance (maximum reduction of 40%) for cavity segmentation, suggesting a good potential to aid in tackling medical data scarcity.

Chapter 4 provides solutions to the challenge of generating plausible heart geometries for synthesizing a CMR image database, including virtual subjects with characteristics of a particular heart pathology. We propose to break down the task of CMR image synthesis into 1) learning the deformation of anatomical content of the ground truth (GT) labels using variational auto-encoders and 2) translating GT labels to realistic images using conditional GANs. We devise different strategies to deform labels in the latent space of the VAE and generate various virtual subjects via three difference approaches, namely i) intra-subject synthesis to improve the through-plane resolution and generate intermediate short-axis slices within a given subject, ii) inter-subject synthesis to generate intermediate heart geometries and appearance between two dissimilar subjects scanned using two different scanner vendors, and iii) pathology synthesis to generate virtual subjects with a target heart disease that affects the heart geometry, e.g. synthesizing a pseudo-pathological subject with thickened myocardium for hypertrophic cardiomyopathy. All mentioned approaches are accomplished via manipulation and interpolation in the latent space of our VAE model trained on GT labels. Furthermore, we propose to model the relationship between 2D slices in the latent space of the VAE through estimating the correlation coefficient matrix between the latent vectors and utilizing it to correlate elements of randomly drawn samples before decoding to image space. This simple yet effective approach results in generating 3D consistent subjects from 2D slice-by-slice generations. We demonstrated that our approach could provide a solution to diversify and enrich an available database of cardiac MR images, resulting in significant improvements in model performance for cardiac segmentation of subjects with unseen heart diseases.

In Chapter 5 we investigate an application of the late-gadolinium enhanced CMR image synthesis for improving the model accuracy and robustness for automated myocardial scar quantification. We generate new pairs of synthetic LGE images by applying morphological operations including elastic deformation, dilation and rotation on the scar geometry. We demonstrate that data augmentation using synthesized LGE images with variable scar shapes further improves the performance of the scar segmentation and quantification.

Another application of the image synthesis for improving model robustness to multi-vendor, multi-center, multi-view and multi-disease CMR images is presented in Chapter 6. To augment and balance out the real data, we utilize conditional

GANs, developed in previous chapters, to generate a large number of realistic-looking images with corresponding labels including variations in contrast, heart appearance and pathology. In particular, we try to identify outlier cases based on the calculated heart volume and synthesize more anatomical variations on them to provide a balanced synthetic data. Extensive experiments and analysis demonstrate the ability of the proposed approach to significantly reduce the number of outliers during segmentation and adapt better to unseen and difficult examples. We show that synthesizing a diverse training dataset, carefully designed to increase the variation of cardiac shapes and appearances during training, plays a significant role in not only boosting the model performance in terms of standard quantitative metrics, but also in improving the automatically derived clinical metrics denoting the function of the heart. This in turn leads to improved stability and reliability of the predictions across both short-axis and long-axis images.

Chapter 7 is a short feasibility study with the purpose of reconciling the two worlds of simulation and synthesis. We present a sim2real translation network to reduce the realism gap between simulated and real data using GANs for unpaired/unsupervised style transfer.

Finally, Chapter 8 concludes this thesis with a discussion of the important findings and an outlook for future research directions.

In short, various studies presented in this thesis demonstrate the benefits of simulating and synthesizing cardiac magnetic resonance images with corresponding labels for data augmentation for boosting the performance and tackling medical data scarcity in the context of developing supervised deep-learning medical image analysis algorithms.

CHAPTER 2

A Framework for Image Simulation

This chapter is based on:

S. Amirrajab, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “A Framework for Simulating Cardiac MR Images with Varying Anatomy and Contrast,” IEEE Transactions on Medical Imaging, 2022.

Which is an extension of:

S. Amirrajab, W. P. Segars, C. Lorenz, J. Weese, M. Breeuwer, “Towards Realistic Cardiac MR Image Simulation; Inclusion of the Endocardial Trabeculae in the XCAT Heart Anatomy,” ISMRM 2020.

S. Amirrajab, Y. Al Khalil, C. Lorenz, J. Weese, M. Breeuwer, “Generation of Realistic and Heterogeneous Virtual Population of Cardiovascular Magnetic Resonance Simulated Images,” ISMRM 2020.

Y. Al Khalil, S. Amirrajab, C. Lorenz, J. Weese, M. Breeuwer, “Heterogeneous Virtual Population of Simulated CMR Images for Improving the Generalization of Cardiac Segmentation Algorithms,” SASHIMI workshop MICCAI 2020.

Abstract

One of the limiting factors for the development and adoption of novel deep-learning (DL) based medical image analysis methods is the scarcity of labeled medical images. Medical image simulation and synthesis can provide solutions by generating ample training data with corresponding ground truth labels. Despite recent advances, generated images demonstrate limited realism and diversity. In this work, we develop a flexible framework for simulating cardiac magnetic resonance (MR) images with variable anatomical and imaging characteristics for the purpose of creating a diversified virtual population. We advance previous works on both cardiac MR image simulation and anatomical modeling to increase the realism in terms of both image appearance and underlying anatomy. To diversify the generated images, we define parameters: 1) to alter the anatomy, 2) to assign MR tissue properties to various tissue types, and 3) to manipulate the image contrast via acquisition parameters. The proposed framework is optimized to generate a substantial number of cardiac MR images with ground truth labels suitable for downstream supervised tasks. A database of virtual subjects is simulated and its usefulness for aiding a DL segmentation method is evaluated. Our experiments show that training completely with simulated images can perform comparable with a model trained with real images for heart cavity segmentation in mid-ventricular slices. Moreover, such data can be used in addition to classical augmentation for boosting the performance when training data is limited, particularly by increasing the contrast and anatomical variation, leading to better regularization and generalization. The database is publicly available at `https://osf.io/bkzhm/` and the simulation code will be available at `https://github.com/sinaamirrajab/CMRI_Simulation` .

2.1 Introduction

Simulation and synthesis have recently received great recognition in the medical imaging community. This has been achieved thanks to the solutions that image generation approaches can provide to medical data challenges such as data scarcity, privacy, expert dependency, and expensive collection procedure. The synergies between different approaches, various applications, challenges, and opportunities have been highlighted in the recent editorial by Frangi *et al.* [1]. With developments in machine learning methods, there is an ever-growing demand for a large heterogeneous medical database that represents enough variability in both anatomical representation and image appearances. Such a diverse database can pave the way for developing, validating, and benchmarking accurate and robust medical image analysis methods that can be employed in routine clinical practice.

In attempts towards generating realistic Cardiac Magnetic Resonance (CMR) images, there has been recent progress in three categories: i) physics-driven image simulation; ii) data-driven image synthesis; and iii) hybrid image generation. In the physics-driven simulation approach, the underlying physics for MR contrast formation, governed by Bloch equations, is implemented and a computerized anatomical model is used to resemble a virtual patient. MR relevant tissue properties are assigned to each label class in the anatomical model to be fed to an image simulator to produce image contrast. The second category uses existing cardiac MR image data to train a generative model that learns the appearance of images and in turn synthesizes similar-looking images. In the final category, the hybrid approach combines patient image data with a biophysical model of the heart to generate images with altered geometry that is informed by mechanical motion parameters. For the purpose of generating diversified images with variable contrast and anatomy, each approach has its distinct advantages and disadvantages.

2.1.1 Physics-Driven Image Simulation

Image simulation is performed by combining a spatio-temporal model representing anatomy of interest and a simulator that encompasses the physics of image formation given a set of controllable parameters. In this category, the Virtual Imaging Platform provides an integrated open-access platform for sharing object models and multi-modality medical image simulation pipelines [4]. However, it consists of only one MR sequence with pre-defined scan parameters on one anatomical model with a simplistic heart geometry and a limited number of surrounding anatomical structures. Among one of the first attempts to generate CMR image data, Tobon-Gomez *et al.* [5] investigate physics-based image simulation using the MRISIM simulator developed by Kwan *et al.* in [6] and anatomical models generated from eXtendec CArdiac and Torso (XCAT) phantoms [7]. The authors put effort into making the simulated images more realistic by modeling the left ventricular papillary muscles and trabeculation. They use real patient data to create a simplified mathematical model of cylindrical objects for the papillary muscles and small discs at random regions adjacent to the ventricular wall for heart trabeculation. More recently, based on the original XCAT anatomical phantom, Wissmann *et al.* in [8] developed a numerical simulation framework suitable for optimizing CMR sampling trajectory, post-processing methods, and reconstruction strategies in the presence of beating and breathing motions, which is referred to as MRXCAT. Despite the modifications to the anatomical model and the simulation approach, the resulting simulated images for both approaches are still far from reality in terms of the detailed structures of the used heart model, the number of neighboring organs visible in the field of view, and the realism of the image contrast and resolution.

2.1.2 Data-Driven Image Synthesis

With the advent of generative modelling and the emergence of various techniques for synthesizing images using available clinical data, new methods have been adopted by researchers in the medical imaging community [9, 10]. In particular, Generative Adversarial Networks (GANs) [2] are at the center of most recently proposed models. To provide labeled data for training a segmentation model, some works proposed multi-modal style transfer method for transferring the image appearance (known as style) of the real images to the anatomical information (known as content) of another imaging modality where the content of the synthetic data comes from the heart annotations of cine CMR images in [11] and [12], or CT cardiac images in [13]. While these methods are able to generate realistic-looking images, they allow limited control over the image synthesis procedure, meaning that neither the underlying anatomy content, nor the local tissue intensity and global image contrast of the generated images can be controlled. Disentangling the anatomy factors from the modality features, Joyce *et al.* in [14] designed separate variational auto encoder (VAE) models [3] for simultaneously learning multi-tissue anatomical model, a deformation model, and an image intensity rendering model. Factorizing the information in the data in this way can provide partial control over generating variable anatomy and overall image appearance. However, the synthetic data may not necessarily represent an accurate anatomy that is required for the downstream supervised tasks. Furthermore, the contrast is not controllable locally and the generated tissue intensity is not based on the physics of MR signal evolution.

2.1.3 Hybrid Image Generation

Combining a biophysical model of the heart with a set of real clinical images in a hybrid approach, Prakosa *et al.* in [15] propose to use a registration algorithm for generating realistic-looking images with controllable cardiac motion. After fusing a surface model of the heart into real cardiac images, the heart geometry is adapted according to a predefined motion model. Along the same direction, in recent work by Duchateau *et al.* [16], an optimized pipeline is proposed for reducing the registration errors and improving the model-to-image adaptation, therefore generating more realistic images. Similarly, Zhou *et al.* in [17] introduce a multi-modality pipeline to generate cine CMR images, tagged CMR, and echocardiography sequences from the same virtual patient. With the aim of augmenting data, similar work was done by Acero *et al.* in [18], who use a heart statistical model of deformation to generate similar looking examples of images from a cardiac MRI database through altering the anatomy. While the generated images using the above-mentioned approaches are realistic in terms of the underlying anatomy, they depend on the availability of real cardiac images. Combining the controllable anatomical model with the imaging features has gained more attention in recent years. Abbasi-Sureshjani *et al.* in [19] and Amirrajab *et al.* in [20] propose to integrate the anatomical information of the XCAT phantoms [7] with modality-specific appearance of real data to synthesize data for creating a virtual database of realistic CMR images with ground truth labels. Although the anatomical variability can be created using the heart model, new image appearances can not be generated. Furthermore, the gray values of the images are not governed by the underlying physics of MRI and controlling image contrast at the tissue level is not yet feasible.

2.1.4 Motivation and Contribution

In this work, we develop a flexible framework suitable for generating a database of heterogeneous cardiac MR images that present variations in acquisition parameters, tissue properties, image contrasts, and anatomical representation. The proposed framework is tailored towards generating a plethora of realistic-looking images with corresponding ground truth labels. We build upon and advance previous works in both areas of anatomical modeling and cardiac MR image simulation. The simulation pipeline consists of three main elements: i) a parameterized anatomical model based on an improved version of the XCAT phantoms; ii) a set of controllable tissue parameters for more than 10 tissue types within various organs; iii) an optimized CMR simulation model to generate images with variable contrast. We save the output of the simulation together with multi-tissue ground truth labels in the NIfTi file format with proper metadata of acquisition parameters.

The main contributions of this chapter can be summarized as follows:

- We enhance the heart model by adding patient-specific detailed structures for the trabeculation anatomy of the left and right ventricles. Moreover, we make use of available anatomical parameters in the XCAT phantom to create virtual subjects with variable organ size, geometry, volume and location.

- We increase the realism of the CMR image simulation by assigning numerous tissue properties to various organs that are visible in the imaging field of view. We utilize physics-based analytical solutions for fast MR contrast computation with controllable acquisition parameters.

- We create a population of virtual patients (25-30 cardiac phases each) with different anatomical characteristics for cine cardiac MR image simulation for functional analysis and make the database with ground truth labels publicly available to the medical imaging research community upon request via `https://osf.io/bkzhm/`

An initial database of CMR images was simulated on virtual patients using the early version of our framework for the recent work of Al Khalil *et al.* in [21]. We showed that such a heterogeneous database of images with variation in both anatomy and appearance can be used for pre-training a deep learning cardiac segmentation model that can generalize better to the variability of real cardiac data and experiments showed that similar performance can be achieved when replacing up to 80% of the real data with simulated data. The initial results indicate the usefulness of a simulated database of CMR images for transfer learning in medical imaging. In this work, we made substantial improvements to our simulation framework by 1) optimizing sequence parameters and image contrast, 2) improving the anatomical model by incorporating patient-specific details of the heart trabeculation, 3) improving the image realism by increasing the number of tissue properties used for simulation, and 4) by modeling the partial volume effects using sampling and filtering in the k-space. Here, we present our framework for cardiac MR image simulation and provide detailed explanation of each module. In addition, we evaluate the usefulness of simulated data in the context of data augmentation for training a neural network for cardiac cavity segmentation. We show that with improved image realism we can directly add the newly simulated data to the real image and demonstrate the benefits of data augmentation and data replacement using the simulated images in this study, which was not possible before due to limited image realism.

The structure of the chapter is as follows. We give a brief overview of the XCAT anatomical phantom in section 2.2.1, the explanation of our approach for incorporating more anatomical details into the XCAT heart model in section 2.2.1, the introduction of the anatomical parameters for creating virtual patients in section 2.2.1, the description of steps involved for cardiac MR simulation in section 2.2.2, the experimental design for evaluating the generated data in a deep-learning setup in section 2.2.3, and qualitative and quantitative analysis of the results in sections 2.3.1 and 2.3.2, followed by discussion and conclusion in 2.4 and 2.5, respectively.

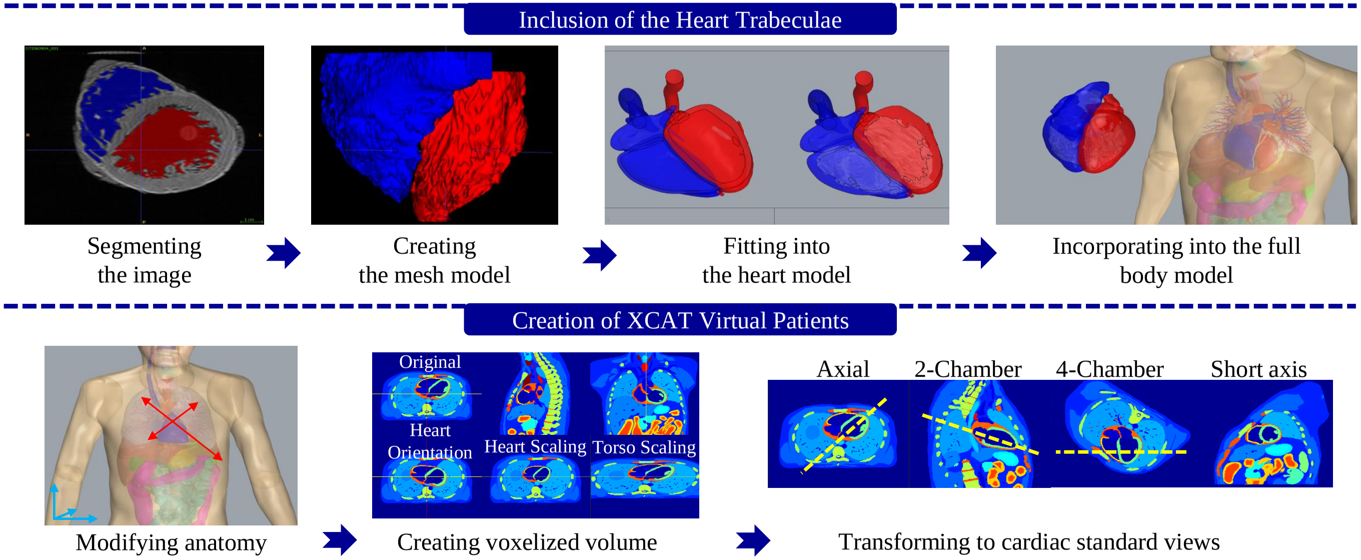

Inclusion of the Heart Trabeculae
Segmenting the image
Creating the mesh model
Fitting into the heart model
Incorporating into the full body model
Creation of XCAT Virtual Patients
Modifying anatomy
Original
Heart Orientation
Heart Scaling
Torso Scaling
Creating voxelized volume
Axial
2-Chamber
4-Chamber
Short axis
Transforming to cardiac standard views

Figure 2.1: Inclusion of the trabeculation structures into the XCAT phantom and creation of virtual patients.

2.2 Proposed Framework

2.2.1 Virtual Subjects

The XCAT Phantom

The 4D eXtended CArdiac and Torso phantom (XCAT) [7] is used as the basis for creating virtual patients for image simulation. The anatomies in the XCAT phantom are based on segmented patient data, modelled as Non-Uniform Rational B-Splines (NURBS) surfaces to accurately capture each structure in the body. Defined anatomies using NURBS surfaces provide a realistic representation of a patient with great flexibility to alter and create models with geometrical deformations. For accurately modeling the regional myocardial contracting and twisting motion of the left and right ventricles, the cardiac motion model was improved using the analysis of tagged CMR images [22]. The analysis of the tagged MRI data resulted in a comprehensive motion model that can produce accurate heart geometries at any given time point during the cardiac cycle and represent the complex geometrical deformation of the beating heart. Note that the motion of the XCAT heart includes the shortening along the long axis of the heart that accounts for the through-plane motion during image simulation.

These features make the XCAT a suitable source for creating times series of varying models for simulating the dynamics of the heart. However, the anatomical details of the trabeculae muscles are missing in the current version of the XCAT phantom which hampers both the realism of the image simulation and the performance of image-based assessment of cardiac function.

Inclusion of Myocardial Trabeculae in the XCAT Heart

Previous studies have explored the importance of the papillary muscles and trabeculae anatomy in analyzing cardiac images. Their inclusion in the left ventricular cavity or left myocardial volumes can have a considerable impact on the final assessment of the cardiac function [23]. Therefore, from a simulation perspective, the construction of a detailed heart model that comprises the mentioned substructures is crucial for quantitative analysis of the images. To model the geometry of the trabeculae of the myocardium, we utilized open access ex-vivo high-resolution 3D MRI data of normal human heart with $256 \times 256 \times 136$ matrix size and $0.4297 \times 0.4297 \times 1\ mm^3$ voxel dimensions [24]. As shown in Figure 2.1, the irregularity of the trabeculae muscular geometry was accurately segmented from the images.

The spatial patterns of the tiny structures of the trabeculae anatomy were accurately captured by manually segmenting the right and left ventricle of the heart slice-by-slice using ITK-SNAP software [25]. It was then converted to a 3D polygon mesh model, while preserving the details of the jagged-like structures, and transformed into the inner surfaces of the XCAT heart chambers. The alignment was done for the end-diastolic phase of the heart and the motion model was applied to the trabeculae mesh model to be altered smoothly during the cardiac cycle. These steps can be seen in Figure 2.1.

Anatomical Parameters

To provide large quantities of varying anatomies in CMR image simulation, we use the pre-defined XCAT parameters to create new subjects. We assign a specific set of anatomical parameters for body size and heart geometry and location with respect to neighboring organs. These parameters include body scaling in different dimensions, orientation angles and translation of the heart within the torso, and LV volumes at end diastole and end systole.

As shown in Figure 2.1 the parameters of the anatomy are modified to create a new virtual patient. The XCAT program outputs a three-dimensional volume of binary labels, shown with different colors, for a desired body area that is defined by field of view, matrix size, and the voxel resolution. The voxelized XCAT anatomy is shown for three orthogonal planes (axial, sagittal and coronal), as well as three examples for changing the anatomical parameters of the axial view. We create an isotropic 4D (3D + time) model for each virtual subject. The slices of the volume are presented in the axial view, while the standard views for assessment of cardiac function using MRI can be different. Since it is common to scan in short-axis view of the heart, rotation and re-slicing is performed on the axial slices following the recommendations provided in the CMR pocket guide [26]. The rotation angles are obtained given the spatial location of the heart within the torso provided by the XCAT metadata information. The generated subjects are used as the input for our CMR image simulation pipeline described in the following section.

2.2.2 Cardiac MR Image Simulation

An overview of the pipeline for simulating CMR images is provided in Figure 2.2. Firstly, MR properties are assigned to each tissue label in the high-resolution voxelized XCAT anatomy (a). Given these parameters and the imaging settings, image contrast is computed using the analytical MR signal model of the desired pulse sequence (b). In c), the simulated contrast data is transformed to high-resolution k-space data by applying a Fourier transformation followed by a sampling operation (e.g. Cartesian sampling grid) to map the k-space data to a coarser grid, given acquisition resolution. Thereafter in d), complex noise is calculated based on the chosen signal-to-noise ratio to be added to the real and imaginary parts of the complex-valued k-space data. Finally the reconstruction operation is carried out to create the final image in e). For dynamic imaging of the heart motion, the same operations from a-e are performed for each time frame of the voxelized anatomy that includes changes in the heart geometry for one cardiac cycle. The reconstructed 4D image data is saved in the NIfTi file format together with the corresponding ground truth labels. In the following section the individual steps of the pipeline are described in more detail.

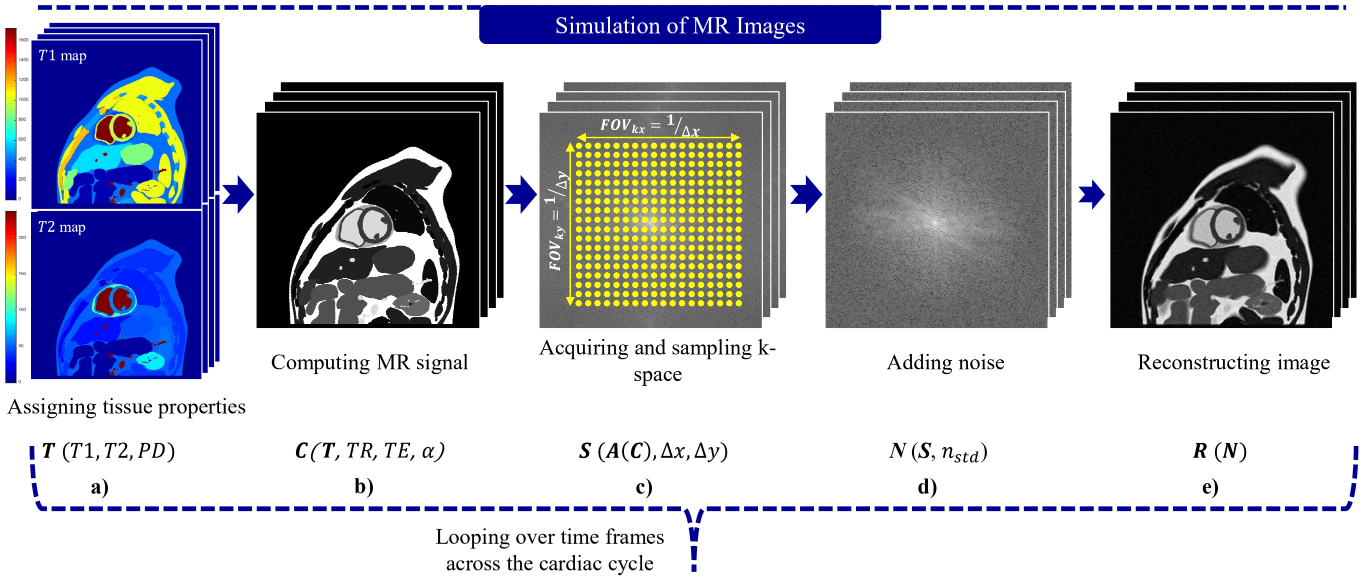

Simulation of MR Images
$T1$ map
1600
1400
1200
1000
800
600
400
200
0
$T2$ map
200
150
100
50
0
Assigning tissue properties
Computing MR signal
$FOV_{kx} = {}^{1}/_{\Delta x}$
$FOV_{ky} = {}^{1}/_{\Delta y}$
Acquiring and sampling k-space
Adding noise
Reconstructing image
$\boldsymbol{T}\,(T1, T2, PD)$
a)
$\boldsymbol{C}(\boldsymbol{T}, TR, TE, \alpha)$
b)
$\boldsymbol{S}\,(\boldsymbol{A}(\boldsymbol{C}), \Delta x, \Delta y)$
c)
$\boldsymbol{N}\,(\boldsymbol{S}, n_{std})$
d)
$\boldsymbol{R}\,(\boldsymbol{N})$
e)
Looping over time frames across the cardiac cycle

Figure 2.2: Simulation pipeline of the cardiac MR images consisting of a) module for assigning MR tissue properties, b) for computing image contrast, c) for data acquisition and sampling k-space, d) for noise addition and e) for image reconstruction.

MR Tissue Properties

The voxelized anatomical models derived from the previous section needs to be complemented with MR tissue properties ($\boldsymbol{T}$) in Figure 2.2 a). It is important to not only assign tissue properties to the heart tissue and blood, but also to the surrounding organs. By adding tissue properties to the organs in the field of view, we expect the simulated images to be a more realistic representation of real images. The T1 and T2 relaxation times are obtained from literature and summarized in Table 2.1. We found that there is a substantial diversity in the normal range of values reported in the literature. Therefore, we combine the statistics to derive one value for mean and one value for standard deviation for each tissue type and use these values to generate random numbers for T1 and T2 relaxation times for each specific virtual patient.

Table 2.1: MR tissue property statistics at 1.5 Tesla used for simulation.

Tissue type	T1 (Mean ± Std) ms	T2 (Mean ± Std) ms
Myocardium	$(977 \pm 42)^{*}(1198.7 \pm 30.3)^{**}$	(55 ± 4)
Blood	(1700 ± 63)	(237 ± 50)
Liver	(581 ± 35)	(48 ± 7)
Kidney	(1080 ± 42)	(86 ± 5)
Spleen	(1057 ± 42)	(79 ± 15)
Body fat	(338 ± 27)	(11 ± 7)
Cartilage	(1168 ± 18)	(27 ± 3)
Skeletal muscle	(1034 ± 87)	(39 ± 5)
Bone	(549 ± 52)	(49 ± 8)
Lung	(1000 ± 82)	(40 ± 8)
Stomach	(765 ± 75)	(58 ± 24)
Intestine	(343 ± 37)	(58 ± 4)
References	[27–35] * inversion-recovery and ** saturation-recovery sequences	

MR Contrast

Image contrast is one of the most important features of any imaging sequence in MRI. It ultimately depends on the selection of the acquisition parameters such as repetition time (TR), echo time (TE), and flip angle (α). The balanced steady-state free precession (bSSFP) sequence has become the most widely used clinical sequence for the cardiac functional assessment using the cine CMR because of producing high contrast between blood and myocardium. We therefore use its analytical solution to, given all parameters, compute the CMR image contrast (Figure 2.2 b). The bSSFP signal ($\boldsymbol{C}$) exhibits a relatively complicated contrast composed of T1 and T2 relaxations that in the absence of any off-resonance is given as follows [36]:

$$\boldsymbol{C} = \frac{PD_0\sqrt{E_2(1-E_1)}sin\alpha}{1-(E_1-E_2)cos\alpha - E_1E_2} \tag{2.1}$$

Where $E_{1,2} = \exp(-\frac{TR}{T_{1,2}})$ and the echo time in the balanced sequence is set to $TE = \frac{TR}{2}$ that is represented as an extra weighting of the signal ($\sqrt{E_2}$).

Note that the simulated MR signal in Equation 2.1 contains the magnitude information under the assumption that the transversal magnetization will not dephase between RF pulses in the ideal cases hence there will be zero or negligible phase accumulation. In the presence of magnetic field inhomogeneity, RF coil sensitivity, tissue susceptibility, motion, diffusion and any other sources of off-resonance, the signal formula should be modified to account for simulation of the phase information [37].

K-space Acquisition and Sampling

MR contrast computed on the high-resolution (HR) input model is transformed to HR complex k-space data as follows:

$$\boldsymbol{A} = \mathcal{F}(\boldsymbol{C}) \tag{2.2}$$

Here $\mathcal{F}(.)$ is the Fast Fourier Transformation operator for transferring the HR simulated contrast to the HR simulated k-space data that has complex values ($\boldsymbol{A}$). This data is then sampled to a desired (lower) resolution matrix using a sampling operator ($\boldsymbol{S}$) given the in-plane ($\Delta x, \Delta y$) acquisition resolution.

The field of view of the simulation is dictated by the size of the input voxelized model. That causes the space between the sampling points at different spatial frequencies to be fixed, given the equation $\Delta kx, y = 1/FOV_{x,y}$. The maximum spatial frequency is therefore derived based on the acquisition resolution given $FOV_{kx,ky} = 1/\Delta x, y$. This is the cut-off for the spatial frequency defining the extent to which we sample the k-space data (Figure 2.2 c). In order to avoid ringing artifact due to sharp truncation in the frequency domain, we apply a Tukey window with $\alpha = 0.5$ which results in smoother reconstructed images.

Noise Addition

Based on the signal-to-noise (SNR) ratio defined by the imaging parameters, a Gaussian complex noise is generated. We add this noise to the real and imaginary part of the simulated k-space data ($\boldsymbol{S}$). The amplitude of the noise signal in the final image depends on the ratio of the magnitude of the simulated signal and the noise standard deviation. It is evaluated by $SNR = \boldsymbol{C}(ROI)/n_{std}$, where $\boldsymbol{C}(ROI)$ is the mean value of the simulated contrast at the region of interest (around the heart) and n_{std} is the standard deviation of the Gaussian distribution, from which random samples are generated. The contrast $\boldsymbol{C}$ depends on the tissue-specific and sequence-specific parameters and is calculated from noise-free k-space data. Given a desired SNR level per simulation, the noise standard deviation is obtained and used to generate a noise complex data ($Noise(n_{std})$) that is added to k-space in Fourier domain. For calculating SNR with the given equation, we assume single-coil data acquisition. As a result of noise averaging, adding extra receiver coils would lead to images with improved SNR. As shown in Figure 2.2 d), the noisy k-space data at acquisition resolution ($\boldsymbol{N}$) is computed as:

$$\boldsymbol{N} = \boldsymbol{S} + Noise(n_{std}) \tag{2.3}$$

Reconstruction

The high-resolution input is considered to account for the continuous nature of the underlying anatomy in the real-world scenario. The final image is reconstructed by performing inverse Fourier transformation $\boldsymbol{R} = \mathcal{F}^{-1}(\boldsymbol{N})$ after adding complex noise (Figure 2.2 e). The simulated images are re-sampled to a uniform grid (e.g. $256 \times 256 \times 13$) and the intensity value for the whole image is scaled to the interval of $[0 - 4095]$, accounting for a 12-bit digital value of the images. Subsequently, the images are saved together with the proper metadata information, making them easily visualizable using standard image viewing software, such as ITK-SNAP and ImageJ [25, 38]. The corresponding ground truth binary labels for all of the tissue types in the simulation are provided alongside the images.

2.2.3 Usefulness of the Simulated Data

We evaluate the usefulness of the simulated CMR images in the context of training a deep convolutional neural network for the task of cardiac cavity segmentation. We investigate three different scenarios of utilizing simulated data, which include: 1) exploring the performance of a segmentation model, trained only using simulated images, on real MR test images, 2) assessing the usability of simulated images as a data augmentation method, and 3) analyzing segmentation performance retention when real data is reduced during training, while the number of simulated data remains the same.

Real Data

We deploy all networks on images acquired from the Automated Cardiac Diagnosis Challenge (ACDC) challenge[1] [39]. The ACDC data-set includes end-systolic (ES) and end-diastolic (ED) images acquired from 100 patients, containing both normal and pathological subjects. Images are acquired by two scanners with different magnetic field strengths and contain expert annotations for left ventricle (LV), right ventricle (RV) blood pool, as well as left ventricular myocardium (MYO).

[1] ACDC data can be found at `https://www.creatis.insa-lyon.fr/Challenge/acdc/`

Out of the available 100 subjects (200 ED and ES MR images), we reserve 90 (180 MR images) for training and 10 (20 MR images) for testing.

Segmentation Network and Training

We adopt a 3D nnU-Net [40] model for a a multi-class segmentation task with several modifications to improve the adaptation from simulated to real MR images during test time. The real cardiac MR images of the ACDC challenge have different in-plane spatial resolution, matrix size, and slice thickness. It is a necessary practice to harmonize that for training a deep-learning segmentation algorithm. All real images used during training and testing are re-sampled to the in-plane resolution of $1.25 \times 1.25mm^2$ and cropped around the heart area to the size of 128×128 pixels while keeping the original slice thickness and number of slices. However, to ensure more variation in the FOV and vary the size of the heart in simulated data, we apply a range of different resolutions when resampling the simulated images. We normalize input images at both the training and inference time to an intensity range from [0,1] using a 99th percentile-based approach. During training, we augment the data on-the-fly by utilizing random horizontal and vertical flips ($p = 0.5$), random rotation by integer multiples of $\frac{\pi}{2}$ ($p = 0.5$), scaling (scale factor $s \in [0.85,1.25]$, $p = 0.3$) and random translations ($p = 0.3$), gamma and brightness transformation ($p = 0.6$), as well as elastic transformations ($p = 0.3$).

At inference time, we apply in-plane resampling, center-cropping and percentile-based normalization on the real images. We additionally apply histogram matching to match the intensity distribution of real images to that of simulated images used during training, using a landmark-based histogram standardization approach described in [41]. Moreover, we apply total variation (TV) denoising on real images, which removes high frequency noise and textural features, but retains sharp edges and outlines of larger tissues. This procedure reduces the bias of the trained network towards tissue texture, which is difficult to realistically simulate. All images are filtered using a strength parameter $\alpha = 15$, which was visually determined to smooth out the texture, while retaining the cavity shape.

Experiments

We explore the usability of simulated data for cardiac cavity segmentation by performing three experiments that outline different aspects of simulated images. To compare the performance of each model, we calculate the Dice similarity metric and Hausdorff Distance (HD) on all slices of each subject in the test set.

Experiment 1: Performance of the Simulated model;
By employing a training and testing procedure described in Section 2.2.3, we first train two segmentation models (**Simulated 1 and Simulated 2**) using

a total of 200 simulated volumes (ED and ES) each for subjects with normal and abnormally thick myocardium, respectively. We create XCAT subjects with thickened myocardium by modifying the NURBS surfaces for the left ventricle. We evaluate the model on 20 ED and ES volumes extracted from the ACDC test set, where all slices are considered during evaluation. We compare the performance of this model to a model trained using real images from the ACDC training set, which is our baseline. All models are evaluated using the Dice score and Hausdorff distance.

Experiment 2: Augmentation using simulated data;
In this experiment, we evaluate the use of simulated data as an augmentation method, whereby we wish to observe whether adding simulated data to the training set of real images has the potential to improve the model's generalization performance and reduce over-fitting. The baseline model for this experiment is trained with 180 ACDC (ED and ES) volumes, extensively augmented using geometric and intensity transformations, as well as transformations affecting the quality of images by adding (Gaussian) noise. The **Augmented models** Aug 1 and Aug 2 are trained by adding 200 simulated volumes (corresponding to Simulated 1 and Simulated 2 data) to the same set of real images used for training the baseline model. The evaluation of both models is performed on all slices of the test set.

Experiment 3: Real data reduction;
In the last experiment, we wish to evaluate the extent to which the simulated data can compensate for the lack of real MR images during training, in scenarios where annotated MR images are limited. Since limitation of expert-annotated data is often a challenge for supervised deep learning algorithms, having the ability to train well-performing models with the help of simulated data is of the utmost importance during model development. To demonstrate this, we systematically reduce the number of real images available during training, while retaining the number of simulated images (200) and evaluate the models on the same set of 20 test volumes as before (all slices included).

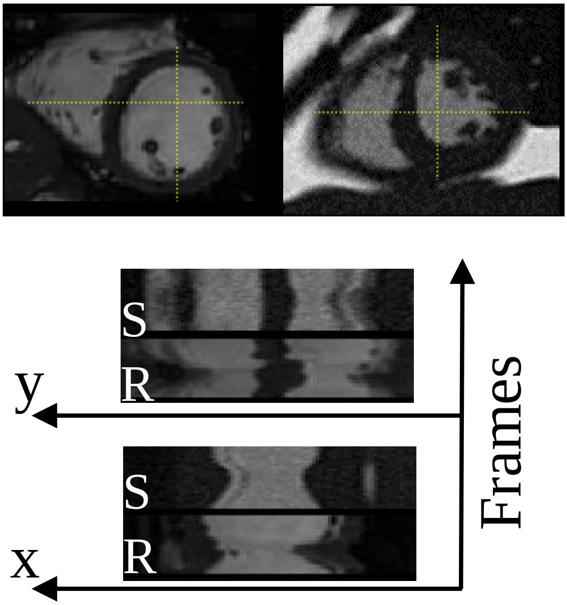


Figure 2.3: Dynamic cine simulation for 25 frames across one cardiac cycle with 1 second period. Time profiles along x and y lines for simulated (S) and real (R) images are shown. The Animated gif is available at `https://bit.ly/3DcU7q3`.

2.3 Results

2.3.1 Qualitative Analysis

Visual comparison of simulated CMR images using the original MRXCAT approach [8] and our proposed framework is shown in Figure 2.4. Both the underlying anatomy and the image contrast have been improved, resulting in increased image quality and realism. As can be observed from the zoom area around the heart, inclusion of the small structures of the trabeculae anatomy in the myocardium has made the XCAT heart more similar to the anatomy of the real human heart. This modelled trabeculation has improved the realism of myocardium-to-blood borders for left and right cavities compared to the previous simulation. Furthermore, increasing the number of tissue types in the image simulation as well as assigning relevant relaxation properties is beneficial to enhance image quality. Using low-pass filtering for partial volume effects in the original MRXCAT approach gave a rather simplified look to the images. By adding the sampling operation in the pipeline we could better account for the effects of partial volume on the smoothness and blurriness of the organ and tissue boundaries.

Our framework can also take the time series of the XCAT models to perform dynamic simulation for the cine CMR imaging study. As can be observed in the movie version of Figure 2.3 (available at `https://bit.ly/3DcU7q3`) the papillary and trabeculae structures inside the left and right ventricles are deformed according to the motion model of the beating heart, which replicates the twisting of the myocardium. The simulated heart motion resembles the real one shown next to it and the time profile of simulated (S) and real (R) image along two perpendicular directions (x; top-down and y; left-right) for all frames show good similarity in terms of myocardial displacement.

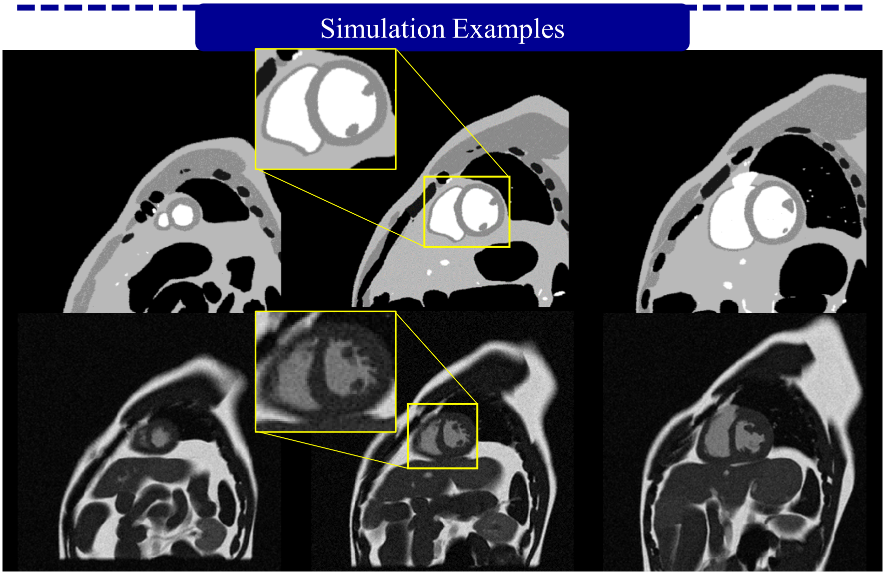


Figure 2.4: Examples of simulation with improved image realism for apex, mid, and base locations of the heart . The results for the original MRXCAT approach and our framework is shown in first row and second row, respectively.

For each virtual subject we have a number of labels to which we assigned tissue properties for image simulation. These input labels provide accurate information about the underlying anatomy and can be utilized as ground truth labels of the simulated images. Different labels (49 tissue types) are represented by different colors in Figure 2.5. The heart-only label map with its different components is shown in Figure 2.5. The combined version representing the heart as 3 simplified classes may be more suitable for training a heart cavity segmentation network.

Generating diversified images with varying parameters can enrich the virtual population. Within our framework, we made numerous parameters available to alter the simulation characteristics. This additional flexibility yields the advantage to perform arbitrary changes in the imaging settings or anatomical features. Two

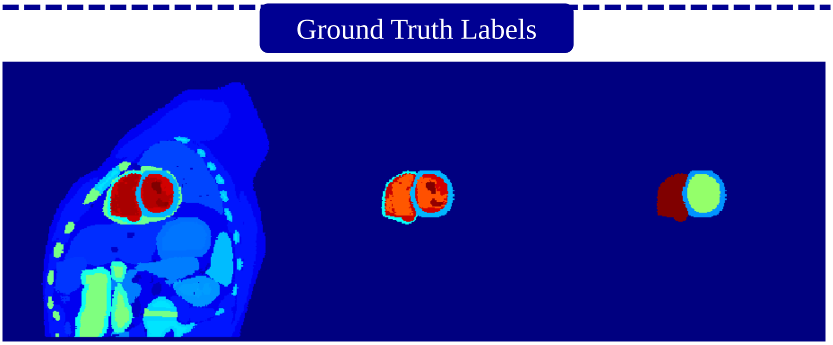


Figure 2.5: Ground truth labels of a simulated cardiac MR image. Full label map with 49 separate tissues shown with different colors (left), the heart tissues (middle), and simplified classes for heart cavity (right).

examples for male and female anatomies are shown in Figure 2.6. It also illustrates some examples of changing image characteristics by modifying the imaging parameters such as flip angle, in-plane resolution, and SNR level. We can observe the alternations on the anatomical features such as location of the heart within torso and scaling of the whole body, as well as changes of the image appearance.

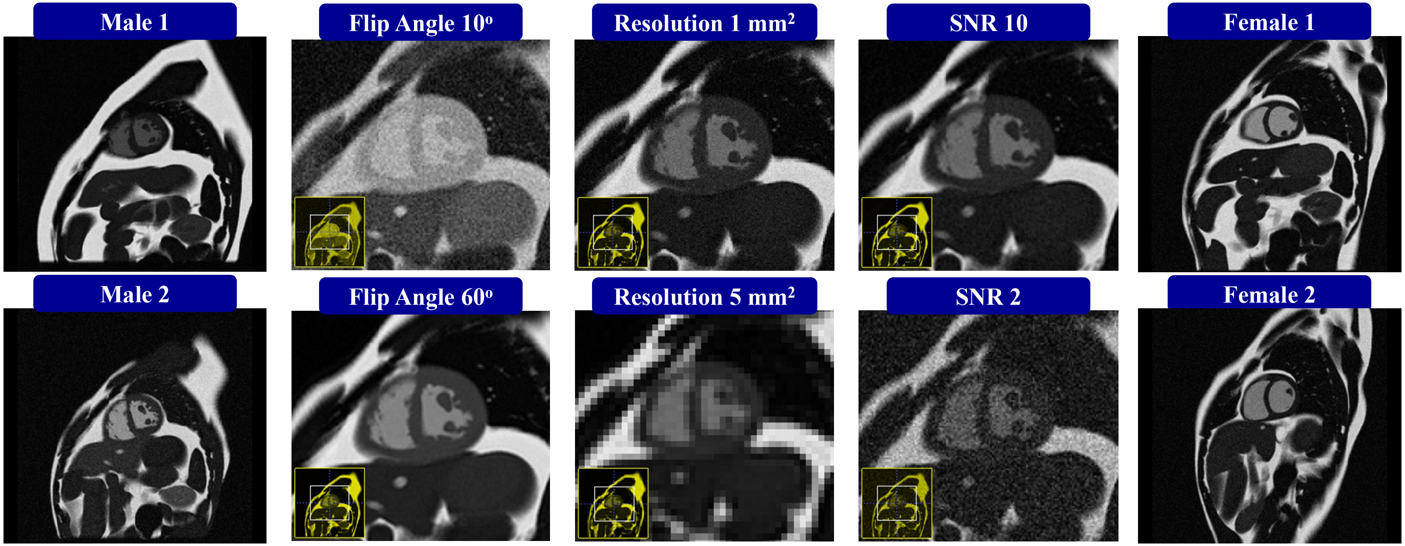


Figure 2.6: Two examples of the simulated images for virtual subject male (first column) and virtual subject female (last column), and examples for varying sequence parameters flip angle (degree), in-plane resolution (mm) and simulated SNR levels. Note that the heart trabeculation was only added to the male subjects.

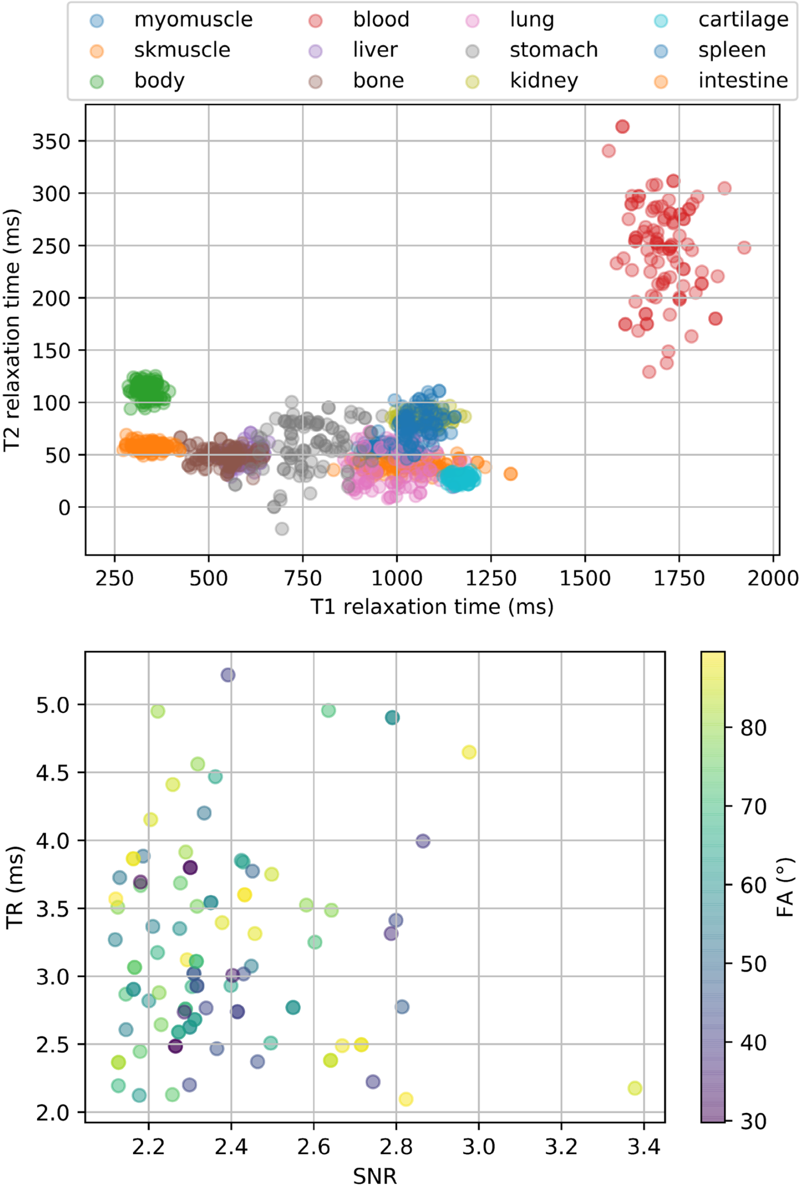


Figure 2.7: The distribution of T1 and T2 relaxation times for tissue properties, repetition time (TR), signal-to-noise ratio (SNR) and flip angle (FA) for sequence parameters used for simulating subjects in the heterogeneous population.

Using the proposed framework, we simulate a population of virtual patients with varying anatomical and contrast characteristics, tissue properties and sequence parameters. The heart model of the virtual male is enhanced by adding an additional layer of the heart trabeculae as explained in section 2.2.1, whereas the virtual female only has the papillary muscles. Figure 2.7 shows the T1 and T2 relaxation times, repetition time (TR), simulated signal-to-noise ratio (SNR) and flip angle (FA) for each generated virtual subjects. For simulations in this study, we use myocardial T1 values of 977 ± 42 ms, which is measured using inversion-recovery techniques [27, 28]. However, saturation-recovery can provide more accurate tissue quantification that can yield higher T1 values as discussed in [42] and these are therefore included in Table 2.1. We observe that selecting slightly higher/lower tissue property values for myocardium has a negligible effect on the simulated image contrast. To visualize the range of anatomical variations, Figure 2.8 depicts left ventricular (LV) volumes at end-diastolic (ED) and end-systolic (ES) phases of the heart for the simulated subjects.

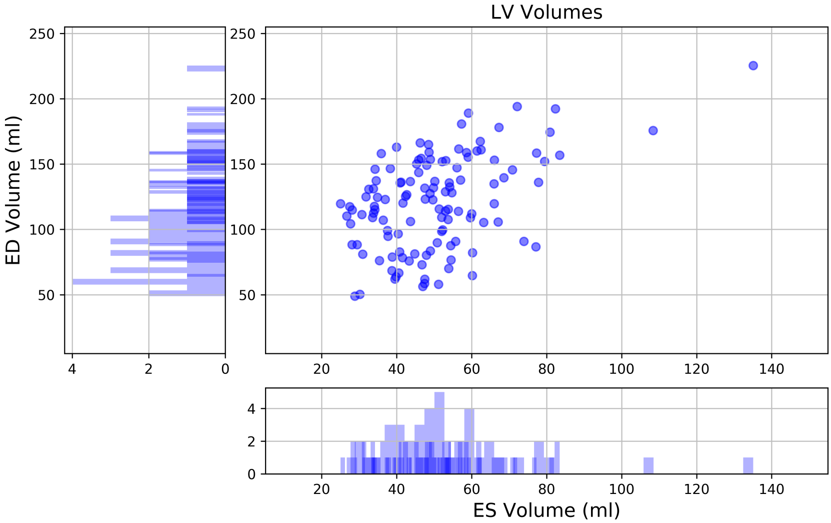


Figure 2.8: The distribution of the left ventricular (LV) volumes at end-diastolic (ED) and end-systolic (ES) phases for simulated subjects.

2.3.2 Quantitative Analysis

Experiment 1: Performance of the Simulated Models

Figure 6.3 showcases some of the predictions for mid-ventricular slices for the segmentation models trained on simulated images only with normal (Simulated 1) and thickened myocardium (Simulated 2), compared to the baseline model trained using real images. The models are evaluated on all slices of testing subjects, resulting in the overall Dice scores and HD scores shown in Figure 2.11. We perform a two-tailed Wilcoxon signed-rank test for the two models with $p < 0.01$, indicating statistical significance between the performance of the two models. By visual observation, we can determine that the performance drop observed in the Simulated 1 is largely due to thicker myocardium appearing in the test set, especially with patients containing pathology. While tissue segmentation in the presence of pathology is generally a challenge when evaluating any segmentation approach, additional challenges stem from the nature of simulation, where most simulated myocardial tissue is thinner due to exclusion of papillary muscles and trabeculae. This also affects the performance of the model when it comes to LV segmentation. To address this limitation, we simulate new subjects with thickened myocardium to train the Simulated 2 model. This newly simulated data help increasing the segmentation performance for thick myocardium present in the testing data, as well as boosting the performance for LV and RV segmentation as can be observed from Figure 2.11. A drop in segmentation performance in comparison with the baseline can stem from the basal and apical slices and can be linked to the complex anatomy of the heart for pathological cases, which is a significant challenge for a network that exclusively learns from the appearance and shape of simulated hearts. Additionally, the performance of the Simulated models is drastically affected by changes in appearance and texture, despite our attempts to minimize these effects through TV filtering. However, these results suggest a strong potential of utilizing a cost-effective artificial data-set for training neural networks that can perform on par with traditional approaches requiring large annotated sets of real MR data.

Experiment 2: Augmentation using simulated data

Figure 2.11 depicts the performance of the Augmented models (Aug 1 and Aug 2) compared to the baseline trained without simulated images. We can observe that both RV and LV segmentation performance improves with the addition of simulated data, with the accuracy of MYO segmentation slightly reducing for Aug 1 model. This is again the effect of thinner myocardial tissue in simulated images in Aug 1, which hampers model performance in the presence of myocardial thickening. To address this issue and improving the performance of the augmented model on the pathological cases with thick myocardium we simulate subjects with abnormally thick myocardium and add them to the training. We observe

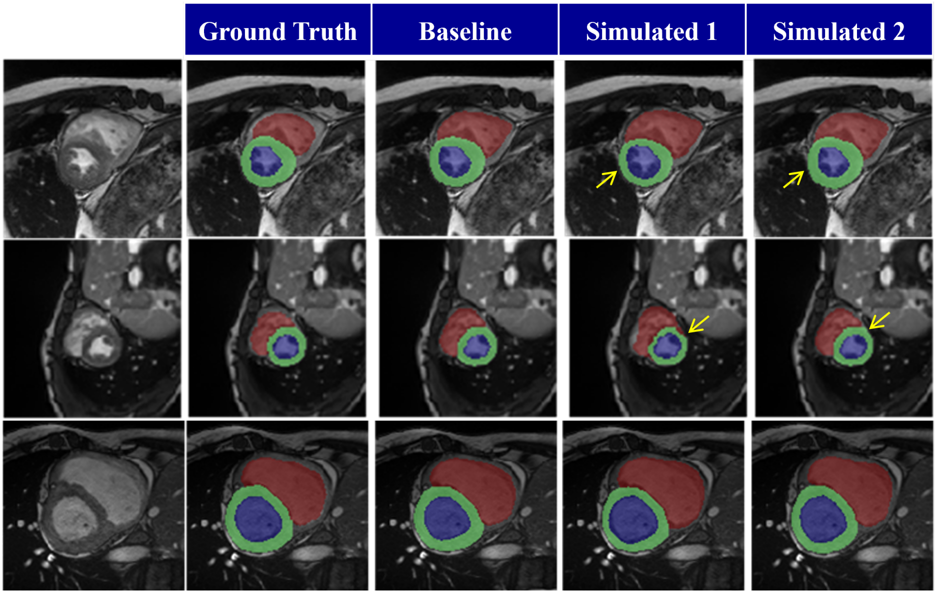


Figure 2.9: Comparison between a segmentation model trained on 180 real images (baseline) and identical model trained completely on 200 simulated images (simulated).

improved performance for the segmentation of the myocardium as well as right and left ventricles in terms of Dice score and substantial reduction in the HD score for Aug 2 model where subjects with thickened myocardium are simulated. The obtained Dice and HD scores for segmentation using the Aug 1 and Aug 2 models are statistically significant with $p < 0.05$. Samples of slices where augmentation with simulated data improved the performance can be observed in the first three columns in Figure 2.10. The majority of visually observed improvement in predictions is typically related to the RV segmentation performance, but it is also notable for LV and myocardium segmentation, especially when simulated images with thickened myocardium is added. We hypothesize that adding simulated data to the model has reduced over-fitting and given more emphasis to cavity shapes during training. Additionally, we find that simulating subjects with thickened myocardium is important to improve the performance of the model on the pathological cases present in the test set.

To further demonstrate the contribution of simulated pathological data to the segmentation performance, we train an additional model augmented with a combination of both Simulated 1 and Simulated 2 data. In total, we add 200 simulated volumes, where 50% of images belong to the Simulated 1 data-set, with the other half extracted from Simulated 2 data. To avoid selection bias when choosing the 50% of total simulated data available, we perform a five-fold cross validation experiment and randomly select 100 images per simulated set for each fold. A model trained under such a setup performs better than the Aug 1 model, whereby we observe a reduction in the number of outliers and improvement in HD scores across all tissues, particularly the LV myocardium. We attribute this to the presence of the Simulated 2 data, which we already show can tackle the appearance of thickened myocardium (Aug 2 model). However, while better than Aug 1, this model under-performs compared with Aug 2, indicating that Simulated 1 images may not introduce enough variability beneficial for improving the segmentation performance of ACDC data. However, this may not be the case for other datasets (such as those containing examples of healthy subjects only).

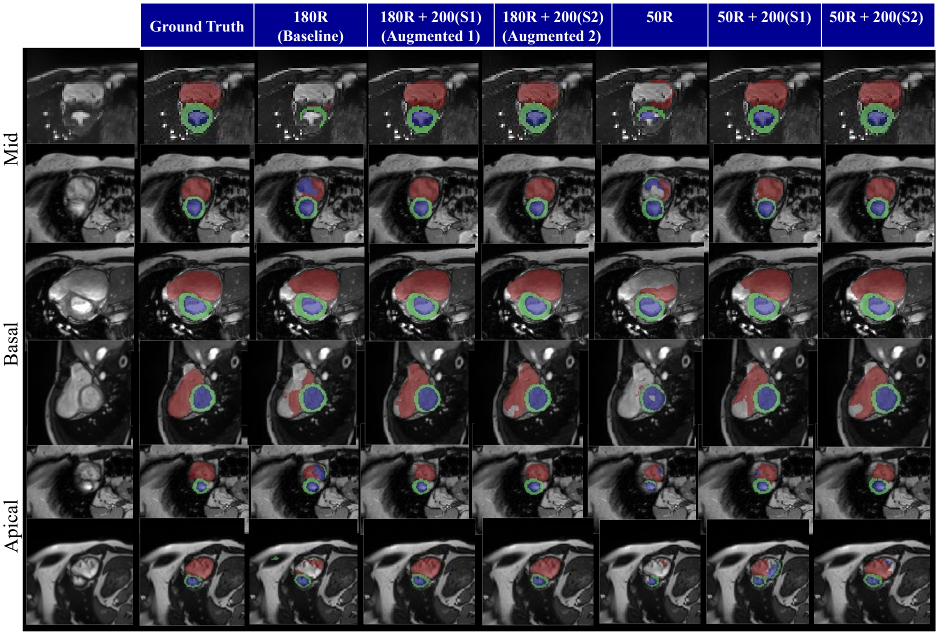


Figure 2.10: Comparison between the baseline segmentation model trained on 180 real images (180R), the augmented 1 model with 200 simulated 1 images (180R + 200(S1)), the augmented 2 model with 200 simulated 2 images (180R + 200(S2)), reduced model trained with 50 real images (50R) and augmented version of that with simulated 1 and 2 images (50R + 200(S1)) and (50R + 200(S2)). Results for mid-ventricular, basal and apical slices of the heart are shown.

Experiment 3: Real data reduction

The effect of reducing real data during training, where we try to imitate the scenario where acquiring annotated real MR data is challenging, can be observed in Figures 2.10 and 2.11. We reduce the number of real MR images available during training to 100, 50 and 25 volumes only, while retaining 200 simulated images. We compare the performance of such models to models trained with only a limited amount of real MR volumes, without the addition of simulated data. As expected, the performance of such models tends to drop for all tissues, with a significant reduction in performance when only 50 and 25 real MR images are used during training. However, if such limited data-sets are aided with simulated images, the performance drop is less significant and in some cases the performance of the augmented model is retained compared to the model trained with the maximum amount of real MR volumes available (compare 180R with 100R+200S in Figure 2.11). In Figure 2.10, we showcase the most extreme improvements, which occur for models trained with 50 real images only, where we observe the 50R model struggling with the segmentation of the RV and producing false positive predictions for the LV. These are successfully compensated by the addition of simulated images during training.

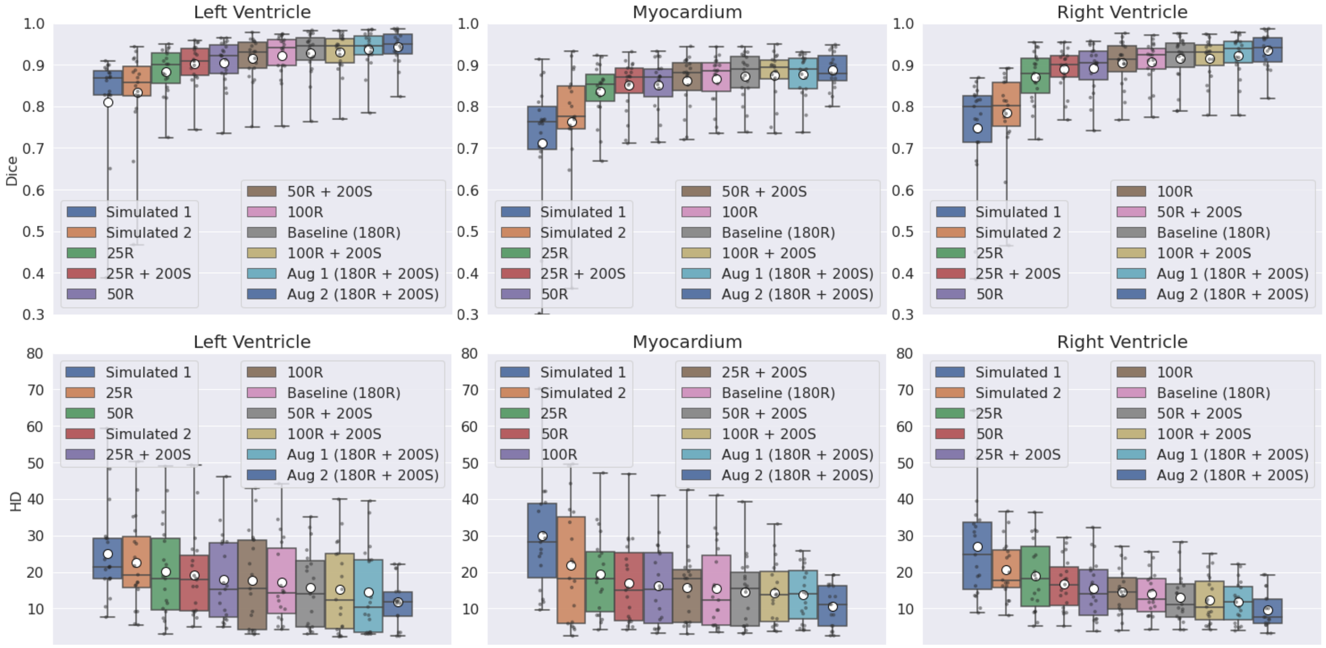


Figure 2.11: Performance evaluation based on Dice score (higher better) and Housdorff Distance (HD) (lower better) for the baseline segmentation model trained on 180 real (180R) images, the augmented models (Aug 1 and Aug 2) with 200 simulated (180R + 200S) images, reduced models trained with 100 real images (100R), 50R, 25R images and augmented versions 100R + 200S, 50R + 200S, and 25R + 200S images. The Simulated 1 and Simulated 2 models are trained using simulated images only with normal and thickened myocardium, respectively.

2.4 Discussion

In this chapter we propose a flexible framework for physics-based CMR image simulation with the purpose of generating a heterogeneous population comprising diversified virtual subjects. We improved the quality of the simulated images by enhancing and optimizing the three main components: i) computerized human anatomical model, ii) magnetic tissue properties, and iii) physics of MR image formation.

Our proposed image simulation pipeline has separated modules that could easily be adapted and replaced to account for more comprehensive experiments. For instance, in the sampling module c, for the fast generation of CMR images, uniform Cartesian sampling for the k-space data can be replaced by more advanced acquisition trajectories. It should be noted that the reconstruction module needs to be modified accordingly, and also the coil sensitivity map may be required for parallel imaging. The analytical description of the MR signal was used for fast generation of the imaging contrast. Alternatively, the signal model could be replaced with the extended phase graph formulation [43] or full numerical simulation of the MR signal using the JEMRIS simulator [44], at the cost of substantially increasing the complexity and computation time of the simulation. The JEMRIS simulation approaches, based on numerically solving the Bloch equations, is considered more suitable for investigating the effects of various pulse sequences design on the spin system, while in the context of generating ample images for deep-learning application, the effects of the parameters on the global contrast of the final images is more relevant.

Similar to the MRXCAT approach [8], the contrast is governed by the analytical solution of the Bloch equations at the steady state of the magnetization for the bSSFP sequence. It was shown that bSSFP sequence is less sensitive to field inhomogeneities when a short TR is used [36]. We use short TR and therefore ignore the presence of any off-resonance effects, which in real world scenario might be due to RF inhomogeneity, imperfect shimming, magnetic susceptibility, and T2* effects. Moreover, the RF slice profile is assumed to be perfectly rectangular, and no spin dephasing is present during the acquisition. Simulating imperfections in MR scanner and designing different RF pulses require numerically solving Bloch equations using software packages such as JEMRIS [44] and MRiLab [45] which found to be extremely time consuming and unsuitable for image database generation.

We consider variations in the sequence parameters (shown in Figure 2.7) to be an import feature of our simulation pipeline that results in producing images with different contrasts and appearances. We observe that different scanner vendors and imaging centers use different sequence parameters to optimize the bSSFP contrast hence they produce varying image appearances. This varying contrasts and imaging features hamper the performance of DL segmentation network as discussed in the M&Ms challenge [46]. We believe by altering sequence parameters we can generate images with diverse appearances to help development of generalizable DL method that can robustly perform on heterogeneous data from different clinical centers, imaging conditions and scanner vendors.

Images with variations in the noise level are also generated in this study and the range of simulated SNR is experimentally identified by inspecting the images. Note that complex noise is added to the complex k-space data to resemble a real scenario where the noise is already present in the MR signal during the acquisition. We found that applying filter to the k-space data and resampling to the grid after reconstruction will change the final SNR. In agreement with what discussed in [47], we found that SNR measurements using two region approaches (background noise and region of interest) do not completely agree with the actual simulated SNR. Our measurements of SNR based on two ROIs from the simulated images resulted in overestimating the simulated SNR level. However, the absolute target SNR value is not crucial here, rather variation in the SNR and the ability to change the noise level is considered an important feature of our framework for generating diversified images for the purpose of DL training.

We created a database of CMR images of virtual patients using the proposed framework to aid the development of data-hungry deep-learning medical image analysis methods. One application using the preliminary version of the framework was demonstrated in [21]. The benefit of using the simulated images for training a cardiac cavity segmentation model was investigated and it was shown that pre-training a deep-learning based segmentation model generalizes better to the inherent variability of real cardiac images.

In addition, we explored the application of the virtual CMR database in the task of cardiac cavity segmentation using supervised deep learning. Our experiments demonstrated that models trained with images simulated in this study solely can already perform comparable cavity tissue (LV, RV and MYO) segmentation in mid-ventricular slices to models trained with real MR images annotated by medical experts. We further show that the use of simulated data can be considered an addition to classical augmentation methods, which are typically limited in producing tissue shapes and appearance different to existing data. We demonstrate that simulated data can compensate for the lack of real data during training and be of significant help in settings where data acquisition is challenging. This indicates that artificial data generated through the proposed framework has the potential to extend the variability of anatomical and contrast features available in training data and consequently, help the network generalize and adapt better to unseen data. Of course, simulating a more realistic appearance and contrast, with significant variation in shape and quality is of the utmost importance for achieving this. The added benefit is the fact that through this framework we can simulate and design a number of pulse sequences, generating simulated data representative of the data available in test sets, which is typically costly and infeasible for typical MR acquisition.

2.4.1 Limitations and Future Work

We simulate single-coil acquisition scenario for fast generation of the images. However, three-dimensional coil sensitivity maps can be simulated using the Biot-Savart law as previously discussed in the MRXCAT [8]. The normalized sensitivity map can be multiplied with the simulated contrast resulting in one image per coil. Similarly, one could use the same multi-coil simulation setup inside our framework. Note that the multi-coil acquisition scenario is more suited for reconstruction purposes and optimization for parallel imaging. We noticed that the addition of multiple coils will substantially increase the simulation time and complexity, making our framework less suitable for generating substantial number of images.

Flowing blood, also known as in-flow effects, can change the blood-to-myocardium contrast and has historically been a source of error in cardiac MR images. The simulations in this work do not account for this effect due to the complexity of computational modeling for the blood flow and time-consuming simulation procedure, requiring careful considerations of RF pulse profile, velocity distributions, and slice thickness. It is, however, possible to extend our framework with such model for the purpose of flow quantification using phase contrast image simulation.

Major cardiac MR imaging artifacts such as respiratory motion and ECG-mistriggering artifacts are not simulated in this study. These artifacts are one of the primary sources of failure of deep-learning (DL) based segmentation models. Recently the authors in [48] and [49] proposed k-space models of motion artifacts for the purpose of making DL algorithms more robust to such artifacts. Particularly in [49] authors proposed a data augmentation strategy to simulate cardiac ECG-mistriggering and breathing artifacts based on k-space data corruption. The same approach can be used to simulate these artifacts on our database. Another concurrent work in [50] showed that our framework can be extended for the application of late gadolinium enhancement (LGE) simulation on a virtual subject with myocardial infarction and included respiratory artifact in the simulation procedure. Precisely, the LGE simulation is performed at various time points across one respiratory cycle from the XCAT model and the data is combined in such a way to resemble the slice-misalignment artifact. The authors investigate the effects of such artifact on the electrophysiology modeling of the heart. Modeling relevant artifacts in cardiac MR imaging and simulating subjects with respiratory motion artifacts, ECG-mistriggering, and ghosting artifacts, as well as investigating their impact on the segmentation remains to be explored in future work.

In attempts to simulate pathological cases in this study, we generate subjects with thickened myocardium for hypertrophic cardiomyopathy by modifying the NURBS surfaces of the XCAT heart model. We demonstrate that the addition of these subjects with abnormally thick myocardium would substantially improve the segmentation performance on pathological cases, achieving best overall score for the Aug 2 model in Figure 2.11. However, the current model lacks subjects with myocardial infarction (potentially thinned heart), congenital heart, and other cardiac diseases. Adding virtual subjects with such diseases would be of great importance and interest for the research community. However, due to the complex nature of the heart modeling for such patients, we believe this could be address as future work. For instance, by anatomical modeling of congenital heart disease in the XCAT phantom, one could use our framework to simulate corresponding images.

We showed that the inclusion of the trabeculation anatomy in the male XCAT subject increased the visual realism of the simulated images. However, the impact of this addition on the segmentation results was not directly quantified in the presented paper. A mix of female subjects without and male subjects with trabeculation were simulated for the Simulated 1 data, whereas for the Simulated 2 data with thickened myocardium we only generated male subjects without the addition of the trabeculation. The results of these experiments suggested that even without the added trabeculation anatomy, the simulated images improved the segmentation performance, and whether or not the addition would bring extra benefits to the segmentation remained to be explored in future studies.

The presented framework was designed to provide simulation of realistic CMR images for the cine study. Extension to other CMR modalities such as late gadolinium enhancement and first-pass perfusion are considered as future research. These additions to the framework could be of great interest for multi-modality studies, especially for expanding the database to increase its applicability for disease classification. We believe this work is a step towards our aim to establish a unified framework for personalized multi-modal cardiac magnetic resonance image simulation.

2.5 Conclusions

In this chapter, we proposed a flexible framework for realistic simulation of cardiac magnetic resonance images with controllable anatomical features, MR tissue properties, acquisition parameters, and image appearance. We generated a virtual population of CMR images with ground truth labels using our proposed framework and made it publicly available to aid development of deep-learning cardiac image analysis algorithms. Furthermore, our usability experiments suggested that augmentation with the simulated population can boost the segmentation performance, and even retain the baseline performance when just 45% of the real data is available.

CHAPTER 3

Conditional GANs for Image Synthesis

This chapter is based on:
S. Amirrajab, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “Label-Informed Cardiac Magnetic Resonance Image Synthesis through Conditional Generative Adversarial Networks,” Computerized Medical Imaging and Graphics, 2022.
Which is an extension of:
S. Amirrajab, S. Abbasi-Sureshjani, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “XCAT-GAN for Synthesizing 3D Consistent Labeled Cardiac MR Images on Anatomically Variable XCAT Phantoms,” MICCAI 2020.
S. Abbasi-Sureshjani, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “4D Semantic Cardiac Magnetic Resonance Image Synthesis on XCAT Anatomical Model,” MIDL 2020.

Abstract

Synthesis of a large set of high-quality medical images with variability in anatomical representation and image appearance has the potential to provide solutions for tackling the scarcity of properly annotated data in medical image analysis research. In this chapter, we propose a novel framework consisting of image segmentation and synthesis based on mask-conditional GANs for generating high-fidelity and diverse Cardiac Magnetic Resonance (CMR) images. The framework consists of two modules: i) a segmentation module trained using a physics-based simulated database of CMR images to provide multi-tissue labels on real CMR images, and ii) a synthesis module trained using pairs of real CMR images and corresponding multi-tissue labels, to translate input segmentation masks to realistic-looking cardiac images. The anatomy of synthesized images is based on labels, whereas the appearance is learned from the training images. We investigate the effects of the number of tissue labels, quantity of training data, and multi-vendor data on the quality of the synthesized images. Furthermore, we evaluate the effectiveness and usability of the synthetic data for a downstream task of training a deep-learning model for cardiac cavity segmentation in the scenarios of data replacement and augmentation. The results of the replacement study indicate that segmentation models trained with only synthetic data can achieve comparable performance to the baseline model trained with real data, indicating that the synthetic data captures the essential characteristics of its real counterpart. Furthermore, we demonstrate that augmenting real with synthetic data during training can significantly improve both the Dice score (maximum increase of 4%) and Hausdorff Distance (maximum reduction of 40%) for cavity segmentation, suggesting a good potential to aid in tackling medical data scarcity.

3.1 Introduction

Medical image segmentation has seen significant progress in recent years thanks to development of automated deep-learning (DL) methods [51,52]. Similarly, with the emergence of deep generative modeling, the medical imaging community has benefited from its solutions for various applications such as image reconstruction, segmentation, registration, and particularly image synthesis [9, 53]. Generative adversarial networks (GANs) demonstrate promising results in generating realistic images for tackling medical data scarcity and patient privacy issues. Medical image synthesis using conditional GANs conditioned on the tissue label map, which tries to translate simplified segmentation mask to realistic image, can be understood as the opposite of the image segmentation. This has a number of important advantages: i) allowing control of the anatomical information (content) of generated images by controlling the segmentation masks, ii) explicit alignment with the nature of medical images in which anatomy features are separated from the style and appearance produced by imaging modality , iii) ability for direct usage of generated images for downstream supervised and unsupervised tasks.

In this chapter, we propose a cardiac magnetic resonance (CMR) image synthesis framework for translating the anatomical information of a segmentation mask to a realistic image. The framework is able to generate a diverse database of CMR images with ground truth labels that can be used for any downstream task. Preserving the semantic information (content) of the anatomy presented in the segmentation mask is important when synthesizing data for training a DL medical image segmentation model. Therefore, we propose to incorporate a specific type of conditional GAN that takes segmentation masks as an extra input to the generator architecture. The segmentation mask corresponding to the input image is fed to each layer of the generator using a conditional normalization layer. This is important for learning the textual features (style) of separate classes and preventing information loss when passed through multiple convolutional layers.

The framework consists of two modules: i) an image segmentation module to provide multi-tissue masks on real CMR images, trained using a physics-based simulated database of CMR images with corresponding labels, and ii) a semantic image synthesis module to translate input segmentation labels to realistic-looking cardiac images, trained using pairs of real CMR images and corresponding labels. The proposed framework works in a supervised manner, requiring real images with paired multi-tissue segmentation masks. Lack of expert annotations for the cardiac MR images poses another challenge for the adaptation of our method. However, we alleviate this challenge by utilizing simulated cardiac MR images for training a multi-tissue segmentation network that provides labels for organs visible in the field of view, such as lung, abdominal organs, musculoskeletal and skin fat tissue. We perform extensive experiments to analyze the effects of different elements of the framework on the quality of the synthesized images, as well as evaluate the effectiveness and usability of such data for training a supervised DL model for the clinical task of cardiac cavity segmentation.

Contributions

The contributions of the chapter are: 1) Proposing an optimized framework for CMR image synthesis trained with pairs of real images and multi-tissue segmentation masks. The framework is comprised of two modules for image segmentation and synthesis. The first module provides multi-tissue segmentation masks for real images. The second module learns the translation from segmentation masks to realistic images, while preserving the anatomical information captured by tissue labels. 2) Proposing a simulation-based approach to provide data for training the multi-tissue segmentation network in module 1), while avoiding the need for pairs of real images and ground-truth multi-tissue masks. We leverage a virtual database of simulated cardiac MR images with corresponding ground truth labels derived from the XCAT phantom (a computerized human anatomical model). 3) Providing detailed assessments of i) the consequences of using a small or large set of segmentation labels, ii) the ability to train a segmentation network solely on synthesized data, iii) the potential of network improvement by augmenting the training set with synthesized data.

In the present chapter, we advance our preliminary works [19,20] substantially: i) we utilize a more realistic database of simulated CMR images for training and use an optimized multi-tissue segmentation model, which provides an enlarged set of labels essential for high-quality synthesis. Furthermore, we experiment with the number of tissue types within the multi-tissue segmentation mask and choose a different combination of tissue classes compared to the previous work; ii) we employ a multi-vendor database of real CMR images to demonstrate the ability of the generator to learn the vendor-specific appearance of the images, without using the style encoder network used in the previous works; iii) we provide a detailed explanation of the network architectures, experiments and evaluation metrics to assess the quality; iv) We provide a detailed assessment of both the image-quality of the synthesized images and the effectiveness and usability of the generated data for two scenarios: 1) training a heart cavity segmentation network completely using synthetic data to evaluate whether we can achieve the same performance when we utilize the same amount of real data; 2) augmenting real data with addition of synthetic data to boost the performance; v) A data reduction experiment is carried out where the number of images for training the image synthesis network is reduced to analyze its effects on the quality of the synthesized images. Additionally, a human visual study is performed for qualitative assessment of image realism.

3.2 Related Work

In this section we briefly highlight some of the recent works in the area of conditional and unconditional image synthesis with emphasis on medical applications. Comprehensive overviews of the generative adversarial networks in medical imaging can be found in [9], [53] and [10].

3.2.1 Generative Adversarial Networks

Generative adversarial networks (GANs) aim to learn the underlying process of data generation by forcing the generator to produce fake samples that are indistinguishable from real images [2]. Image synthesis using GANs has gained major attention thanks to the adversarial strategy for training the generator to synthesize images from a random noise vector. Deep convolutional GAN [54] is a variant for unsupervised learning of image representations that uses convolutional neural networks in the architecture of the generator and discriminator. Such a model has been adopted for unconditional medical image synthesis for various applications such as synthesizing CT lung nodules [55], CT liver and brain lesions [56, 57], MRI brain [58], and skin lesion [59]. Another methodology called progressive growing of GANs [60] is used for synthesizing high-resolution retina fundus and brain MRI data [61]. All of the mentioned work is categorized under the unconditional image synthesis domain, in the sense that a random noise vector is fed to the generator, with no other information for controlling the generation procedure.

3.2.2 Conditional Image Synthesis

Conditional image synthesis, on the other hand, relies on using different types of auxiliary information for guiding the generator to produce results, given an input condition [62]. Among the first works to investigate conditional GANs for the problem of image-to-image translation is pix2pix by [63], which demonstrates a general-purpose approach for synthesizing realistic photos from label maps, edge maps, or black and white images. The framework is supervised in the sense that it needs pairs of input-output images from two domains for training. The input image from the first domain is fed to the generator that adopts a U-Net based architecture for translation to the target domain via adversarial loss. In contrast to L1 loss that leads to blurry images, adversarial loss proves to be more effective in generating high-quality images with fine-grain details. The key success lies behind the ability of the discriminator to learn a trainable loss function to capture the subtle differences between real and fake images. The photo realism and resolution of the pix2pix results have further improved through utilizing a course-to-fine generator with residual blocks and a multi-scale discriminator in the pix2pixHD framework proposed by [64]. Other approaches for unsupervised, unpaired image-to-image translation are used in cases when paired training data from two domains is not available [65–67]. Both paired and unpaired methods are successfully employed in the medical domain for the application of cross modality image synthesis - translating an image from one modality (CT) to another one (MR). MR images are translated to CT images (and vice versa) of the brain for radiotherapy treatment planing [68, 69], as well as the heart for data augmentation [13, 70].

In the present work, we are interested in translating multi-label segmentation masks to realistic medical images, for the application of providing training data for DL-based CMR image analysis. We aim to provide more control over the underlying anatomy of the synthesized images while producing realistic data with high-quality image appearance at the same time. Upon successful training of this label-informed network, the segmentation mask can be manipulated during the test time to generate a new image with altered anatomical characteristics. Such a synthesized image with corresponding (altered) segmentation mask can be used to relieve data scarcity by providing more labeled images for training an analysis model. We analyse both the quality of the synthesized images and the usability of the synthesized data for a clinical task of cardiac cavity segmentation. Experiments are conducted to evaluate the effects of multi-tissue versus cavity-tissue labels, multi-vendor data, and the number of training data on the realism of the generated images. We further perform a small human visual study for scoring the image quality, in addition to calculating the well-known quality metrics.

3.3 Method

An overview of the proposed framework is presented in Figure 3.1. Two consecutive modules, operated for segmenting real cardiac MR images and for synthesizing images, are the main building blocks of our framework. The segmentation module is designed to provide the corresponding multi-tissue labels for the input images, while the synthesis module utilizes such labels with paired images to learn the translation from segmentation to realistic image appearance. Consequently, the synthesis network can generate cardiac images based on the anatomy characterised by the labels.

3.3.1 Image Segmentation Network

We adopt a U-Net architecture [71], completely trained using simulated CMR images, for the task of multi-tissue image segmentation. Several modifications to the original architecture are made to optimize the network, including utilization of Leaky ReLU non-linearity and Batch normalization (BN) after each convolutional layer for stabilizing the training, as well as dropout regularization with a rate of 0.5 to avoid overfitting and improve the generalization. The network consists of five down-sampling and up-sampling blocks, trained with a batch size of 32 2D simulated CMR images, for generating pixel-wise predictions for nine tissue

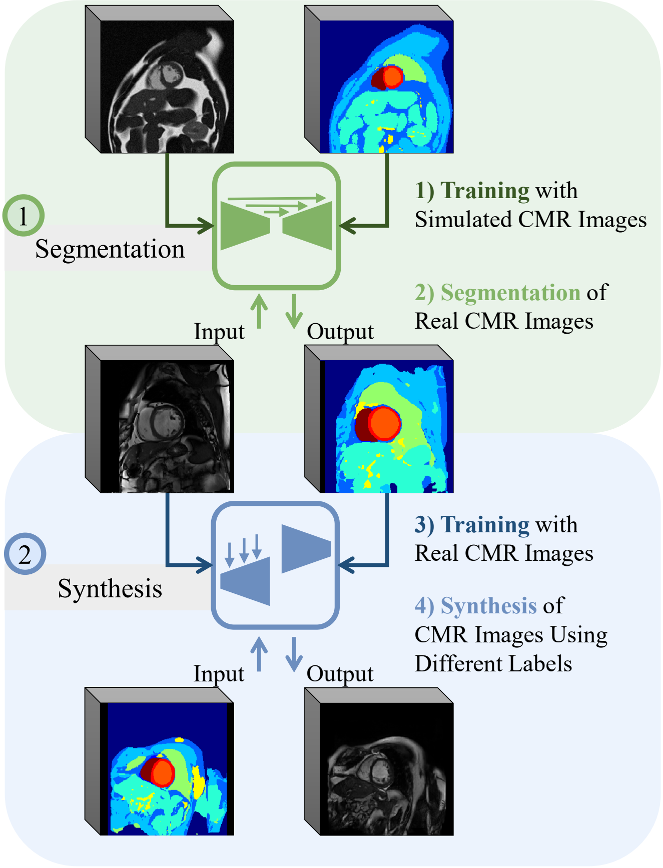


Figure 3.1: The proposed framework including multi-tissue image segmentation and synthesis modules operating offline as follows: **1) training** the segmentation network using only simulated CMR images with ground truth labels for nine tissue types; **2) segmenting** real CMR images to provide corresponding multi-tissue labels for each real image; **3) training** the conditional image generation network using the derived labels alongside with the real images to learn the translation from labels to realistic images; **4) synthesizing**, in turn, images on different labels.

classes, including the background. We use a Focal Tversky loss from [72] for training, optimized using Adam for stochastic gradient descent, with an initial learning rate of 10^{-4}. We apply early stopping when the learning rate drops below 10^{-6} and train the network for a total of 350 epochs using 200 simulated volumes for training.

To alleviate the burden of obtaining real data with corresponding labels for supervised training of the U-Net, we utilize a virtual population of simulated cardiac MR images with accurate ground truth labels for training, as discussed in section 3.4.1. The predicted multi-tissue segmentation masks are used for two purposes: i) providing paired data for training the image synthesis network, and ii) providing input labels for synthesizing new images using the trained generators.

3.3.2 Image Synthesis Network

Preserving the content information of the segmentation labels, which represents the anatomy in the context of medical images, is crucial. Traditionally, this is achieved by feeding the segmentation mask as the input of an encoder-decoder architecture with skip connections, forming a U-Net generator [63, 64]. This approach is sub-optimal for preserving the information of the binary segmentation masks, and common (unconditional) normalization layers such as batch normalization [73] and instance normalization [74] tend to wash away the semantic information of the segmentation mask, as pointed out by [75]. On the contrary, a conditional normalization layer, applied throughout the generator architecture, takes the segmentation map as the condition to compute modulation parameters for de-normalizing the activation in a semantic-preserving manner. A SPatially Adaptive DE-normalization (SPADE) layer is therefore proposed by [75] to replace all the unconditional normalization layers to inject the information of the segmentation map into each feature layer of the network. Conditioning the network using SPADE layers was found to significantly improve the performance of image synthesis compared to directly inputting the segmentation mask to the first layer of the generator. Moreover, the generator can have a much lighter architecture which makes it easier and more stable to train, requiring a fewer number of training data.

Generator Architecture

The generator architecture, inspired by [75], is comprised of several SPADE residual blocks [76], followed by up-sampling layers to increase the spatial dimension of the input activation. The semantic information of the segmentation mask is injected into each residual block using a SPADE conditional normalization layer, as shown in Figure 3.2.

Normalization Layers

Normalization layers play a crucial role in conditioning the generator on the input label. The Batch normalization (BN) [73] layer has been recognized as an effective component for stabilizing the training of modern deep neural networks by normalizing the statistics of feature activation maps after each convolution

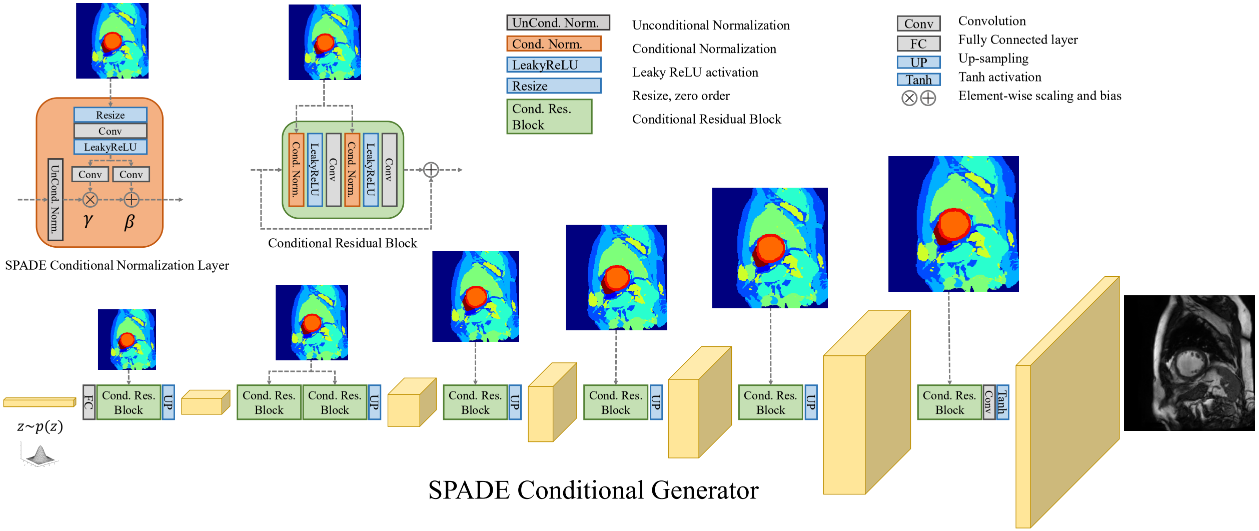

UnCond. Norm.
Cond. Norm.
LeakyReLU
Resize
Cond. Res. Block
Unconditional Normalization
Conditional Normalization
Leaky ReLU activation
Resize, zero order
Conditional Residual Block
Conv
FC
UP
Tanh
Convolution
Fully Connected layer
Up-sampling
Tanh activation
Element-wise scaling and bias
γ
β
SPADE Conditional Normalization Layer
Conditional Residual Block
$z \sim p(z)$
SPADE Conditional Generator

Figure 3.2: Label-conditional generator. Adopted from the architecture proposed by [75], it consists of several conditional residual blocks that take the segmentation map as the condition and inject its semantic information via SPatially Adaptive DE-normalization (SPADE) layers.

followed by a non-linearity layer. The BN first normalizes features in the input mini-batch to its mean and standard deviation, and then uses the scale and bias values for applying the affine transformation. Given the input activation map with the height of H^i and width of W^i for a mini-batch size of N and C^i number of channels ($x^i \in \mathbb{R}^{N \times C^i \times H^i \times W^i}$) at the i^{th} layer of the network, each individual feature channel is normalized as follows:

$$BN_{\gamma,\beta}(x^i) = \gamma \left(\frac{x^i - \mu_c(x^i)}{\sigma_c(x^i) + \epsilon} \right) + \beta, \tag{3.1}$$

where $\gamma, \beta \in \mathbb{R}^C$ are unconditional scale and bias parameters learned during the training. ϵ is small constant added to the denominator to avoid dividing by zero, $\mu_c(x^i), \sigma_c(x^i) \in \mathbb{R}^C$ are the mean and standard deviation calculated across mini-batch and spatial dimensions ($n \in N, h \in H^i, w \in W^i$) for each feature channel ($c \in C^i$) as follows:

$$\mu_c(x^i) = \frac{1}{NH^iW^i} \sum_{n=1}^{N} \sum_{h=1}^{H^i} \sum_{w=1}^{W^i} x^i_{n,c,h,w} \tag{3.2}$$

$$\sigma_c(x^i) = \sqrt{\frac{1}{NH^iW^i} \sum_{n=1}^{N} \sum_{n=1}^{H^i} \sum_{w=1}^{W^i} \left(x^i_{n,c,h,w} - \mu_c(x^i) \right)^2} \tag{3.3}$$

Different from BN, [75] propose to replace the modulation parameters with spatially varying $\gamma^i_{c,h,w}(s^i)$ and $\beta^i_{c,h,w}(s^i)$ dependent on the input segmentation mask (s^i).

Following BN layer, the normalized activation is modulated with spatially-dependent scaling and bias factors calculated using a shallow, two-layer modulation convolutional network. Calculated $\gamma^i_{c,h,w}(s^i)$ and $\beta^i_{c,h,w}(s^i)$ are multiplied and added to the normalized activation function element-wise as follows:

$$SPADE_{c,\gamma_c,\beta_c}(x^i, s^i) = \gamma^i_{c,h,w}(s^i) \left(\frac{x^i - \mu_c(x^i)}{\sigma_c(x^i) + \epsilon} \right) + \beta^i_{c,h,w}(s^i), \tag{3.4}$$

3.3.3 Discriminator Architecture

During the conditional GAN training, the discriminator should also receive the information about the conditional input. In our case, this is the anatomy represented in the segmentation mask. The corresponding ground truth label is concatenated with both the real image and synthesized image (in a channel-wise manner) to construct two pairs of real image-label and synthesized image-label inputs. These two pairs of inputs are then passed to the discriminator in the same batch (as suggested in [75]) to provide two separate predictions for the real image and synthesized image conditioned on the label. Motivated by successful

GAN frameworks, we adopt a multi-scale discriminator for high-resolution image synthesis, to differentiate subtle differences between finer details presented in real and synthesized images [64]. Two identical PatchGAN discriminators [63] with four convolutional layers operate at two image scales that differ by a factor of 2 (see Figure 3.3).

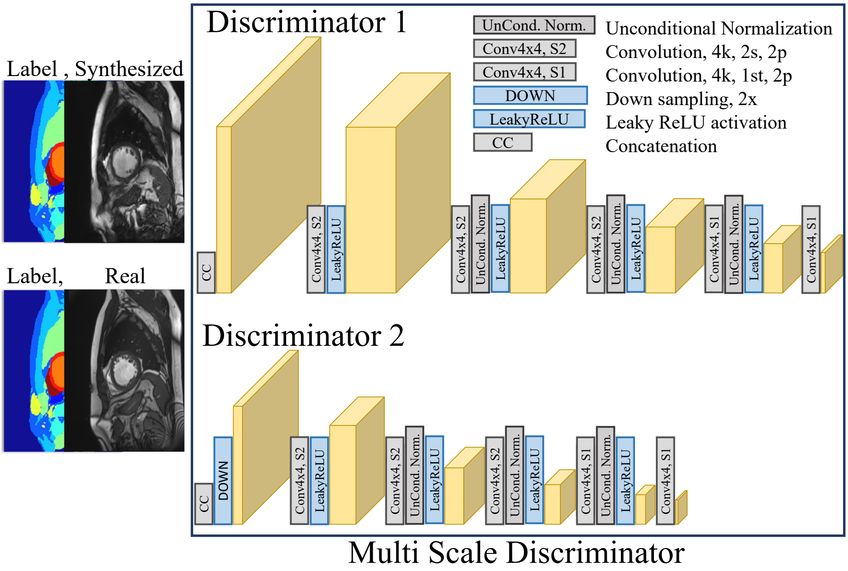


Figure 3.3: Multi-scale discriminator architectures including two discriminators, 1 and 2, having identical network structures but operate at two image scales, full scale and down-sampled by a factor of 2.

3.3.4 Objective Function

In the supervised conditional GANs for translating semantic segmentation maps to realistic images, the training set of paired images x_j with corresponding maps s_j is available as (x_j, s_j) during training. The conditional distribution of real images given the input segmentation maps is learned through alternate optimization of the discriminator's objective function $\mathcal{L}_D$ and the generator's objective function $\mathcal{L}_G$. The so-called Hinge loss, introduced by [77] is given as:

$$\begin{aligned}\mathcal{L}_D = &-\mathop{\mathbb{E}}_{(x,s)\sim p_{data}(x,s)}[min(0, -1 + D(x,s))] \\ &- \mathop{\mathbb{E}}_{s\sim p_{data}(s), z\sim p(z)}[min(0, -1 - D(G(s,z), s)]\end{aligned} \tag{3.5}$$

$$\mathcal{L}_G = -\mathop{\mathbb{E}}_{s\sim p_{data}(s), z\sim p(z)} D(G(s,z), s). \tag{3.6}$$

The GAN is trained via the following modified *minmax* game when using two discriminators D_1 and D_2 operating at different image scales:

$$\min_{G} \max_{D_1, D_2} \sum_{k=1,2} \mathcal{L}_{GAN}(G, D_k) \tag{3.7}$$

Inspired by many successful GANs for image synthesis [64,75,78], we employ an enhanced version of the *adversarial loss* in Equation 3.7 that is improved using a *feature matching loss* based on the discriminator for stabilizing the training of the generator. Feature maps from multiple layers of the discriminator are extracted for real and synthesized images for the *feature matching loss* calculated as follows:

$$\mathcal{L}_{FM}(G, D_k) = \mathop{\mathbb{E}}_{(x,s)\sim p_{data}(x,s)} \textstyle\sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D_k^{(i)}(s,x) - D_k^{(i)}(s, G(s,z)) \right\|_1 \right], \tag{3.8}$$

where D_k^i denotes the i-th layer of the discriminator D_k with the total number of T layers and N_i elements in each layer.

Another well-established practice to overcome the training stability of the GAN models and increase the quality of generated images is to employ a perceptual loss for training the generator, in addition to the adversarial loss [79]. Training with a perceptual loss achieves superior performance for various image transformation tasks [80–82] and its importance has been acknowledged in the domain of medical imaging, as it captures the perceptual quality of the small structures in images [83,84].

Let $\phi_i(x)$ be a VGG19 pre-trained network for extracting feature maps from its i-th layer when passing the image x, M_i be the number of elements, and N be the total number of layers used to calculate the loss. The *perceptual loss* is defined as the mean absolute error between feature representations of the real and generated images, calculated as:

$$\mathcal{L}_P(G) = \sum_{i=1}^{N} \frac{1}{M_i} \left\| \phi^{(i)}(x) - \phi^{(i)}(G(s,z)) \right\|_1 . \tag{3.9}$$

The extracted features in the *perceptual loss* are based on the VGG19 pre-trained model rather than the discriminator model in the *feature matching loss*, therefore independent from discriminator training. The overall objective function, composed of the weighted sum of the above mentioned components with hyper-parameters of $\lambda_1 = 10$ and $\lambda_2 = 10$ for the contribution of each loss term, is minimized during the training process:

$$\min_{G} \left(\left(\max_{D_1, D_2} \sum_{k=1,2} \mathcal{L}_{GAN}(G, D_k) \right) + \lambda_1 \sum_{k=1,2} \mathcal{L}_{FM}(G, D_k) + \lambda_2 \mathcal{L}_P(G) \right). \tag{3.10}$$

3.4 Material and Experiments

Simulated and real CMR images are utilized for training the multi-tissue segmentation and synthesis modules.

3.4.1 Simulated Cardiac MR Image Database

For training the supervised multi-tissue segmentation network of our first module, we require images with labels. We leverage the results from our concurrent project of cardiac MR image simulation and utilize our publicly available generated images provided in the context of the openGTN project[1] including 100 virtual subjects with anatomical and contrast variations. The anatomies of virtual subjects are derived from the 4D XCAT phantoms [7] and the images are simulated via a physics-based simulation tool that implements the Bloch equations for cine study. Each simulated image is provided with its corresponding ground truth label map including all simulated organs and tissue types. We create the multi-tissue segmentation map for the simulated images by combining labels for similar-looking organs to have a simplified representation of anatomies visible in the field of view. The segmentation map is comprised of separate labels for left and right ventricles, myocardium, lung, skeletal muscle, skin fat and abdominal organs. An example can be found in Figure 3.1.

3.4.2 Real Cardiac MR Image Database

For training the image synthesis network, the public dataset from the Multi-Centre, Multi-Vendor & Multi-Disease (M&Ms) challenge[2] [46] is used. The M&Ms data include patients and healthy controls with hypertrophy and dilated cardiomyopathy scanned at clinical centers in three different countries using MR scanners of different vendors, referred to as vendor A1 (Philips scanner center 1), B2 (Siemens scanner center 2), B3 (Siemens scanner center 3). Throughout our experiments, we employ the training subset of data from these two scanner vendors and three clinical centers while the testing subset is completely unseen during the synthesis experiments and it is only used for the final evaluation of the segmentation performance. Expert annotations are only available for left and ventricles, and myocardium, provided for 75 subjects of vendor A1, 50 subjects of vendor B2 and 25 subjects of vendor B3, for end-diastolic (ED) and end-systolic (ES) phases of the heart. We apply our trained multi-tissue segmentation network on all data, but only utilize the data from the first two vendors (A1, B2) for training our image synthesis model, while the last one (B3) is used for evaluation of the synthesis quality of the generated images.

[1] Simulated data can be found at `https://opengtn.eu/`

[2] M&Ms data can be found at `https://www.ub.edu/mnms/`

3.4.3 Data Preparation

Before training the segmentation network, all simulated CMR images are re-sampled to $1.25 \times 1.25 mm$ across short-axis slices, cropped to the same size of 256×256 and normalized with a mean of 0 and standard deviation of 1. Standard data augmentations such as random scaling and rotations, mirroring and horizontal/vertical flips are applied during training. Before training the synthesis network, all real images are pre-processed by re-sampling them to $1.33 \times 1.33\ mm$ in-plane resolution and taking a central crop of the images with 256×256 pixels. The pixel intensity value of images is first clipped between 0 and 4000, assuming 12-bit value, and then normalized between the range of $[-1, 1]$. We only apply random horizontal and vertical flips to images during training with 0.5 probability of the image being flipped. All synthesis models for experiments in this chapter are trained with the same training parameters using 3 NVIDIA TITAN Xp GPUs. To have more accurate labels for three heart classes, we correct the raw predicted labels of the first network by substituting the heart annotations provided by the challenge. Furthermore, we remove the slices above basal and below apical locations of the heart that do not contain any heart labels. Therefore, every slice seen during synthesis contains at least one of the heart labels.

3.4.4 Analysis of the Framework

We conduct experiments to qualitatively (noted as **Ql**) and quantitatively (noted as **Qt**) analyze different elements of the framework and their effects on the final outcome:

Ql - Effects of multi-tissue vs. cavity-tissue labels: We evaluate the need for utilizing the multi-tissue segmentation module with an ablated version of our framework, where we train the image synthesis model with only cavity labels. We compare the results with the network trained with full labels, with identical training parameters.

Ql - Effects of Multi-vendor Data: We train two identical multi-tissue generator using training images from the M@Ms challenge data-set [46], acquired at different clinical centers by different scanner vendors, to explore the possibility of learning a scanner-specific style of CMR images and the generator's ability to synthesize multiple appearance on a given set of labels. The generator trained on data of vendor A1 and B2 are named GenA1 and GenB2, respectively. When the labels from the other data is used for synthesis, we name this cross-vendor synthesis between (e.g. GenA1oB2 is the result of synthesizing using GenA1 model on labels from vendor B2 images). One example is shown in Figure 3.4.

Qt - Synthetic Image Quality Metrics: For quantitative assessment of the synthesis quality, we compute structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and normalized root mean squared error (NRMSE), between the synthesized images using GenA1 and GenB2 models on the common labels of the B3 (n=50) and the corresponding real images, completely unseen during training.

Ql, Qt - Effects of Amount of Data: We perform a data reduction experiment to investigate the effect of the number of images used during the training of the image synthesis module on the quality of the synthesized images at inference time. We train different models using a fraction of real training data and compare the quality of the generated images using multi-tissue or cavity-tissue generators.

3.4.5 Human Visual Scoring of Synthetic Images

The quality and realism of the synthesized cardiac MR images are evaluated using a visual scoring tool. We present to the evaluators the results of the cross-synthesis experiment shown in Figure 3.4 which includes images of GenA1oB2 (n=30) and GenB2oA1 (n=30) together with the real counterpart images of RealA1 (n=20) and RealB2 (n=20) in a randomized order, and the evaluator is unaware whether the displayed image is real or synthesized. To include sufficient heart coverage in each displayed image, the mid ventricular slice of each patient is chosen.

One set of experiments with three phases and three questions with multiple choices is designed. To avoid exhausting the evaluators by one time-consuming experiment, the same experimental setup is repeated with different subsets of images (maximum of 40 images for each round). Training phase includes 20 cardiac MR images to get the evaluator familiarized with the scoring setup, questions, and the types of images they can expect to score. Testing phases 1, 2, and 3 including 40, 40 and 20 cardiac MR images that serve as the main experiments for evaluating and scoring the quality of the cardiac MR images. For each displayed image, we ask three questions with multiple choice options: i) How do you evaluate the overall quality of the image? ii) How do you evaluate the image realism with focus on the heart? with a scoring on a scale of one to five, with one being worst score (very poor quality) and five being best score (very good quality), and iii) How confident are you that the image is real or synthesized? (Synthesized, Maybe synthesized, Cannot tell, Maybe real, Real)

3.4.6 Utilization of Synthetic Data and Experimental Design

We employ synthesized images for training a DL model to segment left and right ventricular cavities and myocardium of the heart. The purpose of these experiments is to investigate two aspects of the synthetic data: i) its ability to mimic the distribution of the real training data and therefore, its effectiveness in training a well-performing segmentation algorithm, which we refer to as a *replacement* experiment and ii) its usability for augmentation purposes, where we quantify the

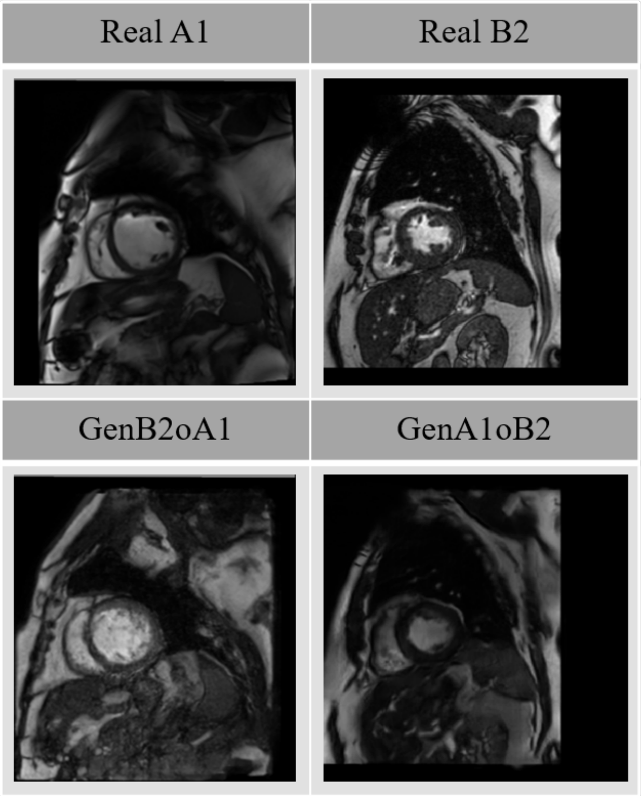


Figure 3.4: Cross-synthesis between vendors A1 and B2. GenB2oA1 image is synthesized using the model trained on vendor B2 data and evaluated on the multi-tissue label derived from images of vendor A1, and vice versa for GenA1oB2.

effect of augmenting the real MR data during training with synthetic data, referred to as an *augmentation* experiment.

To this end, we utilize the results of a cross-vendor experiment explained earlier, which aims to transfer the appearance of the images acquired using the scanner of one vendor on the anatomical masks derived from the subjects scanned using another vendor. Additionally, we vary the heart shape during synthesis by applying random elastic deformation and morphological dilation on heart masks (just before image generation) to introduce anatomical variation in the synthetically generated data. An animated GIF of applying multiple deformations on the heart labels during synthesis is available at `https://bit.ly/3juJE1v`.

The baseline model is a 2D U-Net trained with a combination of real images of vendor A1 and B2 (**Real**). We compare the performance of this model with four identical ones trained with: i) only synthetic data with cavity-tissue labels

(**Synth-Cavity**), ii) only synthetic data with multi-tissue labels (**Synth-Multi**) for the *replacement* experiment, iii) augmenting the real data with cavity-tissue synthesis (**Aug-Synth-Cavity**), and vi) augmenting the real data with multi-tissue synthesis (**Aug-Synth-Multi**) for the *augmentation* experiment.

All networks consist of six downsampling and upsampling convolutional blocks with five max-pooling operations. Each convolutional block contains 3x3 kernel convolutional layers, batch normalization and leaky ReLU activation function. We additionally apply dropout regularization, with a rate of 0.5, after each concatenating operation in the up-sampling path of the network. To increase robustness of the networks, we augment the training set by applying random vertical and horizontal flips (p=0.5), random rotation by integer multiples of $\frac{\pi}{2}$ (p=0.5), random translations (p=0.3) and mirroring (p=0.5), as well as random elastic deformations (p=0.4). We do not apply any contrast transformations, to better inspect the influence of synthetic data. All augmentation operations are applied on the fly during training. At inference time, we only apply normalization and in-plane re-sampling.

To train the network, we use a weighted sum of the categorical cross-entropy and Dice loss. We use Adam for optimization, with an initial learning rate of 5×10^{-5} and a weight decay of $3 \times e^{-5}$. During training, the learning rate is reduced by a factor of 5 if the validation loss does not improve by at least 5×10^{-3} for 50 epochs. We apply early stopping on the validation set to avoid over-fitting and select the model with the highest accuracy. As a post-processing step, we perform connected component analysis on the predicted labels, where we remove all but the largest connected component for each class and therefore, remove large false positive predictions.

3.5 Results

3.5.1 Effects of Multi-Tissue vs. Cavity-Tissue Labels

Utilizing multi-tissue segmentation labels substantially improves the quality of the synthesized images by providing more guidance to the generator. Figure 3.5 depicts the results of utilizing cavity-tissue compared with the use of multi-tissue labels during synthesis, for two synthetic subjects for six slices covering the heart. Although organ labels are not highly accurate for the case of multi-tissue labels, the generator is able to synthesize realistic anatomies. Plausible organs are generated around the heart when we use more labels for image synthesis. Furthermore, we observe from consecutive slices of one subject (An animated GIF is available at `https://bit.ly/3fyOFGt`) that the generated anatomy is consistent in 3D, making results suitable for 3D medical image analysis. Note that the synthesis network is 2D, taking one slice at a time for image generation. Distortions in the generated images for cavity-tissue are highlighted with yellow arrows.

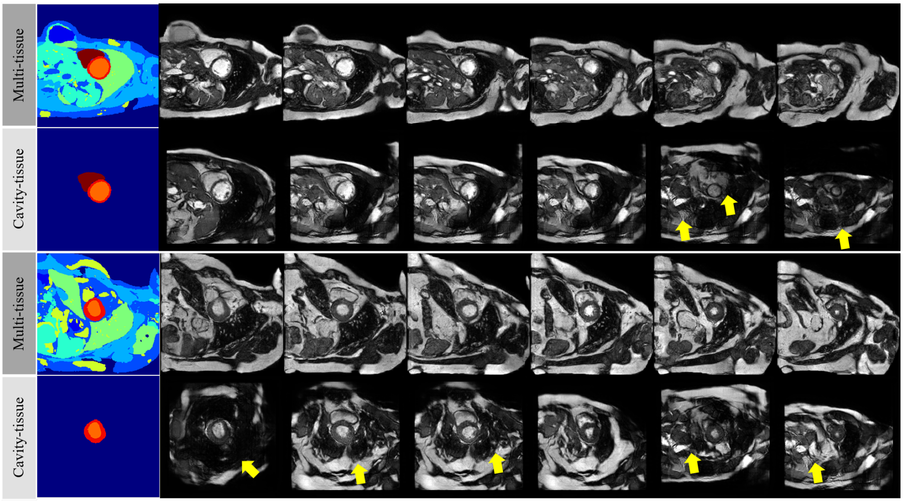


Figure 3.5: Visual comparison of synthesis results for two patients for the cases of using multi-tissue labels and only cavity labels. The first column shows the corresponding segmentation of the second column and the following columns show image slices from basal location to the apex of the heart. Distorted parts of anatomies in the latter ones are depicted with yellow arrows. The image quality and volumetric consistency of the generated anatomy have improved in the synthesized images with multi-tissue labels. An animated GIF version of two examples is available at `https://bit.ly/3fy0FGt`

3.5.2 Effects of Multi-Vendor Data

Figure 3.6 shows the generation results on common labels derived from images of vendor B3, for the generator trained with images of vendor A1 *(GenA1oB3)*, and the generator trained with images of vendor B2*(GenB2oB3)*. From the visual comparison of synthesis results, we observe that the main characteristics of the training images are captured by the generators, allowing them to synthesize multiple appearances on a given label. Vendor-specific characterizations, such as the darkness and blurriness of the images acquired from vendor A1, the sharpness of the edges and the noisiness observed in images from vendor B2, and myocardium-to-blood contrast, are learned by the generators and manifested in the synthesized images.

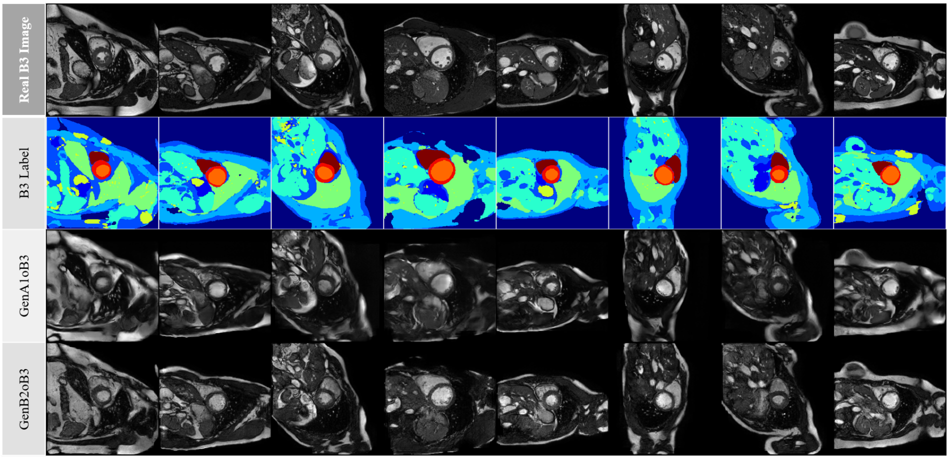


Figure 3.6: Synthesis results on labels derived from vendor B3 data for model trained on data of vendor A1 (GenA1oB3) and the same model trained on data of vendor B2 (GenB2oB3). The difference in the appearance of the synthesized images demonstrates that the generator learns the vendor-specific features of the training images.

3.5.3 Synthetic Image Quality Metrics and Effects of Amount of Data

The images are generated on *Cavity* or *Multi* tissue labels using models trained with combined data of vendor A1 (n=150) and B2 (n=100), *GenA1B2*, separate data of vendor B2, *GenB2*, 50% of vendor B2 data, *Gen50%B2*, 25% of vendor B2 data, *Gen25%B2*, and 12% of vendor B2 data, *Gen12%B2* as depicted in Figure 3.7. The results of the generator utilizing multi-tissue labels achieve the best score for all metrics, suggesting that the better quality is achieved for multi-tissue synthesis, despite the reduction in the training data. Interestingly, we observe that the *Multi-Gen12%B2* generator, which uses only 12 volumes for training, scores significantly higher across all metrics than even the *Cavity-GenA1B2* that uses all 250 real images during training, suggesting the benefit of multi-tissue labels when limited data is available during synthesis.

Figure 3.8 depicts synthesis results on the same subjects generated by the models trained with a fraction of the training data of vendor B2. We observe that the synthesis quality of the multi-tissue generator is retained even when only 12% of the data (12 volumes) is used for training, compared to the generator models trained with the full available data-set. On the other hand, repeating the same experiment on the generator trained with cavity-tissue labels produces images of impaired quality and severe distortion.

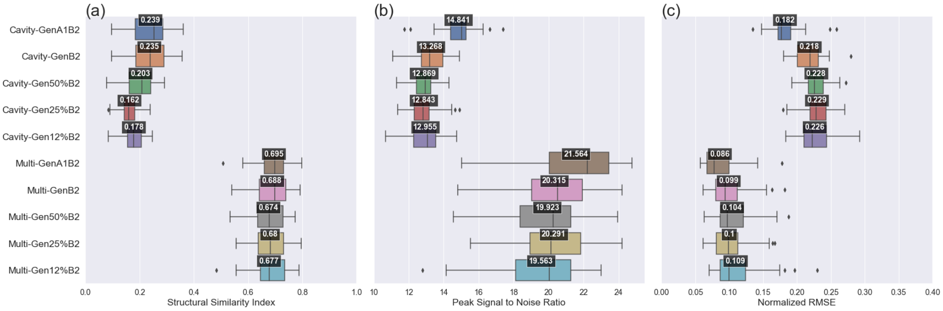


Figure 3.7: Quantitative evaluation metrics for comparing the quality of the synthesized images using models trained with multi-tissue (*Multi-*) and cavity-tissue (*Cavity-*) labels. Structural similarity index (a), peak signal to noise ratio (b) and normalized root mean squared error (c) are calculated between real images and corresponding synthesized counterparts. *GenA1B2* and *GenB2* indicate that the synthesized data are generated by the generator trained on the combination of data from vendor A1(n=150) B2(n=100) and only vendor B2 (n=100), respectively. 50%B2, 25%B2, and 12%B2 indicate models trained with fraction of the data from vendor B2 (n=50, 25, 12). All models are tested on the unseen data of vendor B3 (n=50).

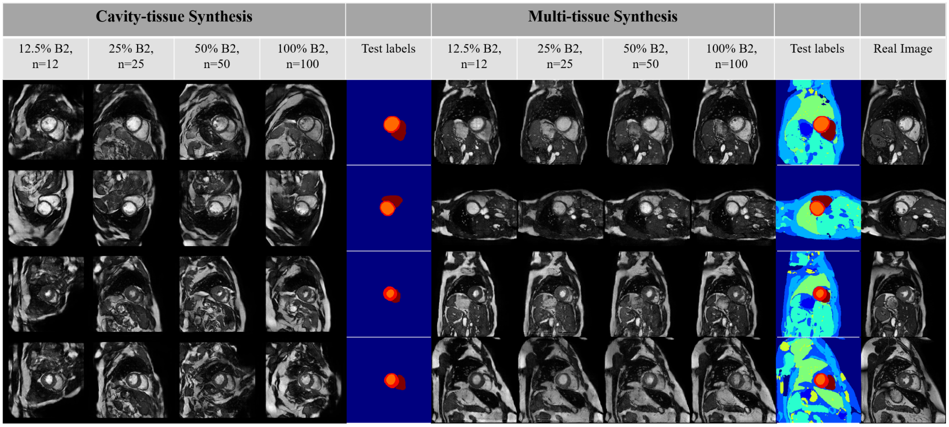


Figure 3.8: The data reduction results for the case of multi-tissue and cavity-tissue generators trained using the data from vendor B2 (center 2, n=100) and tested on the unseen ground truth labels of B3 (center 3, n=50). The corresponding real testing images are shown in the first row.

3.5.4 Human Visual Scoring of Synthetic Data

Two independent raters completed the experiments. Evaluator 1 was an imaging scientist with ample experience in CMR application and processing, who was involved in this image synthesis research and has experience looking at GANs-generated medical images, and evaluator 2 was a CMR clinical expert and consultant cardiologist, who was not involved in this project. Figure 3.9 depicts the scores of the evaluators for testing phases. The evaluation results for two sources of the real (RealA1, RealB2) and two sources of the synthesized images (GenA1oB2, GenB2oA1) are combined. The results for the total of 40 real and 60 synthesized images are shown in this figure. The horizontal axis shows the level of the image quality score, and the vertical axis shows the percentage of the total images rated with that image quality score. e.g. around 70% of the synthetic data compared to around 60% of real images are perceived as *Good* in terms of image realism with focus on the heart (scored 4). Surprisingly, lower overall quality of the real images is scored by the evaluator 1, having more real images with scores of 1 *very poor* and 2 *poor*, while same percentage scored 4 *Good*, and slightly more synthesized scored 3 *Mediocre*. The overall quality of both images are higher for the evaluator 2. From the last question, the evaluator 1 is confident that only just 10% of the synthesized images are *"Synthesized"*, around 55% *"Maybe Synthesized"* and the rest are either unidentifiable or rated as *"Maybe Real"*. For the evaluator 2, interestingly, the majority of the synthesized cases are rated either *"Cannot Tell"* or *"Maybe Real"*. One conclusion to draw is that the synthesized images are on par with real images in terms of visual quality, sometimes perceived as better quality, and they are indistinguishable from the real counterparts even by experienced evaluators.

3.5.5 Utilization of Synthetic Data

We utilize synthetic CMR images with the aim to investigate the effectiveness and utility of such data in training a deep-learning algorithm for segmentation, evaluated by carrying out the *replacement* and *Augmentation* experiments. Hence, we provide a quantitative, application-based evaluation of the usability of synthesized images for a clinical task to segment myocardium (MYO), left ventricle (LV), and right ventricle (RV). All models are tested on the same set of images, acquired from two vendors and unseen during the training and development of the models. We utilize standardized metrices for evaluation of segmentation performance in the literature, namely the Dice coefficient and the Hausdorff Distance (HD) score.

We start by evaluating the *replacement* experiment, where we study the extent to which the data synthesized in this study can be used to train a model for segmenting the heart cavity tissues in real CMR images. By running this experiment, we can additionally quantify the realism of such data and possibly identify points of improvement for future work. The results in this experiment are shown for separated classes in Figure 3.10 a) and b). First, we observe that both models trained with synthetic data exhibit a drop in performance compared

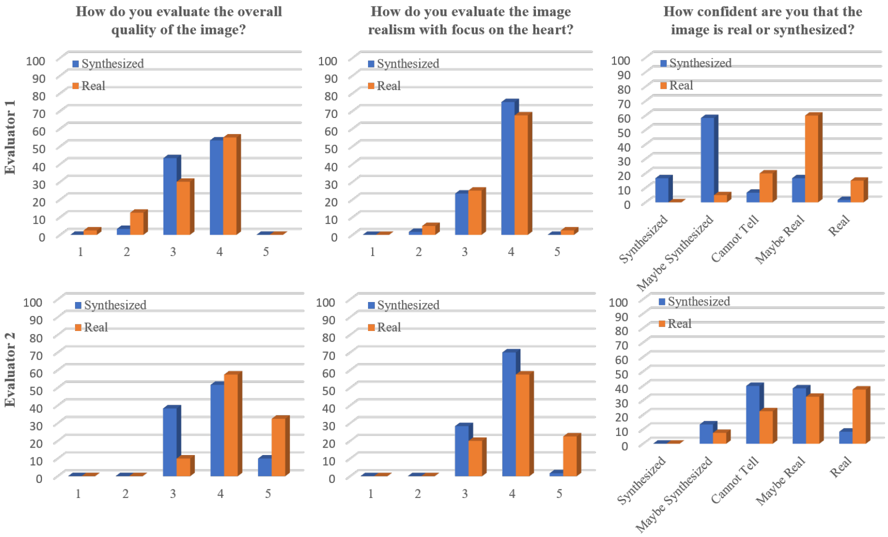


Figure 3.9: The visual quality scoring results for two evaluators (1: Scientist, 2: CMR specialist). The outcome of all testing experiments for two sources of the real (RealA1, RealB2) and two sources of the synthesized images (GenA1oB2, GenB2oA1) are combined. This includes the total of 40 real and 50 synthesized images. The scores correspond to; 1: very poor, 2: poor, 3: mediocre, 4: good, and 5: very good quality.

to the model trained with real data (*Real*). This result was expected due to the nature and characteristics of GAN-based conditional synthesis. However, we see a consistent and significant improvement in the performance of the *Synth-Multi* model compared to the *Synth-Cavity* model, where in some cases *Synth-Multi* performs almost at par with the *Real* model, across both vendors and all tissues. We hypothesize this is due to better image quality achieved through synthesizing with multi-tissue labels, where tissue boundaries are much clearer, with better contrast achieved between the tissues. Moreover, this indicates a better resemblance of the *Synth-Multi* data features to the real data. Using the *Synth-Multi* model, we notice a significant improvement in reducing the appearance of outliers, but also improving the average Dice score across all patients. Most of these outlier predictions tend to be located across basal and apical slices, which can be partly attributed to the removal of unlabeled slices above the base and below the apex of the heart during synthesis. We additionally observe false positive predictions generally appearing across all slices using the *Synth-Cavity* model, where tissue similar in shape, size or appearance to the heart cavity is often falsely identified as a part of the heart. These errors are reduced by utilizing images of better quality for training the *Synth-Multi* model. Being sensitive to false positive predictions, the HD scores for the *Synth-Cavity* model tend to be higher with a higher percentage of outlier cases across all vendors and tissues.

The results for the *augmentation* experiment are shown in Figure 3.10 in rows c) and d). This experiment shows that the addition of synthetic data to the training set containing real data improves the segmentation performance across both vendors and all tissues. This happens for both cases (*Aug-Synth-Cavity* and *Aug-Synth-Multi*). However, adding *Synth-Multi* images to the training is shown to be consistently better than *Synth-Cavity* both in terms of Dice score and HD score, especially when it comes to reducing the variance and outliers of the model performance. This is particularly evident in scores acquired across all vendor B tissues, as well as in vendor A myocardium, which are significantly higher compared to the *Aug-Synth-Cavity* model ($p < 0.01$ according to the Wilcoxon signed-rank test). In fact, visual observations indicate that the *Aug-Synth-Multi* model consistently yields segmentation improvements due to:

- the reduction of false positive and false negative predictions around the base and apex of the heart, respectively;
- better segmentation of cases with a thicker myocardium, often appearing in ES images, which we hypothesize is the influence of wider anatomical and contrast variations introduced to the training through synthetic images;
- consistent improvements in the segmentation of the RV around the base of the heart;
- better segmentation robustness in images with visible acquisition artefacts and noise and better adaptation to cases with poor tissue contrast.

Moreover, models trained on real images only are prone to under-segmentation in the occurrence of hyperintensities, which is largely alleviated with the addition of synthetic data.

Some of the above-mentioned improvements are also present in the *Aug-Synth-Cavity* model. However, the under- and over-segmentation of tissues in apical and basal slices still remains a problem. Despite *Synth-Cavity* images being of lower quality, we hypothesize that an improvement in performance compared to the *Real* model is partly achieved as such images present harder examples during the training procedure, thus improving network regularization and model generalization. A detailed summary of the experiments is presented in Table 3.1.

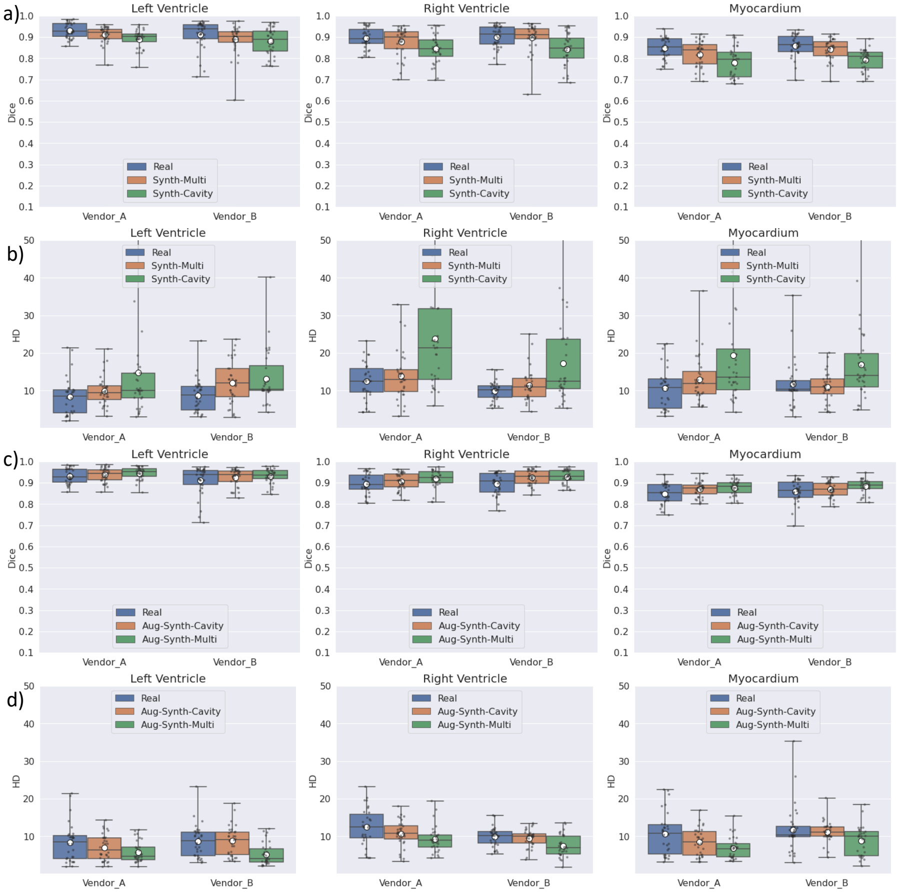


Figure 3.10: Quantitative usability evaluation of the synthesized data for two experiments of data *replacement* (a,b) and data *augmentation* (c,d). Identical 2D segmentation networks are trained using different training data combinations for the task of cardiac segmentation. The baseline model, *Real*, is trained using all available real training examples of vendor A1 and B2 (n=250); *Synth-Cavity* and *Synth-Multi* are models trained completely using synthetic images when cavity-tissue and multi-tissue labels are employed during the GANs training, respectively. *Aug-Synth-Cavity* and *Aug-Synth-Multi* are models trained with augmented data of real with cross-synthesis results. We apply geometrical deformations on heart labels during synthesis to introduce more anatomical variations. The Dice coefficient, *Dice* (higher better), and the Hausdorff distance (pixel), *HD* (lower better) are reported for segmentation performance.

Table 3.1: The performance of segmentation models for data *replacement* and *augmentation* experiments

	Vendor A						Vendor B					
Segmentation models	Dice(LV)	Dice(RV)	Dice(Myo)	HD(LV)	HD(RV)	HD(Myo)	Dice(LV)	Dice(RV)	Dice(Myo)	HD(LV)	HD(RV)	HD(Myo)
Real	0.932	0.894	0.847	8.32	12.44	10.70	0.911	0.896	0.860	8.75	9.89	11.78
Synth-Cavity	0.890	0.845	0.780	14.75	23.86	19.44	0.880	0.842	0.794	13.19	17.22	17.01
Synth-Multi	0.912	0.877	0.819	9.98	13.89	12.92	0.888	0.898	0.842	12.10	11.50	10.91
Aug-Synth-Cavity	0.938	0.907	0.869	6.99	10.72	8.48	0.926	0.926	0.871	8.85	9.51	11.20
Aug-Synth-Multi	**0.945**	**0.920***	**0.877***	**5.83**	**9.22***	**6.83***	**0.932**	**0.929***	**0.883***	**5.24***	**7.47***	**8.85***

Numbers with * have p-value of $p < 0.01$ against the Real model according to the Wilcoxon signed-rank test and numbers with Bold text are the best scores.

3.5.6 Conditional GANs Comparison

We compare the proposed label conditional GANs model (based on the SPADE normalization layers) with pix2pix [63] and pix2pixHD [64] models in terms of synthesized image quality. All three models are trained with the same hyper-parameters using the full data of vendor A1 and B2 with multi-tissue labels and evaluated on unseen data from vendor B3. We can observe from the synthesis results shown in Figure 3.11 that more anatomical distortions are produced by pix2pix and pix2pixHD models (indicated by yellow arrows), especially in the heart area. The SPADE-based generator can better preserve the content of the input labels, resulting in more accurate generation of the heart. The calculated image quality metrics (as described in 3.4.4 **Synthetic Image Quality Metrics**) are shown in Figure 3.12, confirming the superiority of the SPADE-based generator, strongly for the structural similarity index.

3.6 Discussion

3.6.1 Multi-Tissue vs. Cavity-Tissue Labels for Synthesis

For synthesizing realistic cardiac MR images we conduct experiments to investigate the effects of using multi-tissue segmentation compared with using only cavity-tissue labels during synthesis. The quality of the results achieved by the former setup is significantly better, demonstrating superior image quality with fewer distortions and artifacts. We observe substantial benefits of using a rough segmentation mask for partitioning the input image into multi-tissue labels. Firstly, despite the fact that our framework is two dimensional (2D), i.e. it takes 2D labels as the input and produces 2D images, the global consistency for volumetric image generation seems to be improved. Our results are consistent in the slice direction and therefore suitable to generate data for 3D image analysis algorithms. Secondly, we observe that the training time and stability is improved when training the synthesis module with multi-tissue labels, resulting in a fewer number of epochs required and a faster learning procedure. Finally, our experiments show that even with as few as 12 real volumes accompanied with rough multi-tissue labels, we can synthesize realistic images, despite the loss high-frequency details of small

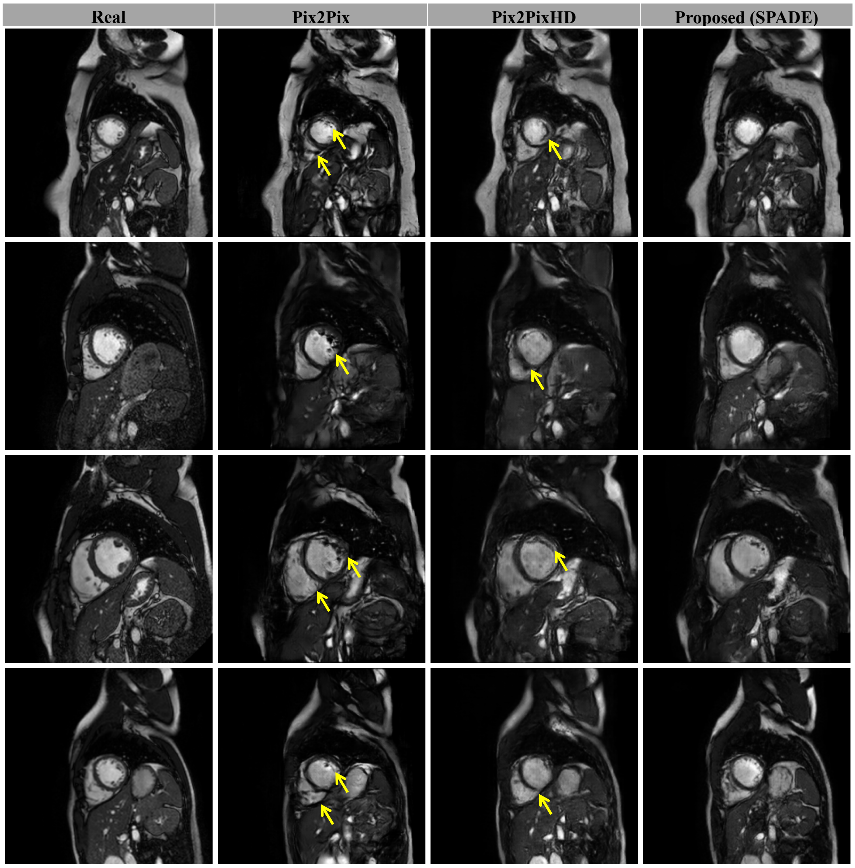


Figure 3.11: Visual comparison between different generator models trained with the same hyper-parameters using the combined data from vendor A1 and B2 and evaluated on the unseen data from vendor B3. Yellow arrows indicate geometrical distortions in the heart area for pix2pix [63] and pix2pixHD [64] models.

structures. From the visual assessment of the results we observe wrong anatomical positioning in some synthesized images on cavity-tissue labels, i.e the cavity-tissue generator is not able to correctly locate anatomical structures, such as lung and abdominal organs surrounding the heart, as seen in Figure 3.8 second and forth rows for the cavity-tissue generator. The correct heart position, with respect to lung and other organs, is preserved for the multi-tissue generator, resulting in synthesis of plausible anatomies. Note that we train our multi-tissue segmentation model using simulated cardiac MRI data while one could replace that module with another approach or manually segment the images.

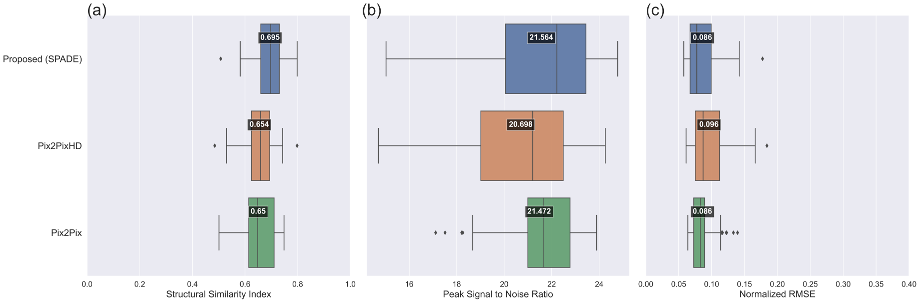


Figure 3.12: Quality metrics for quantitative evaluation of the results for the proposed synthesis model in this study (SPADE), pix2pix [63] and pi2pixHD [64] models. We compare the generated image quality based on a) structural similarity index (higher better), b) peak signal to noise ratio (higher better) and c) normalized root mean square error (lower better) between the generated and real images for unseen data from vendor B3(n=50 subjects).

3.6.2 Synthesis Quality and the Amount of Data

We conduct the data reduction experiment to assess the effects of the number of training images on the synthesis quality, comparing the results of generating on multi-tissue and cavity-tissue labels. The quantitative metrics shown in Figure 3.7 suggest that the quality of the synthesized images on multi-tissue labels is significantly better than synthesized images on cavity-tissue labels, increasing the structural similarity index from 0.239 to 0.695 for the case of utilizing all images from vendor A1 and B2 (total of 250 volumes). Gradually reducing the training data causes drastic degradation in the quality of synthesized images on cavity-tissue labels, whereas it does not severely affect the global quality of the synthesized images on multi-tissue labels. For example, the quality of multi-tissue generated images for the case of using only 12%B2 (n=12) constantly outperforms the cavity-tissue generated ones using all 250 volumes of A1 and B2. This suggests that there are significant advantages of putting effort into obtaining multiple tissue labels when there is a limited amount of data available for training, as shown in Figure 3.7. Furthermore, as seen in Figure 3.8, while in images obtained by the cavity-tissue generator the anatomy of organs and structures are distorted, we observe that only the high-resolution details of the image, such as the papillary muscles and trabeculation of the heart are missing in images generated by the multi-tissue generator for the cases of reducing the number of data during training.

3.6.3 On the Usefulness of the Synthetic Data

Despite recent attempts to synthesize realistic cardiac magnetic resonance images, the analysis of how useful the generated data is for tackling a real world task remains largely unexplored. Our experiments show that there is a performance drop when real data is completely replaced by synthesized data for the task of cardiac cavity segmentation. This implies that some of the features of the real data are not fully captured during synthesis. However, the addition of the synthetic data can substantially improve the model performance, indicated by its benefits for data augmentation. We observe a maximum increase of 4% for Dice and a maximum reduction of 40% for HD in the presented experiments. The HD score is significantly improved when real data is augmented with multi-tissue synthetic data, suggested by the reduction in the outlier predictions and an overall improved segmentation accuracy of tissue boundaries. Based on visual inspection of the segmentation results, aiding the training with multi-tissue synthetic images yields a better segmentation performance around the apex and base of the heart and more accurate segmentation of the right ventricular cavity and myocardium. Our observations suggest that the improvements could be attributed to model robustness to artefacts, noise and poor tissue contrast, but further investigation is needed, which is planned for future work. Despite the fact that we observe better performance when using synthetic multi-tissue data, the cavity-tissue images with lower visual quality are still found to be helpful for data augmentation. This could imply that by seeing somewhat more distorted generated images, the model can learn to deal with some aspects of the data such as unclear tissue boundaries, blurriness, and other imaging artifacts, which are unrepresented in the real training data. Similar results have been reported in [85], where the segmentation performance can be improved with visually low-quality generated cardiac MR images.

3.6.4 Limitations and Future Research

In this work, we synthesize cardiac MR images for cine study with imaging features learned according to the acquisition protocol of the training data. However, multi-modality image synthesis can be achieved by attaching an encoder network to the generator for capturing the modality-specific style of images, e.g. image appearance of the late gadolinium enhancement. As a result, cross-modality image synthesis, defined as domain translation from cine to late enhanced cardiac MR images (or to cardiac CT image), could potentially address the challenge of data availability in a wider medical imaging context. Furthermore, with the ability to manipulate labels for synthesizing a new image, it is possible to generate a population of virtual subjects with heart diseases, as long as the disease can be represented in the tissue mask. We show that the generated image quality increases by utilizing multi-tissue labels even with noisy and inaccurate anatomy. However, to synthesize highly detailed anatomies, one could work on the accuracy of the multi-tissue segmentation network to provide more accurate labels for better guidance of the synthesis network. In fact, the multi-tissue segmentation module could be replaced with any available method to obtain detailed segmentation maps of both the heart and the tissues surrounding the heart. To evaluate the usefulness and effectiveness of synthetic data, we perform experiments for the task of medical image segmentation; however, other medical image analysis tasks can also benefit from the achieved results in this work. Finally, more research should be focused at addressing the synthesis of artifacts and other specific characteristics of MR images that can potentially present a larger feature variability that DL-based models, trained for different medical imaging analysis tasks, can learn from.

3.7 Conclusion

In this chapter, we propose a framework for cardiac MR image segmentation and synthesis with the aim of generating high-quality images with informed anatomical representation and learned imaging features. Leveraging Label-conditioned normalization layers throughout the generator architecture allows for the preservation of content information, while at the same time accurately transferring the image characteristics of real data. Furthermore, one of the main findings of this work is the importance of introducing detailed labels in the form of multi-tissue maps for generation of highly realistic images with accurate anatomies even with a significantly smaller number of training data, compared to utilizing only cavity-tissue labels during training. The effectiveness and usability of synthetic images for the task of cardiac segmentation was evaluated, demonstrating that data augmentation with synthetic data can substantially boost the segmentation performance in terms of both Dice score (maximum increase of 4%) and Hausdorff Distance (maximum reduction of 40%).

CHAPTER 4

Cardiac Pathology Image Synthesis

This chapter is based on:
S. Amirrajab, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “Pathology Synthesis of 3D Consistent Cardiac MR Images Using 2D VAEs and GANs,” submitted, 2022.
Which is an extension of:
S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “Intra- and Inter-Subject Synthesis of Cardiac MR Images Using a VAE and GAN,” ISMRM 2022.
S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “Pathology Synthesis of 3D Consistent Cardiac MR Images Using 2D VAEs and GANs,” SASHIMI workshop MICCAI 2022.

Abstract

We propose a method for synthesizing cardiac MR images with plausible heart pathologies and realistic appearances for the purpose of generating labeled data for the application of supervised deep-learning (DL) training. The image synthesis consists of label deformation and label-to-image translation tasks. The former is achieved via latent space interpolation in a VAE model, while the latter is accomplished via a label-conditional GAN model. We devise three approaches for label manipulation in the latent space of the trained VAE model; i) **intra-subject synthesis** aiming to interpolate the intermediate slices of a subject to increase the through-plane resolution, ii) **inter-subject synthesis** aiming to interpolate the geometry and appearance of intermediate images between two dissimilar subjects acquired with different scanner vendors, and iii) **pathology synthesis** aiming to synthesize a series of pseudo-pathological synthetic subjects with characteristics of a desired heart disease. Furthermore, we propose to model the relationship between 2D slices in the latent space of the VAE via the correlation coefficient matrix to correlate random samples prior to reconstruction. This simple yet effective approach results in generating 3D-consistent subjects from 2D slice-by-slice generations. We demonstrate that such an approach could provide a solution to diversify and enrich an available database of cardiac MR images and to pave the way for the development of generalizable DL-based image analysis algorithms. We quantitatively evaluate the quality of the synthesized data in an augmentation scenario to achieve generalization and robustness to multi-vendor and multi-disease data for image segmentation. Our code is available at `https://github.com/sinaamirrajab/CardiacPathologySynthesis`.

4.1 Introduction

Deep generative modelling has gained attention in medical imaging research thanks to its ability to generate highly realistic images that may alleviate medical data scarcity [9]. The most successful family of generative models known as generative adversarial networks (GANs) [2] and Variational Autoencoders (VAEs) [3] are widely used for medical image synthesis [14, 53]. Many studies have proposed generative models to synthesize realistic and diversified images for brain [86, 87] and heart [13, 88] among other medical applications [83]. However, the generated data are often unlabeled and therefore not suitable for training a supervised deep learning algorithm, for instance, for medical image segmentation.

Despite the benefit of data augmentation and anonymization using synthetic data for brain tumor segmentation [89,90], the application of synthesizing labelled cardiac MRI data remained relatively under-explored with very limited recent attempts to synthesize cardiac images [91]. Recent work by [92] and [20] investigates the effectiveness of using conditional GANs for translating ground truth labels to realistic cardiac MR images that do not require manual segmentation and can be used for training a supervised segmentation model. However, the images are generated on a fixed set of input labels and therefore create similar heart anatomies, very limited to available ground truth labels of the training data, and more importantly, unable to synthesize subjects with cardiac pathology.

Contribution

We propose to break down the task of cardiac image synthesis into 1) learning the deformation of anatomical content of the ground truth (GT) labels using VAEs and 2) translating GT labels to realistic CMR images using conditional GANs. We devise different strategies to deform labels in the latent space of the VAE and generate various virtual subjects via three difference approaches, namely i) **intra-subject synthesis** to improve the through-plane resolution and generate intermediate short-axis slices within a given subject, ii) **inter-subject synthesis** to generate intermediate heart geometries and appearance between two dissimilar subjects scanned using two different scanner vendors, and iii) **pathology synthesis** to generate virtual subjects with a target heart disease that affects the heart geometry, e.g. synthesizing a pseudo-pathological subject with thickened myocardium for hypertrophic cardiomyopathy. All mentioned approached are accomplished via manipulation and interpolation in the latent space of our VAE model trained on GT labels, as demonstrated in Figure 4.1. The synthetic subjects in this study are labeled by design and therefore suitable for medical data augmentation. Furthermore, we propose a method to generate 3D consistent volumes of synthetic subjects by modelling the correlation between 2D slices in the latent space. The relationship between the slices is captured via estimating the covariance matrix calculated for all latent vectors across all slices. The estimated covariance matrix is used to correlate the elements of a randomly drawn sample in the latent space just before feeding it to the decoder part of the VAE. This technique results in a coherent sampling from the latent space and in turn reconstruction of more consistent 3D volume by stacking 2D slices generated from the 2D model.

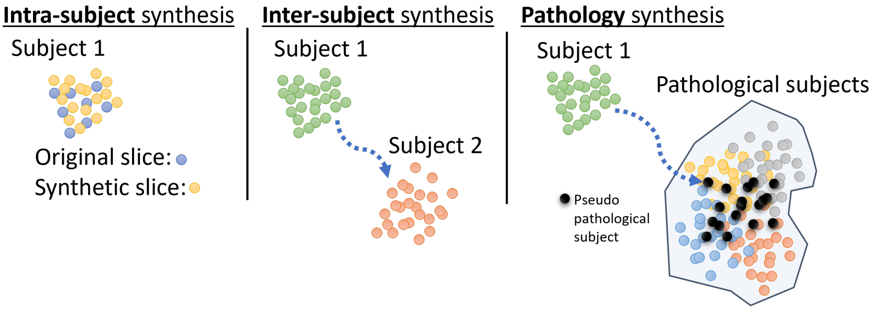


Figure 4.1: Three strategies to traverse and interpolate in the latent space to perform label deformation using the trained VAE model. Each encoded slice of a subject is represented as a dot in the low-dimensional latent space. The number of slices is increased using cubic interpolation in the latent space for intra-subject synthesis, and intermediate latent codes between two subjects (Subject 1 and 2) are generated using linear interpolation for inter-subject synthesis, indicated as dotted blue arrows. Assuming that all pathological subjects can be clustered in a neighboring location of the latent space, the statistics are estimated to draw a sample (pseudo-pathological subject) for pathology synthesis. Interpreting between Subject 1 and the pseudo-pathological subject results in generating subjects with pathological characteristics.

4.2 Methods

4.2.1 Image Synthesis Model

The synthesis model architecture includes a ResNet encoder [76] for extracting the style of an input image and a label conditional decoder based on Spatially Adaptive DE-normalization (SPADE) layers [75]. The SPADE layers preserve the anatomical content of the GT labels. After successful training of the model with pairs of real images and corresponding labels, the generator can translate GT labels to realistic CMR images. To alter the heart anatomy of the synthesized image, we can simply deform the labels. In the previous studies new subjects were synthesized by applying simple transformations such as random elastic deformation, morphological dilation, and erosion on GT labels [93,94]. The effectiveness of both elements of the synthesis model has been demonstrated in other recent work

[93, 94] and the synthetic data generated using this approach, despite generating unrealistic anatomies, boosted the performance of medical image analysis models for tissue segmentation in cine and LGE cardiac MRI data. We utilize the same synthesis network with default training parameters for this study and here we focus on label deformation to generate heart pathology using a VAE model.

4.2.2 Label Deformation Model

We propose a DL-based approach using a VAE model to generate plausible anatomical deformations via latent space manipulation to generate subjects with characteristics of heart pathologies. The VAE model consists of an encoder and a decoder network trained on the ground-truth label masks and tries to learn the underlying geometrical characteristics of the heart present in the labels. The changes in heart geometry can be associated with a specific type of disease. For instance, thickening and thinning of the left ventricular myocardium can be an indicating factor of hypertrophic and dilated cardiomyopathy, respectively. The goal here is to learn the effects of these factors on the heart geometry presented in the GT labels and to explore the latent space of the VAE to generate new labels with plausibly deformed anatomies. Additionally, we model the characteristics of particular heart diseases in the latent space and generate new samples with heart geometries that represent these disease characteristics.

A convolutional VAE model is designed and trained on GT labels of the heart to learn the underlying factors of different heart geometries presented in the database. After training, we encode all data into the latent space using the encoder part of the VAE and perform different operations to manipulate the learned features of the data, in this case the heart geometry. For instance, once labels are encoded into the latent space, we can traverse between two locations by simply interpolating between two latent codes and performing reconstruction to generate new heart geometries with intermediate anatomical shapes. These newly deformed labels are used as an input of the label-conditional GANs for image synthesis.

4.2.3 Approaches to Generate New Subjects

We explore three different approaches to generate new subjects via label deformation using our trained VAE model as shown in Figure 4.1. 1) **Intra-subject** synthesis to increase the number of short-axis slices per subject via interpolation between the latent codes of different slices of one subject. 2) **Inter-subject** synthesis to create subjects with heart geometry and imaging characteristic of between two different looking subjects acquired using two different scanner vendors 3) **Pathology** synthesis to generate pseudo pathological condition of a normal subject to explore the progression of a heart disease and its possible effects on the heart geometry.

Intra- and Inter-subject Synthesis

For intra-subject synthesis, we wish to increase the through-plane resolution of the SA slices for a subject. All slices (ranging between 6-13 slices per subject) are first encoded into the corresponding latent vectors. The latent vectors are then augmented by cubic interpolation to increase to 32 latent vectors, each representing one slice, which are then reconstructed by the decoder network to create labels for all 32 slices. All subjects will consist of 32 slices after the intra-subject synthesis. Note that the first and last slices are kept and only intermediate slices are interpolated.

Inter-subject synthesis aims to generate new examples with intermediate heart anatomy and appearance between two dissimilar subjects. To this end, intra-subject synthesis is first performed to equalize the number of SA slices for each subject in the latent space. Therefore, each encoded subject has 32 latent vectors associated with 32 interpolated slices. Then following the same procedure, intermediate latent vectors associated with in-between heart geometries are created by linear interpolation between the 32 latent vectors of the two encoded subjects. By decoding these newly interpolated subjects using the decoder part of the VAE, the heart geometry of one subject is morphed into the other one. Finally, the deformed labels are fed to the synthesis model for synthesizing new subjects.

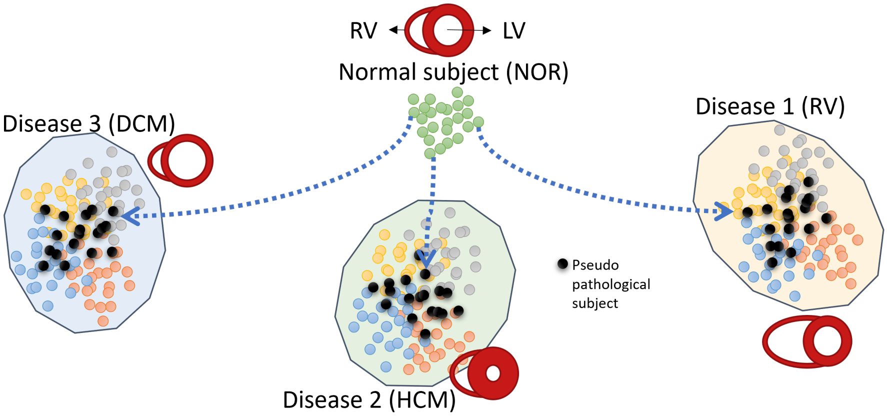


Figure 4.2: Pathology synthesis to generate a normal subject (NOR) with a target pathology such as dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and dilated right ventricle (RV), assuming that these diseases are clustered in the latent space.

Pathology Synthesis

Pathology synthesis is designed to generate subjects with informed characteristics of a heart pathology and its effects on the geometry of the heart, given that the pathology is manifested in the ground truth labels. The assumption here is that subjects with a common pathological class have similar heart characteristics and hence they are encoded to the same area in the latent space of the VAE trained with them. Figure 4.2 depicts a schematic representation of the latent space information of a normal subject (with a normal heart shape) and a number of subjects with the same heart disease, having abnormal heart shapes, grouped in the same location.

Suppose we wish to generate subjects with a target pathology, for instance with characteristics of hypertrophic cardiomyopathy (HCM), potentially thickening of the myocardium. Note that we want to preserve the identity of a normal subject (NOR) and only generate disease characteristics such as thickening of the left myocardium for HCM. To this end, assuming that the disease features can be grouped to a neighboring location in the latent space, we encode all subjects with the desired pathology into the latent space and estimate mean, standard deviation, minimum, and maximum across all subjects for all interpolated slices; $[(\mu, \sigma, min, max)]_{HCM}$. These statistics are calculated on the mean of the posterior distribution which is the output of the encoder. The matrix size for these parameters is $(n_s \times n_z)$, where n_s is the number of interpolated slices (32 in our case) and n_z is the size of the latent vector. Note that we equalize the number of slices for each subject via slice interpolation in the latent space. A sample is drawn from a truncated normal distribution parameterized by these statistics, which we call pseudo-pathology sample; $x_{pHCM} {\sim} TN[(\mu, \sigma, min, max)]_{HCM}$. The sample generated with statistics of all HCM subjects should potentially represent the heart features of a HCM subject: abnormally thick myocardium. We expect to observe an incremental progression of this anatomical feature on a normal heart by performing linear interpolation between a NOR subject and a pseudo-HCM sample.

Modelling 3D Consistency

To model the dependency of variables, the correlation between the dimensions of the latent code for all pathological subjects is measured using the Kendall rank correlation coefficient. The uncorrelated generated sample is then transformed in the latent space according to the overall correlation coefficient $(n_z \times n_z)$ estimated from the training data to account for the relationship between elements of the latent code. The elements of the latent vector are correlated using Cholesky matrix decomposition as explained in the supplementary material. However, the relationship between different slices of one subject has not yet been modelled. This can lead to inconsistent heart geometries in the slice direction as a consequence of slice-by-slice 2D synthesis.

We propose a simple statistical approach to account for the relationship between slices in the latent space. The 2D VAE model is trained as normal while we attempt to take advantage of the correlation between slices of a given subject in the latent space and reconstruct a consistent 3D volume during the inference. In pathology synthesis, we want to perform a linear interpolation between a NOR subject (x_{NOR}) and a random pseudo-pathological sample (x_{pHCM}). Although different slices of the NOR subject are inherently correlated in the latent space, the random sample does not contain any information about the relationship between slices. To model this relationship, we estimate the correlation between slices of the x_{NOR} and construct the associated correlation coefficient matrix ($n_s \times n_s$). Given this matrix, we correlate the slices of the x_{pHCM} using the Cholesky matrix decomposition. The procedure is explained in more detail in the supplementary material.

The interaction between elements of latent vectors as well as the relationship between different slices is modelled to generate more realistic correlated samples in the latent space. We found that both latent correlation matrix ($n_z \times n_z$) and slice correlation matrix ($n_s \times n_s$) are important for consistent synthesis. This simple yet effective approach to sampling better respects the relationship between features presented in the training data and results in generating 3D consistent subjects, despite utilizing 2D models. A similar idea for modelling the distribution of 3D brain MRI data via estimating the correlation in the latent space of a 2D slice VAE has recently been explored in [95].

4.2.4 Data and Implementation

To examine the ability of our inter-subject method to perform cross-vendor and cross-subject synthesis, we utilize CMR images from a pair of subjects scanned using Siemens (vendor A) and Philips (vendor B) scanners provided by the M&Ms-1 challenge [46]. The disease information for each patient is required for our pathology synthesis experiment. For that purpose, we utilize ACDC challenge data [39] including normal cases (NOR) and three disease classes (heart dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and abnormal right ventricle (DRV)). All 150 M&Ms-1 and 100 ACDC subjects are resampled to $1.5 \times 1.5mm$ in-plane resolution and cropped to 128×128 pixels around the heart using the provided ground truth labels. Percentile-based intensity normalization is applied as post-processing and the intensity range is mapped to the interval of -1 and 1.

The input of the VAE model is a one-hot encoding version of the label map including three channels for cardiac classes right ventricle, left ventricle, myocardium, and background. The encoder part of the model includes four convolutional blocks with three convolutional layers each followed by batch normalization (BN) and LeakyReLU activation function. The encoded features are fed to four sequential fully connected layers to output the parameters of a Gaussian distribution

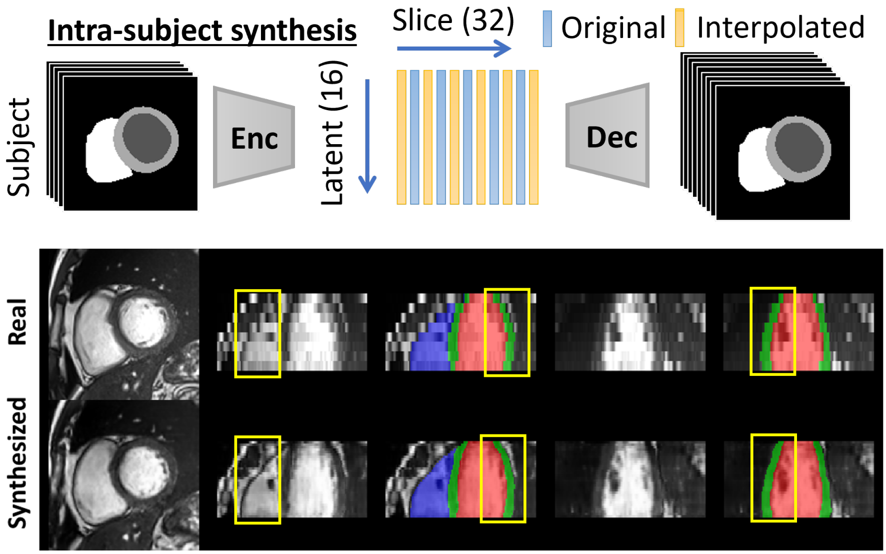


Figure 4.3: Intra-subject synthesis for increasing the resolution of the short axis cardiac MR image stack by interpolating in latent space of GT labels for slices of the same subject and using them for synthesis. The original data contains around 8-10 slices per subject, and we create 32 slices using our intra-subject synthesis.

over the latent representation. The decoder part of the model is comprised of four convolutional blocks each with one up-sampling layer followed by two convolutional layers with BN and LeakyReLU. The last additional block of the decoder includes one convolutional layer followed by BN and another convolution with four channel outputs and Softmax activation function. The VAE model is trained using a weighted combination of cross-entropy loss as the reconstruction loss and Kullback-Leibler divergence (KLD) with a weighting factor of β for regularization of the latent space capacity [96]. We experimentally identify the size of the latent vector ($n_z = 16$) and the weight of KLD ($\beta = 15$) by inspecting the quality of the label reconstruction and the outcome of interpolation.

4.2.5 Usability of Synthetic Data

We generate synthetic data including five pathological versions of each NOR case from ACDC data. The Synth HCM data is generated by interpolating, in the latent space, between each NOR case and pseudo-pathological sample with characteristics of HCM subjects. The same applies to generating Synth DCM data and Synth RV data. Moreover, we interpolate between vendor A and vendor B subjects from the M&Ms-1 challenge to generate Vendor AtoB synthetic data, which has

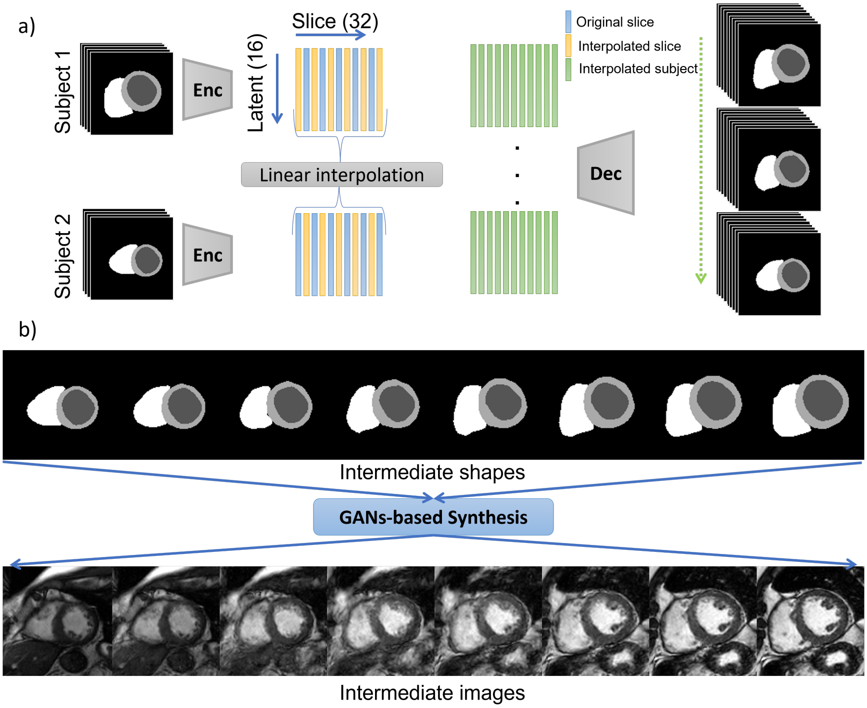


Figure 4.4: Inter-subject synthesis for generating intermediate shapes between two different heart geometries using linear interpolation between two subjects after equalizing the number of slices. The generated new labels are used for image synthesis.

characteristics of in-between vendor A and B data. To visualize the anatomical variation of the synthesized data in comparison with the real data, we calculate the end-diastolic (ED) and end-systolic (ES) volumes for the right ventricle and left ventricle using the ground truth labels. As can be seen from Figure 4.5, there is a considerable similarity between the distribution of the synthesized data and the real data in terms of the calculated volumes.

We quantitatively evaluate the usefulness of the synthetic data for cardiac segmentation in the presence of pathologies and domain shift in CMR image databases. Publicly available data typically suffer from the limited number of pathological cases, with rare diseases less likely to be represented well. Training with such data leads to models that struggle with generalization and adaptation to a wide variety of pathologies appearing among clinical cases. To tackle this, we utilize the synthetic data from our proposed inter-subject and pathology synthesis approaches in order to improve network generalization. To this end, we train four segmentation models with different combinations of real and synthesized data, to produce segmentation maps of three major heart structures - the left ventricle (LV), right ventricle (RV) blood pool and myocardium (MYO);

- 1) **ACDC Real**: a model trained with 200 ED and ES real images acquired from the ACDC training set.
- 2) **ACDC Real + Synth**: a model trained with 200 real ED and ES images from the ACDC training set and augmented with a total of 600 synthesized pathological cases for HCM, DCM, and RV (200 ED and ES images per pathology).
- 3) **ACDC M&Ms Real**: a model trained with 200 real ACDC images and 300 M&Ms-1 training images, acquired from vendors A and B (150 images each which include both ED and ES phases).
- 4) **ACDC M&Ms Real + Synth**: a model trained with real images from ACDC (200) and M&Ms-1 (300) training data augmented with a combination of 600 synthesized pathological cases (as utilized for the augmentation of the **ACDC Real + Synth** model) and 1000 synthesized vendor AtoB images.

To train the above segmentation models, we adapt a 2D nnU-Net [40] for a multi-class segmentation task with several modifications for improving the generalization and adaptation of the model to various data-sets used in this study, as proposed in [97]. In particular, we replace the standard instance normalization layers of the baseline nnU-Net with batch normalization and introduce heavier data augmentation, besides the default transformations used within the nnU-Net pipeline. These include image scaling ($p = 0.3$) with a scaling factor in the range of [0.7-1.4], random rotations within ±60 degrees ($p = 0.7$), random horizontal and vertical flips ($p = 0.3$) and elastic transformations ($p = 0.3$). Moreover, we apply intensity transformations in the form of gamma correction ($p = 0.3$) with the gamma factor ranging within [0.5-1.6], additive brightness transformations ($p = 0.3$) with the brightness factor varying within [0.7-1.3], multiplicative brightness ($p = 0.3$) with a mean of 0 and standard deviation of 0.3 and the addition of Gaussian noise ($p = 0.2$). During the pre-processing step, all data is normalized to an intensity range of [0-1], resampled to $1.5 \times 1.5mm$ in-plane resolution and center-cropped to 128×128 pixels around the heart, which is the same as the training patch size.

We use a combination of Dice and cross-entropy loss for training, optimized using Adam for stochastic gradient descent, with an initial learning rate of 10^{-4} and a weight decay of 3e^{-5}. During training, the learning rate is reduced by a factor of 5 if the validation loss has not improved by at least 5×10^{-3} for 50 epochs. We train all models for a maximum of 1000 epochs, where early stopping is applied when the learning rate drops below 10^{-6}. Please note that we do not apply a cross-validation set-up during training and train all models once, utilizing all available images on four NVIDIA Titan Xp GPUs.

At inference time, we resample all images to $1.5 \times 1.5mm$ in-plane resolution and crop them around the heart area. Since the test images are typically of a larger field-of-view than those we use for training, and we cannot rely on the availability of labels, we apply a heart region detection network, proposed in [94], responsible for obtaining the bounding box encompassing the whole heart. Before segmentation, the cropped images obtained using the predicted bounding boxes are post-processed to be of the size 128×128 pixels and normalized to the intensity range from 0 to 1. Finally, we perform a connected component analysis on the predicted labels and remove all but the largest connected component per class.

We evaluate all four models on the hold-out data (completely unseen) from the M&M-2 challenge with normal subjects (NOR) as well as various cardiac pathologies including Dilated Left Ventricle (DLV), Hypertrophic Cardiomyopathy (HCM), Congenital Arrhythmogenesis (ARR), Tetralogy of Fallot (FALL), Interatrial Communication (CIA), Tricuspidal Regurgitation (TRI), and Dilated Right Ventricle (DRV). This allows us to study the generalization capability of both the baseline models (ACDC Real and ACDC M&Ms Real trained with real images), as well as the models augmented with synthetic data generated in this study, to a wide array of moderate and severe pathological cases, some of which are not present in the training data. We report the segmentation performance in terms of Dice score (Dice) and Hausdorff Distance (HD), which are typically used as the main evaluation metrics in medical image segmentation challenges.

4.3 Results

4.3.1 Intra- and Inter-Subject Synthesis

Figure 4.3 shows the results for slice interpolation in the intra-subject synthesis approach. The subject has originally nine slices and the synthesized version of the subject includes 32 slices. The effects on the through-plane resolution on both the image and on the ground truth labels can be observed from three orthogonal views of the cardiac MR image. Traditional image-based interpolation between slices can potentially result in a severe blurring effect due to the very large slice thickness of short-axis cardiac images. Additionally, the segmentation label masks are not properly preserved after image-based interpolation, especially the through-plane smoothness of the masks may not be achieved.

An example of inter-subject synthesis is given in Figure 4.4. Note that the morphing from one shape to another is only shown for one mid-ventricular slice. The corresponding images are generated using the trained synthesis model.

To examine the ability of our method for cross-vendor and cross-subject synthesis we choose pairs of subjects from Philips and Siemens subjects of M&Ms

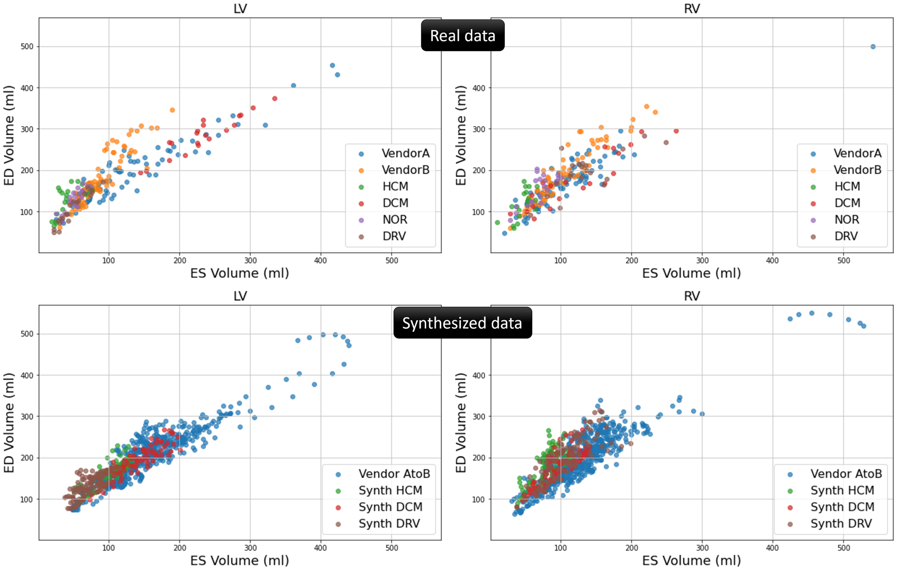


Figure 4.5: Distribution of calculated left and right ventricular volumes (LV and RV respectively) using the ground truth labels for end-diastolic (ED) and end-systolic (ES) phases of the heart for real and synthesized data. Vendor AtoB is the synthetic data for inter-subject synthesis between the data from M&Ms-1 vendor A and vendor B subjects. The synthetic data for pathology synthesis between normal subjects (NOR) and the corresponding diseases such as Hypertrophic cardiomyopathy (HCM), Dilated Cardiomyopathy (DCM), and Right Ventricle Dilation (RV) are respectively denoted by Synth HCM, Synth DCM, and Synth RV

database for inter-subject synthesis. We intend to show not only the morphing between two heart geometries but also to demonstrate the transition from one imaging characteristic to another one. Three examples for different levels of similarities between the heart geometries and image appearances are shown in Figure 4.6. Example 1: for two rather similar heart shapes, example 2: from one heart with normal looking size to one with large left ventricle, and example 3: from another one with large left ventricle to one with large right ventricle. A smooth transition between subjects from vendor A and another from vendor B can be observed from the results.

4.3.2 Pathology Synthesis

The results for pathology synthesis with three target heart diseases namely DCM, HCM, and DRV diseases are shown in Figure 4.7. The characteristic of the

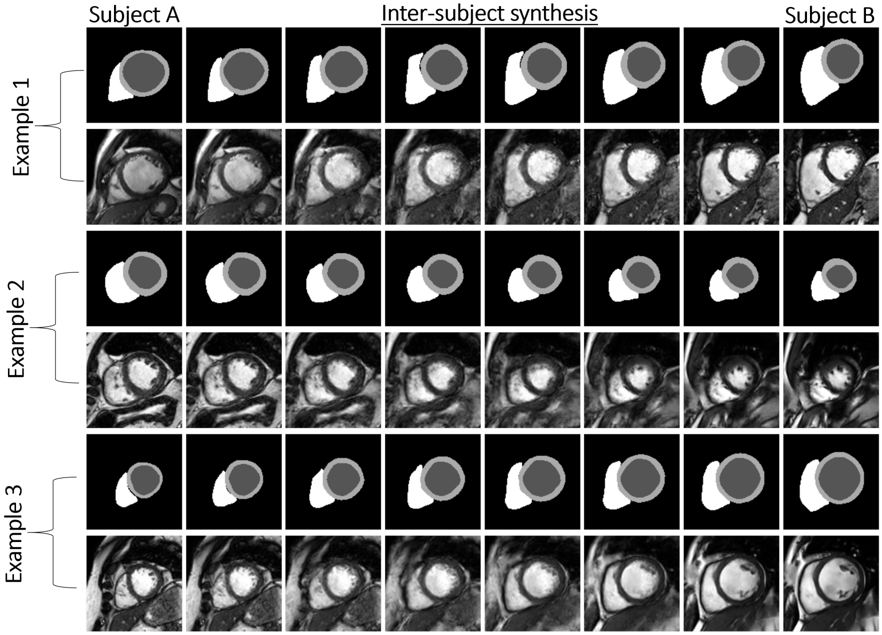


Figure 4.6: Inter-subject synthesis between three pairs of subjects chosen from the subset of vendor A and vendor B data. Subjects with similar heart shapes are chosen for example 2, whereas more dissimilar pairs for examples 1 and 3. Yet, a smooth transition of the shape and appearance can also be observed for examples 1 and 3.

particular heart disease is linearly added to the latent code of a normal subject (NOR). The heart shape characteristic of subjects with DCM, dilation of the left ventricle, is progressively appearing on the NOR subject through interpolation from left to right. The same is observed for thickening of the myocardium in the case of NOR to HCM and dilation of the right ventricle for NOR to DRV. Note that in pathology synthesis, in contrast to inter-subject synthesis, the identity of the NOR subject is not changing while the disease features are manifested on the geometry of the subject's heart and the image appearance stays the same. Interestingly, the detailed structures of the papillary muscles and myocardial trabeculations inside the left and right ventricular blood pool are generated despite not being present in the ground truth labels.

4.3.3 Modeling the Slice Relationship

Our proposed 2D model synthesizes images slice-by-slice with high visual fidelity and realism. However, the synthetic subject that is composed of stacking multiple

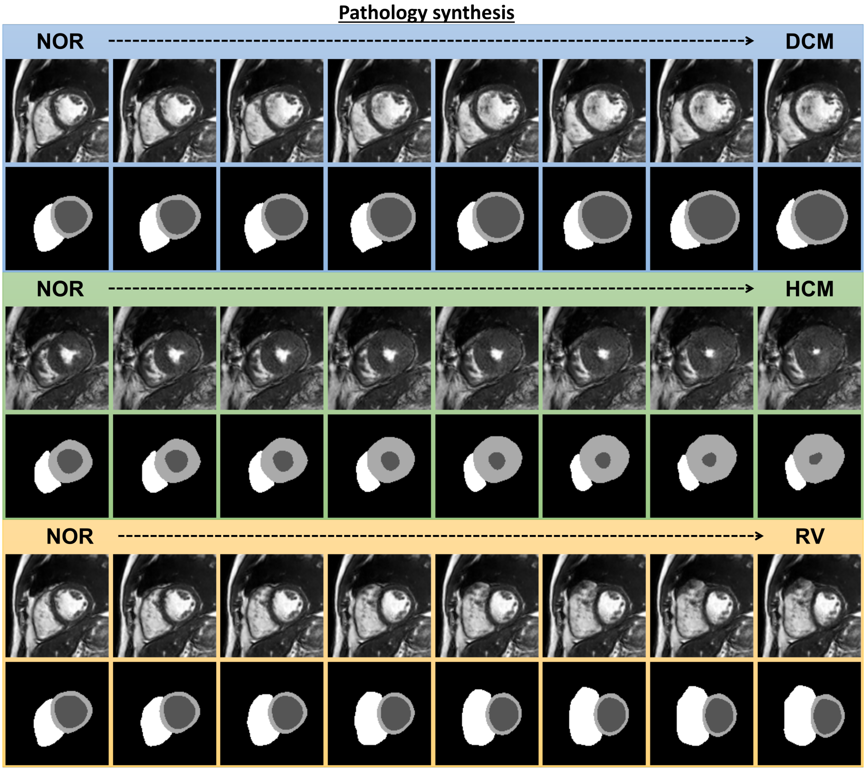


Figure 4.7: Pathology synthesis to generate the transition between a normal subject (NOR) to a target pathology such as dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and dilated right ventricle (RV). The effects of a disease on the heart geometry of a subject are respectively left ventricle dilation, myocardial thickening and right ventricle dilation.

2D slices is not generated coherently by the network when we look at the generated slices from perpendicular directions. The reason is that random samples in the latent space contain no information about the relationship between different slices of one subject, i.e. generated slices are uncorrelated. Synthesis examples with target pathologies and the positive effects of the proposed slice correlation on generating 3D consistent subject are shown in Figure 4.8 with a three-dimensional rendering of the synthesized labels. The irregularities in the slice direction are substantially reduced for the correlated slices for synthesizing different pathological cases. We notice that some real images may originally be hampered by slice misalignment artifacts and our correlated sampling cannot reduce this artifact.

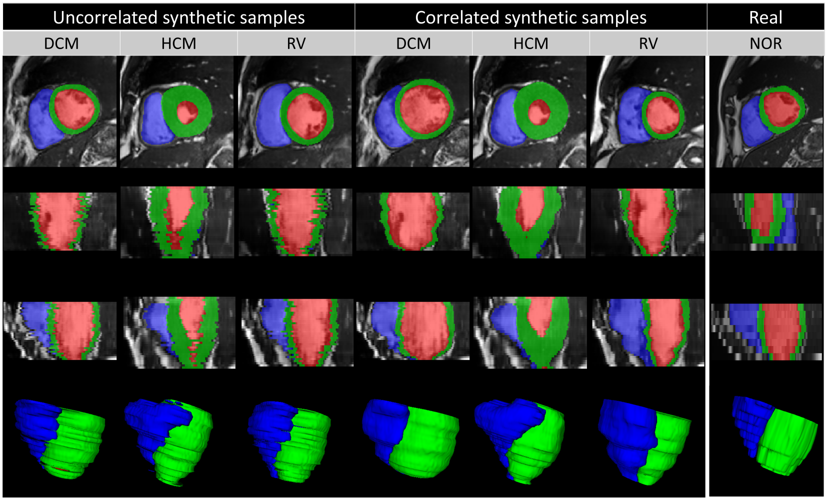


Figure 4.8: Three-dimensional rendering of the labels for uncorrelated and correlated synthesis for different cases of pathology synthesis. The first three columns show the uncorrelated slices and the impact of the inconsistency of the anatomy in the perpendicular views of the short axis slices while the second three columns show the positive effects of correlating samples on reducing the inconsistency and irregularity of the consecutive slices. The last column shows one real example.

4.3.4 Usability of Synthetic Data

Figure 4.9 shows the performance of four segmentation models in terms of Dice and Hausdorff distance (HD) for left ventricle, right ventricle and myocardium segmentation. A significant drop in the performance observed for the ACDC real model (trained using real ACDC data only) suggests a substantial domain shift (distributional shift) between ACDC and M&Ms-2 images, mostly due to the acquisition hardware, protocols, and presence of unseen heart pathologies. While the addition of M&Ms-1 images to the training (ACDC M&Ms model) helps significantly with tackling this domain shift and improving generalization, we also note that the addition of the synthesized data with pathological characteristics substantially improves the segmentation performance and the robustness of the model (ACDC Real + Synth) across all cardiac diseases. This indicates that the pathology synthesis approach can generate realistic images with relevant pathological diversity for training a cardiac segmentation network that generalizes well to even unseen diseases during training. Moreover, it also suggests that synthetic images generated in this study serve as a realistic substitute for real MR images when met with limitations in acquiring more data from the target domain.

Additional benefits are observed when augmenting the training with inter-

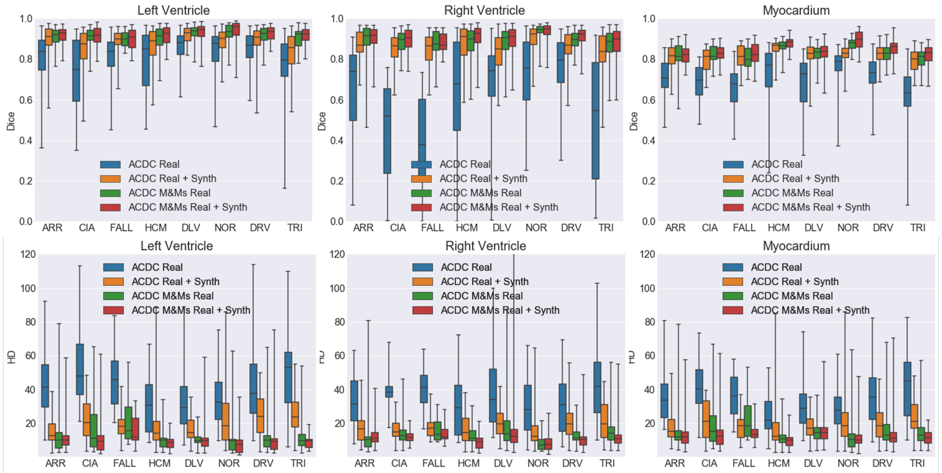


Figure 4.9: Dice scores and Hausdorff Distance (HD) performance on the unseen data from M&Ms-2 challenge with different cardiac pathologies.

subject synthesis results (ACDC M&M-s Real + Synth) in addition to synthesized pathologies. Since inter-subject synthesis is performed on M&Ms-1 data, this further contributes to alleviating the domain shift present between the ACDC and M&Ms images, resulting in the best performance across all three cardiac tissues and varying pathologies. In other words, the best performance in Dice and HD across all cardiac diseases is obtained by the model trained using both pathology synthesis and inter-subject synthesis approaches, indicating the usefulness of the generated images for clinical tasks. We particularly note a significant improvement in HD scores, primarily due to outlier reduction, as well as a decrease in the number of over-segmented and under-segmented tissues. Under-segmentation due to diseased tissue and occlusion is a common problem observed for baseline models, which can be tackled by the augmentation approach proposed in this study, allowing the network to learn from a much higher and diverse pool of examples than just relying on carefully curated and limited real data. We hypothesize that synthesized images contribute to a higher variation in heart tissue shape and appearance, particularly those undergoing changes due to the presence of pathology, which in turn helps with network regularization and generalization.

4.4 Discussion and Conclusion

This study investigated an approach for realistic cardiac magnetic resonance image synthesis with target heart pathologies by separating the task into label deformation using a VAE and image generation using a label-conditional GAN.

We introduced intra-subject synthesis to increase the through-plane resolution of short-axis images and to equalize the number of slices across all subjects. The inter-subject synthesis was designed to perform cross-subject and cross-vendor synthesis by generating subjects that have intermediate heart geometries and appearances between two dissimilar subjects scanned using different vendors. Furthermore, the pathology synthesis was proposed to generate subjects with heart characteristics of a particular disease through sampling in the latent space with statistics of a target pathology and performing linear interpolation between a normal subject and a pseudo-pathological sample in the latent space of the trained VAE.

To tackle one of the important challenges of 3D medical image synthesis, we demonstrated that modelling the correlation between slices in the latent space can be a simple yet effective way to generate consistent 3D subjects from 2D models.

Visualizations of the synthesized images and the distribution of the left and right ventricular volumes on the synthesized data showed encouraging results. Moreover, we devised experiments to quantitatively evaluate the usability of the synthetic data for the development of a generalizable deep learning segmentation network. We found that both generated images by inter-subject and pathology synthesis are extremely useful in improving the generalization and robustness of a deep-learning segmentation model in a challenging clinical environment. The methods proposed in this study could be extended for other applications in medical image synthesis such as brain MR image generation and simulation of lesion progression.

A limitation of our study is the lack of a measure for assessing the 3D consistency of the synthesized subjects. Moreover, the result of the intra-subject synthesis is not quantitatively evaluated on its own as an independent approach for increasing the resolution in the slice direction and the effect on segmentation, apart from qualitative results shown in Figure 4.3. Note that this is a necessary step for inter-subject and pathology synthesis to equalize the number of slices for each subject. Another limitation is that we in fact evaluate the usefulness of the synthetic images for data augmentation, while quantifying the level of image realism for each synthetic subject remains to be explored. For instance, when interpolating from a normal subject to a pathological sample, it is likely, but not guaranteed that the intermediate heart changes reflect what truly happens in disease progression. Similarly, when interpolation between two subjects, not all intermediate shapes reflect what can be found in real life. Despite all that, we examine the benefit of the synthetic images for training a deep learning model which indicates the usefulness of the generated images for data augmentation.

In conclusion, we demonstrated that our approach could provide a solution to diversify and enrich an available database of cardiac MR images, resulting in significant improvements in model performance and generalization for cardiac segmentation of subjects with unseen heart diseases.

Supplementary Material

4.4.1 Cholesky Decomposition and Correlated Samples

In order to simulate correlated variables with a given covariance matrix (C), Cholesky matrix decomposition is used in this study. The Cholesky matrix decomposition is a factorization of a positive-definite symmetric matrix into a product of a lower and upper triangular matrix, L and L^T, respectively.

$$C = LL^T \tag{4.1}$$

Assuming an uncorrelated random sample X with unit covariance matrix of $E(XX^T) = I$, a new random vector can be computed as $Y = LX$ that its covariance matrix is derived:

$$E(YY^T) = E(LX(LX)^T) = E(LXX^TL^T) = LE(XX^T)L^T = LIL^T = LL^T = C \tag{4.2}$$

Note that the expectation is a linear operator; $E(cX) = cE(X)$.

4.4.2 Generating Sample with Pathology Characteristics

For generating a subject with pathological characteristics, a random sample is drown using a truncated normal distribution parameterized by the statistics of the desired pathology, e.g. mean, standard deviation, minimum, and maximum estimated on all subjects with hypertrophic dilated cardiomyopathy (HCM); namely pseudo-pathological sample x_pHCM. These statistics are calculated on the mean of the posterior distribution of features estimated by the encoder part of the VAE. The following steps are followed to correlate the elements of this pseudo-pathological sample cross slice direction and latent dimension:

- Estimate correlation coefficient between latent dimensions across all subjects with desired pathology using Kendall rank correlation coefficient method; $Corr_{zHCM}$ with size $(n_z \times n_z)$ where n_z is the size of the latent vector $(n_z = 16)$

- Calculate the lower triangular matrix L using Cholesky decomposition; L_{zHCM}

- Correlate the latent dimensions of the pseudo pathological sample across the element of latent vector given above formula; $y_{pzHCM} = L_{zHCM}x_{pHCM}$

- Estimate the correlation coefficient between slices of the target normal subject (NOR) we wish to use for interpolation; $Corr_{sNOR}$ with size $(n_s \times n_s)$ where n_s is the number of slices $(n_s = 32)$

- Calculate the lower triangular matrix L using Cholesky decomposition; L_{sNOR}

- Correlate the latent dimensions of the pseudo random sample cross slices given above formula; $z_{pzsHCM} = L_{sNOR} y_{pzHCM}$
- Linearly interpolate between z_{NOR} and z_{pzsHCM} in the latent space
- Reconstruct slices-by-slice the interpolated samples using the decoder part of the 2D VAE
- Compose 3D volume from synthesized 2D slices

The correlation coefficient matrix for all above mentioned steps is shown in Figure 4.10. Correlating latent dimensions found to be as important as correlating slices of subject for generating coherent slices with smoothly changing features.

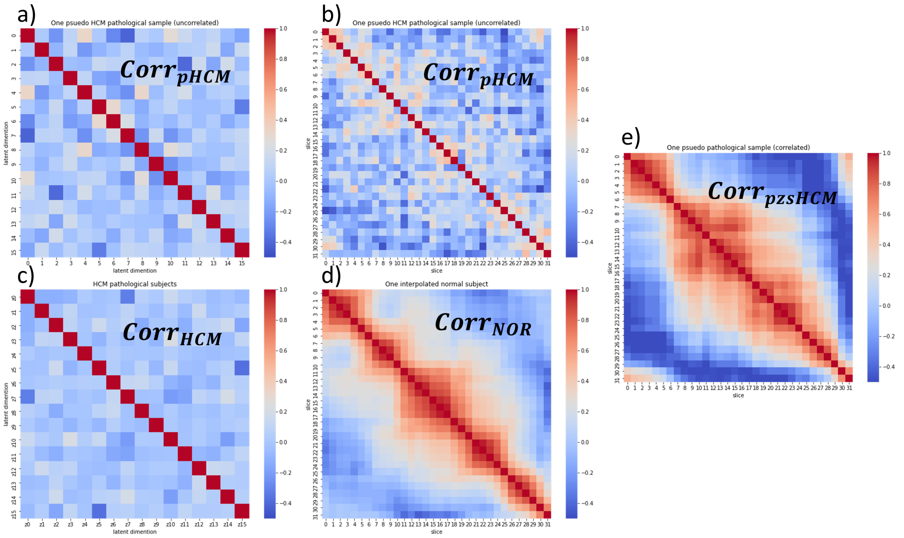


Figure 4.10: Correlation coefficient matrix for a) uncorrelated pseudo-HCM sample across latent dimensions and b) across slices, c) all HCM subjects across latent dimensions, d) one normal subject across slices, and e) the correlated pseudo pathological sample calculated using the Cholesky decomposition.

CHAPTER 5

Image Synthesis for Myocardial Scar Quantification

This chapter is based on:
D.R.P.R.M. Lustermans, S. Amirrajab, M. Veta, M. Breeuwer, C. M. Scannell "Optimized Automated Cardiac MR Scar Quantification with GAN-Based Data Augmentation," Computer Methods and Programs in Biomedicine, 2022.

Abstract

The clinical utility of late gadolinium enhancement (LGE) cardiac MRI is limited by the lack of standardization, and time-consuming post processing. In this work, we tested the hypothesis that a cascaded deep learning pipeline trained with augmentation by synthetically generated data would improve model accuracy and robustness for automated scar quantification. A cascaded pipeline consisting of three consecutive neural networks is proposed, starting with a bounding box regression network to identify a region of interest around the left ventricular (LV) myocardium. Two further nnU-Net models are then used to segment the myocardium and, if present, scar. The models were trained on the data from the EMIDEC challenge, supplemented with an extensive synthetic dataset generated with a conditional GAN. The cascaded pipeline significantly outperformed a single nnU-Net directly segmenting both the myocardium (mean Dice similarity coefficient (DSC) (standard deviation (SD)): 0.84 (0.09) vs 0.63 (0.20), $p < 0.01$) and scar (DSC: 0.72 (0.34) vs 0.46 (0.39), $p < 0.01$) on a per-slice level. The inclusion of the synthetic data as data augmentation during training improved the scar segmentation DSC by 0.06 ($p < 0.01$). The mean DSC per-subject on the challenge test set, for the cascaded pipeline augmented by synthetic generated data, was 0.86 (0.03) and 0.67 (0.29) for myocardium and scar, respectively. The proposed, A cascaded deep learning-based pipeline trained with augmentation by synthetically generated data leads to myocardium and scar segmentations that are similar to the manual operator, and outperforms direct segmentation without the synthetic images.

5.1 Introduction

Late gadolinium enhancement (LGE) cardiac MRI is the reference standard for the non-invasive assessment of myocardial viability, and is widely used in clinical routine [98]. It has been shown to accurately identify areas of myocardial infarction [99], and the size and transmurality of scar regions are important parameters to guide the management of patients [100]. Visual reporting of such parameters is user-dependent, and thus, the robust and accurate quantification of scar would be highly beneficial. If the quantification could be reliably performed automatically, it could also facilitate further adoption of LGE cardiac MRI in clinical practice, particularly in less specialized, low-volume centers.

The standard approach to the quantification of LGE has been the use of a fixed intensity threshold value, usually relative to a reference region. The most common approaches segment scar as being n (typically 5) standard deviations (nSD) above the mean intensity of a remote normal myocardium region or above half of the maximum value of a scar region (full width at half maximum (FWHM)). To date, quantification has primarily been performed in research studies due to the time-consuming manual interaction and lack of reproducibility between operators [101]. Advanced methods for the thresholding of scar regions, that do not require manually drawn reference regions, such as Otsu thresholding [102], or fitting to expected distributions using expectation-maximization [103,104] have also been proposed without achieving clinical adoption. In general, these intensity-based thresholding methods are subject to false positives due to noise and imaging artifacts, and they do not incorporate any spatial context in the thresholding.

More recently, deep learning, using convolutional neural networks (CNNs), has become the state-of-the-art for cardiac MRI segmentation in a wide range of applications [46,105–113], and it has also been applied to LGE segmentation. Fahmy et al. demonstrated accurate scar volume quantification using a 3D U-Net in patients with hypertrophic cardiomyopathy [114] and Zabihollahy et al. proposed a cascaded pipeline which segments the left ventricle (LV) myocardium and scar in two steps using images from multiple planes [115]. There has been further interest in the topic as a result of the open-source dataset made available as part of the automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI (EMIDEC) challenge [116], with several authors investigating the use of cascaded pipelines [117,118] or the incorporation of prior information [119] to improve reliability.

In this work, we developed and evaluated a cascaded deep learning pipeline for automatic myocardium and scar segmentation and quantification in subjects with suspected acute myocardial infarction. In particular, we proposed a cascaded deep learning pipeline, consisting of bounding box detection and myocardium segmentation, followed by scar segmentation, and we investigated the impact of each individual step on the performance pipeline. We further assessed the benefit of including synthetic images, generated by a conditional generative adversarial network (GAN), in the training data (in addition to conventional data augmentation). Both the GAN-based synthetic data and the step of splitting the task into simpler sub-problems are designed to overcome the challenge of the limited amount of training data, and it is hypothesized that both will lead to improved performance with the small amount of training data.

5.2 Materials and Methods

5.2.1 Dataset

The dataset from the EMIDEC challenge was used [116]. This consists of the LGE cardiac MRI scans of 150 patients, out of which 105 are pathological and 45 are normal. The data are divided into training (n=100) and testing (n=50) sets by the challenge organizers. The data acquisition was performed at the University Hospital of Dijon (France) on 1.5T and 3T systems (Siemens Medical Solution, Erlangen, Germany) with a T1-weighted phase sensitive inversion recovery (PSIR) sequence (TR = 3.5 ms, TE = 1.42 ms, TI = 400 ms, flip angle = 20), performed 10 minutes after the administration of a gadolinium-based contrast agent (Gd-DTPA; Magnevist, Schering- AG, Berlin, Germany), at concentration between 0.1 and 0.2 mmol/kg, during a breath-hold. Further details can be found in [116]. The images were manually segmented, in consensus, by two expert operators (a cardiologist with 10 years' experience in cardiac MRI and a physicist with 20 years' experience). For the purpose of this work, the scar and microvascular obstruction (MVO) segmentations are combined in one class label representing the total infarction area.

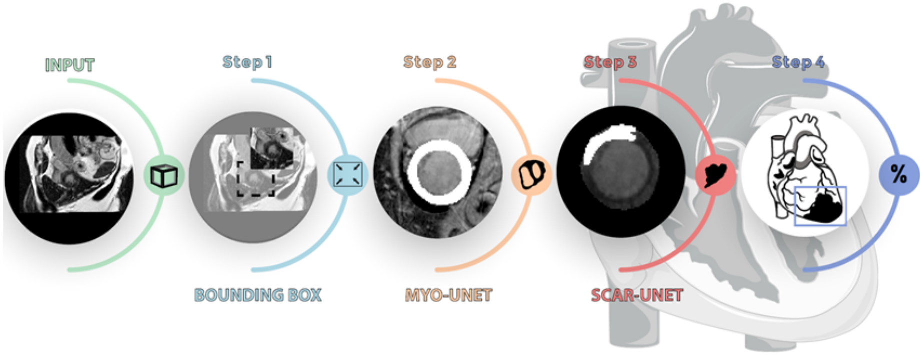


Figure 5.1: The proposed scar quantification pipeline. Firstly, a bounding box is detected around the heart, followed by myocardium segmentation. Subsequently, scar is segmented, if present, and used to compute the scar burden as a percentage of myocardial volume.

5.2.2 Cascaded Pipeline

As shown in Figure 5.1, the proposed cascaded pipeline consists of three main steps: 1) the detection of a bounding box that encompasses the LV cavity and LV myocardium, 2) the segmentation of the myocardium, and 3) the segmentation of scar. In this work, deep learning models are trained sequentially to achieve each of these steps. An ablation study is performed by removing steps in the pipeline to analyze their impact of the performance, and the model performance with the inclusion of the synthetically generated images is compared with that of a model without this data augmentation. The segmentation and computed volumes from the automated analysis are compared to the manual quantification on the EMIDEC challenge test set. The trained models for both segmentation and synthetic data generator are made available at `https://github.com/cianmscannell/lge-quant-emidec`, along with the generated synthetic data.

Bounding Box

The bounding box algorithm used is as proposed in Scannell et al. [109]. This first assumes that there is a fixed bounding box in the center of the image. A CNN is trained to predict, from a LGE image, the transformation of this proposed bounding box so that it covers the LV myocardium and cavity of the image. This is framed as a regression problem to predict four continuous values, the 2D translation of the center of the box and the scaling of the two different sides of the rectangular box. The proposed bounding box is of size 134 × 134 pixels, which is the mean size present in the training set. Due to the shape of the LV, it is sufficient to predict the bounding box on a slice in a basal location. This work uses the second 2D image from the top of the stack to avoid images where the LV cavity and myocardium are not present.

The CNN takes the original images, center-cropped or zero-padded to a size of 256×256 pixels, as input. These input images are min-max normalized, using the $5th$ and $95th$ percentile of intensity values as the pseudo-minimum and maximum, respectively. The architecture consists of four convolutional layers, each layer with two convolutions using 3×3 kernels followed by 2×2 max-pooling. These layers are followed by fully connected layers. Each layer uses batch normalization and rectified linear unit (ReLU) activations, except the output layer which uses a linear activation. A batch size of 32 and L2 regularization on the convolutional kernel parameters was used with a weight of 0.0001. The loss function, the mean squared error between the four predicted and ground-truth translation and scaling values, was optimized with the use of an Adam optimizer and convergence was determined by early stopping. Data augmentation was used, consisting of random combinations of rotation, translation, blurring, scaling and noise added to the images. The parameters of the data augmentation are provided in Supplementary Table 5.3.

Myocardium Segmentation

The segmentation of the myocardium is based on a nnU-Net model, as described by Isensee et al. [40]. The model architecture and training process are used as automatically configured by the nnU-Net software. Briefly, the algorithm automatically optimizes the architecture and hyperparameters of a U-Net model based on a fingerprint of the training dataset. A 2D network is used with leaky ReLU activations and instance normalization. The initial number of feature maps was 32 and doubled each layer until it reached a size of 512. Stochastic gradient descent with Nesterov momentum (μ=0.99) is used as the optimizer with an initial learning rate of 0.01. The loss function is the sum between the cross-entropy and dice loss, as is the default choice for the nnU-Net [40]. After a fixed amount of 1000 epochs, the network with the best validation set performance is chosen. The algorithm also includes on-the-fly data augmentation. The input is the cropped LGE image which is reshaped to 128×128 pixels, and normalized in the same way as described for the bounding box.

Quality control, of the myocardium segmentations is performed in which connected component analysis was used to identify failed (not closed) myocardium segmentations, which were then re-segmented with ten different augmented (by translation) bounding boxes. The 10 predictions are summed and the pixels that are predicted in the myocardium greater than k times are included in the final prediction, where k is the smallest number that yields a closed myocardium segmentation. If no closed myocardium is achieved in this manner, the original prediction is used.

Scar Segmentation

A second nnU-Net model is trained for the scar segmentation. The network input is a 64×64 pixel image, cropped according to the contour gravity center of the myocardium segmentation. The image is also masked with the myocardium segmentation to set intensity values outside the myocardium or LV cavity to 0. The myocardium values are normalized to a signal intensity range between 0 and 1, using the $5th$ and $95th$ percentiles, as before. The LV cavity is set to a fixed signal intensity value of 2.5. A further quality control step is used that removes small regions of predicted scar by removing regions that lead to a predicted scar-to-myocardium volume ratio less than 3%, as scar sizes smaller than this are not feasible in this population.

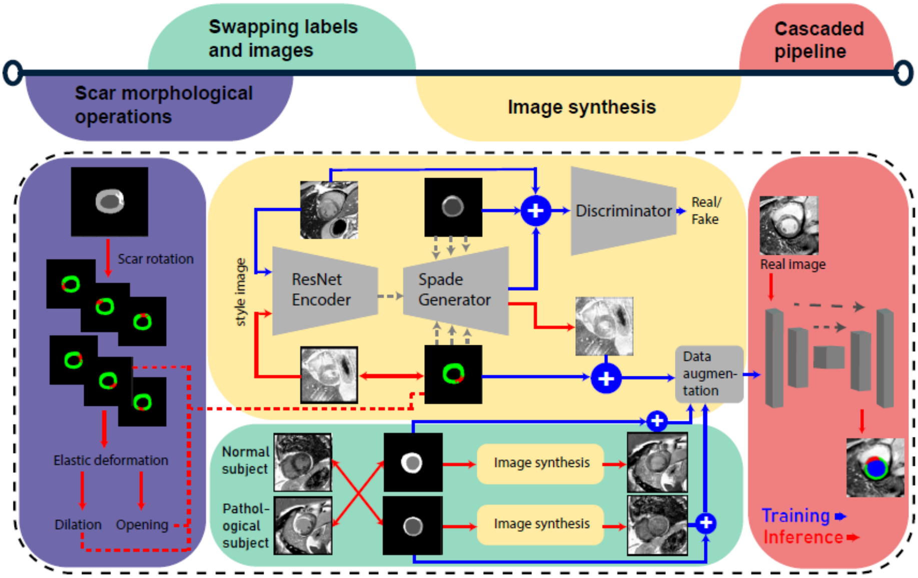


Figure 5.2: A flowchart of the synthetic image generation process. For inference, augmented segmentation labels are input to a trained conditional GAN, accompanied by a style image (red). The label maps are generated by swapping existing labels between pathological and normal subjects (green) and by performing morphological operations on scar labels (elastic deformation, rotation and dilation or opening) (purple). The generated synthetic data are then used as augmentation data for the cascaded pipeline.

5.2.3 GAN-Based Image Synthesis

The image synthesis module, as shown in Figure 5.2, is based on a ResNet-encoder coupled with a segmentation-conditioned GAN that uses SPatially-Adaptive (DE)-normalization layers (SPADE) [75] throughout the generator architecture. The use of SPADE-based conditional GANs for preserving the anatomy of the segmentations for cine cardiac MR image synthesis was investigated previously by Abbasi-Sureshjani et al. [19] and Amirrajab et al. [20] and it was shown that providing multi-class labels to guide the SPADE generator allows the synthesis of realistic images. The generator consists of a series of the residual blocks with SPADE normalization, followed by nearest neighbor up-sampling layers, as in [19]. The SPADE layers normalize the activations with a spatially-varying learned scale and bias that comes from the input segmentation mask. This is done to encode information about the input segmentation in the generated image. In contrast to previous works, a LGE image was input to the ResNet-encoder to extract the style and background anatomical information in the synthesis process. These two steps allow the control of both the underlying anatomy and the image appearance (style) of the synthetic image. The training process used pairs of LGE style images with the corresponding ground-truth segmentations, and is described in more details in [20]. The architecture of the discriminator, the losses, and training settings are kept unchanged from the original work of Park et al. [75].

After training, to generate new pairs of synthetic LGE images with the trained generator, two strategies are used: 1) augmented labels, and 2) swapped labels. For the augmented labels, the segmentation labels from the training set are augmented, by rotation and morphological operations (parameters defined in Supplementary Table 5.4), to create previously unseen shapes and positioning of scar, and input to the generator (shown in the purple box of Figure 5.2), and for the swapped labels, existing segmentation labels from pathological patients are combined with style images from normal patients, and vice versa, to create new patients (shown in the green box of Figure 5.2). The augmentation of the scar segmentations included rotation by a multiple of 60°, elastic deformation, dilation and opening. The augmented and swapped labels are used to generate synthetic data for the myocardium segmentation training, and only the augmented labels are used for the scar segmentation as, due to the masking of the myocardium, changing the background (via the swapped style image) will have no effect.

5.2.4 Evaluation

An ablation study was first performed, trained only with the real patient data, to analyze the impact of the individual steps of the cascaded pipeline, by comparing the myocardium and scar segmentation of:

- a) the full cascaded pipeline
- b) myocardium and scar segmentation in two steps, without the bounding box
- c) directly segmenting the myocardium and scar with one nnU-Net, with the bounding box
- d) directly segmenting the myocardium and scar with one nnU-Net, without the bounding box
- e) the full cascaded pipeline trained with the GAN-based synthetic image data augmentation.

Supplementary Figure 5.7 shows a representation of the four methods (a) to (d). As the cascaded nature of the pipeline can lead to the propagation of errors, when an incorrect myocardium segmentation is used to mask the input to the scar segmentation model, this effect was also studied. In particular, the increase in performance found when replacing the predicted myocardium segmentation with the ground-truth in the cascaded pipeline was analyzed. Secondly, the performance was tested by comparing the cascaded pipeline trained only with the real data (a) and a version with training augmented by synthetic images (e), to assess the impact of adding synthetic data to the real dataset. In order to avoid the possibly confounding effect of the cascaded pipeline, a direct one-step segmentation model was also trained with and without the synthetic data augmentation.

These comparisons were performed on a randomly selected internal testing set (N=20) split from the training set. This analysis was performed using the Dice similarity coefficient (DSC) metric on a per-slice level and the performance of models are compared with the Wilcoxon signed-rank test. The ability to classify slices as either having no scar or have significant ($>$ 15% of the myocardium) was also assessed, with and without the GAN-based synthetic images.

The final evaluation was performed with the full cascaded pipeline, augmented with the GAN-based synthetic data, the model with the best performance on the internal test set, on the 50 test subjects from the EMIDEC challenge, where the mean DSC, Hausdorff distance (HD) and volume difference per-subject was reported. The automated scar quantification was further evaluated with respect to the manual quantification using Pearson correlation and Bland-Altman analyses. We also test the accuracy of classifying patients as scarred or not. The segmentation labels for the test set are not available publicly but the evaluation of these metrics was performed by the challenge organizers.

5.3 Results

5.3.1 Cascaded Pipeline

The DSC values for both myocardium and scar between the manual and automatic segmentations, for the four versions of the pipeline on the internal test set, are shown in Figure 5.3 (a)-(d). The proposed full cascaded pipeline (option (a)) had the highest DSC for both the myocardium (mean DSC (standard deviation (SD)): 0.84 (0.09)) and scar (0.72 (0.34)) segmentations. This was significantly higher than all other approaches including a single nnU-Net directly segmenting both the myocardium and scar (option (d): myocardium DSC: 0.63 (0.20), $p < 0.01$ and scar DSC: 0.46 (0.39), $p < 0.01$), the direct myocardium segmentation followed by scar segmentation (option (b): myocardium DSC: 0.79 (0.15), $p < 0.01$ and scar DSC: 0.68 (0.37), $p = 0.02$), and the two-step pipelines of bounding box followed by direct myocardium and scar segmentation (option (c): myocardium DSC: 0.80 (0.12), $p < 0.01$ and scar DSC: 0.68 (0.35), $p < 0.01$). These results are summarized in Table 5.1.

Table 5.1: The mean and standard deviation (SD) DSC for the myocardium and scar segmentation with each of the model and training configurations.

	Myocardium		Scar	
	Mean	SD	Mean	SD
Option a)	0.84	0.09	0.72	0.34
Option b)	0.79	0.15	0.68	0.37
Option c)	0.80	0.12	0.68	0.35
Option d)	0.63	0.20	0.46	0.39
Option e)	0.85	0.09	0.78	0.28

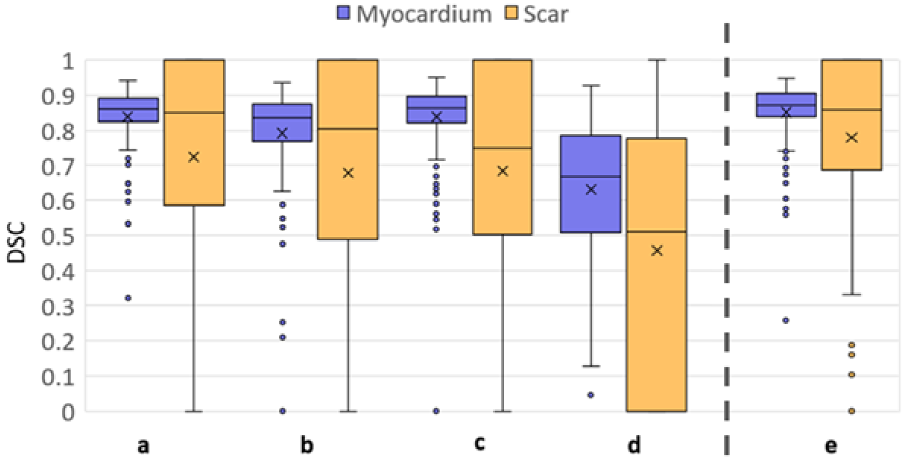


Figure 5.3: The distributions of the DSC values for myocardium (purple) and scar (yellow) segmentations on the internal test set, for the five trained versions of the cascaded pipeline. These are (option (a)) the proposed full cascaded pipeline, (b) myocardium and scar segmentation in two steps, without the bounding box, (c) directly segmenting the myocardium and scar with one nnU-Net, with the bounding box, (d) directly segmenting the myocardium and scar with one nnU-Net, without the bounding box, and (e) the full cascaded pipeline trained with the synthetic data augmentation. The X mark indicates the mean value.

5.3.2 GAN-Based Synthetic Image Data Augmentation

Figure 5.4 shows GAN-based synthetic images generated from one real patient by rotating, elastically deforming, and morphologically opening the original segmentation mask. The DSC values for the full cascaded pipeline trained with the synthetic data are shown in Figure 5.3 (e). The mean (SD) DSC between the manual and automatic myocardium segmentation increased by 0.01 (from 0.84

Table 5.2: The mean (SD) DSC for the scar segmentation for both the cascaded pipeline and a single nnU-Net model with and without the GAN-based data augmentation.

	Without synthetic data	with synthetic data	Difference	p-value
Cascaded pipeline	0.72 (0.34)	0.78 (0.28)	0.06	< 0.01
Single nnU-Net	0.68 (0.35)	0.71 (0.33)	0.03	< 0.01

(0.09) to 0.85 (0.09)) with the addition of the synthetic training data on the internal test set. The inclusion of the synthetic data for training resulted in a 0.06 increase in the scar DSC (from 0.72 (0.34) to 0.78 (0.28)), a statistically significant difference ($p < 0.01$), as well as a decrease in the SD. Moreover, on the internal test set, the inclusion of the GAN-based synthetic images improved the classification of slices as scarred or not from 87% (122/141) correct to 94% (132/141). The identification of slices with significant levels of scar (>15% of the myocardium) increased from 86% (49/57) to 97% (55/57). For the direct one-step segmentation the DSC is also increased from 0.80 (0.12) to 0.84 (0.11) for the myocardium and 0 .68 (0.35) to 0 .71 (0.33) for scar (both $p < 0.01$). Replacing all the predicted myocardium segmentations with the manual ground-truth label only gives a modest improvement in DSC for the scar segmentation from 0.78 to 0.80, also on the internal test set.

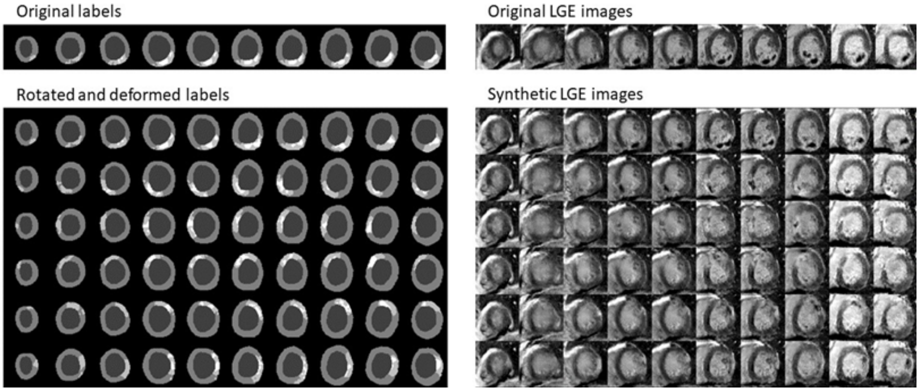


Figure 5.4: The GAN-based synthetic LGE images generated with a set of rotated and deformed input labels from a single patient (original LGE images and labels shown in the top row) for all slices apex (left) to base (right).

5.3.3 Challenge Test Set

The proposed cascaded pipeline trained with synthetic data augmentation was evaluated on the EMIDEC challenge test set (N=50 subjects). Figure 5.5 shows segmentations for three representative patients from the test set showing both good and bad performance (note that the ground-truth segmentations for the test set are not available for comparison). The mean (SD) DSC, HD and volume difference per-subject was 0.86 (0.03), 15.7 (11.9) mm and 11.5 (8.4) cm^3 respectively for myocardium segmentation. Five out of 358 myocardium segmentations failed to generate a closed shape and were identified by the quality control procedure, and a correction was attempted. Secondly, the model showed a mean (SD) DSC and volume difference between the manual and automatic scar regions of 0.67 (0.29) and 41.0 (5.8) cm^3 per-subject, and the difference in scar volume relative to the volume of the myocardium was 3.41% (4.8%). Furthermore, the model classified patients as scarred or not with an accuracy of 94% (47 out of 50 subjects).

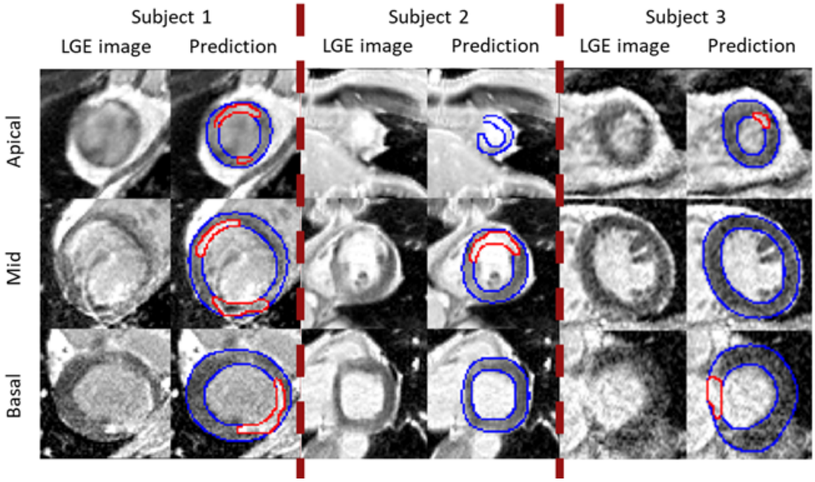


Figure 5.5: The segmentations of three representative patients selected from the test set, showing the LGE image and the predicted scar (red) and myocardium (blue) segmentations. It can be seen that in subject 1 (left), in the mid-slice of subject 2 (middle), and apex slice of subject 3 (right) scar is accurately identified by the models. However, for subject 2, the myocardium segmentation in the apical slice is inaccurate, and in the basal slice of subject 4 the LV outflow tract is incorrectly identified as scar. Note that the manual ground-truth for the test set is not publicly available.

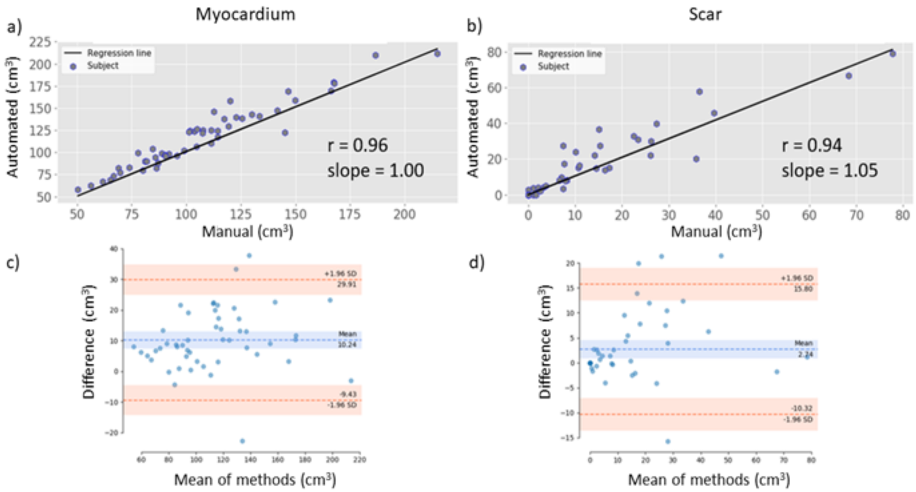


Figure 5.6: The top row shows a scatterplot of the manual versus automatic per-subject segmentation volumes for the myocardium and scar segmentation ((a) and (b)) with the Pearson correlation coefficient (r) and line of best fit, with the slope reported. (c) and (d) shows the Bland-Altman analysis for the myocardium volume (c) and scar volume (d).

Figure 5.6 compares the computed volumes for the automated and manual segmentations in a scatterplot, the Pearson correlation coefficient is 0.96 and 0.94 for myocardium and scar, (a) and (b) respectively. Per-subject, Bland-Altman analysis on the LV myocardial volume (c) showed an agreement between the manual and automatic quantified volumes of the proposed cascaded pipeline with bias of 10.24 cm^3 and limit of agreement 19.67 cm^3. Additionally, a good agreement between the manual and the proposed automatic quantified scar volume with a bias of 2.74 cm^3 and limit of agreement of 13.06 cm^3 was shown (d).

5.4 Discussion

In this work, two approaches are studied to learn from small datasets for the segmentation of scar from LGE cardiac MRI: the splitting of the task into smaller sub-problems and the use of synthetic data to increase the amount of available data. It is thought that the simpler sub-problems can be solved more effectively with the limited amount of available data and then applied in a cascaded pipeline to improve performance. In particular, a cascaded model was proposed that used three consecutive neural networks to identify the left ventricle, delineate the left ventricular myocardium and segment regions of myocardial infarction. The pipeline was trained based on manual segmentations of publicly available

LGE cardiac MR images from the automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI (EMIDEC) challenge. Additionally, a segmentation-conditional GAN was proposed that uses SPADE layers coupled by a ResNet encoder to synthesize realistic LGE images on given augmented labels, for the purpose of data augmentation.

The proposed cascaded pipeline outperformed direct segmentation consistently in both the left ventricular myocardium and scar segmentation by a mean DSC increase of 0.21 per-slice on the internal test set. The three-step pipeline also improved over the combinations of two-step pipelines. Furthermore, the performance was improved by the synthetic image augmentation, with a mean DSC increase of 0.06 for the scar segmentation. These reported DSC results were comparable to the inter- and intra-observer agreement found by Lalande et al. [116], intra-observer 0.84 and 0.76, inter-observer 0.83 and 0.69 for myocardium and scar segmentation, respectively. The impact of the synthetic data augmentation was also studied without the potentially confounding effects of the cascaded pipeline. That is, for the direct segmentation of scar and myocardium using a single 2D nnU-Net, models were trained with and without the synthetic data augmentation. As also found for the cascaded pipeline, for the direct segmentation the model trained with synthetic data augmentation significantly outperformed the model trained without the synthetic data.

A potential disadvantage of the cascaded approach is that errors can be propagated through the steps of the pipeline so that if, for example, an error is made in segmenting the myocardium, it will impact the subsequent scar segmentation. To test the effect of this error propagation, the model for scar segmentation was applied using the ground-truth myocardial segmentations and compared to using the predicted myocardium segmentations, on the internal test set. There is a small increase in mean DSC for the scar segmentation from 0.78 to 0.80 indicating that the negative impact of the cascaded pipeline is minimal and the cascaded approach still significantly outperforms the alternatives. This result confirms the findings of the original challenge, where cascaded pipelines were seen to perform well [120].

Our mean myocardium (0.86) and infarction (0.67) DSC scores compare favorably with the challenge results [120], with the DSC scores only being bettered by a single participant. This winning solution of Zhang reported a mean DSC of 0.88 for the myocardium and 0.71 for the infarction regions [118]. Zhang proposed a two-step system in which the coarse segmentation output of an initial 2D model was then input to a further post-processing 3D model to improve the 3D spatial consistency of the segmentations. Our proposed cascaded pipeline with GAN-based synthetic data augmentation performs better than all other challenge participants. For the purely 2D segmentation methods, our proposed pipeline represents a new state-of-the-art. Although it is not studied in this work, an extra post-processing step, similar to Zhang et al, has the potential to improve on this [118]. One of the typical disadvantages of using a 3D model in this application is that there are much less 3D images for training than 2D slices, and our initinal experiments with a 2.5D model did not improve results. However, as was shown in this work, it is possible to use synthetic images to augment the training dataset, and this is a possible future line of research to exploit the 3D nature of the data.

The current work uses rotations with dilation and opening of scar to augment the segmentation labels to input to the synthetic data generator. This work could be extended to use more complex patterns of scar and increase the robustness of the model. For example, patients with hypertrophic cardiomyopathy (HCM) often have complex patchy scar patterns and this could be simulated to allow training with a synthetic cohort of HCM patients without having to manually generate the training labels. Since the trained generator synthesizes the images based on a given LGE style image, different style images, from a difference acquisition sequence for example, could also be used to generate a more diverse training set.

The thorough evaluation of the impact of the synthetic data in this work using a challenging dataset with varying levels of contrast, noise, and artifacts indicates that the benefit is also likely to be more generally applicable to different applications and is also similar to that found in previous studies [56, 121, 122]. In addition to the use of GAN-based synthetic data, an advantage can also be seen to a cascaded pipeline, where the overall task is split into more manageable sub-problems, and together these approaches can be exploited to lower the manual annotation burden of deep learning.

5.5 Limitations

The major limitation of this work is that it used a homogenous cohort of selected patients, from the EMIDEC challenge. These images were acquired at a single center, using scanners from a single vendor and uniform imaging protocols. Therefore, the trained model may not generalize to different clinical cohorts due to the "domain-shift" (the varying levels of signal, noise and contrast, differing scan planning, and diverse disease patterns). Although methods are being developed to account with this [123], in future work, the model would need to be tested on images from different patient cohorts, scanners, and centers prior to clinical deployment.

Our approach of using a 2D model treats each imaging slice independently and does not take advantage of the 3D relations between the slices. Indeed, we observe sub-optimal performance in the apical (e.g Figure 5.5) and basal slices, as is commonly found for cardiac MRI segmentation ([46]), due to the more complex anatomy, thinner myocardium in the apical slices or LV outflow tract in the basal slices. In future work, 3D or long axis information could be incorporated to constrain the segmentations to be more spatially consistent. Potential extensions of the GAN-based image synthesis could also focus on generating more complex cases for training a more robust segmentation model.

This work segmented the scar and MVO regions as only one region of infarction and did not separate the MVO regions. The pipeline could be adapted, in future work, to consider the MVO regions separately. It also only focused on the identification of infarctions and the pipeline could also be extended beyond the segmentation to also classify patients' disease [124]. This could potentially incorporate spatial information of the scar, as well as other relevant clinical biomarkers to improve the classification [120].

5.6 Conclusion

In a population of patients with suspected acute myocardial infarction, our results demonstrate that a cascaded deep learning-based pipeline trained with augmentation by synthetically generated data leads to myocardium and scar segmentations and quantitative volume values that are similar to the manual operator. The three-step cascaded pipeline was shown to significantly outperform direct segmentation with a mean DSC increase of 0.26 per slice. Additionally, the inclusion of GAN-based synthetic images as data augmentation further improved the performance and yielded a further mean DSC increase of 0.06 per-subject for scar segmentation.

Supplementary Material

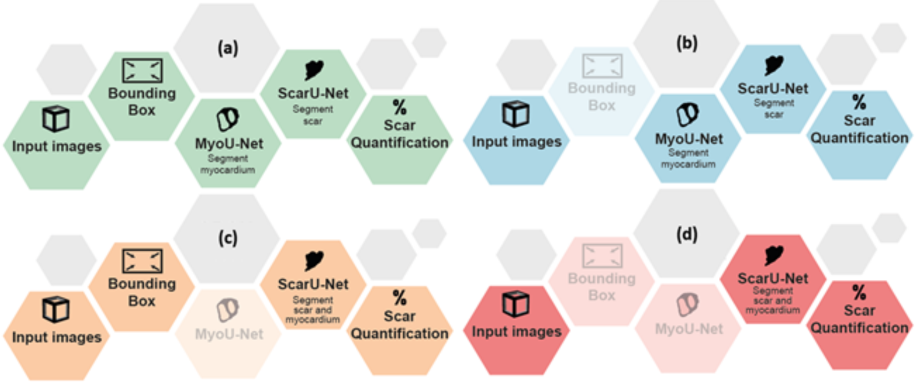


Figure 5.7: A schematic representation of the ablation test, showing the different combinations of steps tested in the pipeline.

Type of Augmentation	Value
Gaussian Noise	μ= 0.1, σ= 0.1
Gaussian Blur	σ= 1.5
Shear	U(-20, 20)
Rotation	U(-90, 90)
Translation (independent in x and y direction)	$\pm U$(0.14, 0.21)
Scale	U(0.5, 1.5)

Table 5.3: The parameters used for the data augmentation in the bounding box training. U(a, b) denotes that the parameter value was randomly sampled from a uniform distribution on the interval [a, b]

Type of Augmentation	Value
Rotation	0, 60, 120, 180, 240, 300
Elastic deformation [1]	α= 50, σ=5
Morphological opening	-
Morphological dilation	-

Table 5.4: The parameters used for the augmentation of ground-truth labels for image synthesis.

CHAPTER 6

Image Synthesis for Reducing Segmentation Failures

This chapter is based on:

Y. Al Khalil*, S. Amirrajab*, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "Reducing Segmentation Failures in Cardiac MRI via Late-feature Fusion and GAN-based Augmentation," submitted, 2022.

Which is an extension of:

Y. Al Khalil*, S. Amirrajab*, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "Late Fusion U-Net with GAN-based Augmentation for Generalizable Cardiac MRI Segmentation," M&Ms-2 challenge MICCAI 2021.

Abstract

Cardiac magnetic resonance (CMR) image segmentation is an integral step in the analysis of cardiac function and diagnosis of heart related diseases. While recent deep learning-based approaches in automatic segmentation have shown great promise to alleviate the need for manual segmentation, most of these are not applicable to realistic clinical scenarios. This is largely due to training on mainly homogeneous datasets, without variation in acquisition, which typically occurs in multi-vendor and multi-site settings, as well as pathological data. Such approaches frequently exhibit a degradation in prediction performance, particularly on outlier cases commonly associated with difficult pathologies, artifacts and extensive changes in tissue shape and appearance. In this work, we present a model aimed at segmenting all three cardiac structures in a multi-center, multi-disease and multi-view scenario. We propose a pipeline, addressing different challenges with segmentation of such heterogeneous data, consisting of heart region detection, augmentation through image synthesis and a late-fusion segmentation approach. Extensive experiments and analysis demonstrate the ability of the proposed approach to tackle the presence of outlier cases during both training and testing, allowing for better adaptation to unseen and difficult examples. Overall, we show that the effective reduction of segmentation failures on outlier cases has a positive impact on not only the average segmentation performance, but also on the estimation of clinical parameters, leading to a better consistency in derived metrics.

6.1 Introduction

Accurate segmentation of cardiovascular magnetic resonance (CMR) images is an essential step for heart structure and function assessment, as well as a reliable diagnosis of major cardiovascular diseases [108]. In current-day clinical practice this procedure is typically performed manually or semi-automatically, requiring significant input and correction from clinicians. However, recent developments in automating this task have achieved a remarkable performance. These include approaches ranging from more classical techniques based on statistical shape models or cardiac atlases to newer deep learning (DL) based models, which have gradually outperformed previous state-of-the-art methods [111]. However, most DL methods proposed in the literature have been trained and evaluated using images acquired from single clinical centers, utilizing similar imaging protocols and hardware. Consequently, such models exhibit a significant drop in performance when evaluated on unseen, out-of-distribution data such as abnormal and pathological cases not included in the training set, often characterized by a considerable amount of outliers [125–127]. While typically defined as erroneous or low quality samples, we refer to outliers as data samples exhibiting rare conditions, under-represented in the training data, which often occur at the deployment time. A rare occurrence of such classes during training negatively affects model's ability to adapt to their appearance at test time, leading to a significant degradation in prediction performance and generalization ability.

6.1.1 Challenges of CMR Segmentation

A change in acquisition parameters causes cardiac MR images to exhibit a great variability in terms of contrast, texture and noise. On the other hand, variation in patient characteristics causes significant divergence in tissue shapes and sizes. Such diversity of imaging characteristics is further intensified by changes in scanner models or vendors. The appearance of pathology has a significant influence on the ventricle morphological variation, resulting in unique tissue shapes and contrast, often under-represented in datasets available for training. Ventricular remodeling further causes changes in heart mass, geometry, function and respiratory wall motion. Segmentation is most commonly hindered by the appearance of dilated left and right ventricles, causing increased wall thickness and regional wall abnormalities. Diseases such as tetrology of fallot and defects in inter-atrial communication induce challenges such as the overriding aorta, right ventricular outflow tract obstruction and pulmonary stenosis. Moreover, segmentation difficulties are caused by gray level inhomogeneities in the blood flow, as well as the presence of papillary muscles and trabeculations, which exhibit the same intensity levels as the myocardium.

Finally, segmentation complexity is largely affected by the slice level of the image, where apical and basal slices are more difficult to segment compared to mid-ventricular slices. Due to low MRI resolution, sizes of small structures at the apex and base are often incorrectly estimated due to the vicinity of the atria. Moreover, while the short-axis image orientation is typically used to develop segmentation algorithms due to its efficiency for analysing both ventricles, it is not fully optimized for the right ventricle [128–130].

6.1.2 Related Work

Recent attempts to handle issues with model robustness and a large number of outliers propose training with images acquired from multiple large cohorts; however, these works do not explicitly evaluate the trained models on completely unseen cohorts from other centers, nor directly address the domain shift between training and unseen cohorts [39, 105, 106, 131]. Other research focused on image and latent space augmentation techniques on models trained and evaluated mostly using single cohorts, with some examples of performance evaluation on a limited number of unseen cohorts [132–135]. However, such approaches are limited by different standards in annotation operating procedures, experiments conducted on private data, as well as the need for a training set sufficiently large to model immense variability across subjects. Subsequently, these models may perform significantly worse in clinical settings due to difficulties adapting to a more heterogeneous subject population, typically containing a significant number of outlier cases.

Besides data augmentation, other techniques incorporating modifications in model architectures have been proposed to improve the robustness of DL models [105]. Solutions such as transfer learning [136] have been successful, but are limited by the requirement to perform fine-tuning for each specific domain. Domain adaptation approaches [137], focused on extracting domain invariant features, discriminative enough for the task at hand, have shown variable, but promising results for a number of image analysis applications [138, 139]. However, many studies still report inconclusive results about the positive effects of domain adaptation on out-of-domain, unseen data [140]. An attempt to address the issue of generalization in CMR segmentation is a recently organized M&Ms challenge, providing a benchmark for testing CMR segmentation algorithms on data from different centers and scanner vendors [46]. Many approaches presented throughout this challenge demonstrated improvements in domain adaptation, adversarial training, disentangled representation learning and augmentation to improve model generalization. However, few studies additionally focus on the performance of such methods in the presence of diseased tissue, as well as on tackling outlier cases hindering the overall performance of the proposed approaches.

Recent developments in the area of generative adversarial networks (GANs) have paved the way towards a number of interesting medical imaging applications, from style transfer to utilizing GAN-like architectures for classification or segmentation [9, 53, 83, 141]. However, methods focused on medical image synthesis have captured the most attention due to their ability to generate realistic-looking medical images, thus having a potential to increase and vary the available training data [13, 86–88]. While a lot of research so far has focused on improving the quality of image synthesis, a small amount of work evaluates their applicability across different medical image analysis tasks. Moreover, the application of GAN-synthesized images to address algorithm robustness in the presence of out-of-domain images, as well as on data undergoing variations in size, shape and contrast induced by the presence of pathology, has rarely been explored.

6.1.3 Our Contributions

Motivated by the observed heterogeneity in cardiac MR images, we propose a multi-stage pipeline aimed at improving segmentation robustness and handling outliers in multi-vendor, multi-center, multi-view and multi-disease cardiac MRI data.We identify the main aspects causing a domain shift between images of different cardiac pathologies, acquired from different sources, and hypothesize that these properties can be simulated by applying a series of steps proposed in this work. To reduce the effect of variations in the FOV and heart size, we introduce a heart region detection module, trained to constrain the amount of visible background tissue and centralize the heart in the image. To address variations in contrast, heart appearance and shape, we utilize conditional GANs to generate a large number of highly realistic and diverse images with corresponding labels. We particularly focus on generating sufficient pathological examples to balance the ratio between pathological and normal cases. In addition, we handle variations in contrast by performing a series of intensity transformations, aimed at emphasizing tissue shape. To regularize the segmentation performance and produce a robust model that is able to handle difficult, outlier cases, we introduce a late fusion approach to training.

Throughout our experiments, we (i) systematically analyze the effect of the proposed pipeline on the model robustness across publicly available M&Ms-2 challenge data[1] [46] and show that the proposed technique is able to improve the performance on outlier cases;(ii) demonstrate the importance of outlier reduction on the overall segmentation performance on pathological data, as well as clinically-relevant parameters;(iii) handle the limitation in data availability by introducing a conditional image synthesis module, able to generate highly realistic and diverse images with corresponding labels; (iv) address the mis-segmentation of pathological tissue altering heart tissue intensity levels and shape by utilizing a variational auto-encoder to increase the diversity of pathological examples in the training set and (v) show that the proposed pipeline has the ability to adapt to completely unseen, out-of-domain datasets.

[1] More information about the M&Ms-2 challenge and the data provided can be accessed at `https://www.ub.edu/mnms-2/`

6.2 Materials and Methods

6.2.1 Method Overview

The overview of the proposed pipeline is shown in Figure 6.2, consisting of 1) heart region detection module, 2) label-conditional image synthesis with a variational autoencoder (VAE) for label deformation and 3) late fusion-based segmentation module utilizing transformed versions of input images during training. We apply the proposed method to both short-axis and long-axis cardiac MR images. In the following sections, we introduce each component of the pipeline, as well as the data used for training and validation.

6.2.2 Data

The M&Ms-2 challenge data is comprised of 360 patients with a variety of right-ventricle (RV) and left-ventricle (LV) pathologies, as well as a control group, distributed as shown in Table 6.1. The data is acquired using different 1.5T and 3.0T scanners from three different manufacturer vendors (Siemens, GE, and Philips), undergoing variations in contrast and anatomy (see Figure 6.1). The in-plane resolution of the provided images varies between 0.78 to 1.57 mm, with slice thickness ranging from 8.6 to 14 mm, resulting in a total number of slices varying between 9 to 13 slices per each short-axis image.

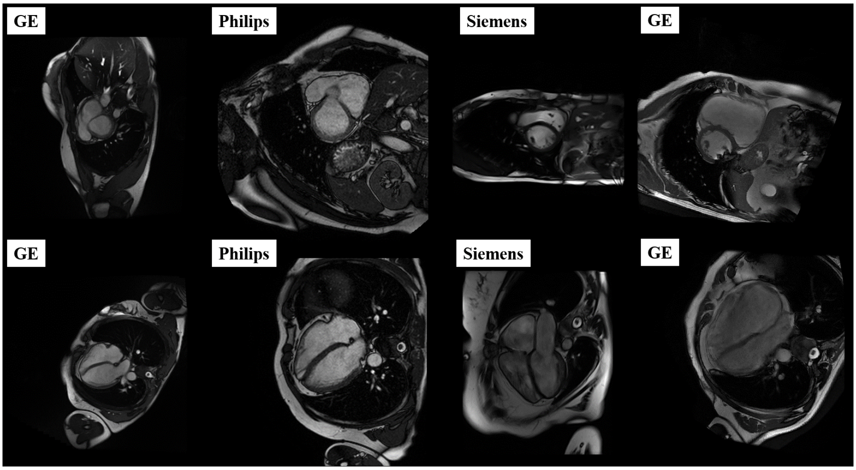


Figure 6.1: Variations in field-of-view, image contrast and appearance, anatomy, and pathology for SA and LA images in the training set.

The training subset includes 160 cases with expert annotations for RV and LV blood pool, as well as the LV myocardium (MYO). The short-axis and long-axis view is provided for each patient. The training set contains five different types of LV and RV pathologies, as well as healthy subjects. The validation set contains 40 cases with 10 cases of pathologies not present in the training set. The final algorithm is evaluated on a separate test set containing 160 cases as outlined in Table 6.1. We use the provided validation set for testing, increasing the size of the testing set to a total of 200 patients (or 400 ED and ES SA/LA images) while the evaluation and the development of the algorithm is exclusively done on the training set alone. All images were annotated by two annotators according to the same standard operating procedure (SOP) used for the ACDC MICCAI 2017 challenge [39], while maintaining consistency between short and 4 chambers long-axis in basal and apical regions.

Table 6.1: Distribution of the M&Ms-2 challenge data per pathology. Note that all values represent the total number of studies, taken at both ED and ES phases.

Pathology	Training	Validation	Testing
Dilated Right Ventricle (RV)	0	5	25
Tricuspidal Regurgitation (TRI)	0	5	25
Tetralogy of Fallot (FALL)	20	5	10
Interatrial Communication (CIA)	20	5	10
Congenital Arrhythmogenesis (ARR)	20	5	10
Dilated Left Ventricle (LV)	30	5	25
Hypetrophic Cardiomiopathy (HCM)	30	5	25
Normal (NOR)	40	5	30

6.2.3 Heart Region Detection Module

Cardiac MR images acquired at different sites and from varying scanner vendors, typically undergo changes in the acquisition protocol, resulting in images of varying resolution and FOV. This leads to varying heart sizes across different scans, where the heart often takes up only a small portion of the image compared to the background. Experiments show that neural networks trained on images of varying FOV, without ensuring that there is an equal representation of different heart sizes in the training set, often confuse background tissue for cardiac tissue and lead to a large number of false positive predictions.

To address this, we train a regression-based convolutional neural network (CNN), proposed in [94], to automatically detect a bounding box encompassing the heart in both SA and LA images. The detected bounding box is then used for cropping the full FOV images at inference time. The CNN is trained in a supervised manner, with labels obtained from ground truth masks available in the training set by computing the smallest bounding box that fits the entire heart in the FOV and expanding it by 25 voxels to include some background tissue. Before generating the training labels, we resample all SA images to a median spatial resolution of $1.25 \times 1.25 \times 10mm^3$ and all LA images to a spatial resolution of $1.25 \times 1.25mm^2$.

The inputs to the network are 1000 2D (256 × 256) mid-cavity SA slices extracted from the training data-set and all LA slices, normalized to intensity values in the range of [0,1]. The cropped SA and LA images using the predicted bounding box are post-processed to the size of 128×128 voxels and 176×176 voxels, respectively. While the trained network is applied on test images during evaluation, training images are cropped manually using the available ground truth labels. A detailed description of the architecture and training procedure is available in Supplementary Material 6.6.

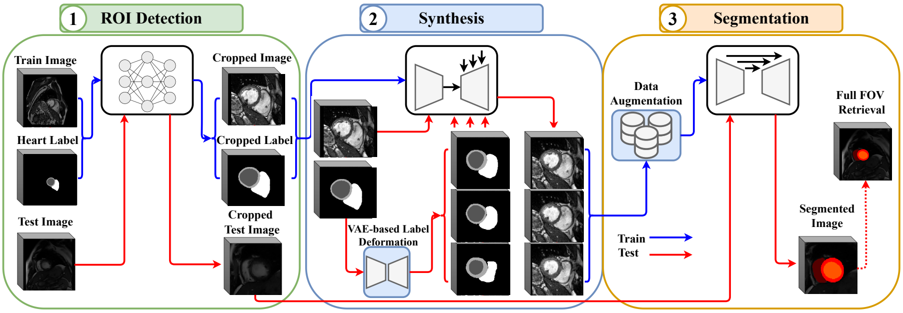


Figure 6.2: Proposed pipeline including the ROI detection module (left), image synthesis module (middle) with VAE-based label deformation, and image segmentation module (right).

6.2.4 Synthesis Module

The synthesis module encompasses two models; i) image synthesis via label-conditional GANs and ii) label deformation via latent space manipulation in VAEs. The conditional GANs translate the ground truth labels to realistic images while the VAEs produces new labels with anatomically plausible deformations.

Conditional Image Synthesis

The image synthesis model is comprised of a ResNet-based [76] style encoder coupled with a label-conditional generator that uses spatially adaptive normalization layers (SPADE) [75] throughout the network architecture. The ResNet encoder is designed to extract style information of the input image and provide it to the generator that preserves the content of the input label map via the conditional SPADE normalization layers. The input image is first fed to the ResNet encoder including a set of convolutions, downsampling and residual blocks followed by two fully connected layers to extract the style information in the bottleneck. This information is then passed to the generator that consists of six SPADE residual blocks, each including SPADE normalization layers that utilize corresponding segmentation mask of the input image for modulating the activation [75].

Previous works in [19, 20] have shown the effectiveness of using SPADE-based generators in translating input segmentation labels to realistic CMR images. In contrast to their work, our approach alleviates the need for providing multi-tissue segmentation masks for high-quality synthesis by adding the ResNet style encoder network. Moreover, to introduce anatomical variations, random elastic deformation and morphological dilation are applied on the segmentation masks in a previous work [94]. Despite showing the benefit of label deformation through morphological operations, the heart anatomy of the synthesized subjects is not necessarily anatomically plausible. Here we propose a VAE-based label deformation approach that would provide us with more plausible heart deformations via label encoding and latent space manipulation.

VAE-Based Label Deformation

Instead of elastically deforming the labels as in [94], we propose a deep-learning based label deformation using Variational Autoencoders (VAEs) aiming to learn the underlying factors of heart geometries from the ground truth labels. The VAE model encodes the shape information of the heart in a compressed manner in the latent space during training. We add random perturbations to the latent code of the original label and then perform label reconstruction by feeding the manipulated latent code to the decoder network. This manipulation of the latent code changes the heart geometry of the reconstructed label. The rationale behind this approach is that we attempt to directly manipulate the learned geometrical features of the input label in the latent space rather than randomly deform the labels in the image space.

The input of the VAE model is a one-hot encoding version of the label map including four channels for cardiac classes and background. The encoder part of the model includes four convolutional blocks with three convolutional layers each followed by batch normalization (BN) and LeakyReLU activation function. The encoded features are fed to four sequential fully connected layers to output the parameters of a Gaussian prior over the latent representation. The decoder is comprised of four convolutional blocks each with one up-sampling layer followed by two convolutional layers with BN and LeakyReLU. The last additional block of the decoder includes one convolutional layer followed by BN and another convolution with four channel outputs and Softmax activation function. The VAE model is trained using a weighted combination of cross-entropy loss as the reconstruction loss and Kullback-Leibler divergence (KLD) with a weighting factor of β for regularization of the latent space capacity [96]. We experimentally identify the size of the latent vector ($n_z = 16$) and weight of KLD ($\beta = 15$) by inspecting the quality of the label reconstruction and the outcome of latent manipulation.

6.2.5 Synthesis Strategy

Two identical image synthesis models are trained using LA and SA cardiac MR images. To augment and balance the data using these trained synthesis models, the following strategies are devised. For each vendor-specific subset, the outlier cases are identified based on the end-diastolic or end-systolic volume for the RV calculated using the ground truth label of the SA images. These outlier cases, separated from the rest of the population, are used for image synthesis. For balancing the ratio between outlier cases and the rest of the population, we add different perturbations in the form of Gaussian noise to the corresponding latent space of each label to manipulate labels. This is done in a way that we eventually create roughly 2000 synthesized cases including 50% outliers and 50% the rest of cases. We follow the same strategy for the data from each scanner vendor.

The same strategy is not optimal for LA images as we observe anatomical distortions when noise is added to the latent space of the LA slice. We hypothesize this might be due to having a limited number of LA slices compared to SA stacks and consequently not learning a rich latent space for coherent sampling and accurate reconstruction. Instead, we interpolate between the latent codes from end-diastolic and end-systolic phases of the same subject and feed the interpolated latent codes to the decoder to reconstruct the intermediate shapes. We additionally apply elastic deformation to create more anatomical variations for the LA images.

6.2.6 Segmentation Module

Contrast Transformations to Enhance Heart Shape Features

One of the main challenges of deploying a segmentation algorithm on heterogeneous data is its performance in the presence of extensive contrast and intensity variations, arising from a combination of different protocols, signal weighting techniques and hardware used for acquisition. Applying image appearance transformations during training introduces a diversity in image contrasts, prevents overfitting and focuses the model optimization toward the fundamental geometry and shape of the target tissue.

We select a set of six contrast transformations per image, each fed into a separate encoding path during training of the late fusion model. First, we match the intensities of images to those representative of each scanner vendor by utilizing histogram standardization [142]. To that end, we generate a standardized set of image histogram landmarks per vendor, used as a reference for matching the histograms of each image at both training and testing time. Next, we apply Total Variation (TV) based denoising [143] to discard high frequency image components and emphasize tissue shape in both training and testing images. The scale of the TV filter is controlled by changing the smoothing parameter α, where $\alpha \in [0.1, 15]$. To additionally emphasize tissue edges and flatten the image, while retaining the general appearance, we apply a combination of solarization and posterization.

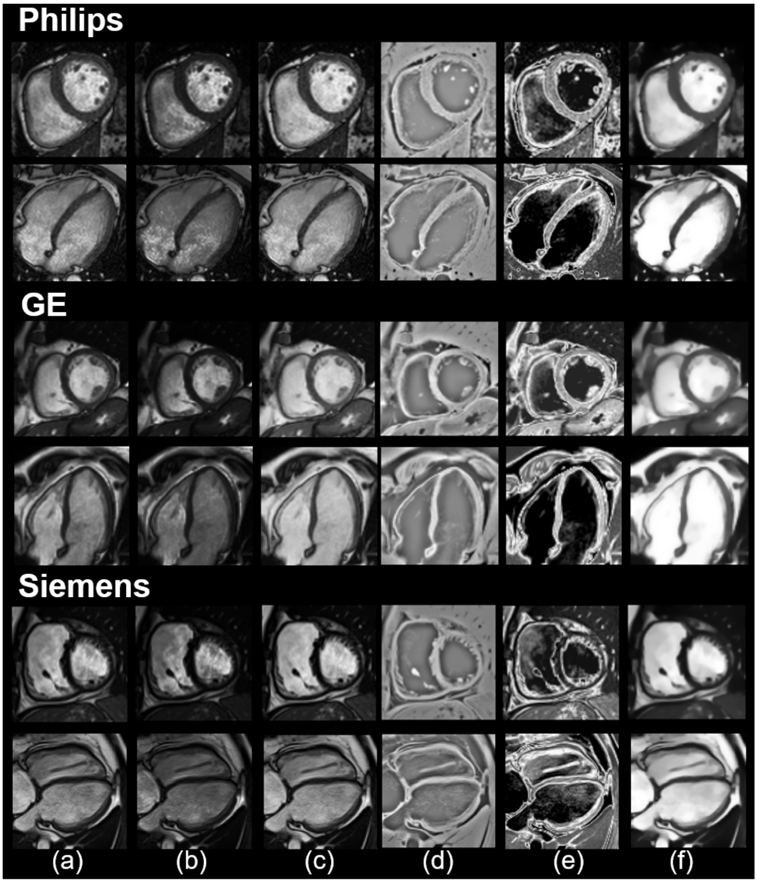


Figure 6.3: Examples of contrast-transformed SA and LA training images per vendor (Philips, GE and Siemens). Transformations applied include: (a) histogram standardization to GE images, (b) histogram standardization to Philips images, (c) histogram standardization to Siemens images, (d) a Laplacian operator, (e) a combination of solarization and posterization and (f) TV-based filtering.

Finally, we calculate the Laplacian of the image, to highlight the regions of rapid intensity changes and outline the shapes of major objects in the image. The effect of each transformation can be observed in Figure 6.3, resulting in a sequence of six augmented images.

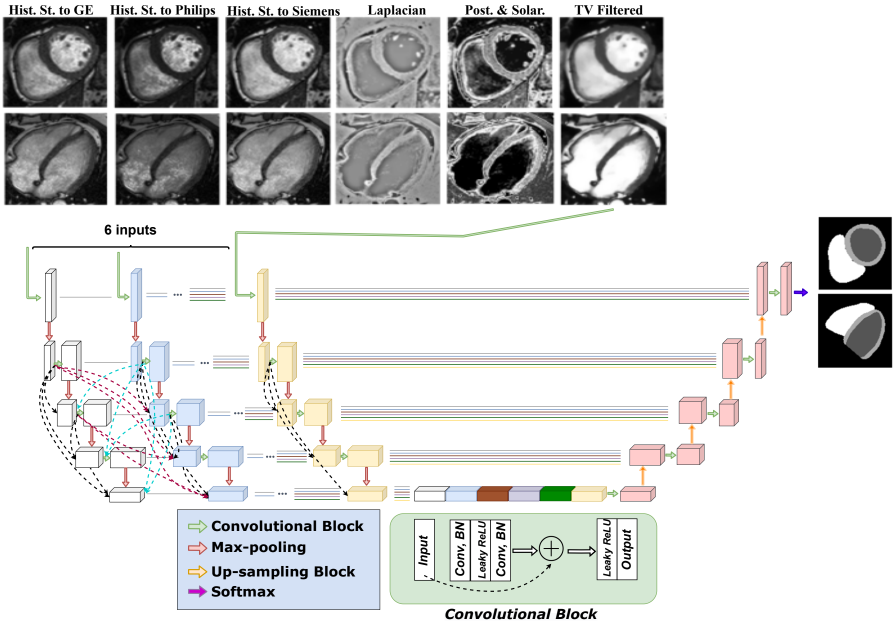


Figure 6.4: Late fusion multi-encoder U-Net proposed in this work, used for the segmentation of LA and SA MR images. Dotted lines represent some of the inter- and intra-stream dense connections applied at each layer. At both training and testing time, the network processes 6 transformed variations of the input image, shown at the top of the figure, to produce the final segmentation LA and SA segmentation maps.

Network Architecture

Most existing CNN techniques targeting robustness across multi-site and multi-vendor images typically employ a diverse set of transformation methods on the available data. These techniques usually follow an early-fusion strategy, integrating information extracted from various data transformations from the original space of low-level features, merged at the input of the network. Similar approaches have also been proposed for applications utilizing multi-modal imaging data [144–146]. However, research has shown that the detection of inter-relations between the low-level features of different modalities is a challenging process, particularly due to the non-linear nature of these relationships and different statistical properties these modalities exhibit [147, 148].

Instead of simply combining multi-modal data using early fusion, recent methods propose deep architectures to effectively fuse higher-level information from different modalities (late fusion), with the assumption that the extracted high-level representations are more complementary to each other [149–152]. Inspired by such late fusion approaches, we modify the nnU-net [40] architecture to include multiple encoder layers processing each transformed image fed at the input in a separate path. The extracted features by each encoder are then fused at the bottleneck, allowing the network to learn complementary information between different transformations of each image and a better representation of their interrelationships.

Furthermore, we extend the standard convolutional layers into a convolutional block, consisting of two convolutional and linear units, with batch normalization (BN) applied between each convolution and leaky rectified linear unit (ReLU). We use a short residual connection to sum the input at each covolutional block with the output coming from the second convolutional layer, followed by leaky ReLU to produce an output. Each encoding path consists of five convolutional blocks, with four max-pooling layers.

To improve the modeling of relationships between different streams and promote the learning of highly complex, but more discriminative features, we adopt hyper-dense connections between multiple streams, as proposed in [150, 153, 154]. This helps with information and gradient propagation through the entire network and has a regularizing effect, reducing the risk of overfitting and improving generalization. As shown in Figure 6.4, the outputs from previous layers across different streams are concatenated at the input of each subsequent layer per stream. We additionally shuffle the feature maps from densely connected layers by concatenating them in a different order per branch and layer, which is shown to have strong regularizing effects [155]. To illustrate this, let x_l denote the output of the l^{th} layer and F_l the mapping function, such as a convolution layer with a non-linear activation or a complete convolutional block. Typically, the output of the l^{th} layer in CNNs is derived by passing the output of the previous layer, x_{l-1}, through a mapping function:

$$x_l = F_l(x_{l-1}). \tag{6.1}$$

In a densely connected network, this can be extended to

$$x_l = F_l([x_{l-1}, x_{l-2}, x_{l-3}, \ldots, x_0]), \tag{6.2}$$

indicating that all previous feature outputs are concatenated in a feed-forward fashion. By introducing inter-stream hyper-dense connections and feature shuffling, the output of the l^{th} layer in a given stream s, x_l^s, where we consider two streams only, can be defined as

$$x_l^s = F_l^s\left(\phi_l^s\left(x_{l-1}^1, x_{l-1}^2, x_{l-2}^1, x_{l-2}^2, \ldots, x_0^1, x_0^2\right)\right), \tag{6.3}$$

where ϕ_l^s represents a feature map permutation function. Thus, the output of the l^{th} layers in a two-stream network can be represented as

$$\begin{aligned} x_l^1 &= F_l^1([x_{l-1}^1, x_{l-1}^2, x_{l-2}^1, x_{l-2}^2, \ldots, x_0^1, x_0^2]) \\ x_l^2 &= F_l^2([x_{l-1}^2, x_{l-1}^1, x_{l-2}^2, x_{l-2}^1, \ldots, x_0^2, x_0^1]). \end{aligned} \tag{6.4}$$

Training Procedure

All SA and LA images used for training are first resampled to a median pixel spacing of $1.25 \times 1.25 \times 10mm^3$ and a spatial resolution of 1.25×1.25, respectively. This is followed by a 98^{th} percentile normalization to an intensity range from $[0, 1]$. All training images are further cropped to reduce the FOV to the region of interest (heart), as described in Section 6.2.3, while the heart region detection module is used at inference time. This results in all SA and LA images cropped to the size of 128×128 voxels and 176×176 voxels, respectively. Finally, all images are processed to form a set of six different contrast-transformed images (see Section 6.2.6), at both training and testing time.

After pre-processing, each encoding path is fed with batches of 60 128×128 images for training the SA segmentation model and batches of 20 176×176 images for the LA model. To further increase robustness during training, we employ data augmentation in the form of random vertical and horizontal flips ($p = 0.5$), random rotation by integer multiples of $\frac{\pi}{2}$ ($p = 0.5$), random scaling with a scale factor of s $\in$ $[0.8, 1.2]$ (p=0.2), mirroring ($p = 0.3$) and random elastic deformations ($p = 0.3$). All augmentations are applied on the fly during training.

To train the network, we use a weighted sum of the categorical cross-entropy and Dice loss. We employ Adam optimizer to train the proposed model, with an initial learning rate of 1×10^{-4} and a weight decay of $5 \times e^{-5}$. The training of all models converges in 500 epochs, where the initial learning rate is reduced by a factor of 5 if the validation loss does not improve by at least 5×10^{-3} for the last 50 epochs. We apply early stopping on the validation set and select the model with the highest accuracy, to avoid overfitting. We train each model (LA and SA) using a five-fold cross-validation on the training set and use them as an ensemble to produce final predictions on the validation and test sets. The implementation of the model was done in Pytorch and all experiments were performed on an Nvidia TITAN XP GPU with 12 GBs of RAM.

Post-processing

We perform a connected component analysis on the predicted labels and remove all but the largest connected component per class, which handles most false positive predictions. Since test images are both resampled and cropped, we first restore the original size using the cropping parameters predicted by a heart region detection module and perform bilinear upsampling to recover the original resolution.

6.3 Experiments

Experiment setup: We train the proposed pipeline on all images provided as a part of the M&Ms-2 training data, consisting of 70, 64 and 26 studies acquired from Siemens, Philips and GE scanners, respectively, and augment the training set with synthetic images generated using the method described in Section 6.2.4. All studies consist of LA and SA images, at both end-diastolic and end-systolic phases, whereby we train two separate networks per image view (LA vs. SA images). The segmentation performance of the proposed pipeline is compared to the **baseline** model, which is a single-channel nnU-Net [40] combined with heart region detection module. The model is trained on all available real training images from the dataset in a 5-fold cross-validation setup, using the standard augmentation set-up as proposed in [40], without any additional synthetic images.

Overall analysis: The obtained results are evaluated on the unseen test set, containing images acquired across all 3 scanners that have not been previously utilized for the training of any pipeline component. We assess the performance both qualitatively and quantitatively, in terms of standard metrics, such as the Dice score and Hausdorff distance. This is further supported by deriving clinical indicators, such as ventricular volumes and ejection fraction to further assess the benefits of the proposed approach. Detailed discussion is provided in Section 6.4.2.

Analysis per pathology: Since the major focus of this work is accurate segmentation of cardiac tissue in patients affected by geometrical and textual complexities appearing due to cardiac pathology, we evaluate the proposed pipeline across different diseases available in the test set. In total, we report the results on 160 patient studies, grouped per disease, as well as the normal subjects (seen in Table 6.1), for both SA and LA images, available in Section 6.4.3.

We perform additional evaluation on out-of-domain data, namely the short-axis ACDC and M&Ms-1 challenge data (Supplementary Material 6.6) to study the robustness of the proposed method. Detailed results are discussed in Section 6.4.4. Finally, to study the impact of different pipeline components on segmentation performance, we conduct an ablation experiment by removing one or several elements of the proposed method. All models are evaluated across the whole test set in terms of the Dice score for both SA and LA images, with results available in Section 6.4.5.

6.4 Results

6.4.1 Image Synthesis Results

As discussed in Section 6.2.5, we balance out the number of outlier and normal cases in the final synthetic data by applying a different number of deformations (by adding Gaussian noise to the latent space) on each group. Randomly generated examples from each group are shown in Figure 6.5.

Figure 6.66.6 shows the RV and LV volumes at ED and ES phases for the real and synthetic data distribution to inspect how the synthetic population changes the heart cavity distribution of subjects. Each subject is represented as a point with different shape and color for its corresponding scanner vendor. We can

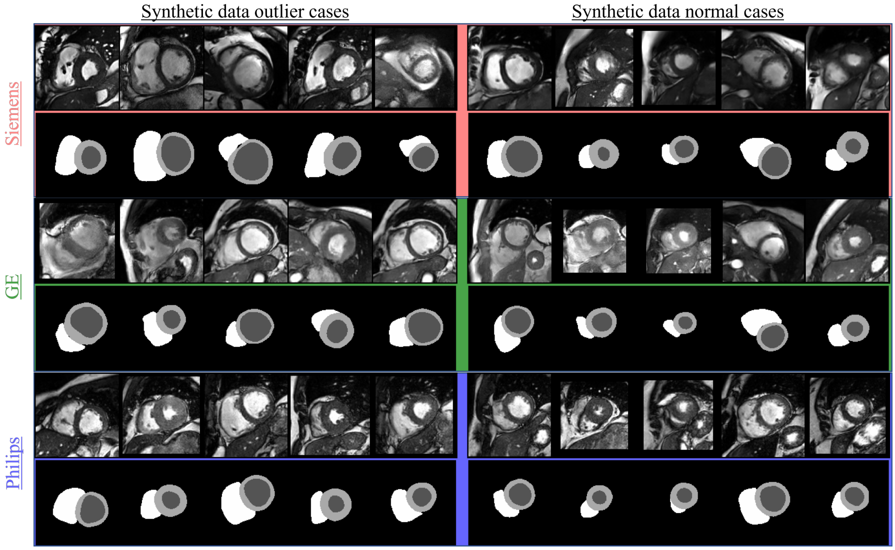


Figure 6.5: Random synthetic examples for outlier and normal cases with corresponding labels, stratified into different scanner vendors for short-axis slices.

observe a gap between subjects in the real data distribution (indicated with black oval) that is filled with generated subjects in the synthetic data distribution as a result of deforming labels and synthesizing more subjects for each real subject. Moreover, the number of samples near the mean of the distribution are increased, resulting in a densely populated area covered by synthetic subjects with normal ranges of ventricular volumes.

Table 6.2: Segmentation performance comparison between the baseline and the proposed model in this work, evaluated on short-axis (SA) and long-axis (LA) test images, across all cardiac tissues. Numbers listed in the table are the means and standard deviations of Dice (DSC) and Hausdorff Distance (HD) scores. DSC and HD values indicated in bold are those which are significantly higher compared to the baseline performance, according to the Wilcoxon signed-rank test for p<0.01.

View	**Method**	**Dice**			**HD**		
		LV	MYO	RV	LV	MYO	RV
SA	Baseline	0.941 (0.05)	0.881 (0.06)	0.923 (0.07)	9.36 (9.22)	13.93 (12.76)	11.71 (10.84)
	Proposed	**0.959 (0.02)**	**0.907 (0.04)**	**0.938 (0.03)**	**6.42 (4.38)**	**9.37 (5.88)**	**8.62 (6.07)**
LA	Baseline	0.947 (0.07)	0.871 (0.08)	0.902 (0.08)	5.04 (4.87)	7.42 (7.21)	7.78 (7.18)
	Proposed	0.958 (0.03)	**0.901 (0.03)**	**0.924 (0.04)**	4.07 (2.09)	**5.27 (3.31)**	**5.81 (3.42)**

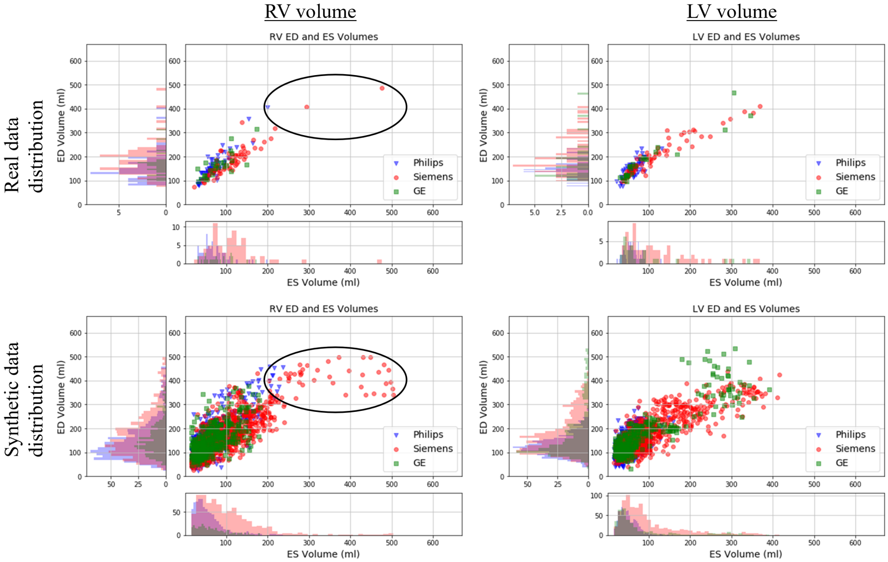


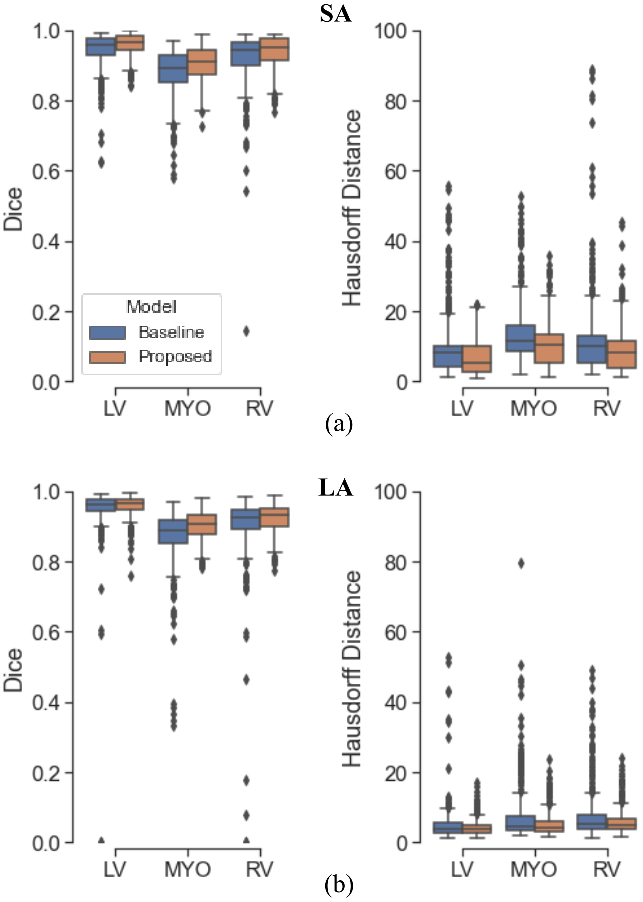
Figure 6.6: Distribution of the RV and LV volumes at ED and ES for real and synthetic data. Each subject is represented as a marker (with different colors and shapes indicating the corresponding scanner vendor).

6.4.2 Overall Segmentation Results

Table 6.2 shows the quantitative results in terms of Dice and HD scores obtained by the proposed pipeline compared to the baseline model, across SA and LA images available in the test set. The obtained results suggest significant improvements in segmentation performance across most tissues, except for the LV in LA images, which we ascribe to relative consistency of LV shape over the long-axis view. However, visual observation suggests improvements in patients with dilated left ventricle (LVD) and hypertrophic cardiomyopathy (HCM), which both cause changes in LV shape and appearance. Additional observation of score distribution, depicted in Figure 6.7, suggests a significant reduction in the number of outlier predictions by the proposed approach across both SA and LA views.

This has a particular impact on the segmentation of the right-ventricular (RV) blood-pool and myocardium (MYO), whereby visual observation of delineations implies that the existing outliers predicted using the baseline mostly relate to false positive predictions, specifically in relation to over-estimation of both the LV and RV blood-pool, which further causes the under-segmentation of the myocardium.

Further inspection suggests that over-segmentation mainly occurs in the basal region of the heart, whereby the baseline model falsely predicts the presence of the RV and other tissues, particularly at the boundary of the pulmonary artery



(a)

(b)

Figure 6.7: Segmentation performance of the proposed and baseline models on (a) SA and (LA) images available in the test set, across three cardiac tissues (LV, MYO and RV). Performance is reported in terms of Dice and Hausdorff distance (HD) scores.

and the right atrium. Under-segmentation by the baseline commonly occurs at the apex of the heart, where endocardium appears smaller and tissue boundaries are less well-defined. The presence of dense papillary muscles at the apex often causes further difficulties for accurate segmentation. The observations obtained by visual inspection are confirmed by quantitative evaluation performed across heart regions, shown in Figure 6.8.

Compared to mid-ventricular slices there is an evident performance drop around both the base and the apex of the heart, which is noticeably improved by the

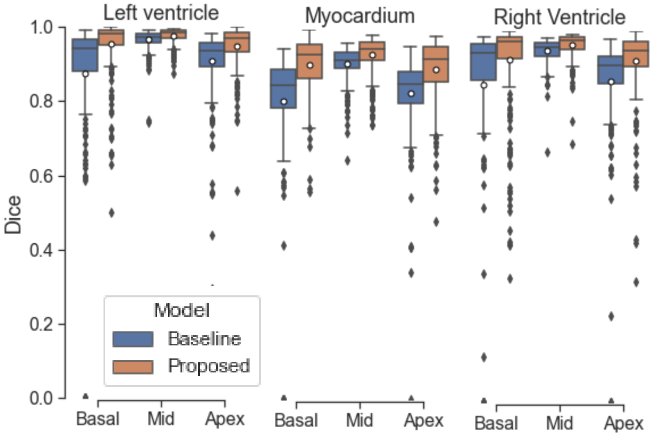


Figure 6.8: Segmentation performance of proposed and baseline models in SA images across basal, mid-ventricular and apical heart regions, calculated per cardiac tissue.

proposed approach. We hypothesize the improvement largely stems from augmentation with synthetic data, where we focus on including a vast array of examples with variable appearance of tissues in the basal and apical regions of the heart, as well as simulating the effects of heart pathology on cardiac tissue appearance and the presence of artifacts. However, a moderate performance drop when segmenting both ends of the heart compared to the mid-ventricular region, obtained by the proposed approach, suggests that under- and over-segmentation at the base and the apex of the heart is still not a completely resolved problem.

To gain more insight into the value and importance of outlier reduction achieved by the proposed segmentation method, we evaluate three automatically derived clinical parameters with reference to manually derived ones using the available ground truth segmentation masks. Namely, we derive the RV (Figure 6.9) and LV (Figure 6.10) end-diastolic (EDV) and end-systolic (ESV) volumes, as well as the ejection fraction (EF) from segmentations obtained by both the proposed and baseline model across the entire test set.

Correlation plots of baseline and proposed ED and ES RV volumes in Figure 6.9 show significant improvements in both EDV and ESV correlation acquired from the proposed method, which leads to better agreement between manual and automatically quantified EF compared to the baseline. Similar effects are observed in Figure 6.10 for all three clinical parameters related to the LV. These results can largely be attributed to a smaller number of outlier predictions, which in turn decrease the difference between the calculated ED and ES volumes compared to those derived from ground-truth labels. Moreover, the remaining outliers are still relatively close to the acceptable range of deviation, which reduces their overall

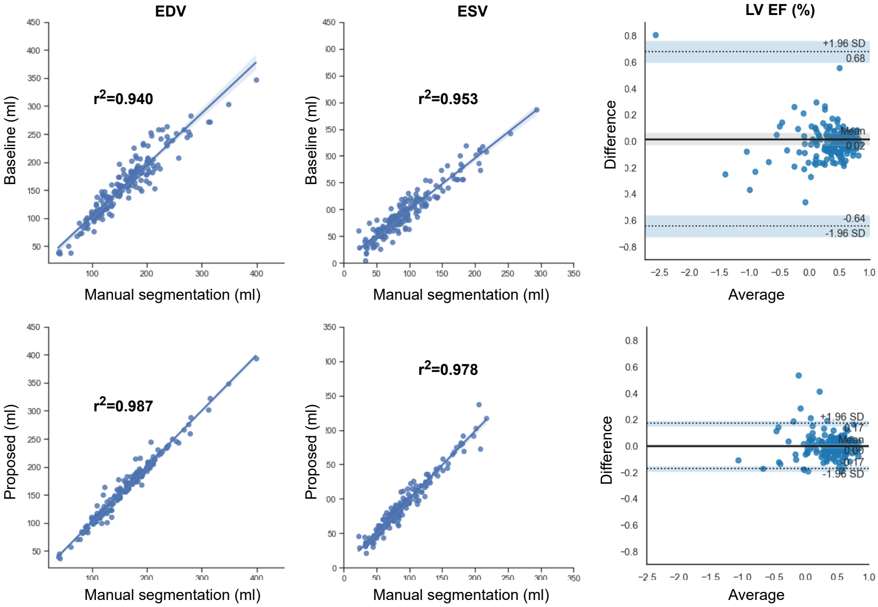


Figure 6.9: Correlation and Bland-Altman (rightmost) plots of right ventricular (RV) functional parameters generated from manual and automatically predicted segmentation masks using the baseline (first row) and proposed (second row) models.

impact on calculated ED and ES volumes, as well as the ejection fraction.

6.4.3 Analysis Per Pathology

To gain additional insight into the performance of the segmentation methods analyzed in this study, we stratify the quantitative analysis per pathology, as shown in Figure 6.11. Figure 6.11 depicts Dice scores achieved by the baseline and proposed methods, extracted per tissue across SA images.

Overall, we note consistent improvements in segmentation performance across all tissue, with more prominent gains in the case of myocardium (MYO) and right-ventricular (RV) blood-pool segmentation. In fact, statistically significant improvements in MYO segmentation, according to the paired Wilcoxon signed-rank test ($p < 0.01$), are obtained for SA cases undergoing defects related to arrhythmogenic cardiomyopathy (ARR), tetrology of fallot (FALL), dilated left ventricle (LVD), dilated right ventricle (RVD), tricuspidal regurgitation (TRI), as well as on healthy patients (NOR). Likewise, statistically significant increase in Dice scores for RV segmentation are observed for patients suffering from inter-atrial

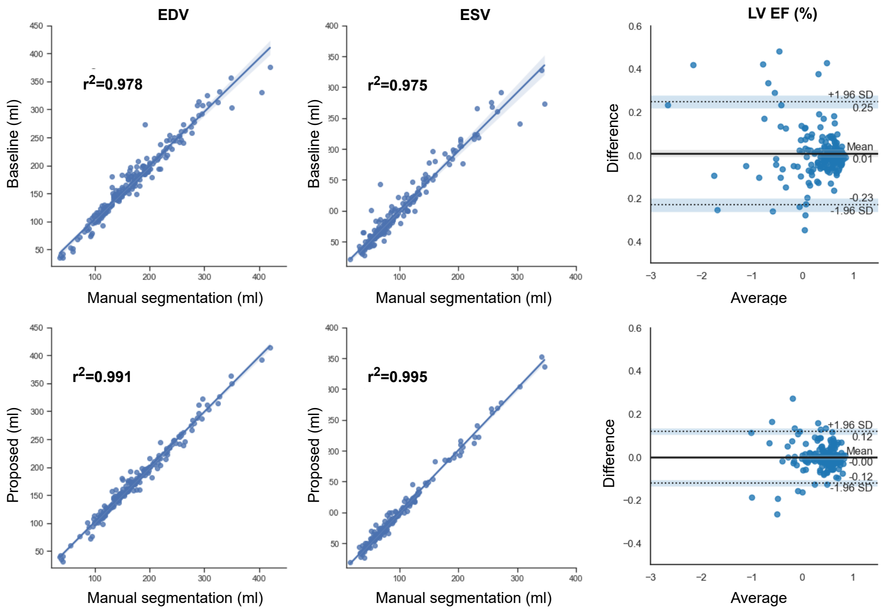


Figure 6.10: Correlation and Bland-Altman (rightmost) plots of left ventricular (LV) functional parameters generated from manual and automatically predicted segmentation masks using the baseline (first row) and proposed (second row) models.

communication (CIA), tetrology of fallot (FALL), dilated left ventricle (LVD) and dilated right ventricle (RVD). However, segmentation over LV shows only slight improvements, mostly related to outlier reduction, with statistically signifant differences observed for patients with defects in tetrology of fallot (FALL), dilated left ventricle (LVD), as well as dilated right ventricle (RVD).

The scores obtained on RVD and TRI cases are of a particular interest, as these are completely unseen during training, suggesting that the proposed method has the ability to compensate for unseen diseases. The improvement in segmentation of patients undergoing RVD and TRI further leads to enhanced derivation of clinical parameters, as seen in Figure C.2 and Figure C.3 for both the LV and RV, respectively (Supplementary Material 6.6). Similar trends are observed in LA images (see Figure C.1) across RV And MYO segmentation, with more moderate improvement across the LV.

Figure 6.12 shows some challenging segmentation cases across different pathologies present in the test set for both SA images and their LA counterparts.

Patients undergoing arrhythmogenic cardiomyopathy often exhibit right ventricular dilatation and scarring in the myocardial area. This is commonly reflected

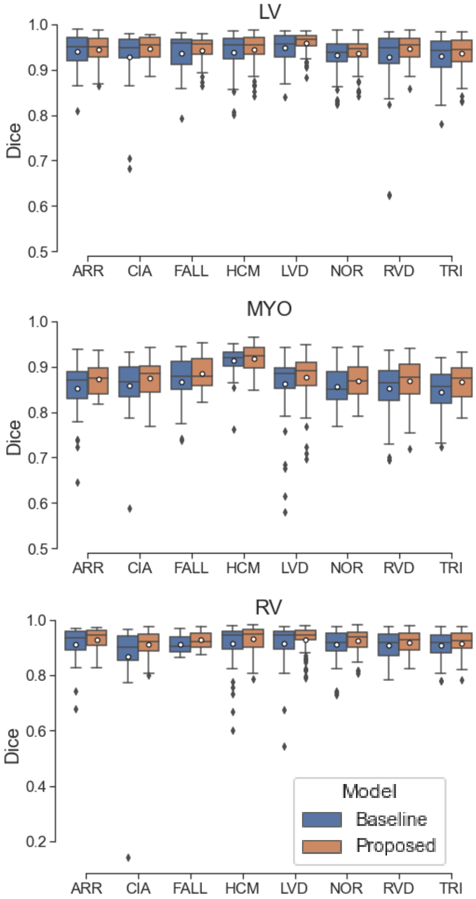


Figure 6.11: Average segmentation performance of the baseline and proposed models across SA test images, derived per disease, in terms of Dice score per tissue.

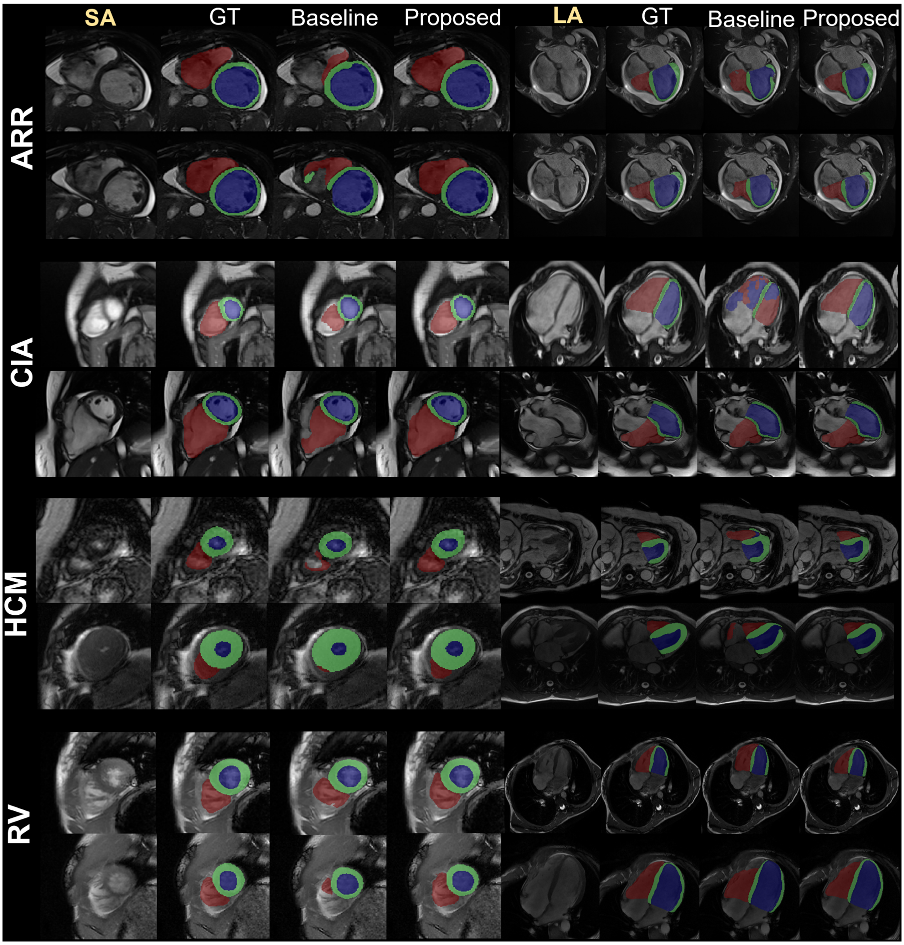


Figure 6.12: Qualitative visualization of segmentation results in challenging cases undergoing different cardiac disease and pathology, outlining the improvement i segmentation when using the proposed pipeline compared to the baseline. Each row presents a single patient, where we showcase one SA slice, as well as the LA view of the heart corresponding to the same patient. Model predictions are compared to the ground truth, shown in the column marked as GT.

by difficulties in segmenting both the RV and MYO, as shown in Figure 6.12, which can be tackled by utilizing the proposed pipeline. Similar can be observed for patients suffering from interatrial communication defects (CIA), where RV dilatation is typical. Hypertrophic cardiomyopaty (HCM) is often found in the middle septum at the midventricular level, as well in the inferior region at the apical region [156], as shown in Figure 6.12. In these cases, the baseline model often under-segments the RV, but also struggles with over-segmenting the myocardium. Finally, a dilated right ventricle presents another typical case of under-segmentation, especially in slices towards the base and apex of the heart. However, the proposed pipeline shows noticeable improvement in handling such examples.

6.4.4 Evaluation on External Datasets

To demonstrate the robustness of the proposed pipeline, we perform an additional evaluation on a completely different set of out-of-domain CMR images. These include data acquired from ACDC [39] and M&Ms-1 [46] challenges, which we use to directly test both the baseline and the proposed models and report the results in terms of Dice and Hausdorff distance scores. We do this without additionally re-training or adapting the models to new data.

A description of both datasets is available in Supplementary Material 6.6. Since only the training data from the ACDC Challenge dataset is available publicly, consisting of ED and ES images from 100 subjects, we utilize this as our test set. However, the evaluation on the M&Ms-1 data is done on the actual test data provided by the challenge organizers, consisting of 80 ED and ES subjects (a total of 160 images) from four different vendors. It is important to note that the evaluation of this experiment is only performed for SA images, since both challenges do not contain LA data.

While there is an evident domain shift between the ACDC and M&Ms-1 data compared to the M&Ms-2 data used for the training of the complete pipeline, we still observe a significant improvement when utilizing the proposed multi-modal approach and augmenting the training set with synthetic images, as seen in Table 6.3. A large portion of the ACDC dataset contains pathological cases, where we observe significant improvements in performance, particularly when segmenting RV and MYO. However, we hypothesize that additional improvements in performance could be achieved if the training was adapted to those datasets specifically, especially the synthesis module, as the current approach is specifically tailored to M&Ms-2 data.

6.4.5 Ablation Study

We perform an ablation study to understand the value of different pipeline components on segmentation performance. Therefore, we train the following models: (i) **U-Net + BB**, a regular single encoder nnU-Net with Bounding Box (BB)

Table 6.3: Segmentation performance comparison between the baseline and the proposed model in this work, evaluated on short-axis (SA) images acquired from the M&Ms-1 [46] and ACDC [39] challenges (Supplementary Material 6.6). The evaluation is performed over all three cardiac tissues, in terms of Dice and Hausdorff Distance (HD) scores. Values indicated in bold are those which are significantly higher compared to the baseline performance, according to the Wilcoxon signed-rank test for p<0.01.

	Model	**Dice**			**HD**		
		LV	MYO	RV	LV	MYO	RV
M&Ms-1	Baseline	0.908 (0.05)	0.799 (0.05)	0.873 (0.07)	12.04 (8.6)	16.04 (9.1)	13.77 (8.1)
(n=160)	Proposed	**0.925 (0.03)**	**0.821 (0.04)**	**0.901 (0.04)**	**7.81 (3.8)**	**11.34 (5.6)**	11.84 (6.2)
ACDC	Baseline	0.955 (0.03)	0.868 (0.03)	0.922 (0.05)	7.69 (5.2)	9.51 (7.5)	16.49 (15.9)
(n=200)	Proposed	0.962 (0.01)	**0.891 (0.02)**	**0.934 (0.03)**	**5.62 (3.3)**	**7.61 (5.2)**	**11.45 (4.9)**

detection, corresponding to the baseline model; (ii) **U-Net + BB+ IT**, a model similar to (i) but augmented with the same set of intensity transformations (IT) as in the late-fusion model; (iii) **U-Net + BB + IT + Synth**, a model similar to (ii) with added synthetic data (Synth) at training time, generated as described in Section 6.2.4; (iv) **LF-U-Net + BB**, a dense late fusion (LF) approach combined with bounding box detection and (v) **LF-U-Net + BB + Synth**, a dense late fusion approach proposed in this chapter. All models are trained using the procedure in Section 6.2.6.

The obtained results across the entire M&Ms-2 test data, for both SA and LA images, are outlined in Table 6.4. We start observing significant improvements in performance with the addition of synthetic images, generally related to patients with dilated right and left ventricles, hypertrophic cardiomyopathy and arrhythmogenic cardiomyopathy. However, we do not observe any improvement in segmentation among healthy patients, which we hypothesize is due to the fact we focus the augmentation process on diseased patients and abnormal heart shapes. On the other hand, introducing a late-fusion approach, combined with hyper-dense connections, demonstrates some performance improvement in those cases. In LA images, the addition of synthetic images has a significant effect on right ventricle segmentation, where visual observation suggests improvements on patients with severe changes in RV shape due to underlying pathologies.

The late fusion model used in this study leads to more refined segmentations of the LV and MYO in SA and LA images, respectively with consistent improvements in images with visible artifacts, as well as in cases with low contrast between tissues. Adding synthetic data to the late-fusion model (**LF-U-Net + BB + Synth**) yields further improvements, mostly around the RV area, as well as the myocardium. Augmentation with synthetic data tends to reduce the amount of variation between the predictions, leading to better reliability and stability of segmentations. This is particularly manifested when evaluating the trained models across patients with pathologies unseen during training, such as the tricuspidal regurgitation (TRI) and the dilated right ventricle. Similar results can be observed across LA images, where the proposed model tackles both under- and over-segmentation across all tissues, noticeable in single-encoder models. This is particularly noticeable when observing the delinations over the LV and MYO. Largest improvements in performance are obtained across the patients suffering from dilated RV and TRI - the unseen cases during training, suggesting that both synthetic data and better modeling of relationships between differently transformed images aid with tackling the changes in both heart shape and appearance due to the presence of pathological tissue.

Table 6.4: Segmentation performance comparison between the baseline and the proposed model in this work, as well as the models trained with different elements of the proposed pipeline, according to the ablation experiment described in Section 6.4.5. Each model is evaluated on original short-axis (SA) and long-axis (LA) test images, across all cardiac tissues. Numbers listed in the table are the means and standard deviations of Dice score. Dice values indicated in bold are those which are significantly higher compared to the baseline performance, according to the Wilcoxon signed-rank test for p<0.01.

Method	**Short-Axis (SA)**			**Long-Axis (LA)**		
	LV	MYO	RV	LV	MYO	RV
U-Net + BB (Baseline)	0.941 (0.05)	0.881 (0.06)	0.923 (0.07)	0.947 (0.07)	0.871 (0.08)	0.902 (0.08)
U-Net + BB + IT	0.945 (0.04)	0.886 (0.05)	0.925 (0.05)	0.949 (0.07)	0.874 (0.06)	0.907 (0.07)
U-Net + BB + IT + Syn	0.950 (0.04)	**0.891 (0.06)**	**0.931 (0.04)**	0.951 (0.06)	0.879 (0.05)	**0.911 (0.08)**
LF U-Net + BB	**0.953 (0.03)**	**0.898 (0.05)**	**0.934 (0.05)**	0.954 (0.05)	**0.885 (0.04)**	**0.919 (0.06)**
LF U-Net + BB + Syn (Proposed)	**0.959 (0.02)**	**0.907 (0.04)**	**0.938 (0.03)**	**0.958 (0.03)**	**0.901 (0.03)**	**0.924 (0.04)**

6.5 Discussion

In this work, we propose a pipeline designed to tackle the segmentation of pathological CMR images across multiple views (SA and LA) and sources. We provide a comprehensive analysis of the proposed pipeline and compare its performance to the baseline model, widely used in the literature (nnU-Net). The obtained results demonstrate the ability of the proposed pipeline to reduce the performance gap between the outlier cases, riddled by artifacts and shape deformation caused by underlying pathologies, and cases similar to those available at training time.

While outlier cases are typical and commonly found across many medical imaging tasks, they are more prominent in pathological data, owing to limitations in representation and number of cases. This particularly affects data-hungry deep learning algorithms, known to fail on cases poorly represented during training. However, the method proposed in this work can tackle such cases more effectively, leading to a notable decrease in the number of outliers during segmentation. This in turn has a significant impact on not only the average segmentation performance, but also the derivation of clinically relevant metrics characterizing heart function. In fact, we demonstrate significant improvements in levels of agreement and bias reduction for the left- and right-ventricular ejection fraction across all cases available at inference time. Outlier reduction has shown to be consistent throughout all pathologies and cardiac tissues. Further investigation suggests that outliers occurring in this dataset belong to images with large differences in appearance and contrast compared to the majority of images available at training time, as well as those containing higher levels of noise and artifacts. However, cases with severe tissue deformation and occlusion due to the presence of pathologies represent the most challenging cases during segmentation, largely addressed by the proposed approach.

We further show that using conditional GANs holds a lot of promise for the generation of missing data and addressing data scarcity, which is particularly emphasized when dealing with pathological patients. In fact, careful generation of images varying in contrast and tailored to different cardiac diseases can significantly improve model generalization and reduce class imbalance, common in regular datasets that are skewed towards non-pathological images. We demonstrate the impact of the generated synthetic images on the distribution of heart cavity samples in the training set, which we hypothesize increases the representation of challenging cases during training and leads to a more stable segmentation performance, even in the presence of unseen diseases. The proposed synthesis approach can be adapted to any type of scarce data, such as rare pathologies, whereby only a small subset of such data is needed for training a synthesis module that can expand the training set with artificial patients of varying appearance.

Although data augmentation with more extreme intensity transformations has recently shown to positively influence the regularization and generalization of DL-based methods, as well as reduce overfitting, we show that combining such transformations using a multi-path approach aids the network with learning complementary information and fosters better data representation, enhancing the networks' discriminative power. Moreover, enhancing the flow of information between the multi-path layers through dense connections shows further benefits in obtaining a more accurate segmentation. Visual observation suggests that this improvement is particularly related to the segmentation of small structures (such as the tissue at the apex of the heart) or at region boundaries, where single encoder networks tend to struggle at differentiating between tissues.

Performance analysis across different pathologies in both SA and LA images reveals additional insights about the behavior of different models evaluated in this study. We observe that baseline models are prone to over-segmentation, particularly in the basal regions, where they falsely predict the presence of either the RV or the whole heart. This is usually caused by blood movement-related artifacts, low tissue contrast or occlusion due to specific diseased tissue. Additional difficulties appear in cases where the myocardial muscle does not completely enclose the blood pool and exhibits variability in shape, becoming non-circular. Furthermore, we note that the proposed method tends to exhibit more significant improvements in terms of the HD scores, which is primarily the result of outlier reduction. Visual observation suggests that cases exhibiting high HD scores when evaluated with a baseline model, contain false positive predictions in the areas outside of the heart, often consisting of tissues similar in appearance and shape to cardiac tissue. Additional errors contributing to segmentation inaccuracies obtained by the baseline include areas with weak or missing edges, artifacts and low signal-to-noise ratio.

In general, we note considerable improvements using the proposed method across most pathological cases, with better adaptation to unseen cases. The obtained results are comparable and even outperform those reported in the M&Ms-2 challenge, ranging from 0.83 to 0.93 and 0.8 to 0.92 in Dice score for RV segmentation across SA and LA images, respectively[2] [157–170]. Moreover, augmenting the training set with highly diverse data and introducing a more efficient way to extract meaningful features from data leads to improved performance on out-of-domain datasets (Section 6.4.4). This shows that despite training the model on a completely different set of images, the proposed modules aid in adaptation to the existing domain shift that commonly occurs between different datasets. Finally, we demonstrate the effects on performance when one or more modules are removed from the proposed pipeline and identify the largest sources of improvement in the ablation study (see Section 6.4.5). We demonstrate that each element of the proposed pipeline adds value to the overall segmentation improvement, but significant differences start appearing when utilizing augmentation with synthetic data, as well as the late fusion approach with dense connections.

[2] Evaluation over LV and MYO is not included in the evaluation procedure of the M&Ms-2 challenge and is not reported by any participants.

6.5.1 Limitations and Future Work

Despite the reported improvement in segmentation performance and outlier reduction, the proposed model still has several limitations. The performance drop on ES slices remains higher compared to the ED slices, which further affects the calculation of clinical parameters, such as the ejection fraction. Moreover, basal regions are prone to under-segmentation, followed by a drop in accuracy around the apex of the heart, mainly due to its small size compared to the rest of the cavity. While we manage to partly handle some of these issues, they consistently remain the biggest sources of errors, which is in agreement with findings reported by similar work in the literature on other datasets. This implies that special attention should be placed on addressing these regions, which we plan to focus on in future work. Additionally, the provided LA images could help with extracting the inter- and intra-view information from the complementary SA and LA images, allowing for both the localization of the basal plane and possibly better segmentation of the basal slices. Thus, integration of the proposed modules into a truly multi-view approach would be one of the main aspects to focus on in our future work.

Furthermore, while we extensively analyze the impact of the proposed pipeline on a wide array of pathological data with varying sources of acquisition, we would further benefit from assessing its confidence and identifying possible prediction uncertainties under difficult settings. In the same line, extending this study on other open-source datasets, as well as to clinical settings, would allow us to further identify the necessary points of improvement, particularly when performing the evaluation on other unseen cardiac pathologies. Although we focus this work on handling the variation in cardiac tissue shape and appearance in the presence of various diseases, we note that the segmentation performance on healthy patients does not show significant improvements. Thus, ensuring that the model generalizes well to both healthy and diseased tissue is another major focus of our future experiments.

To balance out the number of pathological and normal cases from the M&Ms-2 challenge, we identify outlier subjects by calculating the right-ventricle volume using the ground truth labels and taking into account the mean and standard deviation values. However, this method for outlier detection may not necessarily cover all pathological cases available in the dataset, as just the volume may not be indicative of a cardiac disease. Instead, having access to labels for each pathology, one can synthesize more subjects for a particular disease in such a way to obtain a balanced number of different diseases present in the data.

6.6 Conclusions

In this work, we propose a pipeline including three distinct modules to handle different challenges of multi-vendor and multi-disease cardiac MR images for the task of increasing the segmentation robustness and outlier reduction. We demonstrate the ability of the proposed approach to balance the segmentation of outlier cases, typically related to increased levels of artifacts and shape deformations induced by the presence of pathologies, with those more commonly represented in the training set. Synthesizing a diverse training dataset, carefully designed to increase the variation of cardiac shapes and appearances during training, plays a significant role in not only boosting the model performance in terms of standard quantitative metrics, but also in improving the automatically derived clinical metrics denoting the function of the heart. This in turn leads to improved stability and reliability of the predictions across both short-axis and long-axis images. Such observations are additionally confirmed on completely unseen images, extracted from other publicly available datasets, whereby we observe both outlier reduction and better adaptation to the presence of the domain shift between datasets. Future work includes more precise synthesis of pathological cases, conditioned on the pathology type, as well as utilizing the availability of LA images to inform the positioning of the basal plane for more accurate segmentation.

Supplementary Material

Heart Region Detection

As presented in Section 6.8, the first stage of our pipeline is a heart region detection module, consisting of a regression-based neural network that locates and extracts the heart in both SA and LA images. Please note that this model is mainly used at inference (test) time when ground truth labels are not available and need to be predicted. We use a simple CNN designed for a regression task, where the output consists of 6 continuous values representing the bounding box surrounding the heart. The inputs to the network are pre-processed 2D (256×256) mid-cavity SA slices extracted from the training set and all LA slices, respectively. The proposed CNN consists of five convolutional layers, followed by two fully-connected layers with a linear activation. Each convolutional layer uses 3×3 kernels, followed by a 2×2 max-pooling layer. Batch normalization and leaky ReLU activations are used in each layer, except for the output. Dropout with the probability of 0.5 is used in the fully connected layers. The network is trained for 2000 epochs with a batch size of 32 and early stopping (assessed from the validation accuracy), by minimizing the mean squared error between the computed and the actual transformation (estimated from the ground-truth) using the Adam optimizer. We start with an initial learning rate of 0.001, decreased by a factor of 0.5 every 250 epochs. All image dimensions and scaling/displacement parameters are normalized to generate translations in the range from -1 to 1.

After prediction, all the parameters are de-normalized to reflect the original image scale. On-the-fly data augmentation is applied to the training images, consisting of random translation, rotation, scaling, vertical and horizontal flips, contrast augmentation and addition of noise. At inference time, we again use mid-cavity slices from the SA test images to obtain the adjustment parameters of the ROI (not needed for LA). The predicted bounding boxes on mid-cavity slices of SA images are then propagated through the whole 3D volume, from which these slices were extracted. This procedure is not applied for LA images, where direct detection is possible (both ED and ES LA images consist of a single slice only). The obtained cropped SA and LA images using the predicted bounding box are post-processed to be of the size 128×128 voxels and 176×176 voxels, respectively.

External, Out-of-Domain Data Used for Evaluation

To demonstrate the impact of the proposed pipeline on a completely different set of out-of-domain CMR images, we additionally evaluate the method on data acquired from ACDC [39][3] and M&Ms-1 challenges [46]. A detailed evaluation and comparison to the baseline model is provided in Section 6.4.4, while the description of both testing sets is provided below. Please note that the evaluation is performed only on SA images, as both of the out-of-domain data-sets do not contain any LA images.

1. *Automated Cardiac Diagnosis Challenge (ACDC):*

   The ACDC data-set includes end-systolic (ES) and end-diastolic (ED) images acquired from 150 patients, containing both normal and pathological subjects. Images are acquired by two scanners with different magnetic field strengths and contain expert annotations for left ventricle (LV), right ventricle (RV) blood pool, as well as left ventricular myocardium (MYO). Out of the 150 subjects (300 ED and ES MR images), 100 are reserved as a training set, while the remaining are used for testing. The 100 training subjects are available to download with their respective annotations, while the annotations belonging to the testing data are not available to public. For this reason, we select the entire training set, consisting of 100 subjects (200 ED and ER images) for testing in this experiment.

2. *Multi-center, multi-vendor and multi-disease cardiac image segmentation challenge (M&Ms-1) data:*

   The M&Ms-1[4] challenge data-set consists of 350 images from a mix of healthy controls and patients with hyptertrophic and dilated cardiomyopathies. All patients were scanned in clinical centers across three different countries (Spain, Germany and Canada) using four different MRI scanner vendors (Siemens, Philips, General Electric-GE and Canon). The provided training set contains 150 annotated patient scans from two different scanner vendors (Philips and Siemens, 75 each) and 25 un-annotated scans from a third vendor (GE). The in-plane resolution of the training images varies between 1.18 to 1.72 mm, with slice thickness ranging between 9.2 to 10.0 mm. Annotations have been provided by experienced clinicians at both end-diastolic and end-systolic phases, including contours for the left (LV) and right ventricle (RV) blood pools, as well as the left ventricular myocardium

[3] ACDC data-set can be accessed at `https://www.creatis.insa-lyon.fr/Challenge/acdc/`

[4] M&Ms-1 data can be acquired at `https://www.ub.edu/mnms/`

(MYO). This amounts to 300 annotated and 50 un-annotated images, taking both phases into consideration.

For testing we utilize the additional images provided by the M&Ms challenge as a separate test-set. These consist of an additional 50 studies from each of the vendors provided, as well as another 50 studies from a vendor unseen during training (Canon), with in-plane resolution ranging from 0.68 to 1.8 mm. We refer to different scanner vendors contained in the test set, namely Philips, Siemens, GE and Canon scanners, as vendors A, B, C and D, respectively.

Quantitative Analysis and Derived Clinical Parameters

To demonstrate the impact of the proposed pipeline on the cardiac tissue segmentation, we evaluate its performance compared to the baseline model across different pathological cases available in the data-set. The results related to LA images are shown in Fig. 6.13. In addition, we showcase the impact of outlier reduction on automatically derived clinical parameters with reference to those derived from the manually outlined ground truth, across two unseen types of pathologies, which are not available during training. Figures 6.14 and 6.15 visualize the correlation and Bland-Altman analysis of left ventricular (LV) and right ventricular functional parameters, respectively, which include the end-diastolic (EDV) and end-systolic (ESV) volumes, as well as the ejection fraction (EF) derivation across patients with dilated right ventricle (RVD) and tricuspidal regurgitation (TRI).

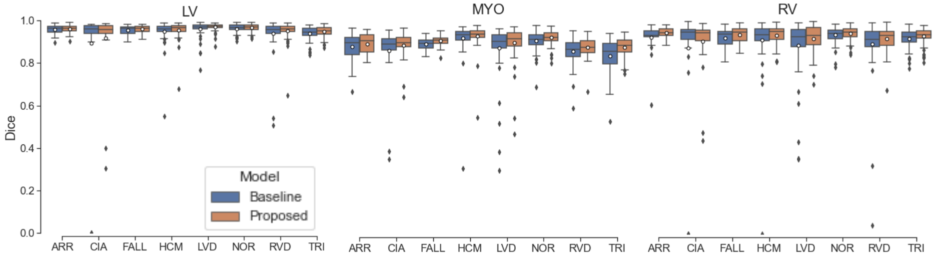
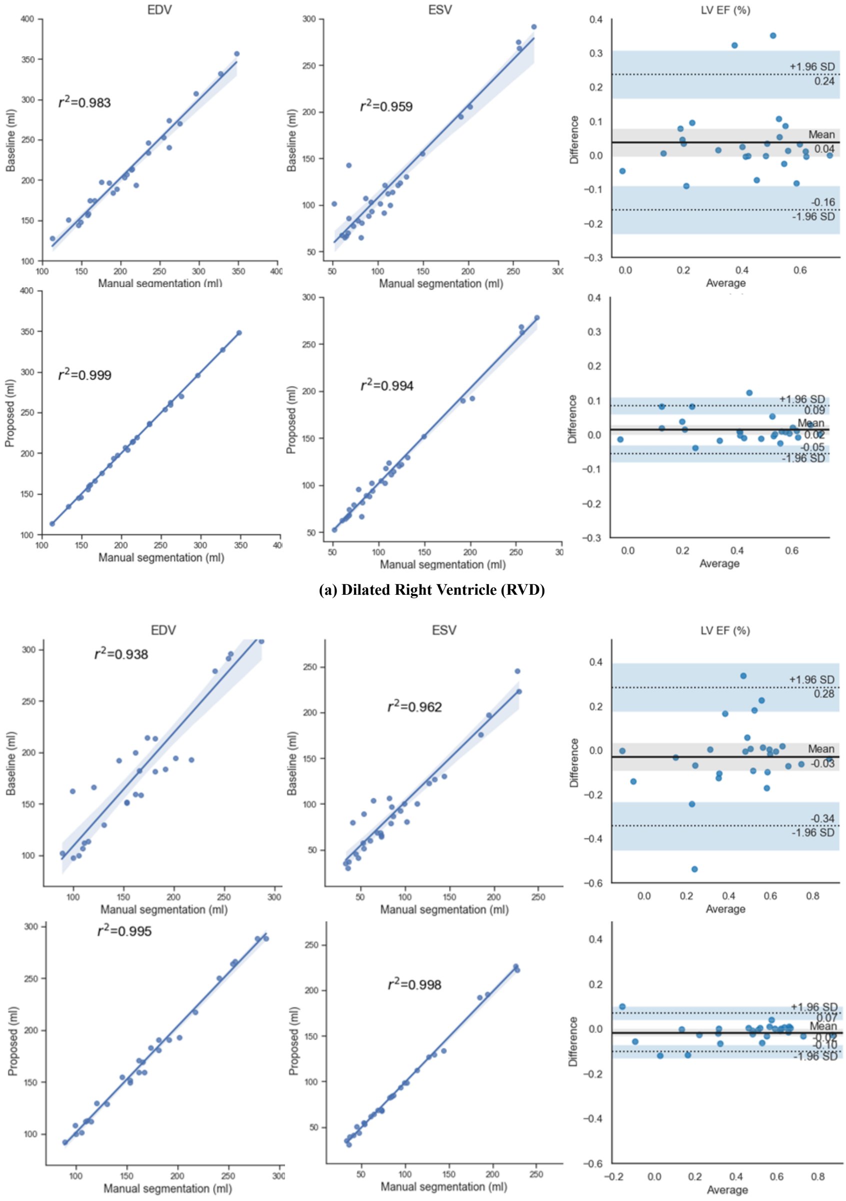


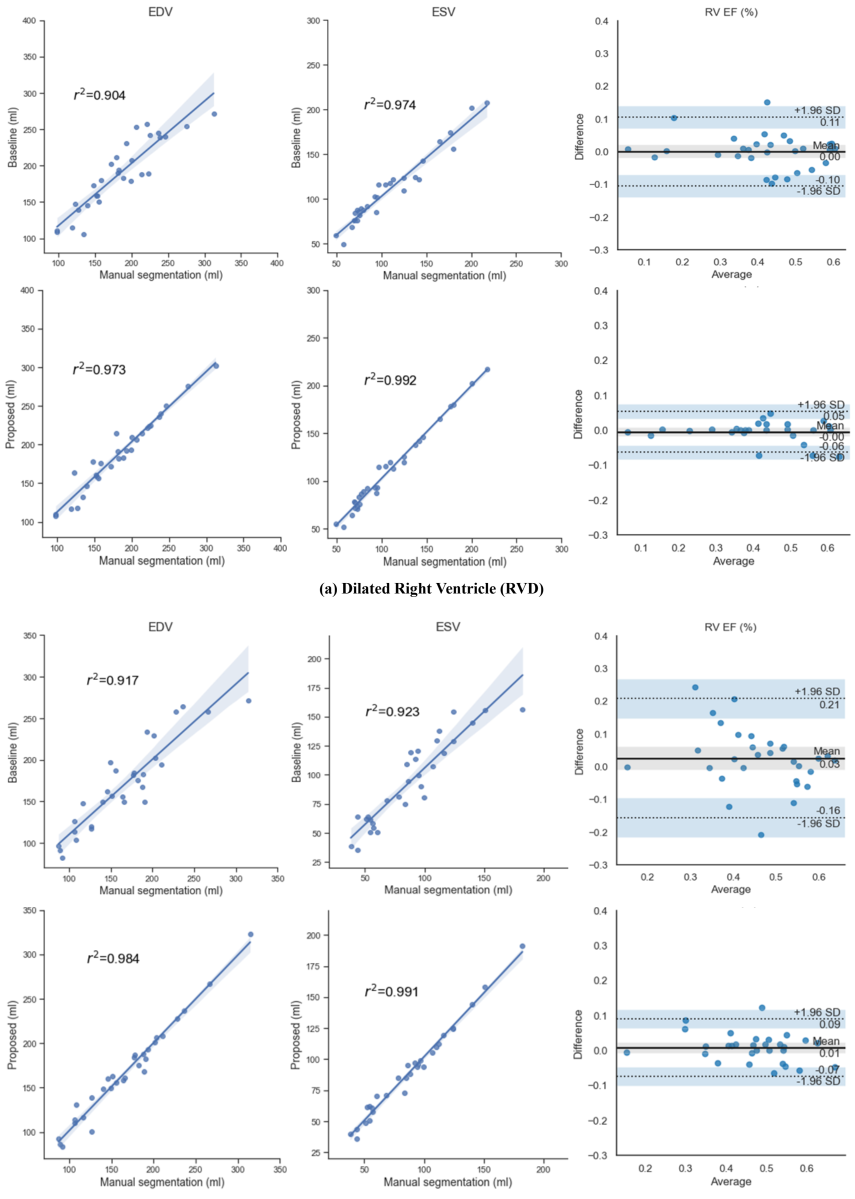
Figure 6.13: Average segmentation performance of the baseline and proposed models across LA test images, derived per disease, in terms of Dice score per tissue.



(a) Dilated Right Ventricle (RVD)

(b) Tricuspidal Regurgitation (TRI)

Figure 6.14: Correlation and Bland-Altman plots of left ventricular (LV) functional parameters generated from manual and automatically predicted segmentation masks using the baseline and proposed models. LV end-diastolic volume (EDV), LV end-systolic volume (ESV) and the LV ejection fraction (LV EF) are derived for patients undergoing (a) dilated right ventricle (RVD) and (b) tricuspidal regurgitation (TRI).



(a) Dilated Right Ventricle (RVD)

(b) Tricuspidal Regurgitation (TRI)

Figure 6.15: Correlation and Bland-Altman plots of rigt ventricular (RV) functional parameters generated from manual and automatically predicted segmentation masks using the baseline and proposed models. RV end-diastolic volume (EDV), RV end-systolic volume (ESV) and the RV ejection fraction (RV EF) are derived for patients undergoing (a) dilated right ventricle (RVD) and (b) tricuspidal regurgitation (TRI).

CHAPTER 7

Simulation to Real Translation

This chapter is based on:
S. Amirrajab, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "sim2real: Cardiac MR Image Simulation-to-Real Translation via Unsupervised GANs," ISMRM 2021, in preparation for a journal publication.

Abstract

There has been considerable interest in the MR physics-based simulation of a database of virtual cardiac MR images for the development of deep-learning analysis networks. However, the employment of such a database is limited or shows suboptimal performance due to the realism gap, missing textures, and the simplified appearance of simulated images. In this work we 1) provide image simulation on virtual XCAT subjects with varying anatomies, and 2) propose sim2real translation network to improve image realism. Our usability experiments suggest that sim2real data exhibits a good potential to augment training data and boost the performance of a segmentation algorithm.

7.1 Introduction

A cohort of virtual cardiac magnetic resonance images (CMR) can be simulated to aid the development and adaptation of data-hungry deep-learning (DL) based medical image analysis methods. Recent studies have shown the effectiveness of image simulation in the context of training a DL model for CMR image segmentation [21, 171]. Although such models provide accurate anatomical information, their performance is still suboptimal as a result of the realism gap, missing texture and simplistic appearance of the simulated images. This holds especially for models trained completely on simulated images and evaluated on real ones. Generative adversarial networks (GANs) [2], on the other hand, promise to synthesize realistic examples, as demonstrated by applications for multi-modal medical image translation [13,172,173]. However, GAN-generated images may not necessarily represent plausible anatomy.

The purpose of the current research is to reconcile the two worlds of simulation and synthesis, as defined in [1], and take advantage of recent developments in the field of computer vision to reduce the realism gap between simulated and real data using GANs for unpaired/unsupervised style transfer. The contributions are two-fold: 1) Physics-based simulation of cardiac MR images on a population of XCAT subjects 2) GANs-based image-to-image translation for style (texture) transfer from real images. The framework is named sim2real translation.

7.2 Material and Method

The 4D XCAT phantom [7] is utilized as the basis of the anatomical model for creating virtual subjects by carefully adjusting available parameters for altering the geometry of the human anatomy. We employ our in-house CMR image simulation framework based on the analytical Bloch equations to generate varying image contrast on the labels of the XCAT virtual subjects [21].

An unsupervised GAN model based on contrastive learning, known as CUT [174], is used for the task of unpaired translation between the real and the simulated images to transfer the realistic style (texture) from real images to simulated ones while preserving the anatomical information (content). Contrastive learning encourages encoded features of two patches from the same location in the real and translated images to be similar yet different to other patches. Compared to other unpaired translation frameworks such as CycleGAN [65], CUT is a one-sided network with a much lighter generator architecture hence requiring few data for training. The content of the simulated image is preserved through a multilayer patch-wise contrastive loss added to the adversarial loss, as shown in Figure 7.1.

The M&Ms challenge data [46] are used as the source of real cardiac MR images. To explores the effects of multi-vendor data, we utilize a subset of the data consisting of 150 subjects with a mix of healthy controls and patients with a

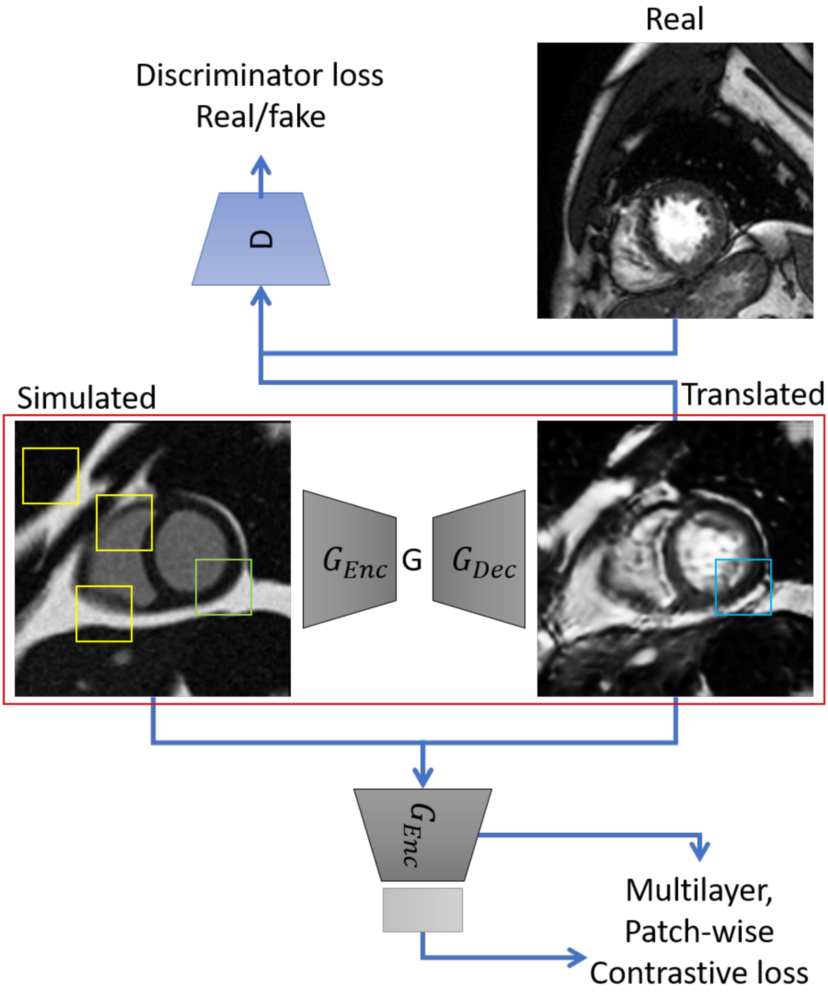


Figure 7.1: simulation-to-real (sim2real) translation using contrastive learning for unpaired translation model proposed in [14]. The style translation in achieved by the discriminator loss, while the content of anatomical information in the simulated data is preserved by the multilayer, patch-wise contrastive loss

variety of heart conditions scanned using Siemens (Vendor A) and Philips (Vendor B). We extract four mid-ventricular slices at end diastolic (ED) and end systolic (ES) phases for each subject. All subjects are resized, centre cropped to the size of 128 x 126, and normalized to the range of [0, 1]. The same pre-processing is applied on the simulated images, despite the fact we use the available ground truth labels of the simulated data to find a bounding box around the heart and crop accordingly instead of centre cropping.

Two identical sim2real models are trained using the data from vendor A and vendor B (sim2real A, and sim2real B) to investigate the network's ability to transfer vendor-specific appearance images on simulated ones. We calculate the widely-used Fréchet Inception Distance (FID) score [175] between feature vectors calculated for real and translated images to evaluate the similarity between the simulated database and its respective real data, before and after translation.

Additionally, we evaluate the usefulness of our sim2real data in aiding a DL segmentation model for the task of cardiac segmentation. We utilize a nnUNet [40], trained to segment the left ventricle (LV), right ventricle (RV), and the left ventricular myocardium (MYO). First, we train a model using 150 sim2real images with the style of vendors A and B and compare it to a model trained on 150 real images. We additionally train a model with a mixed set of real and sim2real data to observe the applicability of generated data for data augmentation.

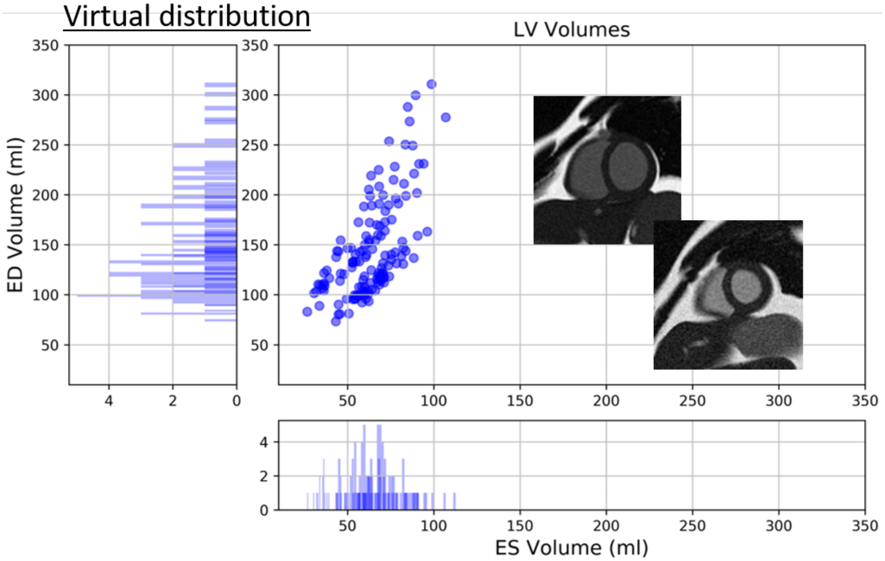


Figure 7.2: End-diastolic and end-systolic volume of left ventricle for simulated virtual subjects and two examples of cardiac MR image simulation.

7.3 Results

Two examples of simulated images and statistics of the XCAT virtual subjects' distribution in terms of left ventricular volumes are depicted in Figure 7.2.

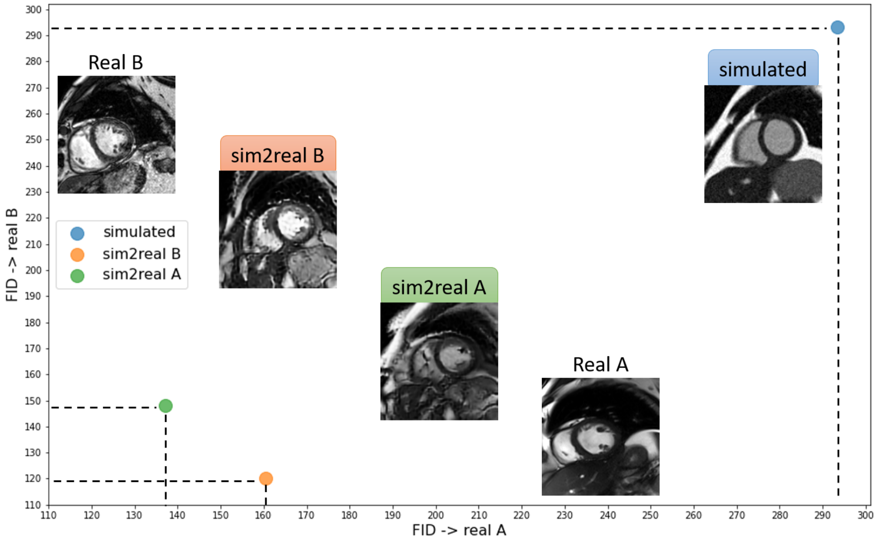


Figure 7.3: Fréchet inception distance (FID) score, which lower value means more similarity, between simulated and real data from vendor A (FID-> real A), simulated and real data from vendor B (FID-> real B), style transfer from vendor A on simulated denoted as sim2real A, and style transfer from vendor B on simulated denoted as sim2real B. One real example for each domain is also shown.

The FID score is computed between the simulated data, sim2real A data, sim2real B data and the data from vendor A and vendor B. The lower value for the FID score suggests more realistically generated images and thus higher similar feature statistics to real database. The results are shown in Figure 7.3. As expected, the original simulated data has a high FID score on both real A and real B data. Generally, the sim2real model substantially reduces the FID between the simulated data and real images, indicating improvement in image realism. Moreover, the vendor-specific imaging features are captured by the network and transferred to the simulated images. One real example from each vendor and each sim2real translation is shown for visual comparison.

The segmentation performance of three different models can be observed in Table 7.1, presenting the evaluation of all models on a separate test set from the M&Ms challenge. The results suggest that the model trained with sim2real images already adapts well to real data, exhibiting a slight drop in performance compared to the model trained with real data. Additionally, we observe that augmenting the training with sim2real data has a positive impact on segmentation accuracy (Dice score), particularly for the LV.

Table 7.1: Segmentation performance of 2D nnUNet models trained with only simulated data (row 1), only real data (row 2) and a mix of real and simulated data (row 3), where N indicates the total number of images used for training. All models are evaluated on the unseen test set from vendors A and B in terms of the average Dice score and Hausdorff Distance (HD) per three cardiac tissues. Results outlined in red indicate the best performing model per tissue.

Training		**Testing**											
		Vendor A						**Vendor B**					
Real	**Simulated**	**LV**		**RV**		**MYO**		**LV**		**RV**		**MYO**	
		Dice	HD	Dice	HD	Dice	HD	Dice	HD	Dice	HD	Dice	HD
N/A	N=160	0.887	9.25	0.851	12.45	0.801	14.72	0.871	10.38	0.861	11.21	0.831	12.11
N=160	N/A	0.901	8.19	0.878	9.35	0.863	9.88	0.893	9.31	0.872	10.67	0.849	9.76
N=160	N=160	0.915	7.85	0.882	10.21	0.872	12.32	0.911	7.28	0.874	10.85	0.851	10.21

7.4 Discussion and Conclusion

In this work, we created a database of virtual cardiac MR images simulated on the XCAT anatomical phantom and investigated the effectiveness of an unsupervised GAN for the task of simulation-to-real translation, named sim2real. We attempted to reduce the realism gap between the simplified image simulation and complex realistic image textures. Our sim2real model could learn the vendor-specific imaging features and map them onto the simulated images, resulting in reduction of the FID scores which can be attributed to more similarity between the simulated and real databases. Our usability experiments suggested that sim2real data exhibits a good potential to augment real training data, particularly in scenarios where data is scarce.

CHAPTER 8

Discussion and Future Research

There is an ever growing need for high-quality medical data with annotations to adopt data-hungry state-of-the-art supervised deep neural networks for automated medical image analysis. Collecting, annotating, and sharing medical data with desired quality and variation may not always be feasible yet essential for development of such methods. To facilitate the development of such methods, the main theme of this research was to investigate and propose different solutions for the generation of large quantities of artificial images with ground truth labels.

In this thesis, we developed several frameworks and pipelines for cardiac magnetic resonance image simulation and synthesis for the purpose of generating substantial numbers of images with corresponding labels that can be utilized for developing automated image segmentation methods based on deep neural networks.

In the physics-driven image simulation category, a flexible image simulation framework grounded on the physics of magnetic resonance imaging was developed in Chapter 2 for simulating realistic images with variable anatomical and imaging characteristics. We built upon and advanced previous models for human anatomies and image simulation. The simulation framework included three main modules for 1) creating anatomical models based on the XCAT phantom, 2) assigning tissue properties to more that 10 tissue types and organs, and 3) optimizing simulated sequences to generate images with varying contrast. The benefit of the simulated database was shown for training a deep neural network for cardiac image segmentation, suggesting substantial performance increase when the simulated data is added to the real data, as well as retaining the baseline performance when only 45% of the real data is used and the rest is compensated with the simulated images.

While the framework was designed to generate cine cardiac images, we demonstrated that with a small modification of the sequence module in the framework, late-gadolinium enhanced (LGE) imaging could be performed for the application of simulating a subject with myocardial scars with different levels of respiratory motion artifacts at multiple resolutions. As discussed in reference [50], we investigated the effects of slice misalignment artifacts on image-based electrophysiological (EP) modeling of the heart with a defined myocardial scar. The XCAT anatomical model was extended with a customized cylindrical shaped scar placed around the right coronary artery inside the myocardium of the left ventricle. The simulation pipeline was extended to perform LGE CMR experiments to generated images with and without respiratory motion artifacts to estimate the effects of artifacts on the accuracy of EP outcome.

In the data-driven image synthesis category, a framework was developed in Chapter 3, grounded on recent conditional generative adversarial networks for label-to-image translation for generating high-quality and diverse cine cardiac magnetic resonance images. The synthesis framework consisted of 1) a multi-tissue segmentation module trained using the pairs of simulated images with corresponding labels generated using the framework described in Chapter 2, and

2) a conditional synthesis module to learn the translation from multi-tissue labels to realistic images, where the anatomical content is preserved through conditional normalization layers. We evaluated different design choices of the framework such as the benefit of using multi-tissue labels and its positive effects on the quality of the synthetic images compared to utilizing cavity-tissue labels. Our experiments on the usability of the synthetic data suggested that although there is a drop in performance when the real data is replaced with synthetic counterparts, we observed a substantial increase in model performance when we augment the real data with simulated ones, achieving a maximum increase of 4% for Dice score (higher better) and a maximum reduction of 40% for Hausdorff Distance score (lower better).

Label deformation was found to be an important aspect of label-conditional image synthesis to introduce anatomical variations in the synthetic data. We applied image-based random elastic deformations to change the heart geometry during synthesis for reference [94]. However, such deformations resulted in generating images with anatomically implausible shapes, which decreased the realism of the images, despite being useful for training. For synthesizing images with plausible heart shapes, we developed a latent space-based manipulation method using variational auto-encoders in Chapter 4. We devised different strategies to perform interpolation in the latent space of a trained model to generate virtual subjects with a target heart pathology that affects the heart geometry. We demonstrated that data augmentation with with our approach could provide a solution to diversify and enrich an available database of cardiac MR images, resulting in significant improvements in model performance and generalization for cardiac segmentation of subjects with unseen heart diseases.

To reconcile the two worlds of simulation and synthesis, in a short feasibility study described in Chapter 7, we attempted to reduce the realism gap between simulated and real data using GANs for unpaired/unsupervised style transfer and proposed a framework named sim2real translation.

Other ways to deal with limited data; The presented work throughout this thesis concerned generating artificial data to deal with medical data scarcity. There are other approaches to deal with insufficient training including transfer-learning [176, 177], domain adaptation [178] and data augmentation [179]. While transfer-learning in the form of fine-tuning a portion of pre-trained networks has shown significant improvements in the tasks involving natural images, it is limited in the medical imaging domain due to the lack of properly pre-trained models developed on large sets of medical data [180]. Domain adaptation addresses the development of models that can generalize to known target domains whose annotations are unknown or limited during training. However, the assumption is that examples from the target domain are available, which is not often the case with medical data [180, 181]. On the other hand, observed domain shift properties can be imitated by applying a variety of data augmentation approaches in the image space, which has been shown across many fields [133, 135, 182].

Simulation versus synthesis; Both simulation and synthesis have advantages and disadvantages. For instance, in image simulation we have more control over the image generation process such that we can modify different parameters at different levels to change the anatomy, tissue contrast, noise level, and imaging resolution. The results of image simulation, however, are less realistic than the outcome of synthesis, with rather simplified appearances and lack of highly-detailed textures in simulation. Synthesis results are visually more realistic, they resemble features of training data, while less control over the generation of different contrasts and less accurate ground truth labels since they depend on the expert delineations.

In the context of data augmentation, data-driven synthesis approaches can generate data with characteristics of the training images, which require no further processing to add them to the training, whereas substantial data preparation is needed on simulated images due to the gap between simulated and real data. The simplified appearance of the simulation results can be attributed to two major reasons related to the simulation pipeline and the use of the XCAT phantom. In the simulation, tissue heterogeneity and texture, system imperfections such as B0 and B1 inhomogenieties, tissue susceptibility, motion and other physiological artifacts are not included. Extending the simulation framework to generate imaging artifacts related to cardiac MRI such as respiratory and ECG mistriggering artifacts can be a direction for future research.

Despite our efforts to improve the heart anatomy of the XCAT phantom by adding trabeculae structures, the XCAT still lacks many important small details of organs. Furthermore, the ability to generate multiple subjects is somewhat limited to changing the scaling factor of the torso and slight rotation and translation of the heart within the chest. Drastic changes to the position of organs would destroy the integrity of the whole anatomy, causing organ overlap and distortion of the organ shapes. Moreover, the motion of the XCAT heart is not based on electriophysiological modeling. More advanced heart models, such as living heart project, can provide accurate information of the heart motion link to electrical and mechanical properties [183]. The proposed simulation framework would benefit from highly-detailed and accurate human anatomical models to increase the realism of the simulated images.

When simulation is preferred over synthesis; The main theme of this thesis is to generate data for training a deep-learning segmentation algorithm, where we demonstrate the benefit of both approaches. In another application, to investigate the effects of respiratory motion induced slice misalignment artifacts on image-based electrophysiological (EP) modeling of the heart with a myocardial scar, we perform accurate image simulation at multiple acquisition resolutions on a modified version of the XCAT with a defined scar geometry and location [50]. The XCAT anatomical model is extended with a customized cylindrically shaped scar placed around the right coronary artery inside the myocardium of the left ventricle. The accurate ground truth labels for different image acquisition setups are essential to validate the final outcome of the EP modelling. When the effects of sequence parameters or imaging artifacts are under investigation, the physics-based simulation is preferred over synthesis. The same holds for the application of generating high and low-resolution pairs of images for training a deep-learning based super-resolution network. In other words, we can benefit from the physics-driven simulation to provide training data for the development of data-driven synthesis approaches.

When synthesis is preferred over simulation; We explore data-driven image synthesis when we have access to some examples for training generative models and want to generate new samples to augment and increase the sample diversity. For instance, we investigated the usability of synthetic data to improve the generalization and adaptation of a segmentation network to data-sets collected from various sites and scanners [184]. Deficits in generalization to real-world data-sets with moderately different characteristics (distribution-shifts) represent one of the most common hurdles appearing due to scarce data. The adoption of deep learning methods in clinical settings is significantly hampered by these deficiencies. However, realistically generated synthetic could address these deficits, particularly when it comes to anonymization, protection of patient information and decreasing the cost of data collection. When the aim is to generate images with domain-specific features and characteristics of the training data (e.g. style of different scanner vendors), data-driven synthesis is preferred.

A major bottleneck to generating realistic and diverse cohorts of virtual subjects is the necessity to have an accurate computerized human anatomical model. We used the 4D XCAT phantoms for this research and modified the heart model to include details of myocardial trabeculations. Despite being one of the most flexible and accurate anatomical models for medical imaging research, the XCAT phantom has many limitations and shortcomings for generating diverse virtual subjects that would resemble anatomical variabilities of a cohort of real subjects. Future research could focus on investigating algorithms to combine electrophysiological models of the heart with the XCAT phantom for both simulation and synthesis approaches to generate subjects with complex dynamic heart motions for normal and regionally hampered motions. The anatomical and organ modeling, which includes taking into account details of the human organs and complex movements due to breathing and beating motions is essential for generating realistic images with plausible and accurate anatomies. Image-based modeling of the heart and adaptation of the XCAT heart shapes to real subjects would be particularly interesting for personalized patient-specific simulation and synthesis.

We believe this research would play a crucial role in the development of imaging simulation and synthesis platforms for studies on cohorts of artificially generated virtual subjects and the development, optimization, and benchmarking of a number of medical image analysis algorithms, including, but not limited to deep-neural networks for automated image segmentation.

Bibliography

[1] A. F. Frangi, S. A. Tsaftaris, and J. L. Prince, “Simulation and synthesis in medical imaging,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 673–679, 2018.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, *et al.*, “Generative adversarial networks,” *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[3] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.

[4] T. Glatard, C. Lartizien, B. Gibaud, R. F. Da Silva, *et al.*, “A virtual imaging platform for multi-modality medical image simulation,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 1, pp. 110–118, 2012.

[5] C. Tobon-Gomez, F. Sukno, B. Bijnens, M. Huguet, and A. Frangi, “Realistic simulation of cardiac magnetic resonance studies modeling anatomical variability, trabeculae, and papillary muscles,” *Magnetic Resonance in Medicine*, vol. 65, no. 1, pp. 280–288, 2011.

[6] R.-S. Kwan, A. C. Evans, and G. B. Pike, “MRI simulation-based evaluation of image-processing and classification methods,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1085–1097, 1999.

[7] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, “4D XCAT phantom for multimodality imaging research,” *Medical Physics*, vol. 37, no. 9, pp. 4902–4915, 2010.

[8] L. Wissmann, C. Santelli, W. P. Segars, and S. Kozerke, “MRXCAT: Realistic numerical phantoms for cardiovascular magnetic resonance,” *Journal of Cardiovascular Magnetic Resonance*, vol. 16, no. 1, pp. 1–11, 2014.

[9] X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: A review,” *Medical Image Analysis*, vol. 58, p. 101552, 2019.

[10] N. K. Singh and K. Raza, “Medical image generation using generative adversarial networks: A review,” *Health Informatics: A Computational Perspective in Healthcare*, pp. 77–96, 2021.

[11] C. Chen, C. Ouyang, G. Tarroni, J. Schlemper, *et al.*, "Unsupervised multi-modal style transfer for cardiac MR segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 209–219, Springer, 2019.

[12] C. Ma, Z. Ji, and M. Gao, "Neural style transfer improves 3D cardiovascular MR image segmentation on inconsistent data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 128–136, Springer, 2019.

[13] A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, "Adversarial image synthesis for unpaired multi-modal cardiac data," in *International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 3–13, Springer, 2017.

[14] T. Joyce and S. Kozerke, "3D medical image synthesis by factorised representation and deformable model learning," in *International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 110–119, Springer, 2019.

[15] A. Prakosa, M. Sermesant, H. Delingette, S. Marchesseau, *et al.*, "Generation of synthetic but visually realistic time series of cardiac images combining a biophysical model and clinical images," *IEEE Transactions on Medical Imaging*, vol. 32, no. 1, pp. 99–109, 2012.

[16] N. Duchateau, M. Sermesant, H. Delingette, and N. Ayache, "Model-based generation of large databases of cardiac images: synthesis of pathological cine MR sequences from real healthy cases," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 755–766, 2017.

[17] Y. Zhou, S. Giffard-Roisin, M. De Craene, S. Camarasu-Pop, *et al.*, "A framework for the generation of realistic synthetic cardiac ultrasound and magnetic resonance imaging sequences from the same virtual patients," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 741–754, 2017.

[18] J. Corral Acero, E. Zacur, H. Xu, R. Ariga, *et al.*, "SMOD-data augmentation based on statistical models of deformation to enhance segmentation in 2D cine cardiac MRI," in *International Conference on Functional Imaging and Modeling of the Heart*, pp. 361–369, Springer, 2019.

[19] S. Abbasi-Sureshjani, S. Amirrajab, C. Lorenz, J. Weese, *et al.*, "4D semantic cardiac magnetic resonance image synthesis on XCAT anatomical model," in *Medical Imaging with Deep Learning*, pp. 6–18, PMLR, 2020.

[20] S. Amirrajab, S. Abbasi-Sureshjani, Y. Al Khalil, C. Lorenz, *et al.*, "XCAT-GAN for synthesizing 3D consistent labeled cardiac MR images on anatomically variable XCAT phantoms," in *International Conference on*

Medical Image Computing and Computer-Assisted Intervention, pp. 128–137, Springer, 2020.

[21] Y. A. Khalil, S. Amirrajab, C. Lorenz, J. Weese, and M. Breeuwer, "Heterogeneous virtual population of simulated CMR images for improving the generalization of cardiac segmentation algorithms," in *International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 68–79, Springer, 2020.

[22] W. P. Segars, D. S. Lalush, E. C. Frey, D. Manocha, *et al.*, "Improved dynamic cardiac phantom based on 4D NURBS and tagged MRI," *IEEE Transactions on Nuclear Science*, vol. 56, no. 5, pp. 2728–2738, 2009.

[23] J. W. Weinsaft, M. D. Cham, M. Janik, J. K. Min, *et al.*, "Left ventricular papillary muscles and trabeculae are significant determinants of cardiac MRI volumetric measurements: effects on clinical standards in patients with advanced systolic dysfunction," *International Journal of Cardiology*, vol. 126, no. 3, pp. 359–365, 2008.

[24] P. A. Helm, H.-J. Tseng, L. Younes, E. R. McVeigh, and R. L. Winslow, "Ex vivo 3D diffusion tensor imaging and quantification of cardiac laminar structure," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 54, no. 4, pp. 850–859, 2005.

[25] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, *et al.*, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[26] B. Herzog, J. Greenwood, and S. Plein, "CMR pocket guide," 2013.

[27] D. Dabir, N. Child, A. Kalra, T. Rogers, *et al.*, "Reference values for healthy human myocardium using a T1 mapping methodology: results from the international t1 multicenter cardiovascular magnetic resonance study," *Journal of Cardiovascular Magnetic Resonance*, vol. 16, no. 1, pp. 1–12, 2014.

[28] M. Granitz, L. J. Motloch, C. Granitz, M. Meissnitzer, *et al.*, "Comparison of native myocardial T1 and T2 mapping at 1.5 T and 3T in healthy volunteers," *Wiener Klinische Wochenschrift*, vol. 131, no. 7, pp. 143–155, 2019.

[29] K. Chow, J. A. Flewitt, J. D. Green, J. J. Pagano, *et al.*, "Saturation recovery single-shot acquisition (SASHA) for myocardial T1 mapping," *Magnetic Resonance in Medicine*, vol. 71, no. 6, pp. 2082–2095, 2014.

[30] S. Matsumoto, S. Okuda, Y. Yamada, T. Suzuki, *et al.*, "Myocardial T1 values in healthy volunteers measured with saturation method using adaptive recovery times for T1 mapping (SMART1Map) at 1.5 T and 3 T," *Heart and Vessels*, vol. 34, no. 11, pp. 1889–1894, 2019.

[31] J. A. Luetkens, R. Homsi, A. M. Sprinkart, J. Doerner, *et al.*, "Incremental value of quantitative CMR including parametric mapping for the diagnosis of acute myocarditis," *European Heart Journal-Cardiovascular Imaging*, vol. 17, no. 2, pp. 154–161, 2016.

[32] B. Baeßler, F. Schaarschmidt, C. Stehning, B. Schnackenburg, *et al.*, "A systematic evaluation of three different cardiac T2-mapping sequences at 1.5 and 3T in healthy volunteers," *European Journal of Radiology*, vol. 84, no. 11, pp. 2161–2170, 2015.

[33] M. Barth and E. Moser, "Proton NMR relaxation times of human blood samples at 1.5 T and implications for functional MRI," *Cellular and Molecular Biology*, vol. 43, no. 5, pp. 783–791, 1997.

[34] G. J. Stanisz, E. E. Odrobina, J. Pun, M. Escaravage, *et al.*, "T1, T2 relaxation and magnetization transfer in tissue at 3T," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 54, no. 3, pp. 507–512, 2005.

[35] C. M. De Bazelaire, G. D. Duhamel, N. M. Rofsky, and D. C. Alsop, "MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 T: preliminary results," *Radiology*, vol. 230, no. 3, pp. 652–659, 2004.

[36] K. Scheffler and S. Lehnhardt, "Principles and applications of balanced SSFP techniques," *European Radiology*, vol. 13, no. 11, pp. 2409–2418, 2003.

[37] K. Scheffler, "A pictorial description of steady-states in rapid magnetic resonance imaging," *Concepts in Magnetic Resonance: An Educational Journal*, vol. 11, no. 5, pp. 291–304, 1999.

[38] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH image to ImageJ: 25 years of image analysis," *Nature Methods*, vol. 9, no. 7, pp. 671–675, 2012.

[39] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[40] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[41] A. Madabhushi and J. K. Udupa, "New methods of MR image intensity standardization via generalized scale," *Medical Physics*, vol. 33, no. 9, pp. 3426–3434, 2006.

[42] P. Kellman and M. S. Hansen, "T1-mapping in the heart: accuracy and precision," *Journal of Cardiovascular Magnetic Resonance*, vol. 16, no. 1, pp. 1–20, 2014.

[43] M. Weigel, "Extended phase graphs: dephasing, RF pulses, and echoes-pure and simple," *Journal of Magnetic Resonance Imaging*, vol. 41, no. 2, pp. 266–295, 2015.

[44] T. Stöcker, K. Vahedipour, D. Pflugfelder, and N. J. Shah, "High-performance computing MRI simulations," *Magnetic Resonance in Medicine*, vol. 64, no. 1, pp. 186–193, 2010.

[45] F. Liu, J. V. Velikina, W. F. Block, R. Kijowski, and A. A. Samsonov, "Fast realistic MRI simulations based on generalized multi-pool exchange tissue model," *IEEE Transactions on Medical Imaging*, vol. 36, no. 2, pp. 527–537, 2016.

[46] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martin-Isla, *et al.*, "Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.

[47] O. Dietrich, J. G. Raya, S. B. Reeder, M. F. Reiser, and S. O. Schoenberg, "Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 26, no. 2, pp. 375–385, 2007.

[48] R. Shaw, C. H. Sudre, T. Varsavsky, S. Ourselin, and M. J. Cardoso, "A k-space model of movement artefacts: application to segmentation augmentation and artefact removal," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2881–2892, 2020.

[49] I. Oksuz, B. Ruijsink, E. Puyol-Antón, J. R. Clough, *et al.*, "Automatic CNN-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning," *Medical Image Analysis*, vol. 55, pp. 136–147, 2019.

[50] E. Kruithof, S. Amirrajab, M. J. Cluitmans, K. Lau, and M. Breeuwer, "Influence of image artifacts on image-based computer simulations of the cardiac electrophysiology," *Computers in Biology and Medicine*, vol. 137, p. 104773, 2021.

[51] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, *et al.*, "Medical image analysis using convolutional neural networks: a review," *Journal of Medical Systems*, vol. 42, no. 11, pp. 1–13, 2018.

[52] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[53] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, *et al.*, "GANs for medical image analysis," *Artificial Intelligence in Medicine*, p. 101938, 2020.

[54] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[55] M. J. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 240–244, IEEE, 2018.

[56] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, *et al.*, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.

[57] G. Zhang, K. Chen, S. Xu, P. C. Cho, *et al.*, "Lesion synthesis to improve intracranial hemorrhage detection and classification for CT images," *Computerized Medical Imaging and Graphics*, vol. 90, p. 101929, 2021.

[58] C. Bermudez, A. J. Plassard, L. T. Davis, A. T. Newton, *et al.*, "Learning implicit brain MRI manifolds with deep learning," in *Medical Imaging 2018: Image Processing*, vol. 10574, p. 105741L, International Society for Optics and Photonics, 2018.

[59] C. Baur, S. Albarqouni, and N. Navab, "MelanoGANs: high resolution skin lesion synthesis with GANs," *arXiv preprint arXiv:1804.04338*, 2018.

[60] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[61] A. Beers, J. Brown, K. Chang, J. P. Campbell, *et al.*, "High-resolution medical image synthesis using progressively grown generative adversarial networks," *arXiv preprint arXiv:1805.03144*, 2018.

[62] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[63] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, 2017.

[64] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, *et al.*, “High-resolution image synthesis and semantic manipulation with conditional GANs,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8798–8807, 2018.

[65] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232, 2017.

[66] M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” *arXiv preprint arXiv:1703.00848*, 2017.

[67] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 172–189, 2018.

[68] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, *et al.*, “Deep MR to CT synthesis using unpaired data,” in *International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 14–23, Springer, 2017.

[69] Y. Liu, A. Chen, H. Shi, S. Huang, *et al.*, “CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy,” *Computerized Medical Imaging and Graphics*, vol. 91, p. 101953, 2021.

[70] Z. Zhang, L. Yang, and Y. Zheng, “Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9242–9251, 2018.

[71] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.

[72] N. Abraham and N. M. Khan, “A novel focal tversky loss function with improved attention U-Net for lesion segmentation,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 683–687, IEEE, 2019.

[73] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning*, pp. 448–456, PMLR, 2015.

[74] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[75] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2337–2346, 2019.

[76] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[77] J. H. Lim and J. C. Ye, "Geometric GAN," *arXiv preprint arXiv:1705.02894*, 2017.

[78] P. Zhu, R. Abdal, Y. Qin, and P. Wonka, "Sean: Image synthesis with semantic region-adaptive normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5104–5113, 2020.

[79] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1511–1520, 2017.

[80] C. Ledig, L. Theis, F. Huszár, J. Caballero, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, 2017.

[81] C. Wang, C. Xu, C. Wang, and D. Tao, "Perceptual adversarial networks for image-to-image transformation," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4066–4079, 2018.

[82] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," *arXiv preprint arXiv:1602.02644*, 2016.

[83] K. Armanious, C. Jiang, M. Fischer, T. Küstner, *et al.*, "MedGAN: Medical image translation using GANs," *Computerized Medical Imaging and Graphics*, vol. 79, p. 101684, 2020.

[84] Y. Liang, D. Lee, Y. Li, and B.-S. Shin, "Unpaired medical image colorization using generative adversarial network," *Multimedia Tools and Applications*, pp. 1–15, 2021.

[85] S. Karimi-Bidhendi, A. Arafati, A. L. Cheng, Y. Wu, *et al.*, "Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases," *Journal of Cardiovascular Magnetic Resonance*, vol. 22, no. 1, pp. 1–24, 2020.

[86] G. Kwon, C. Han, and D.-s. Kim, “Generation of 3D brain MRI using auto-encoding generative adversarial networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 118–126, Springer, 2019.

[87] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, *et al.*, “Image synthesis in multi-contrast MRI with conditional generative adversarial networks,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.

[88] M. Rezaei, “Generative adversarial network for cardiovascular imaging,” in *Machine Learning in Cardiovascular Medicine*, pp. 95–121, Elsevier, 2021.

[89] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, *et al.*, “Medical image synthesis for data augmentation and anonymization using generative adversarial networks,” in *International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 1–11, Springer, 2018.

[90] B. Yu, L. Zhou, L. Wang, J. Fripp, and P. Bourgeat, “3D cGAN based cross-modality MR image synthesis for brain tumor segmentation,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 626–630, IEEE, 2018.

[91] Y. Skandarani, A. Lalande, J. Afilalo, and P.-M. Jodoin, “Generative adversarial networks in cardiology,” *Canadian Journal of Cardiology*, 2021.

[92] Y. Skandarani, N. Painchaud, P.-M. Jodoin, and A. Lalande, “On the effectiveness of GAN generated cardiac MRIs for segmentation,” *arXiv preprint arXiv:2005.09026*, 2020.

[93] D. R. Lustermans, S. Amirrajab, M. Veta, M. Breeuwer, and C. M. Scannell, “Optimized automated cardiac mr scar quantification with gan-based data augmentation,” *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107116, 2022.

[94] Y. Al Khalil, S. Amirrajab, J. Pluim, and M. Breeuwer, “Late fusion U-Net with GAN-Based augmentation for generalizable cardiac MRI segmentation,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 360–373, Springer, 2021.

[95] A. Volokitin, E. Erdil, N. Karani, K. C. Tezcan, *et al.*, “Modelling the distribution of 3D brain MRI using a 2D slice VAE,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 657–666, Springer, 2020.

[96] I. Higgins, L. Matthey, A. Pal, C. Burgess, *et al.*, “beta-vae: Learning basic visual concepts with a constrained variational framework,” *5th International Conference on Learning Representations*, 2016.

[97] P. M. Full, F. Isensee, P. F. Jäger, and K. Maier-Hein, "Studying robustness of semantic segmentation under domain shift in cardiac MRI," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 238–249, Springer, 2020.

[98] C. M. Kramer, J. Barkhausen, C. Bucciarelli-Ducci, S. D. Flamm, *et al.*, "Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update," *Journal of Cardiovascular Magnetic Resonance*, vol. 22, no. 1, pp. 1–18, 2020.

[99] O. P. Simonetti, R. J. Kim, D. S. Fieno, H. B. Hillenbrand, *et al.*, "An improved MR imaging technique for the visualization of myocardial infarction," *Radiology*, vol. 218, no. 1, pp. 215–223, 2001.

[100] R. J. Kim, E. Wu, A. Rafael, E.-L. Chen, *et al.*, "The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction," *New England Journal of Medicine*, vol. 343, no. 20, pp. 1445–1453, 2000.

[101] J. Schulz-Menger, D. A. Bluemke, J. Bremerich, S. D. Flamm, *et al.*, "Standardized image interpretation and post-processing in cardiovascular magnetic resonance-2020 update," *Journal of Cardiovascular Magnetic Resonance*, vol. 22, no. 1, pp. 1–22, 2020.

[102] Q. Tao, J. Milles, K. Zeppenfeld, H. J. Lamb, *et al.*, "Automated segmentation of myocardial scar in late enhancement MRI using combined intensity and spatial information," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 586–594, 2010.

[103] A. Hennemuth, A. Seeger, O. Friman, S. Miller, *et al.*, "A comprehensive approach to the analysis of contrast enhanced cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 11, pp. 1592–1610, 2008.

[104] H. Engblom, J. Tufvesson, R. Jablonowski, M. Carlsson, *et al.*, "A new automatic algorithm for quantification of myocardial infarction imaged by late gadolinium enhancement cardiovascular magnetic resonance: experimental validation and comparison to expert delineations in multi-center, multi-vendor patient data," *Journal of Cardiovascular Magnetic Resonance*, vol. 18, no. 1, pp. 1–13, 2016.

[105] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, p. 65, 2018.

[106] Q. Tao, W. Yan, Y. Wang, E. H. Paiman, *et al.*, "Deep learning–based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study," *Radiology*, vol. 290, no. 1, pp. 81–88, 2019.

[107] R. P. Lim, S. Kachel, A. D. Villa, L. Kearney, *et al.*, "CardiSort: a convolutional neural network for cross vendor automated sorting of cardiac MR images," *European Radiology*, pp. 1–14, 2022.

[108] T. Leiner, D. Rueckert, A. Suinesiaputra, B. Baeßler, *et al.*, "Machine learning in cardiovascular magnetic resonance: basic concepts and applications," *Journal of Cardiovascular Magnetic Resonance*, vol. 21, no. 1, pp. 1–14, 2019.

[109] C. M. Scannell, M. Veta, A. D. Villa, E. C. Sammut, *et al.*, "Deep-learning-based preprocessing for quantitative myocardial perfusion MRI," *Journal of Magnetic Resonance Imaging*, vol. 51, no. 6, pp. 1689–1696, 2020.

[110] B. Ruijsink, E. Puyol-Antón, I. Oksuz, M. Sinclair, *et al.*, "Fully automated, quality-controlled cardiac analysis from CMR: validation and large-scale application to characterize cardiac function," *Cardiovascular Imaging*, vol. 13, no. 3, pp. 684–695, 2020.

[111] C. Chen, C. Qin, H. Qiu, G. Tarroni, *et al.*, "Deep learning for cardiac image segmentation: a review," *Frontiers in Cardiovascular Medicine*, vol. 7, p. 25, 2020.

[112] Y. Zhu, A. S. Fahmy, C. Duan, S. Nakamori, and R. Nezafat, "Automated myocardial T2 and extracellular volume quantification in cardiac MRI using transfer learning–based myocardium segmentation," *Radiology: Artificial Intelligence*, vol. 2, no. 1, 2020.

[113] E. Ferdian, A. Suinesiaputra, K. Fung, N. Aung, *et al.*, "Fully automated myocardial strain estimation from cardiovascular MRI-tagged images using a deep learning framework in the UK Biobank," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, 2020.

[114] A. S. Fahmy, U. Neisius, R. H. Chan, E. J. Rowin, *et al.*, "Three-dimensional deep convolutional neural networkgs for automated myocardial scar quantification in hypertrophic cardiomyopathy: a multicenter multivendor study," *Radiology*, vol. 294, no. 1, p. 52, 2020.

[115] F. Zabihollahy, M. Rajchl, J. A. White, and E. Ukwatta, "Fully automated segmentation of left ventricular scar from 3D late gadolinium enhancement magnetic resonance imaging using a cascaded multi-planar U-Net (CMPU-Net)," *Medical Physics*, vol. 47, no. 4, pp. 1645–1655, 2020.

[116] A. Lalande, Z. Chen, T. Decourselle, A. Qayyum, *et al.*, "EMIDEC: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI," *Data*, vol. 5, no. 4, p. 89, 2020.

[117] J. Ma, "Cascaded framework for automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI," *arXiv preprint arXiv:2012.14556*, 2020.

[118] Y. Zhang, "Cascaded convolutional neural network for automatic myocardial infarction segmentation from delayed-enhancement cardiac MRI," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 328–333, Springer, 2020.

[119] Z. Chen, A. Lalande, M. Salomon, T. Decourselle, *et al.*, "Myocardial infarction segmentation from late gadolinium enhancement MRI by neural networks and prior information," in *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2020.

[120] A. Lalande, Z. Chen, T. Pommier, T. Decourselle, *et al.*, "Deep learning methods for automatic evaluation of delayed enhancement-MRI. the results of the EMIDEC challenge," *Medical Image Analysis*, vol. 79, p. 102428, 2022.

[121] C. Bowles, L. Chen, R. Guerrero, P. Bentley, *et al.*, "GAN augmentation: Augmenting training data using generative adversarial networks," *arXiv preprint arXiv:1810.10863*, 2018.

[122] Y.-B. Tang, S. Oh, Y.-X. Tang, J. Xiao, and R. M. Summers, "CT-realistic data augmentation using generative adversarial network for robust lymph node segmentation," in *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950, pp. 976–981, SPIE, 2019.

[123] C. M. Scannell, A. Chiribiri, and M. Veta, "Domain-adversarial learning for multi-centre, multi-vendor, and multi-disease cardiac MR image segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 228–237, Springer, 2020.

[124] A. Lourenço, E. Kerfoot, I. Grigorescu, C. M. Scannell, *et al.*, "Automatic myocardial disease prediction from delayed-enhancement cardiac MRI and clinical information," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 334–341, Springer, 2020.

[125] K. Yasaka and O. Abe, "Deep learning and artificial intelligence in radiology: Current applications and future directions," *PLoS Medicine*, vol. 15, no. 11, p. e1002707, 2018.

[126] V. M. Bashyam, J. Doshi, G. Erus, D. Srinivasan, *et al.*, "Deep generative medical image harmonization for improving cross-site generalization in deep learning predictors," *Journal of Magnetic Resonance Imaging*, vol. 55, no. 3, pp. 908–916, 2022.

[127] R. Attar, M. Pereanez, A. Gooya, X. Alba, *et al.*, "Quantitative CMR population imaging on 20,000 subjects of the UK Biobank imaging study: LV/RV quantification pipeline and its evaluation," *Medical Image Analysis*, vol. 56, pp. 26–42, 2019.

[128] C. Petitjean and M. A. e. a. Zuluaga, "Right ventricle segmentation from cardiac MRI: a collation study," *Medical Image Analysis*, vol. 19, no. 1, pp. 187–202, 2015.

[129] S. Marchesseau, J. X. Ho, and J. J. Totman, "Influence of the short-axis cine acquisition protocol on the cardiac function evaluation: a reproducibility study," *European Journal of Radiology Open*, vol. 3, pp. 60–66, 2016.

[130] P. Yilmaz, K. Wallecan, W. Kristanto, J.-P. Aben, and A. Moelker, "Evaluation of a semi-automatic right ventricle segmentation method on short-axis MR images," *Journal of Digital Imaging*, vol. 31, no. 5, pp. 670–679, 2018.

[131] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 111–119, Springer, 2017.

[132] C. Chen, K. Hammernik, C. Ouyang, C. Qin, *et al.*, "Cooperative training and latent space data augmentation for robust medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 149–159, Springer, 2021.

[133] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 113–123, 2019.

[134] S. Jeong and S. Lee, "Biased extrapolation in latent space for imbalanced deep learning," in *International Workshop on Machine Learning in Medical Imaging*, pp. 337–346, Springer, 2021.

[135] J. Xu, M. Li, and Z. Zhu, "Automatic data augmentation for 3D medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 378–387, Springer, 2020.

[136] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[137] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, *et al.*, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[138] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, “Simultaneous deep transfer across domains and tasks,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4068–4076, 2015.

[139] N. K. Dinsdale, M. Jenkinson, and A. I. Namburete, “Unlearning scanner bias for MRI harmonisation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 369–378, Springer, 2020.

[140] J. C. Acero, V. Sundaresan, N. Dinsdale, V. Grau, and M. Jenkinson, “A 2-step deep learning method with domain adaptation for multi-centre, multi-vendor and multi-disease cardiac magnetic resonance segmentation,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 196–207, Springer, 2020.

[141] J. J. Jeong, A. Tariq, T. Adejumo, H. Trivedi, *et al.*, “Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation,” *Journal of Digital Imaging*, pp. 1–16, 2022.

[142] L. G. Nyúl, J. K. Udupa, and X. Zhang, “New variants of a method of MRI scale standardization,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143–150, 2000.

[143] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[144] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, *et al.*, “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[145] W. Zhang, R. Li, H. Deng, L. Wang, *et al.*, “Deep convolutional neural networks for multi-modality isointense infant brain image segmentation,” *NeuroImage*, vol. 108, pp. 214–224, 2015.

[146] X. Zhuang, L. Li, C. Payer, D. Štern, *et al.*, “Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge,” *Medical Image Analysis*, vol. 58, p. 101537, 2019.

[147] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, “A review of multimodal medical image fusion techniques,” *Computational and Mathematical Methods in Medicine*, vol. 2020, 2020.

[148] N. Srivastava and R. R. Salakhutdinov, “Multimodal learning with deep boltzmann machines,” *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[149] M. Aygün, Y. H. Şahin, and G. Ünal, “Multi modal convolutional neural networks for brain tumor segmentation,” *arXiv preprint arXiv:1809.06191*, 2018.

[150] J. Dolz, C. Desrosiers, and I. B. Ayed, “IVD-Net: Intervertebral disc localization and segmentation in mri with a multi-modal unet,” in *International Workshop and Challenge on Computer Methods and Clinical Applications for Spine Imaging*, pp. 130–143, 2018.

[151] D. Nie, L. Wang, Y. Gao, and D. Shen, “Fully convolutional networks for multi-modality isointense infant brain image segmentation,” in *2016 IEEE 13Th International Symposium on Biomedical Imaging (ISBI)*, pp. 1342–1345, IEEE, 2016.

[152] G. van Tulder and M. de Bruijne, “Learning cross-modality representations from multi-modal images,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 638–648, 2018.

[153] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

[154] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, *et al.*, “HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 5, pp. 1116–1126, 2018.

[155] T. Zhang, G.-J. Qi, B. Xiao, and J. Wang, “Interleaved group convolutions,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4373–4382, 2017.

[156] Y. Amano, M. Kitamura, H. Takano, F. Yanagisawa, *et al.*, “Cardiac MR imaging of hypertrophic cardiomyopathy: techniques, findings, and clinical relevance,” *Magnetic Resonance in Medical Sciences*, vol. 17, no. 2, p. 120, 2018.

[157] X. Sun, L.-H. Cheng, and R. J. Geest, “Right ventricle segmentation via registration and multi-input modalities in cardiac magnetic resonance imaging from multi-disease, multi-view and multi-center,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 241–249, Springer, 2021.

[158] T. W. Arega, F. Legrand, S. Bricq, and F. Meriaudeau, “Using MRI-specific data augmentation to enhance the segmentation of right ventricle in multi-disease, multi-center and multi-view cardiac MRI,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 250–258, Springer, 2021.

[159] L. Li, W. Ding, L. Huang, and X. Zhuang, "Right ventricular segmentation from short-and long-axis MRIs via information transition," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 259–267, Springer, 2021.

[160] C. Galazis, H. Wu, Z. Li, C. Petri, *et al.*, "Tempera: Spatial transformer feature pyramid network for cardiac MRI segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 268–276, Springer, 2021.

[161] S. Jabbar, S. T. Bukhari, and H. Mohy-ud Din, "Multi-view SA-LA Net: A framework for simultaneous segmentation of RV on multi-view cardiac MR images," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 277–286, Springer, 2021.

[162] S. Queirós, "Right ventricular segmentation in multi-view cardiac MRI using a unified U-Net model," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 287–295, Springer, 2021.

[163] M. J. Fulton, C. R. Heckman, and M. E. Rentschler, "Deformable bayesian convolutional networks for disease-robust cardiac MRI segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 296–305, Springer, 2021.

[164] Z. Gao and X. Zhuang, "Consistency based co-segmentation for multi-view cardiac MRI using vision transformer," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 306–314, Springer, 2021.

[165] D. Liu, Z. Yan, Q. Chang, L. Axel, and D. N. Metaxas, "Refined deep layer aggregation for multi-disease, multi-view & multi-center cardiac MR segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 315–322, Springer, 2021.

[166] M. Beetz, J. Corral Acero, and V. Grau, "A multi-view crossover attention U-Net cascade with fourier domain adaptation for multi-domain cardiac MRI segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 323–334, Springer, 2021.

[167] M. Mazher, A. Qayyum, A. Benzinou, M. Abdel-Nasser, and D. Puig, "Multi-disease, multi-view and multi-center right ventricular segmentation in cardiac MRI using efficient late-ensemble deep learning approach," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 335–343, Springer, 2021.

[168] K. Punithakumar, A. Carscadden, and M. Noga, "Automated segmentation of the right ventricle from magnetic resonance imaging using deep convolutional neural networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 344–351, Springer, 2021.

[169] L. Tautz, L. Walczak, C. Manini, A. Hennemuth, and M. Hüllebrand, "3D right ventricle reconstruction from 2D U-Net segmentation of sparse short-axis and 4-chamber cardiac cine MRI views," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 352–359, Springer, 2021.

[170] F. Galati and M. A. Zuluaga, "Using out-of-distribution detection for model refinement in cardiac image segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 374–382, Springer, 2021.

[171] C. G. Xanthis, D. Filos, K. Haris, and A. H. Aletras, "Simulator-generated training datasets as an alternative to using patient data for machine learning: an example in myocardial segmentation with MRI," *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105817, 2021.

[172] D. F. Bauer, T. Russ, B. I. Waldkirch, C. Tönnes, *et al.*, "Generation of annotated multimodal ground truth datasets for abdominal medical image registration," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 8, pp. 1277–1285, 2021.

[173] C.-B. Jin, H. Kim, M. Liu, W. Jung, *et al.*, "Deep CT to MR synthesis using paired and unpaired data," *Sensors*, vol. 19, no. 10, p. 2361, 2019.

[174] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *European Conference on Computer Vision*, pp. 319–345, Springer, 2020.

[175] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[176] V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Medical Image Analysis*, vol. 54, pp. 280–296, 2019.

[177] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, *et al.*, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 516–524, Springer, 2017.

[178] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7167–7176, 2017.

[179] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data augmentation for brain-tumor segmentation: a review," *Frontiers in Computational Neuroscience*, vol. 13, p. 83, 2019.

[180] L. Zhang, X. Wang, D. Yang, T. Sanford, *et al.*, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.

[181] A. Choudhary, L. Tong, Y. Zhu, and M. D. Wang, "Advancing medical imaging informatics by deep learning-based domain adaptation," *Yearbook of Medical Informatics*, vol. 29, no. 1, p. 129, 2020.

[182] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016.

[183] B. Baillargeon, N. Rebelo, D. D. Fox, R. L. Taylor, and E. Kuhl, "The living heart project: a robust and integrative simulator for human heart function," *European Journal of Mechanics-A/Solids*, vol. 48, pp. 38–47, 2014.

[184] Y. Al Khalil, S. Amirrajab, C. Lorenz, J. Weese, *et al.*, "On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images," *Medical Image Analysis*, p. 102688, 2022.

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor, Marcel Breeuwer, for his support and guidance throughout my PhD research. He has been a constant source of encouragement and motivation, and I could not have completed this work without his expert guidance. Marcel has not only been a knowledgeable and scientifically-minded supervisor, but also very friendly, kind, and caring in personal relationships. You have always been easy to talk to and get along with, I enjoyed early lunches on Thursdays and off-work chats. Most importantly, I admire the freedom and flexibility that you gave me during my PhD research. Your trust in my ability to explore different ideas, as well as your attention to learning and experimenting, have been instrumental in my academic and personal growth. I am truly grateful for your mentorship and support throughout this journey.

I am incredibly grateful to have had Josien Pluim as my co-supervisor during my research. I especially appreciated her amazing suggestions and comments on my scientific papers. Josien has a real talent for giving super useful feedback, which helped me improve my writing and be more efficient and concise with my ideas. Thank you, Josien, for your exceptional support and wisdom.

I would like to extend my sincerest gratitude to my industrial supervisors, Jurgen Weese and Cristian Lorenz from Philips Research Hamburg. Their critical thinking and scientific rigor have been invaluable in shaping my research. Through our weekly progress meetings, I learned to approach my work from a bottom-up perspective, considering the conclusion and message first to build a cohesive story and designing experiments accordingly. I am deeply grateful for their mentorship.

My sincere appreciation goes to Amedeo Chiribiri for his excellent teaching and guidance during the CMR course at KCL. Amedeo's extensive knowledge and expertise in cardiac MRI have been invaluable in helping me develop a deeper understanding of this field. Additionally, I want to recognize Tony Stoecker for his invaluable advice at the beginning of my research project. During my secondment in his group at DZNE Bone in April 2019, Tony provided me with clear and concise explanations of the concepts underlying JEMRIS for MR simulation. Furthermore, His insightful suggestions to use analytical solutions for Bloch equations instead of JEMRIS simulation were a key turning point in my research.

I'm grateful to the OpenGTN supervisory team for their support and guidance throughout my research. Thank you Tom Geraedts for valuable insights into Philips pulse sequence design, Jouke Smink for scanning my heart, which ultimately contributed to the publication of my work in IEEE TMI paper, and Liesbeth Geerts for accommodating me at the MR R&D Clinical Science depart-

ment in Best which provided me with valuable industrial exposure.

I would like to express my gratitude to Mitko Veta for trusting in my abilities and recommending me for the Spectralligence postdoc position. I also appreciate Mitko's kindness in lending me some renovation tools for my DIY projects, who needs a contractor when you've got Mitko and a YouTube tutorial!

Thank you Cian Scannell for your encouragement and fruitful discussions about the use cases of synthetic data for cardiac image analysis. I also extend my thanks to Evianne Kruithof and Didier Lustermans with whom I had the opportunity to collaborate on their graduation projects, resulting in two journal publications.

Special thanks to Samaneh Abbasi for her openness in discussing ideas and effective collaboration on using conditional GANs for image synthesis, which significantly impacted my research and resulted in multiple publications.

To my dear fellow OpenGTN researchers, including Yasmina Al Khalil and Aymen Ayaz, I am incredibly grateful for your encouragement, feedback, and willingness to lend a helping hand throughout my research. I would like to give special thanks to Yasmina for her insights and countless hours of stimulating discussions, which have led to never-ending experiment ideas. I have thoroughly enjoyed working with her, and I am grateful to have her as my research buddy. Let's continue to collaborate in the future. To my math wizard, Andrey Zhylka for being such a kind and generous human being and bringing countless presents during the four years we shared our office at Gemini-Zuid 2.111. I really enjoyed the Kibbeling on Tuesday markets with you!

I want to give a shoutout to my wonderful colleagues in the IMAGE/e group, including Seb (the peanut butter linux geek), Dennis (always asking why), Evi (hilarious traveller), and Iris (curious and attentive). Thank you for the countless laughs, diverse conversations, and delightful coffee breaks that have made this journey so enjoyable.

I am beyond lucky to have met some of the coolest people ever during my time in the Netherlands. Huge thanks to my amazing neighbors at Pastoor for introducing me to the wildest game nights, especially the infamous Mafia game. I'll never forget the time I got locked out of my apartment in my pajamas and had to turn to my trusty neighbor Razie Moji for help! Shoutout to Faranak, Amir, Roozbeh, Parinaz, Amirreza, Sahar, Alireza, Maryam, Abolfazl, Sara, Haleh, Majid, Ali, Neusha, Mahya G for being the perfect distraction from the sometimes overwhelming world of scientific research. You guys rock!

I want to give a big shoutout to my party crew! Aylar, Tohid, Amin, you guys always know how to have a good time and make me laugh, even by making fun of me! And a special thanks to Sadaf o Peyman for the novel and unforgettable experiences we've shared. Cheers to many more wild nights together!

Big ups to my "Refighe Garmabeh o Gollestan" from Iran! I'm talking about the folks who keep me laughing, inspired, and always in a good mood. To Masood, Franak, Ali, Armita, Meysam, Roya, Saeed, Aida.

I would like to express my gratitude to my dearest parents, Mozhgan and Ali, and my lovely sister, Setare. Without their love and invaluable support, my academic journey would not have been possible. Thank you, Tofigh, Soheila, Aida, and Barbari, for being a part of this journey and for your constant encouragement.

I couldn't possibly end this acknowledgement without expressing my heartfelt gratitude and undying love to the most important person in my life, although words cannot fully express the depth of my love for you. I am fortunate to have had you by my side every moment since we where only 18 years old. You are an absolute blessing, Saina. Thank you for your unconditional love, support, wisdom, and care. You are the shining light that has guided me through life's challenges and the reason I have made it this far, thank you for bringing joy, happiness, and meaning to my life. I love you more than words could ever say, my beloved Nafas to whom this thesis is dedicated.

About the Author

Sina Amirrajab was born on April 29, 1991, in Dezful, a city with scorching summers in the south of Iran. Sina received his Bachelor's degree in Electrical Engineering from the University of Guilan in Rasht (2009-2013), an evergreen land in the north of Iran bordering on the Caspian Sea, and his Master's degree in Biomedical Engineering from the Amirkabir University of Technology in Tehran (2014-2017), the capital. After completing his undergraduate degree, Sina worked as an electrical research and development engineer at a company in Tehran from 2013 to 2015, where he was responsible for electronic PCB design and product quality control for industrial devices. During his Master's degree, he specialized in magnetic resonance imaging (MRI) artifacts in the presence of metallic objects in the scanner, gaining expertise in the physics of MR imaging and different techniques of image simulation. In 2016, Sina began collaborating with a start-up company as an MR researcher to establish a quality assurance procedure for MRI based on the ACR accreditation program, which was implemented for the first time in Iran. As a member of a team of three, he was responsible for performing quality control and acceptance testing for newly purchased MRI scanners for institutes and hospitals. Since July 2018, Sina has been a PhD candidate in the Medical Image Analysis group within the Department of Biomedical Engineering at Eindhoven University of Technology. He has been working on the topic of Simulation and Synthesis for Cardiac MR Image Analysis as part of the openGTN project supported by the European Union in the Marie Curie Innovative Training Networks (ITN) fellowship program. During his PhD, Sina has been under the academic supervision of Prof.dr.ir. Marcel Breeuwer and Prof.dr Josien Pluim from the TU/e and the industrial supervision of dr. Jürgen Weese and dr. Cristian Lorenz from Philips Research. The results of his research are presented in this thesis.

List of Publications

Journals

[J1] **S. Amirrajab**, Y. A. Khalil, C. Lorenz, J. Weese, J. Pluim and M. Breeuwer, "A Framework for Simulating Cardiac MR Images with Varying Anatomy and Contrast," in IEEE Transactions on Medical Imaging, 2022.

[J2] E. Kruithof, **S. Amirrajab**, M.J.M. Cluitmans, K.D. Lau, M. Breeuwer, "Influence of Image Artifacts on Image-Based Computer Simulations of the Cardiac Electrophysiology," Computers in Biology and Medicine, 2021.

[J3] **S. Amirrajab***, Y. Al Khalil*, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "Label-informed Cardiac Magnetic Resonance Image Synthesis through Conditional Generative Adversarial Networks," Computerized Medical Imaging and Graphics, 2022.

[J4] Y. Al Khalil*, **S. Amirrajab***, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "On the Usability of Synthetic Data for Improving the Robustness of Deep Learning-based Segmentation of Cardiac Magnetic Resonance Images," Medical Image Analysis, 2022.

[J5] D.R.P.R.M. Lustermans, **S. Amirrajab**, M. Veta, M. Breeuwer, C. M. Scannell, "Optimized Automated Cardiac MR Scar Quantification with GAN-Based Data Augmentation," Computer Methods and Programs in Biomedicine, 2022.

[J6] E. Kruithof, **S. Amirrajab**, K.D. Lau, M. Breeuwer, "Simulated Late Gadolinium Enhanced Cardiac Magnetic Resonance Imaging Dataset from Mechanical XCAT Phantom Including a Myocardial Infarct," Data Brief, 2021.

[J7] Y. Al Khalil*, **S. Amirrajab***, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "Reducing Segmentation Failures in Cardiac MRI via Late-Feature Fusion and GAN-Based Augmentation," submitted, 2022.

[J8] **S. Amirrajab**, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "Pathology Synthesis of 3D Consistent Cardiac MR Images Using 2D VAEs and GANs," submitted, 2022.

Conference proceedings

[C1] S. Abbasi-Sureshjani*, **S. Amirrajab***, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "4D Semantic Cardiac Magnetic Resonance Image Synthesis on XCAT Anatomical Model," MIDL 2020.

[C2] **S. Amirrajab***, S. Abbasi-Sureshjani*, Y. Al Khalil*, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "XCAT-GAN for Synthesizing 3D Consistent Labeled Cardiac MR Images on Anatomically Variable XCAT Phantoms," MICCAI 2020.

[C3] **S. Amirrajab**, W. P. Segars, C. Lorenz, J. Weese, M. Breeuwer, "Towards Realistic Cardiac MR Image Simulation; Inclusion of the Endocardial Trabeculae in the XCAT Heart Anatomy," ISMRM 2020.

[C4] E. Kruithof, **S. Amirrajab**, M.J.M. Cluitmans, K.D. Lau, M. Breeuwer, "Influence of Image Artifacts on Image-Based Electrophysiological Simulations Using Simulated XCAT Phantom MR Images," ESMRMB 2020.

[C5] **S. Amirrajab**, C. Lorenz, J. Weese, M. Breeuwer, "A Multipurpose Numerical Simulation Tool for Late Gadolinium Enhancement Cardiac MR Imaging," ESMRMB 2020.

[C6] **S. Amirrajab**, Y. Al Khalil, C. Lorenz, J. Weese, M. Breeuwer, "Generation of Realistic and Heterogeneous Virtual Population of Cardiovascular Magnetic Resonance Simulated Images," ISMRM 2020.

[C7] **S. Amirrajab**, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "sim2real: Cardiac MR Image Simulation-to-Real Translation via Unsupervised GANs," ISMRM 2022.

[C8] **S. Amirrajab**, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, "Intra- and Inter-Subject Synthesis of Cardiac MR Images Using a VAE and GAN," ISMRM 2022.

[C9] Y. Al Khalil*, **S. Amirrajab***, C. Lorenz, J. Weese, M. Breeuwer, "Heterogeneous Virtual Population of Simulated CMR Images for Improving the Generalization of Cardiac Segmentation Algorithms," SASHIMI workshop MICCAI 2020.

[C10] Y. Al Khalil*, **S. Amirrajab***, J. Pluim, M. Breeuwer, "Late Fusion U-Net with GAN-Based Augmentation for Generalizable Cardiac MRI Segmentation," M&MS-2 challenge 2021.

[C11] S. D. Harrevelt*, Y. Al Khalil*, **S. Amirrajab***, J. Pluim, M. Breeuwer, A. Raaijmakers, "Field Strength Agnostic Cardiac MR Image Segmentation," MIDL 2022.

[C12] **S. Amirrajab***, C. Lorenz, J. Weese, J. Pluim, M. Breeuwer, “Pathology Synthesis of 3D Consistent Cardiac MR Images Using 2D VAEs and GANs,” SASHIMI workshop MICCAI 2022.

[C13] **S. Amirrajab***, Y. Al Khalil*, J. Pluim, M. Breeuwer, C. M. Scannell, “Cardiac MR Image Segmentation and Quality Control in the Presence of Respiratory Motion Artifacts Using Simulated Data,” CMRxMotion challenge MICCAI 2022.

* Equal contribution