Fast Fully Automatic Segmentation of the Severely Abnormal Human Right Ventricle from Cardiovascular Magnetic Resonance Images using a Multi-scale 3D Convolutional Neural Network

Archontis Giannakidis[*†¶], Konstantinos Kamnitsas[‡], Veronica Spadotto[§*], Jennifer Keegan[*†], Gillian Smith[*], Ben Glocker[‡], Daniel Rueckert[‡], Sabine Ernst[*], Michael A. Gatzoulis[*], Dudley J. Pennell[*†], Sonya Babu-Narayan[*‖], David N. Firmin[*†‖]

[*]*NIHR Cardiovascular Biomedical Research Unit, Royal Brompton Hospital, London, SW3 6NP, UK*
[†]*National Heart & Lung Institute, Imperial College London, London, SW3 6NP, UK*
[‡]*Department of Computing, Imperial College London, London, SW7 2AZ, UK*
[§]*Department of Cardiac, Thoracic and Vascular Sciences, University of Padua, Padua, 35128, Italy*

[¶]*Corresponding author: Email: A.Giannakidis@rbht.nhs.uk, Tel: +44(0)20-7351-8819, Fax: +44(0)20-7351-8816*
[‖]*Joint senior author*

***Abstract*—Cardiac magnetic resonance (CMR) is regarded as the reference examination for cardiac morphology in tetralogy of Fallot (ToF) patients allowing images of high spatial resolution and high contrast. The detailed knowledge of the right ventricular anatomy is critical in ToF management. The segmentation of the right ventricle (RV) in CMR images from ToF patients is a challenging task due to the high shape and image quality variability. In this paper we propose a fully automatic deep learning-based framework to segment the RV from CMR anatomical images of the whole heart. We adopt a 3D multi-scale deep convolutional neural network to identify pixels that belong to the RV. Our robust segmentation framework was tested on 26 ToF patients achieving a Dice similarity coefficient of 0.8281±0.1010 with reference to manual annotations performed by expert cardiologists. The proposed technique is also computationally efficient, which may further facilitate its adoption in the clinical routine.**

***Keywords*-3D convolutional neural network; segmentation; right ventricle; cardiavascular magnetic resonance; tetralogy of Fallot; deep learning**

I. Introduction

Congenital heart defects (CHD) are the most common type of birth abnormalities [1]. Among the various CHD, tetralogy of Fallot (ToF) is the most prevalent cyanotic anomaly, with one baby being born with ToF every five hours in the United States [2], [3]. ToF is made up of the following four flaws: a subaortic ventricular septal defect, narrowing of the pulmonary outflow tract, overriding aorta, and right ventricular hypertrophy. Although palliative and subsequent reparative surgery have revolutionised the survival prospects of the ToF population, their care is extremely challenging due to common complications. The growing ToF patient group requires lifelong follow-up.

The knowledge of the anatomy of the right ventricle (RV) is critical in ToF management, as it may support the entire clinical workflow from diagnosis and risk stratification to therapy planning and interventions [4]. Cardiovascular magnetic resonance (CMR) is increasingly being heralded [5] as the imaging modality of choice for the evaluation of the right ventricular anatomy in ToF patients. It allows enhanced visualization of cardiac structures without exposing the body to ionizing radiation.

Currently, the clinical routine to annotate the RV in CMR images is a manual process performed by an experienced cardiologist who relies on visual inspection. However, this process is time-consuming and laborious. In addition, it is subject to high intra- and inter-observer errors. Alternatively, automatic segmentation methods could help relieve the cardiologists' workload, as well as improve the reliability of the outcome. However, the RV shape in ToF patients eludes any standardization and may assume various morphologies depending on previous surgical treatment and/or other patho-physiological conditions [6]. There is also a high variation in CMR image intensity and quality. All these render the automatic RV segmentation in ToF patients a challenge.

Deep learning is a rapidly growing trend in general data analysis that currently drives the artificial intelligence boom. Convolutional neural networks (CNNs) [7], in particular, have convincingly outperformed the state-of-the-art in several computer vision applications, raising expectations that they might be applied in other domains, such as medical image analysis. Being inspired by the biological processes in the brain, CNNs consist of a series of inter-connected layers of artificial neurons. Their deep architecture allows for extracting a set of highly discriminating features at multiple levels of abstraction. Contrary to other traditional supervised machine learning techniques that rely on human ingenuity to manually devise features, CNNs use an automatic data-driven process to learn features concurrently with the training of the classifier.

In this paper, we propose a framework that adopts a 3D multi-scale deep CNN to perform the challenging task of automatically identifying voxels that belong to the RV of ToF patients. To the best of our knowledge, this is the first study to fully automatically segment the RV of ToF

patients from 3D high spatial resolution CMR images.

II. The Method

We adopt a 3D multi-scale CNN architecture [8] that is 11-layers deep and consists of two pathways to segment the RV from the whole heart images. Next, we briefly describe the network's architecture and its justification.

The proposed network is 3D which, contrary to the most commonly used 2D CNNs, makes optimal use of the volumetric image data. The network is implemented as fully-convolutional using the Parametric Rectified Linear Unit (PReLU) non-linearity [9]. To tackle the slow inference associated with 3D CNNs, the network's input size ($= 25^3$) is designed to be greater than the CNN's receptive field ($= 17^3$), so that the dense-inference may be exploited [10]. This strategy allows the simultaneous prediction of V ($= 9^3$) voxels in one pass of the network. Therefore, the computational load is significantly reduced through avoiding to repeat convolutional operations on the same voxels of patches that overlap. At the same time, the effective training batch size increases by V (without concurrently increasing the computational load) which is preferred as it improves the accuracy of the estimation [8].

The intentional use of larger inputs also increases the flexibility when sampling input segments. This may be exploited to significantly improve the segmentation performance by using the dense training proposed in [8]. According to this hybrid training scheme, the training batches are formed from training images with 50% possibility being centred on a foreground (RV) or background voxel, thus alleviating the class-imbalance present in the data.

If B segments are used to form a training batch, then the CNN's parameters $\mathbf{\Theta}$ (such as weights and biases) may be estimated by minimizing the following cross-entropy-based cost function

$$J(\mathbf{\Theta}; \mathbf{I}_s, \mathbf{c}_s) = -\frac{1}{B \cdot V} \sum_{s=1}^{B} \sum_{v=1}^{V} \log(p_{c_s^v}(\mathbf{x}^v)) \qquad (1)$$

where $\mathbf{I}_s$ and $\mathbf{c}_s$ are the $s-$th segment of the batch and the true labels of its predicted voxels, respectively. c_s^v is the true label of the $v-$th voxel, $\mathbf{x}^v$ is the corresponding position in the classification feature maps, and $p_{c_s^v}$ is the output of the softmax function. The Stochastic Gradient Descent (SGD) may be used to solve (1), while multiple optimization steps over the various batches lead to convergence.

The adopted network has a deeper architecture. This puts it at an advantage over shallower CNNs, as deeper architectures allow for greater discriminative powers due to the additional non-linearities. In addition, they escape more easily from local minima [11]. However, deeper CNNs are more prone to overfitting. To address this problem, the employed network makes use of small 3^3 kernels, which is an implicit way of regularization [12]. The smaller kernels are faster to convolve with as well as contain less weights.

Combining local and larger contextual information in the decision process has been shown to improve segmentation results. In order to efficiently achieve this merging, the selected network uses a dual pathway architecture that processes the input images at multiple scales simultaneously. The first pathway operates on the original images capturing the finest details, whereas the second pathway operates on down-sampled images (with a factor three) learning higher level features. To preserve the dense inference characteristic of the CNN, the feature maps of the last convolutional layer of the second pathway are upsampled to match the dimensions of the last convolutional layer of the first pathway. Then, the two feature maps are concatenated together.

Finally, the employed network includes two more hidden layers for combining the multi-scale features before the final classification layer, resulting in a deep network of 11 layers in total. The kernel size for the last three layers is 1^3. Figure 1 shows an overview of the network architecture.

III. Evaluation on clinical data

A. Clinical data

Twenty-six patients with ToF underwent CMR imaging after admission to Royal Brompton Hospital. The study was approved by the local (UK) research ethics committee. Written informed consent from all research participants was obtained. All data were acquired on a Siemens Avanto 1.5 Tesla scanner (Siemens Medical Systems, Erlangen, Germany). The roadmap acquisition consisted of a navigator-gated non-selective [13] 3D balanced steady state free precession sequence (TE = 1 ms, TR = 2.3 ms; GRAPPA $\times 2$) acquired in the coronal plane. Data (72 slices at 1.6 $\times$ 1.6 $\times$ 3.2 mm, reconstructed to 144 slices at 0.8 $\times$ 0.8 $\times$ 1.6 mm) were acquired over 100 ms in a patient-specific mid-diastolic or systolic pause. The imaging sequence incorporated chemical-shift fat-suppression, T2-preparation and CLAWS [14] (continuously adaptive windowing strategy) respiratory motion control. The RV chamber was manually annotated on all roadmaps by CMR experts.

B. Pre-processing, network configuration and training

The original images were downsampled by a factor of two on speed grounds, and also based on the reasoning that this would have minimal influence on the results. The image intensities were normalized to have zero mean and unit variance [15]. We applied dense training, with the image segments being extracted with equal probability centred on the RV and other background. The training data was enriched by adding images reflected along the saggital axis. Ten segments were used to form a training batch. To refrain the forward (neuron activations) and backwards (gradients) propagated signal from exploding or vanishing, the kernel weights were initialized by sampling from the normal distribution $N(0, \sqrt{2/n_l^{input}})$, where n_l^{input} is the number of weights through which a neuron of layer l is connected to its input [9]. To stop the internal covariate shift from hindering the weight convergence at deeper layers, the Batch Normalization technique [16] to all hidden layers was adopted. The network was regularized

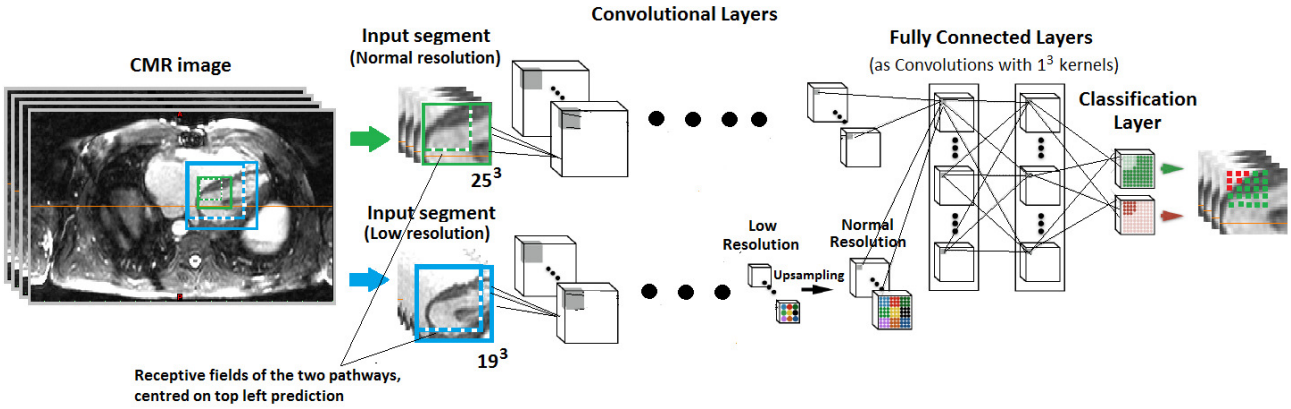

Figure 1. The proposed architecture for fully automatic segmentation of the right ventricle. The 3D CNN has two convolutional pathways. The neurons of the last layers of the two pathways have receptive fields equal to 17^3 voxels. The inputs of the two pathways are centered at the same image location, but the second segment is extraced from a down-sampled version of the image (down-sampling factor = 3). The second pathway processes context in an actual area of size 51^3 voxels.

using dropout [17] with a 50% rate on the last two layers. For the training, the RMSProp optimizer [18] and Nesterov momentum [19] were used. The network was evaluated with two-fold cross-validation on the 26 subjects. Training one-fold took approximately 24 hours, and inference can be done within one minute on an NVIDIA GTX Titan X GPU.

C. Results

To evaluate the accuracy of the proposed segmentation technique, the Dice similarity coefficient and the absolute volume difference (both with reference to the manual annotations) were used. The results are summarized in Table I. Examples for qualitative assessment are provided in Fig. 2.

Table I
EVALUATION OF THE PROPOSED CNN-BASED SEGMENTATION TECHNIQUE USING THE DICE SIMILARITY COEFFICIENT AND THE ABSOLUTE VOLUME DIFFERENCE (BOTH WITH REFERENCE TO THE MANUAL ANNOTATIONS). RESULTS OF THE TWO-FOLD CROSS-VALIDATION ARE EXPRESSED AS MEAN ± STANDARD DEVIATION.

Dice similarity coefficient	Absolute volume difference (%)
0.8281±0.1010	12.6864±12.9872

IV. DISCUSSION & CONCLUSION

Only recently has the image interpretation process begun to benefit from artificial intelligence. We have presented a fully automatic approach for segmenting the severely abnormal RV of ToF patients from CMR images using a 3D multi-scale deep CNN. The deep learning-based method was evaluated with two-fold cross-validation on 26 subjects. The results we acquired from this pilot study (Dice score = 0.8281±0.1010) show that the proposed technique has potential for automating the annotation of this peculiar anatomical structure. The achieved score is superior to the RV segmentation performance (Dice score = 0.80) reached by other machine learning techniques [20], even though the latter relied on manual engineering and were applied to a patient group that is less variable/challenging than the adult congenital heart disease cohort of this study. The outcomes of this paper can be of great value in particular for institutions (such as ours) that receive great numbers of ToF patients. The proposed technique is also computationally efficient, which may further facilitate its adoption in the clinical routine. In general, the fast and robust evaluation of this widely varied anatomy may help to develop personalized preventive and therapeutic regimens for ToF. In addition, it may promote the establishment of novel anatomical biomarkers.

Future work will involve running the experiments without down-sampling the original images by a factor of two. In addition, we will use more datasets, the annotation of which is currently in progress. Another task we plan to explore is the refinement of the segmentation results using a 3D fully connected conditional random field (CRF) [8]. This post-processing step is expected to ”clean-up” the CNN results and achieve more structured predictions, though the CRF configuration might be a challenging task. For this study, the CNN was trained using the whole images. A future task will be to perform the CNN training by excluding areas outside the heart that contribute little to the learning process. To this end, we will initialize our segmentation framework by performing automatic anatomical landmark localization [21], [22]. Ultimately, our goal is to test the proposed framework in a multi-class classification context [where the task of interest will be the whole heart segmentation (four chambers)], and, then, to juxtapose its performance against other state-of-the-art whole heart segmentaton techniques [23], [24].

ACKNOWLEDGMENT

This work was supported by funds from the NIHR Cardiovascular Biomedical Research Unit of Royal Brompton & Harefield NHS Foundation Trust and Imperial College London. Sonya Babu-Narayan is supported by a British Heart Foundation Intermediate Clinical Research Fellowship. Konstantinos Kamnitsas is supported by the Imperial College London PhD Scholarship Programme. The authors

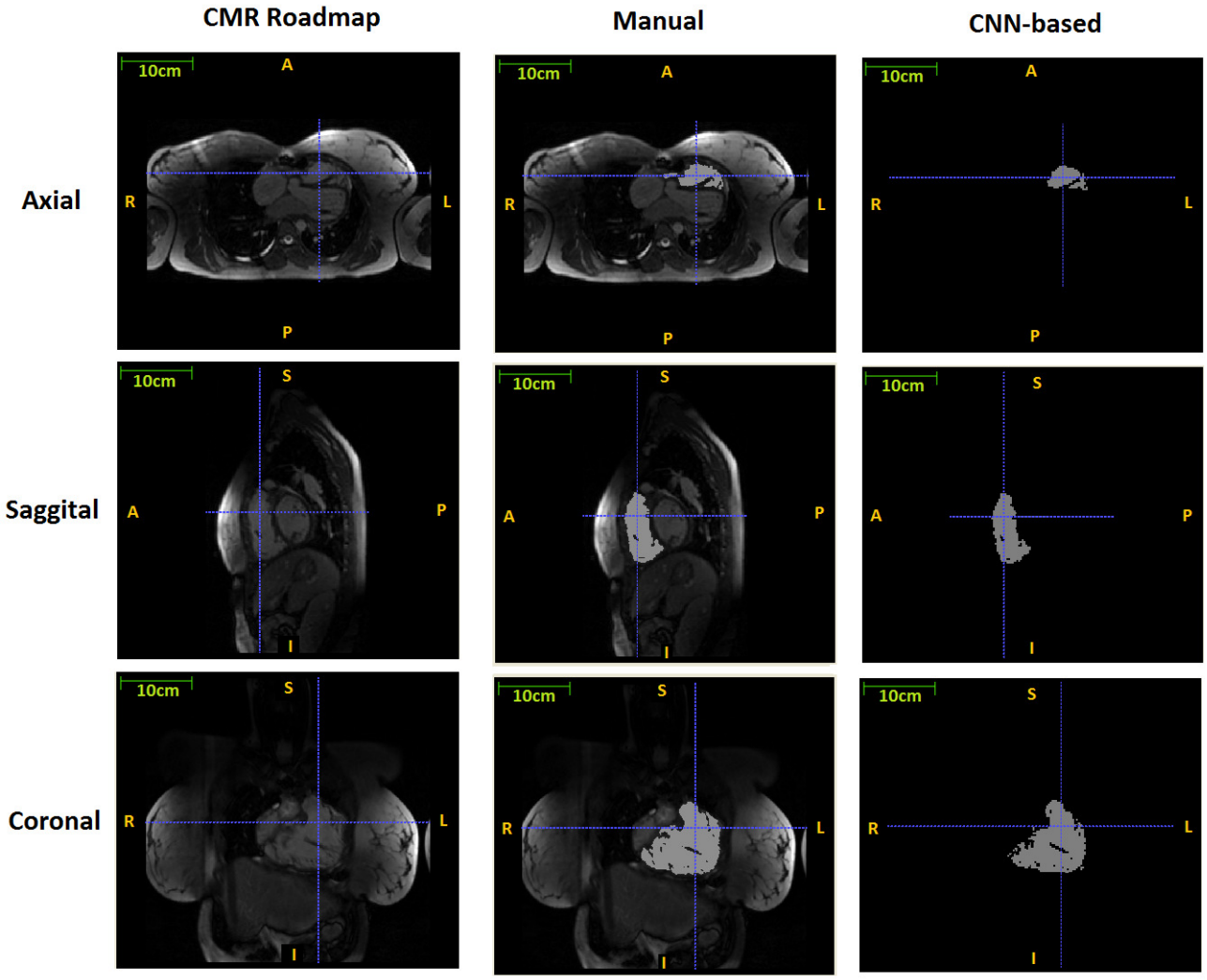

Figure 2. Qualitative evaluation of the proposed RV segmentation technique for one ToF patient. First column: The CMR images. Second column: The manual annotations super-imposed on the images. Third column: The segmentation masks acquired by the fully automatic CNN-based approach.

would like to thank Steve Collins and Yasin Karanfil for their help. The views expressed are those of the author(s) and not necessarily those of the NHS and the NIHR.

References

[1] S. Mendis, P. Puska, and B. Norrving, Global Atlas on cardiovascular disease prevention and control, World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization, 2011.

[2] C. Apitz, G. D. Webb, and A. N. Redington, Tetralogy of Fallot, Lancet. 2009;374(9699):1462–1471.

[3] S. E. Parker, C. T. Mai, M. A. Canfield, R. Rickard, Y. Wang, R. E. Meyer, P. Anderson, C. A. Mason, J. S. Collins, R. S. Kirby, A. Correa, and National Birth Defects Prevention Network, Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006, Birth Defects Res A Clin Mol Teratol. 2010;88(12):1008–1016.

[4] P. A. Davlouros, K. Niwa, G. Webb, and M. A. Gatzoulis, The right ventricle in congenital heart disease, Heart. 2006;92(Suppl 1): i27-i38.

[5] R. M. Wald, A. M. Valente, and A. Marelli, Heart failure in adult congenital heart disease: Emerging concepts with a focus on tetralogy of Fallot, Trends in Cardiovascular Medicine. 2015;25(5):422–432.

[6] C. Bussadori, G. Di Salvo, F. R. Pluchinotta, L. Piazza, G. Gaio, M. G. Russo, and M. Carminati, Evaluation of Right Ventricular Function in Adults with Congenital Heart Defects, Echocardiography. 2015;32(Suppl 1):S38–52.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, In: Advances in Neural Information Processing Systems, 2012, pp:1097–1105.

[8] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, 2016, arXiv preprint arXiv:1603.05959.

[9] K. He, X. Zhang, S. Ren, and J. Sun, Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision. 2015, pp. 1026–1034.

[10] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, 2013. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229.

[11] A. Choromanska, M. Henaff, M. Mathieu, G. B. Arous, and Y. LeCun, The Loss Surfaces of Multilayer Networks. 2015. Aistats 38, 192–204.

[12] K. Simonyan, and A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014. arXiv preprint arXiv:1409.1556.

[13] M. S. Krishnam, A. Tomasian, V. S. Deshpande, et al. Non-contrast 3D SSFP MR angiography of the whole chest using non-selective RF excitation over a large field of view: comparison with single-phase 3D contrast-enhanced MRA, Invest Radiol. 2008;43(6):411–420.

[14] P. Jhooti, J. Keegan, and D. N. Firmin, A fully automatic and highly efficient navigator gating technique for high-resolution free-breathing acquisitions: Continuously adaptive windowing strategy, Magn Reson Med. 2010;64(4):1015–1026.

[15] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, What is the best multi-stage architecture for object recognition? In: Computer Vision, 2009 IEEE 12th International Conference on. IEEE, pp. 2146–2153.

[16] S. Ioffe, and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. 2015. arXiv preprint arXiv:1502.03167.

[17] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

[18] T. Tieleman, and G. Hinton, 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning.

[19] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, 2013. On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th international conference on machine learning (ICML-13). pp. 1139–1147.

[20] C. Petitjean, M. A. Zuluaga, W. Bai, J. N. Dacher, D. Grosgeorge, J. Caudron, S. Ruan, I. B. Ayed, M. J. Cardoso, H. C. Chen, D. Jimenez-Carretero, M. J. Ledesma-Carbayo, C. Davatzikos, J. Doshi, G. Erus, O. M. O. Maier, C. M. S. Nambakhsh, Y. Ou, S. Ourselin, C. W. Peng, N. S. Peters, T. M. Peters, M. Rajchl, D. Rueckert, A. Santos, W. Shi, C. W. Wang, H. Wang, J. Yuan. Right ventricle segmentation from cardiac MRI: A collation study. Med Image Anal. 2015;19(1):187–202.

[21] O. Oktay, W. Bai, R. Guerrero, M. Rajchl, A. de Marvao, D. O'Regan, S. A. Cook, M. P. Heinrich, B. Glocker, and D. Rueckert, Stratified Decision Forests for Accurate Anatomical Landmark Localization in Cardiac Images, IEEE Transactions on Medical Imaging, 2016, (in press).

[22] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, and D. Comaniciu. 3D deep learning for efficient and robust landmark detection in volumetric data. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 565–572. Springer, 2015.

[23] M. A. Zuluaga, M. J. Cardoso, M. Modat, and S. Ourselin. Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion. In Functional Imaging and Modeling of the Heart (FIMH), pages 174–181. Springer, 2013.

[24] X. Zhuang, J. Shen, Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI, Med Image Anal. 2016;31:77–87.