Anatomical and Functional Analysis of Cardiac MRI scans using Deep Learning



Hao Xu

Christ Church College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2020

Abstract

Cardiac magnetic resonance (CMR) is the current gold standard imaging technique to assess myocardium anatomy and regional and global function of the heart. The accuracy and robustness of CMR imaging has been validated using manual segmentation. However, the quality of manually traced contours is critical for the assessment. Manual segmentation is tedious and time consuming and suffers from inter- and intra-observer variability. Therefore, much research work has been focused on automating the segmentation of CMR images. The rapid developments and applications of deep learning algorithms have dramatically improved the accuracy of automated segmentation methods and open the possibilities for alternative approaches.

In this thesis, we first investigate and show the importance of the segmentation method for calculating cardiac anatomy and function parameters using deep neural networks. We then develop and validate our segmentation neural network using both human and mouse CMR images for both short-axis (SAX) slices and long-axis (LAX) slices in order to utilise all available information for more meaningful and accurate evaluation. We then develop our three-dimensional (3D) neural network for reconstructing ventricular surfaces using sparse contours extracted from both SAX and LAX two-dimensional (2D) slices, with or without spatial consistency between them. Finally, we integrate our segmentation and reconstruction networks into a fully automated pipeline for the assessment of mouse heart using CMR images and achieved promising results by validating against our database with both normal and infarct hearts coming from two imaging centres.

Acknowledgments

I would like to express my gratitude to my supervisors Professor Vicente Grau and Professor and Jurgen E. Schneider, for their support and guidance throughout the project. They are always positive and encouraging and give me a great freedom of exploring the field. It has been a great experience working with them and I have enjoyed every moment even the challenging times.

I would also like to thank Professor Alistair Young, Lee-Anne Stork, Dr Ernesto Zacur and Joanna Koch-Paszkowski for their help in each stage of the project. Their knowledge, professional experiences and inspiring thoughts also open my mind and allow me to view the problem from different angles.

Many thanks to my parents, my wife, and all the families who supported and trusted me all along. Special thanks to my lovely cats for giving me a warmer home.

Contents

Chapter I. Introduction 1

1.1 Motivation 2

1.2 Contributions 3

1.3 Thesis Outline 4

1.4 List of Publications 7

Chapter II. Background 8

2.1 Cardiac Physiology 9

2.1.1 Human Heart 9

2.1.2 Mouse Heart 13

2.2 Cardiac Imaging 14

2.2.1 Echocardiography 15

2.2.2 Cardiac CT 16

2.2.3 Cardiac MR for Human Heart 17

2.2.4 CMR for Mouse Hearts 20

2.3 Quantification of Cardiac Function using CMR 21

2.4 Medical Image Analysis for Cardiac Function Estimation 23

2.4.1 Thresholding 24

2.4.2 Edge-based Method 25

2.4.3 Region-based Method 26

2.4.4 Pixel-based Classification 26

2.4.5 Deformable and Active Model 27

2.4.6 Atlas-based Method 28

2.4.7 Deep Learning in Cardiac Function Analysis 29

2.5 Reconstruction of LV Shape 39

2.5.1 Guide-point Modelling 40

2.5.2 3D Shape Reconstruction Network 42

2.6 Conclusions 44

Chapter III. Cardiac Function Estimation of the Left Ventricle 46

3.1 Introduction 47

3.2 Dataset 49

3.3 Method 52

3.3.1 Multi-mode Network 52

3.3.2 Calculation of LV indices from Segmentation 55

3.3.3 Data Augmentation 59

3.4 Experiments and Results 60

3.4.1 Implementation 60

3.4.2 Training Data Experiments 60

3.4.3 Analysis of Segmentation Results 63

3.5 Conclusions 68

Chapter IV. Segmentation Neural Network Development 71

4.1 TDM for Cardiac Cine-MR 72

4.1.1 Introduction 72

4.1.2 Dataset and Method..... 72

4.1.3 Experiments and Results..... 73

4.2 Comparison of Segmentation Networks..... 84

4.2.1 Introduction..... 84

4.2.2 Dataset and Method..... 84

4.2.3 Experiments and Results..... 88

4.3 Segmentation of Mouse CMR..... 91

4.3.1 Introduction..... 91

4.3.2 Dataset and Method..... 92

4.3.3 Experiments and Results..... 99

4.3.4 Segmentation of LAX slices of mouse hearts..... 101

4.4 Conclusions..... 103

Chapter V. A Neural Network for Reconstruction of 3D Shapes from Sparse Contours 106

5.1 Introduction..... 107

5.2 Dataset..... 110

5.3 Method..... 114

5.3.1 Method Overview..... 114

5.3.2 Baseline Neural Network..... 116

5.3.3 Adversarial Loss..... 118

5.3.4 Isosurfacing..... 120

5.4 Experiment and Results..... 120

5.4.1 Implementation 120

5.4.2 Statistical Shape Model 120

5.4.3 DETERMINE Dataset 124

5.4.4 Visual Inspection of Reconstructed Surfaces 131

5.5 Conclusions 133

Chapter VI. LV Function Analysis of the Mouse Heart 136

6.1 Introduction 137

6.2 Dataset 141

6.2.1 GPM Analysis 143

6.2.2 Data Preparation for 2D Image Segmentation 144

6.2.3 Data Preparation for 3D Shape Reconstruction 147

6.3 Method 151

6.3.1 Pipeline Overview 151

6.3.2 Segmentation Neural Network 152

6.3.3 Reconstruction Neural Network 153

6.3.4 Cardiac Function Estimation 154

6.4 Experiments and Results 155

6.4.1 Implementation 155

6.4.2 Intermediate Evaluation 156

6.4.2 LV Functional Metrics 158

6.5 Conclusions 167

Chapter VII. Conclusions and Future Work 171

7.1 Summary 172

7.2 Future Work 176

7.2.1 Short-term Future Work 176

7.2.2 Long-term Future Work 178

References 180

List of Figures

Figure 2. 1 Circulation of blood reproduced from [74]..... 9

Figure 2. 2 Phases and volume changes of cardiac cycle reproduced from [74]. 10

Figure 2. 3 Internal anatomy of the left atrium and ventricle reproduced from [54]. 12

Figure 2. 4 Comparison between mid-cavity slice of Sham and Infarct mice hearts reproduced from [36]. **S** indicates the septum, and **i** indicates the infarct site. 13

Figure 2. 5 Schematic diagram showing the cardiac anatomy in SAX, 2-CH and 4-Ch planes reproduced from [8] 14

Figure 2. 6 Examples of 2D echocardiogram images reproduced from [70]. The images are 2-CH (left) and 4-CH (right) views of the LV. The LV cavity is darker, and myocardium is brighter. The images are noisy with low contrast. 16

Figure 2. 7 Examples of CCTA images reproduced from [33]. The images are 2-CH (left) and 4-CH (right) views. LA and LV are visible in 2-CH view and LA, LV, RA, and RV are visible in 4-CH view. Unlike Figure 2.6, the cavities are brighter than myocardium. 17

Figure 2. 8 Examples of cardiac cine CMR images. The images are 2-CH (left) and 4-CH (right) views. LA and LV are visible in 2-CH view and LA, LV, RA, and RV are visible in 4-CH view. Like Figure 2.7, the cavities are brighter than myocardium while the contrast is not as strong. 18

Figure 2. 9 Examples of tagging MR images of SAX and LAX slices in ED and ES frames reproduced from [22]. The intersections of the tag lines of the SAX slices, and the intersections of the tag lines and the endocardial and epicardial borders are indicated

as red points. 19

Figure 2. 10 Examples of mouse CMR images for normal and infarct heart reproduced from [107]. Top rows show SAX images in ED (left) and ES frames (right) and the bottom rows show 4-CH LAX images in ED (left) and ES frames (right). Arrows indicate the MI site, which is extremely thin..... 21

Figure 2. 11 The 17-segment model shown in standard imaging planes reproduced from [8]. Basal segments are labelled as 1 to 6, mid-cavity segments are labelled as 7 to 12, apical segments are labelled as 13 to 16, and the apex segment is labelled as 17.... 22

Figure 2. 12 Bull's-eye display of the 17 segments reproduced from [8]..... 23

Figure 2. 13 AlexNet architecture reproduced from [68]. 31

Figure 2. 14 FCN architecture reproduced from [79]..... 32

Figure 2. 15 U-Net architecture reproduced form [102]. 32

Figure 2. 16 3D U-Net architecture reproduced form [28]..... 33

Figure 2. 17 Overview of generalised adversarial network framework. Both generator G and discriminator D are trained to minimize the difference between generated distribution G(x) and original data distribution Y. 35

Figure 2. 18 Overview of Cycle-GAN reproduced from [146]. Two generators, tow discriminators and two cycle-consistency are combined to form the Cycle-GAN mapping between X and Y..... 36

Figure 2. 19 Network architecture overview reproduced from [94]. 38

Figure 2. 20 Examples of model and guide points reproduced from [139]. White points are the guide points, and the mesh is the 3D finite element model. The LV cavity is

rendered as the grey mesh. 41

Figure 2. 21 ACNN baseline segmentation and super resolution network architecture reproduced from [91]. 42

Figure 2. 22 T-L network overview reproduced from [91]..... 43

Figure 2. 23 Network overview reproduced from [116]..... 44

Figure 3. 1 Graphical overview of different types of LV indices quantification methods. The methods on the left (as orange solid lines) and right/centre (as blue and green solid lines) correspond to DR and SB, respectively. The method on the right, including the dashed line on the bottom right corner, represents a novel modification of standard methods in which the LV indices are not directly calculated from the segmentation results, but predicted using an additional neural network..... 47

Figure 3. 2 The first frames of 20 subjects from the dataset. The quality of images and anatomies vary widely across the dataset. The hearts are normalised to a similar orientation and size. 50

Figure 3. 3 Regional LV indices illustration reproduced from [133]. A, AS, IS, I, IL and AL represent anterior, anteroseptal, inferoseptal, inferior, inferolateral and anterolateral respectively, and the six regional wall thicknesses are estimated for each of the segment. Dimensions 1 – 3 are the directional dimensions of the LV cavity between AS and IL, IS and AL, and A and I respectively. 51

Figure 3. 4 Neural network architecture of the multi-mode network. 54

Figure 3. 5 Contour points illustration. The red, green and blue points are endocardial,

middle and epicardial contour points. D1 and D2 represent the shortest distance from a middle contour point to the nearest endocardial and epicardial contour points. The point O is the centroid of the endocardial border of the binary mask. 57

Figure 3. 6 Illustration of phase estimation. The orange and green stars represent the ED and ES frames, and the corresponding frames are labelled 0s and 1s for diastolic and systolic respectively. 58

Figure 3. 7 Training data segmentation results for visual inspection. The red contours are the endocardial contours and the blue contours are the epicardial contours. 64

Figure 3. 8 Training data reference contours. The red contours are the endocardial contours and the blue contours are the epicardial contours. The segmentation results in Figure 3.7 are similar to the reference contours..... 65

Figure 3. 9 Box plots of dice metric values for areas enclosed by endocardial and epicardial contours and myocardial area..... 67

Figure 3. 10 Worst performing case of our segmentation method. The reference contours are completely off. 67

Figure 4. 1 Neural network architecture of Seg-TDM Mode. Multiple frames of the images are segmented simultaneously with our Seg-TDM Mode. 73

Figure 4. 2 Visualisation of endocardial surfaces for three representative cases (a), (b) and (c) are the iso-surfaces formed of three endocardial contours in time, generated from manual binary segmentations (top row) and the predicted binary segmentation using the Seg-TDM Model (bottom row).The red arrows indicate sudden changes of manual

segmentations in time, and the blue arrows point at corresponding places in Seg-TDM Mode results. 76

Figure 4. 3 Visualisation of epicardial surfaces for three representative cases (a), (b) and (c) are the iso-surfaces formed of three epicardial contours in time, generated from manual binary segmentations (top row) and the predicted binary segmentation using the Seg-TDM Model (bottom row). The red arrows indicate sudden changes of manual segmentations in time, and the blue arrows point at corresponding places in Seg-TDM Mode results. 76

Figure 4. 4 Manual segmentation and Seg-TDM segmentation comparison for case (a). 78

Figure 4. 5 Manual segmentation and Seg-TDM segmentation comparison for case (b). 79

Figure 4. 6 Manual segmentation and Seg-TDM segmentation comparison for case (c). 80

Figure 4. 7 Testing data (1 - 15) segmentation result visual inspection. The red contours are the endocardial contours and the blue contours are the epicardial contours. 82

Figure 4. 8 Testing data (16 - 30) segmentation result visual inspection. The red contours are the endocardial contours and the blue contours are the epicardial contours. 83

Figure 4. 9 Comparison between challenge datasets. (a) is an example from the MICCAI 2018 Left Ventricle Full Quantification Challenge dataset, which is pre-processed so that the heart is centred with the same orientation. (b) and (c) are examples from the MICCAI 2019 Left Ventricle Full Quantification Challenge dataset, which have different sizes and orientations of the heart. 85

Figure 4. 10 Randomly chosen 20 subjects from the dataset; the first frames are displayed. The position and orientation of the LV varies within the dataset. 86

Figure 4. 11 Neural network architecture of the U-Net variation. 87

Figure 4. 12 Neural network architecture of the VGG-16 variation. 87

Figure 4. 13 SAX slice images of the ED frame. The images are from basal to apical slices. The images are noisy compared to clinical CMR images. 93

Figure 4. 14 SAX slice images of the ES frame. The images are from basal to apical slices. The images are noisy compared to clinical CMR images. 93

Figure 4. 15 LAX slice images of both ED and ES frames. The images from left to right are: 2-CH ED, 2-CH ES, 4-CH ED and 4-CH ES. The images are noisy compared to clinical CMR images. 94

Figure 4. 16 Examples of LV reference contours: basal slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 95

Figure 4. 17 Examples of LV reference contours: basal slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 95

Figure 4. 18 Examples of LV reference contours: mid-cavity slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 96

Figure 4. 19 Examples of LV reference contours: mid-cavity slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 96

Figure 4. 20 Examples of LV reference contours: apical slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 96

Figure 4. 21 Examples of LV reference contours: apical slice, ES frame. The red contours

are the endocardial contours and the blue contours are the epicardial contours. 97
Figure 4. 22 Examples of LV reference contours: 2-CH slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 97
Figure 4. 23 Examples of LV reference contours: 2-CH slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 97
Figure 4. 24 Examples of LV reference contours: 4-CH slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 98
Figure 4. 25 Examples of LV reference contours: 4-CH slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours. 98
Figure 4. 26 Segmentation neural network architecture. 99
Figure 4. 27 Projected segmentation layers. Black, grey, and white pixels are corresponding to background, myocardium and blood pool respectively. 102

Figure 5. 1 Motion artefact examples with different level of misalignment. The red arrows indicate the positions of the misalignment, and from (a) to (c) the misalignment is less obvious. More details are shown in (d) and (e), where the green and orange lines indicate the cross-sections of the myocardium for each intersecting slices. 108
Figure 5. 2 Examples of reference LV shapes from the statistical shape model. The purple, blue and green surfaces represent LV epicardial, LV endocardial and RV endocardial surfaces. A great variation of plausible shapes is covered within the generated datasets. 111
Figure 5. 3 Examples of generated LV shapes from the DETERMINE data. The purple and

golden surfaces represent the epicardial and endocardial surfaces of the LV. The size and shape of the LV and the thickness of myocardium varies within the database. 113

Figure 5. 4 Method overview demonstrated using the statistical shape model. The illustration shows an individual slice, but all the datasets and operations are in 3D. The same architecture is implemented as both the baseline networks and reconstruction networks (the generator, which is trained simultaneously with adversarial loss function), and therefore the testing phases of them are identical...115

Figure 5. 5 Schematic representation of reconstruction neural network architecture. 117

Figure 5. 6 Schematic representation of discriminator neural network architecture.....119

Figure 5. 7 Data preparation and training progress overview. Red, green, blue, and light grey voxels are RV cavity, LV cavity, LV myocardium, and background. The dark grey voxels are foreground in binary volumes. 121

Figure 5. 8 Illustration of the effect of rigid transformation. The purple surfaces represent the reconstructed LV myocardium and the yellow surfaces represent the reference LV myocardium. The reconstructed surface aligns better with the reference surface after the rigid transformation. 122

Figure 5. 9 Validating loss of U-Net (or generator of GAN) in training progress. The experiment starts with training the U-Net with spatially consistent input data, which is the same in both the blue and orange lines. The second stage of the blue line illustrates the training progress after adding the adversarial loss to the consistent U-Net. The second stage of the orange line illustrates the training progress of taking the consistent U-Net parameters to initiate the training of the U-Net with misaligned input

data. The final stage of the orange line illustrates the training progress after adding the adversarial loss to the misaligned U-Net. The red arrows show a sudden increase of the loss function when the network is fine-tuned with a different type of dataset. 125

Figure 5. 10 Visualisation of the reconstructed LV surfaces for a spatially consistent input volume. The golden and purple regions represent LV cavity and LV myocardium in the input voulume. The predicted and reference surfaces are in green. 131

Figure 5. 11 Visualisation of the reconstructed LV surfaces for a misaligned input volume. The golden and purple regions represent LV cavity and LV myocardium in the input voulume. The predicted and reference surfaces are in green. 132

Figure 6. 1 Overview of the mouse CMR image segmentation method. We first segment the SAX slices using a variation of U-Net. The three slices represent the basal, mid-cavity and apical slices. A typical SAX stack usually consists of 10 to 15 slices. . 138

Figure 6. 2 In the second step, we project the segmentation of the SAX slices onto the LAX slice planes (usually two slices for each case: a 2-CH view and a 4-CH view.) This projected segmentation layer assists the segmentation of the LAX slices in the next step. 138

Figure 6. 3 Finally, we concatenate the image and projected segmentation layers and use our modified U-Net to segment the LAX slices. This completes the segmentation method. 139

Figure 6. 4 SAX slices (base to apex) and 4-CH LAX slice images of the ED frame.

Compared to Figure 4.13 to Figure 4.15, the number of SAX slice is smaller, and the contrast is better overall, while segmenting epicardial contours remains challenging. 142

Figure 6. 5 SAX slices (base to apex) and 4-CH LAX slice images of the ES frame. Compared to Figure 4.13 to Figure 4.15, the number of SAX slice is smaller, and the contrast is better overall, while segmenting epicardial contours remains challenging. 143

Figure 6. 6 Examples of LV reference contours (a) CIM model surfaces from which the contours were calculated, ED and ES frames. The red and blue surfaces are endocardial and epicardial surfaces, which are corresponding to the red (endocardial) and blue (epicardial) contours in Figure 6.7 to Figure 6.12..... 145

Figure 6. 7 Examples of LV reference contours (b): basal slice, ED frame..... 145

Figure 6. 8 Examples of LV reference contours (c): basal slice, ES frame. 145

Figure 6. 9 Examples of LV reference contours (d): mid-cavity slice, ED frame..... 146

Figure 6. 10 Examples of LV reference contours (e): mid-cavity slice, ES frame..... 146

Figure 6. 11 Examples of LV reference contours (f): apical slice, ED frame. 146

Figure 6. 12 Examples of LV reference contours (g): apical slice, ES frame. 146

Figure 6. 13 Examples of reference LV shapes from the Oxford cohort. The purple and golden surfaces represent the epicardial and endocardial surfaces of the LV. The size and shape of the LV and the thickness of myocardium varies within the database.148

Figure 6. 14 Examples of reference LV shapes from the Tubingen cohort. The purple and golden surfaces represent the epicardial and endocardial surfaces of the LV. The size

and shape of the LV and the thickness of myocardium varies within the database.149

Figure 6. 15 Examples of input sparse volumes. For the Oxford cohort, three combinations (both 2-CH and 4-CH views, 2-CH view only and 4-CH view only) of LAX slices are shown from (b) to (d). For the Tubingen cohort, only one LAX slice is available as shown in (a). The colours are the same as Figure 6.13 and Figure 6.14. 150

Figure 6. 16 Pipeline for LV anatomy and functional analysis of mouse heart. The SAX slice images are segmented with our variation of U-Net, and the results are used to produce the projections of SAX segmentation on the LAX image plane(s). The projection layer(s) are then combined with the corresponding LAX image(s) and passed through our LAX segmentation U-Net giving the LAX segmentation. The segmentation results from all slices are used to produce the sparse input volume for 3D shape reconstruction, and the prediction from our variation of 3D U-Net is finally converted to LV cavity and myocardium for anatomical and functional analysis. The point P represents the segmentation projection layer for LAX segmentation. 151

Figure 6. 17 Illustration of the segmentation network architecture. 153

Figure 6. 18 Illustration of architecture of neural network for LV shape reconstruction. 154

Figure 6. 19 Illustration of misalignment of sparse volume with suboptimal segmentation results. All combinations of SAX and LAX are shown, with both 2-CH and 4-CH views, 2-CH view, or 4-CH view. The colours are the same as Figure 6.13 and Figure 6.14. 157

Figure 6. 20 Bland-Altman plot and correlation plot for LV functional metrics: EDV.. 160

Figure 6. 21 Bland-Altman plot and correlation plot for LV functional metrics: ESV. . 161

Figure 6. 22 Bland-Altman plot and correlation plot for LV functional metrics: SV..... 162

Figure 6. 23 Bland-Altman plot and correlation plot for LV functional metrics: EF..... 163

Figure 6. 24 Bland-Altman plot and correlation plot for LV functional metrics: LVM. 164

List of Tables

Table 3. 1 Average results of compared methods. Values are shown as mean error ± standard deviation, followed by the correlation coefficient shown in brackets. 61

Table 4. 1 Average results of compared methods. Values are shown as mean error ± standard deviation, followed by the correlation coefficient shown in brackets. 74

Table 4. 2 Average results of compared methods. Values are shown as mean error ± standard deviation. The best results of overall metrics are shown in bold. For details about the A, B, C and D methods, refer to main text. 89

Table 5. 1 Mean and standard deviation of Dice score and HD of the statistical shape model experiments 122

Table 5. 2 Mean and standard deviation of Dice score and HD of DETERMINE dataset experiments 126

Table 5. 3 Mean and standard deviation of regional HD of DETERMINE dataset experiments 130

Table 6. 1 Percentage absolute error compared to intra- and inter-observer variability. 166

Chapter I. Introduction

In this chapter, I provide an overview of the thesis including the motivation, the contributions and the outline of the thesis.

1.1 Motivation

Cardiac magnetic resonance (CMR) is the current gold standard imaging technique to assess myocardium anatomy and regional and global function of the heart [61]. The accuracy and robustness of CMR imaging has been validated using single- and multi-centre datasets [61, 83, 113]. However, the quality of manually traced contours is critical for the assessment. Manual segmentation is tedious and time consuming and suffers from inter- and intra-observer variability. Therefore, much research work has been focused on automating the segmentation of CMR images [98].

From a medical imaging point of view, the argument above recognizes the importance of CMR imaging as the reference modality for non-invasive assessment of cardiac structure and function, supported by manual segmentation. From an image analysis perspective, it also points out the weaknesses of manual segmentation in order to address the importance of developing automatic segmentation methods for CMR images. The critical point of the underlying logic is the segmentation method.

For decades, there has been a significant gap between the performance of automated segmentation methods and manually traced contours, and thus the problem has remained an area of active research. However, the rapid developments and applications of deep learning algorithms are gradually closing the gap and the difference is becoming indistinguishable [15], which changes the focus of the question [26]. Alternative to segmentation-based methods, in which cardiac anatomy and function parameters are calculated from segmented contours, deep learning models have been proposed to directly regress these parameters for cardiac function estimation [5]. In this

thesis, we first investigate different approaches for cardiac function quantification using CMR images to show the importance of using a segmentation method (compared to direct regression) and develop our own segmentation neural networks for both human and mouse hearts.

Manual contouring is a commonly used segmentation method in clinical practice and is usually considered the gold-standard for the testing of automatic segmentation methods. Short-axis (SAX) slices are generally used for quantitative assessment of cardiac function, while long-axis (LAX) slices are not generally involved in these calculations. Therefore, most segmentation and cardiac functional analysis methods are focusing on two-dimensional (2D) SAX slices. In the second half of this thesis, we develop our method in three-dimensions (3D) and include LAX slices for segmentation and cardiac functional analysis.

1.2 Contributions

The contributions of this thesis are grouped into the following sections.

1. We investigated different types of quantitative analysis methods of the left ventricle (LV), including segmentation-based, direct regression, and a mixture of both under a single neural network framework.
2. We developed a segmentation method for human and mouse CMR images using a neural network. We investigated the impact of introducing temporal dynamic modelling and compared two widely used networks. For mouse heart LAX segmentation, we modified our network to incorporate information from

the SAX to further improve the segmentation accuracy.

3. We developed a neural network for reconstructing 3D shapes from sparse contours. The network is fully 3D and can accommodate CMR slices with different combinations and orientations. We investigated the effect of slice misalignment by testing the network on both spatially consistent and misaligned contours. We further investigated the impact of including LAX slices and introducing an adversarial loss.
4. We combined our segmentation and reconstruction networks to form a fully automatic pipeline for analysing the LV function of the mouse LV. Our pipeline takes CMR images as input and produces 2D and 3D segmentations as intermediate results. We then calculated the LV volumes and mass using the 3D segmentation. The performance of our pipeline achieves similar results to manual analysis when accounting inter- and intra-observer variability.

1.3 Thesis Outline

The remaining sections of the thesis are summarised as below.

Chapter 2 We introduce background knowledge of cardiac imaging and cardiac image analysis (including cardiac physiology, cardiac imaging and functional analysis), as well as the state-of-the-art deep learning algorithms for medical image analysis.

Chapter 3 We investigate two alternative approaches for the evaluation of cardiac function metrics. Direct prediction of volumes and mass from images (as compared to calculation from segmentation masks) is starting to show reasonable outcomes with the

introduction of deep neural networks, while segmentation methods have also improved dramatically with deep neural network applications. With a direct comparison between the two families of methods on a benchmark dataset, the MICCAI 2018 Left Ventricle Full Quantification Challenge dataset, our segmentation-based approach shows a better performance compared to our direct regression approach, and our best performing network also shows a competitive performance in the challenge by taking the third place.

Chapter 4 We summarise further development on our segmentation method, including investigating the importance of temporal dynamic modelling (TDM), comparing widely used network architectures, and designing a targeted network for segmenting mouse cardiac cine-MR images with both SAX and LAX slices. TDM is shown to be beneficial for sequential image analysis; however, due to the limitation of manual segmentation on individual frames, the measured improvement of introducing TDM in our experiment is small. For network architecture comparison, we implement variations of both U-Net and VGG-16 networks for the MICCAI 2019 Left Ventricle Full Quantification Challenge dataset, observing that the overall performance of the U-Net is better. Mouse CMR images are more challenging for segmentation comparing to human CMR images, and nevertheless the performance of our network on mid-cavity SAX mouse CMR images is comparable to human LV quantification challenges. The limited number of LAX slices constrains the performance of the network, and we propose to utilise information from the SAX slices for LAX segmentation and see clear improvements.

Chapter 5 We develop a novel approach to reconstruct 3D ventricular shapes from

sparse contours extracted from 2D CMR slices. SAX slices are currently used the most for reconstructing 3D shape of the ventricles, however, LAX slices have great advantage when reconstructing basal and apical regions compared to SAX slices, as well as correcting misalignments between SAX slices. Our method combines both SAX and LAX slices into a sparse 3D volume, and we train a 3D neural network to predict the dense volumetric representation of the ventricles, from which the LV shapes are reconstructed. We develop and validate our method using a statistical shape model and a clinical database for both spatially consistent and misaligned slices. We also investigate the effects of including LAX slices and introducing an adversarial loss for the reconstruction accuracy. For spatially consistent and misaligned slices, both including LAX and introducing the adversarial loss show better overall performance, while the improvements are more obvious for misaligned inputs.

Chapter 6 We investigate the performance of our fully automated pipeline for assessment of mouse LV using CMR images. The pipeline consists of two stages: segmenting 2D CMR images and reconstructing 3D structures from the segmentation results. In the segmentation stage, information from SAX slices is used to assist LAX segmentation as discussed in Chapter 4 to improve the segmentation accuracy. In the reconstruction stage, mouse CMR images are assumed to be spatially consistent and therefore our U-Net model is used to balance the performance and computational cost based on our experiments in Chapter 5. The quantitative analysis of our pipeline shows similar performance compared to inter- and intra-observer variability of manual segmentation in measuring LV volumes and mass of mouse heart.

Chapter 7 summarises the work presented in the thesis and discusses the potential future work.

1.4 List of Publications

- **Xu, H.**, Schneider, J.E. and Grau, V., 2018, September. Calculation of anatomical and functional metrics using deep learning in cardiac MRI: comparison between direct and segmentation-based estimation. *In International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 402-411). Springer, Cham.
- **Acero, J.C.**, **Xu, H**., Zacur, E., Schneider, J.E., Lamata, P., Bueno-Orovio, A. and Grau, V., 2019, October. Left Ventricle Quantification with Cardiac MRI: Deep Learning Meets Statistical Models of Deformation. *In International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 384-394). Springer, Cham.
- **Xu, H**., Zacur, E., Schneider, J.E. and Grau, V., 2019, June. Ventricle surface reconstruction from Cardiac MR slices using deep learning. *In International Conference on Functional Imaging and Modeling of the Heart* (pp. 342-351). Springer, Cham.

Chapter II. Background

In the thesis, we focus on assessing the left ventricle of the heart, and therefore we start by introducing the anatomy of the heart, including both the human heart and mouse heart. We use cine cardiac MR images to analyse the left ventricle. By comparing to echocardiography and cardiac CT, we address the advantages of cardiac MR for anatomical and functional analysis of the heart. Deep learning has shown promising results for medical image analysis and has been widely applied to cardiac MR images. Compared to traditional image processing algorithms, deep learning algorithms provide accurate segmentation and have robust performance with noisy images, which is especially important for analysing mouse cardiac MR images. The state-of-the-art deep learning methods mainly focus on short-axis slices and two-dimensional analysis. To effectively incorporate long-axis slices, we introduce a fully 3D network for reconstructing the left ventricle shapes.

In this chapter, the background of the thesis will be introduced in three sections, including cardiac physiology, cardiac imaging and functional analysis, and medical image analysis algorithms for cardiac function estimation.

2.1 Cardiac Physiology

2.1.1 Human Heart

The heart is a muscular pump driving the circulation of blood as shown in Figure 2.1. The right ventricle (RV) pumps blood through the lungs and the left ventricle (LV) pumps blood through the rest of the body. Blood comes back to the left atrium (LA) and the right atrium (RA), which serve as reservoirs to fill the ventricles [74].

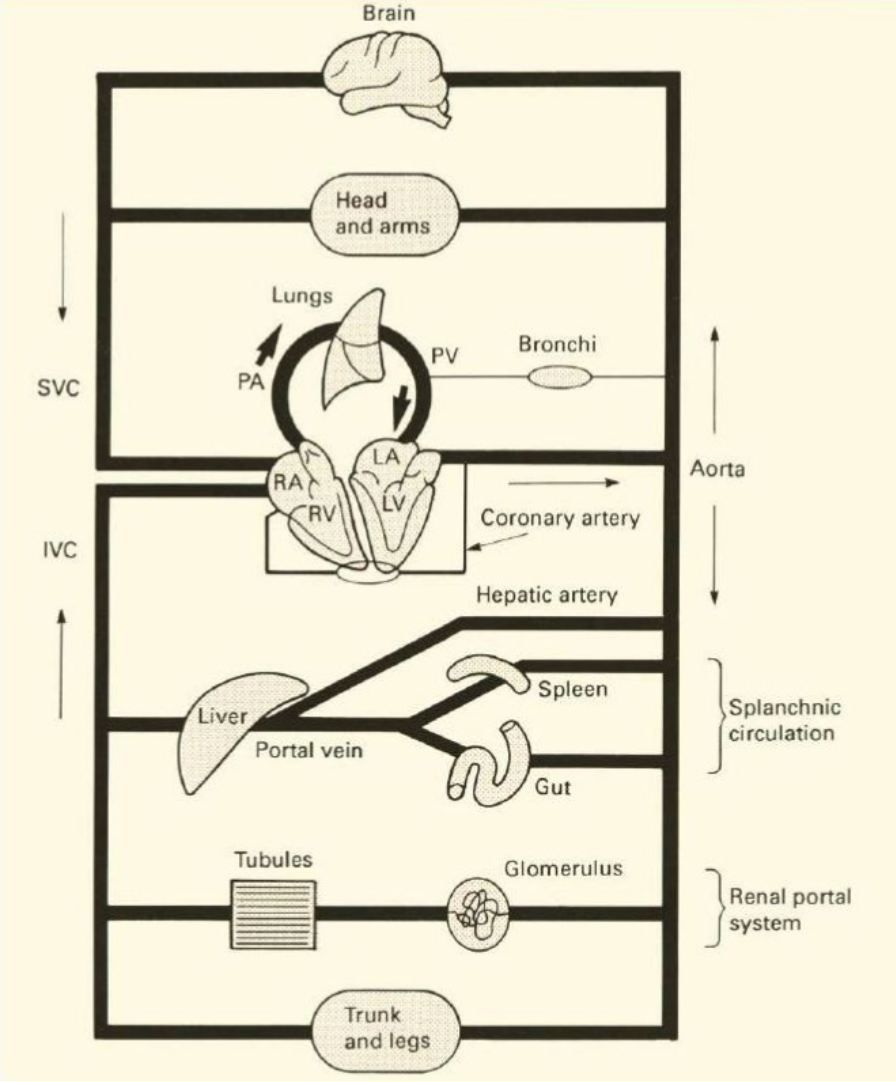


Figure 2. 1 Circulation of blood reproduced from [74]

In healthy hearts, both ventricles contract simultaneously and eject the same

volume of blood, however, the LV produces a much higher pressure to drive the blood through systemic circulation resulting in thicker muscular wall (or myocardium) comparing to the RV [74].

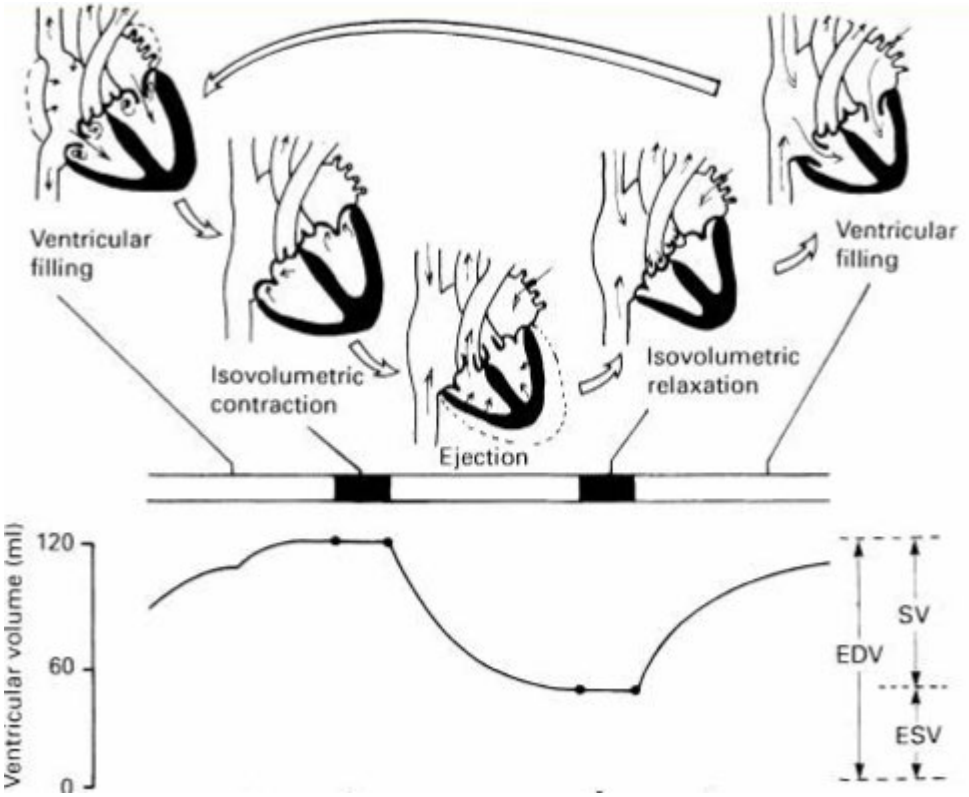


Figure 2. 2 Phases and volume changes of cardiac cycle reproduced from [74].

The cardiac cycle is formed by the sequential contractions of atria and ventricles with pressure and volume changes, and can be divided into four phases: ventricular filling, isovolumetric contraction, ejection and isovolumetric relaxation, as shown if Figure 2.2. Ventricles are in diastole (or relaxation) when refilling the ventricles, as well as the atria for the first two-thirds of the filling phase, and in the last third of the filling phase the atria contract to boost additional blood into the ventricles, which is important during exercise. The ventricle volume at the end of the filling phase is called the end-diastolic volume (EDV) and the corresponding pressure is called the end-diastolic

pressure (EDP). The ejection phase begins when ventricular pressure is above arterial pressure and the outflow valves are open. Active ventricular contraction stops at around two-thirds of the ejection-phase and a slow outflow continues due to momentum of the blood. The volume of ejected blood is called the stroke volume (SV), and the volume of residual blood in the ventricle is called end-systolic volume (ESV) with the corresponding pressure called end-systolic pressure (ESP). The ejection fraction (EF) is calculated as below:

$$EF = SV/EDV$$

which indicates how well the heart is functioning and is an important factor in the diagnosis and follow-up of multiple heart pathologies including heart failure. Isovolumetric phases take place during the brief period when both inflow and outflow valves are closed [38, 74].

Trabeculae carneae and papillary muscles are characteristic tissues present inside the ventricles that assist muscular contraction and relaxation, as shown in Figure 2.3. Trabeculae carneae are muscular ridges and columns located around the apical region of the ventricles, which strengthen the myocardial walls in order to increase the contraction force. Papillary muscles attach to the leaflets and the trabeculae carneae, in order to prevent the prolapse of the leaflets up into the atrium and prepare for the contraction of the ventricle [54].

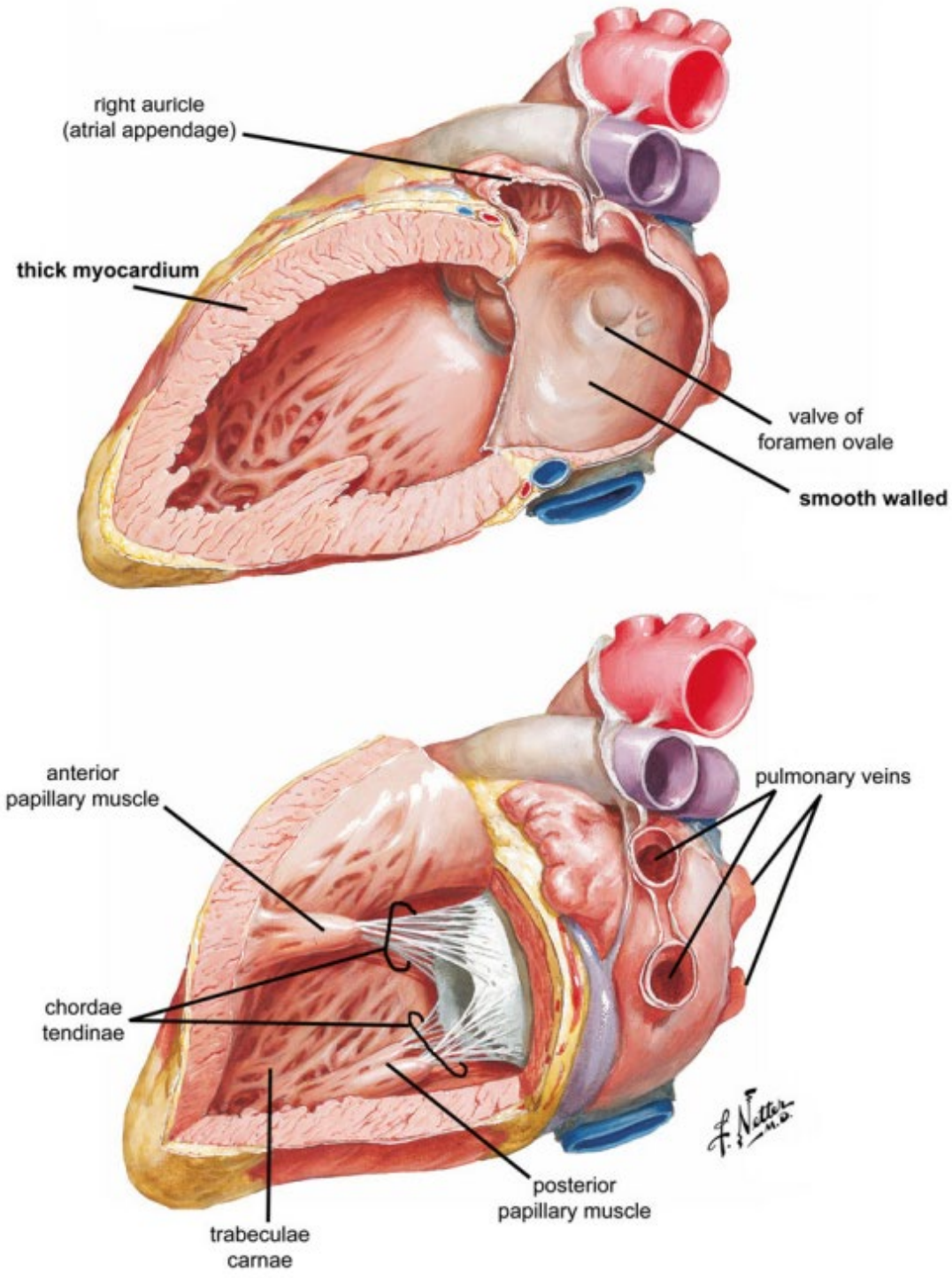


Figure 2. 3 Internal anatomy of the left atrium and ventricle reproduced from [54].

2.1.2 Mouse Heart

The mouse has shown its value for mammalian genetic modifications to mimic and analyse human cardiovascular pathology [36]. The ventricular structure of the mouse heart is similar to that of the human heart, while the differences between trabeculations and myocardium in mice are not clearly distinguishable [125]. For an adult mouse, the weight of the heart is between 150 mg to 180 mg, giving a heart to body ratio of around 5.5 mg/g, where the mouse body weight is approximately 0.05% of an adult human on average [36]. The energy required by the mouse heart per unit time and body mass [20] and the heart rate of mouse heart (ex vivo around 350 beats/min [47] and in vivo around 500 beats/min [32, 67]) are substantially higher than those of the human heart.

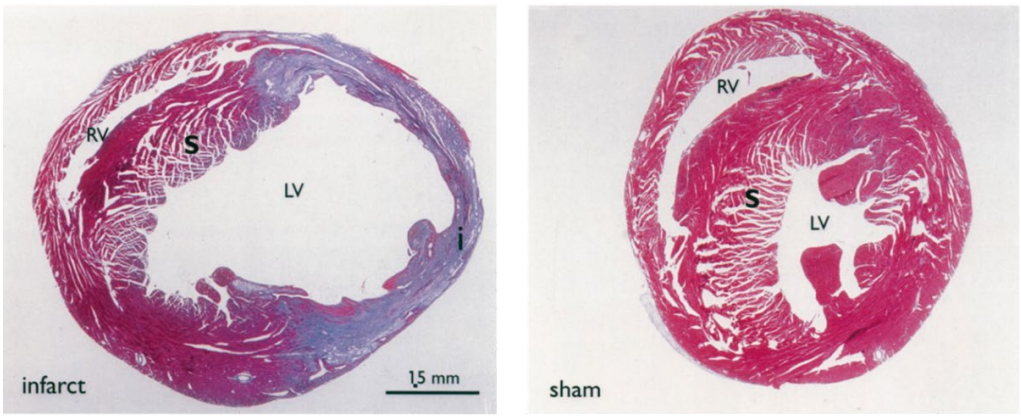


Figure 2. 4 Comparison between mid-cavity slice of Sham and Infarct mice hearts reproduced from [36]. ***S*** *indicates the septum, and* ***i*** *indicates the infarct site.*

Manipulations in cardiovascular system in other experimental animals can be adapted to mouse in theory, while in practice the size of the heart and the experience of the operator are the limiting factors [36]. Left ventricular hypertrophy can be established by banding the transverse aorta, and the increasing heart to body weight ratio is shown in [101]. Ligation of the left coronary artery can be used to introduce

myocardial infarction (MI) [81] contributing to our understanding of the functional and structural responses of MI, and a comparison between Sham and Infarct mice hearts is shown in Figure 2.4.

2.2 Cardiac Imaging

Echocardiography, nuclear cardiology, cardiac magnetic resonance (CMR) and cardiac computed tomography (CCT) are commonly used techniques in cardiovascular imaging. In this work we focus on analysing and evaluating the structure and pumping function of the heart, and therefore in the following sections a brief introduction to the related imaging modalities will be given discussing their advantages and limitations in practice.

For all cardiac imaging modalities, it is recommended to define, orient and display the heart using the long axis (LAX) of the LV, and select planes oriented at 90º relative to the LAX [8]. Short axis (SAX) view, vertical LAX view and horizontal LAX view are widely used in all imaging modalities, and SAX plane, two chamber (2-CH) plane and four chamber (4-CH) planes approximate these views respectively, as shown in Figure 2.5.

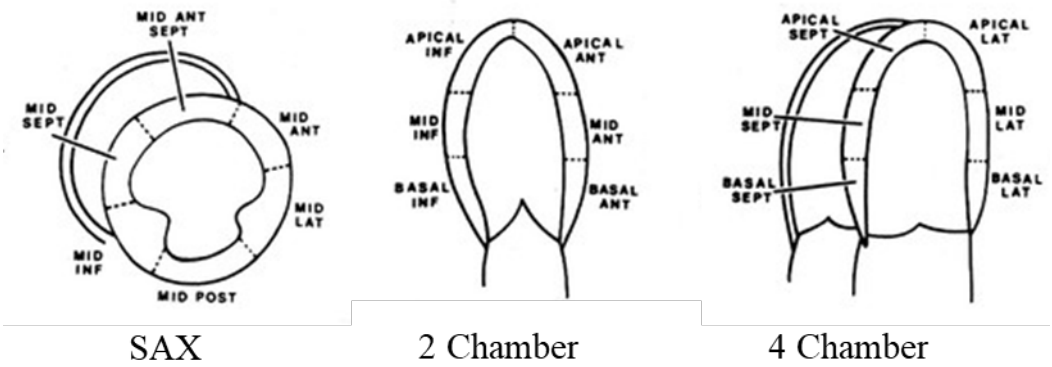


Figure 2. 5 Schematic diagram showing the cardiac anatomy in SAX, 2-CH and 4-Ch

planes reproduced from [8]

For mouse hearts, due to the similarity between ventricle structures compared to human hearts, similar imaging planes are adapted for assessment of the function of the LV including a stack of SAX planes, 2-CH and 4-CH LAX planes [108].

2.2.1 Echocardiography

Echocardiography is a widely available, low-cost, non-invasive and real-time imaging modality with no ionising radiation, and two-dimensional (2D) echocardiography in particular is currently the most commonly used imaging modality to answer questions about structural and functional normality of the ventricles, while the accuracy is limited by the geometric assumptions, the operator and the acoustic window, which allows sound waves to penetrate to the underlying heart and make the scan [7, 67, 109]. Three-dimensional (3D) echocardiography has been used in clinical practice and shown advantages comparing to 2D echocardiogram. 3D echocardiography can be used to estimate both ventricular volumes and requires no geometric assumptions, while quantification of the RV is challenging for 2D echocardiography [12]. When evaluating ventricular volumes, cardiac mass and ejection fraction, 3D echocardiography also shows comparable results to assessments from cardiac cine MR images, which is the current gold standard in non-invasive cardiac imaging [33, 88, 90]. The additional dimension also allows navigation and cropping of the 3D dataset to extract views that 2D echocardiography cannot produce, and similarly, re-aligned planes on 3D images guarantees the identification of maximum longitudinal axis of the ventricles [12].

However, 3D echocardiography also has more limitations compared to 2D echocardiography, including specific training required for operators to acquire and visualise images, breath holding of the patients, and good acoustic window [12]. Examples of 2D echocardiogram images are shown in Figure 2.6, and we can see that the contrasts are weak, and the images are noisy.

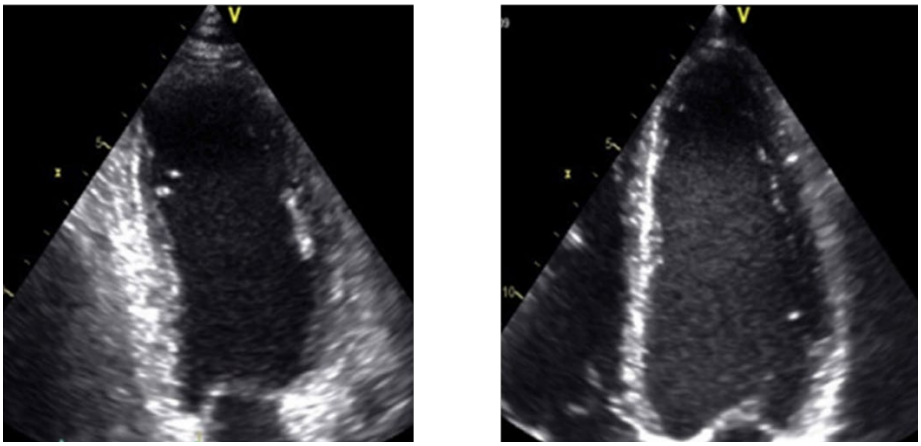


Figure 2. 6 Examples of 2D echocardiogram images reproduced from [70]. The images are 2-CH (left) and 4-CH (right) views of the LV. The LV cavity is darker, and myocardium is brighter. The images are noisy with low contrast.

2.2.2 Cardiac CT

Cardiac CT provides images with high spatial resolution for imaging small structures like coronary arteries and achieves high temporal resolution for motion-free imaging of the heart, producing spatially consistent volumetric images. Calcium scoring and coronary CT angiography (CCTA) are the most important clinical techniques for CCT, and the same dataset acquired for coronary artery imaging can be reconstructed to produce images with multiple frames (typically 20) per cardiac cycle and used for cardiac function analysis, including assessing of atrial and ventricular volumes,

myocardial wall thickness, tracking wall motion, and calculating EF [33]. Examples of cross-sectional views of CCTA images are shown in Figure 2.7, and the contrasts between different structures are strong. However, due to the radiation risk associated with CCT, criteria are outlined for its appropriate use [126].

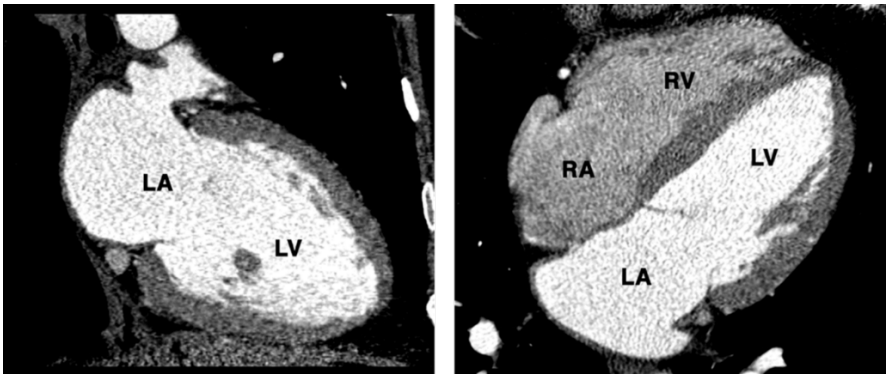


Figure 2. 7 Examples of CCTA images reproduced from [33]. The images are 2-CH (left) and 4-CH (right) views. LA and LV are visible in 2-CH view and LA, LV, RA, and RV are visible in 4-CH view. Unlike Figure 2.6, the cavities are brighter than myocardium.

2.2.3 Cardiac MR for Human Heart

CMR is currently the gold standard imaging modality to assess myocardial anatomy, regional and global function of the heart. Spin echo sequences are mostly used for anatomic imaging and tissue characterization, and current standard cardiac MR sequences are steady-state free precession (SSFP), which provide the best contrast between the blood pool and myocardium and are routinely used for cardiac function imaging [61]. To capture the entire LV, a total number of 10 SAX slices are needed on average, with a slice thickness around 10 mm [16].

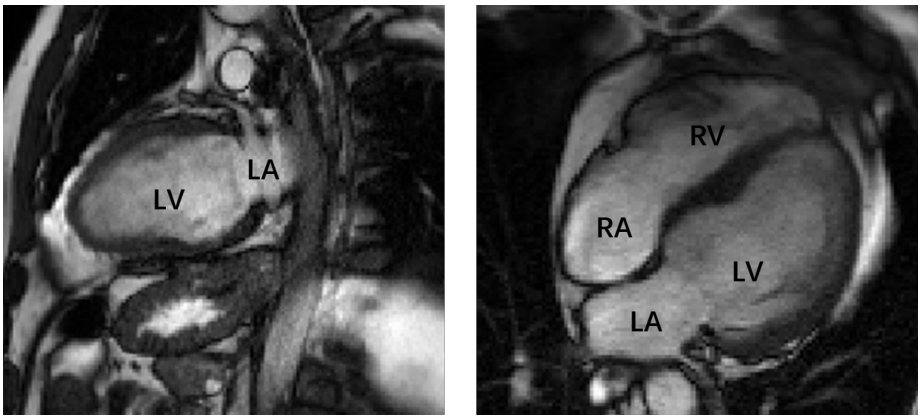


Figure 2. 8 Examples of cardiac cine CMR images. The images are 2-CH (left) and 4-CH (right) views. LA and LV are visible in 2-CH view and LA, LV, RA, and RV are visible in 4-CH view. Like Figure 2.7, the cavities are brighter than myocardium while the contrast is not as strong.

Comparing Figure 2.6, Figure 2.7, and Figure 2.8, CCTA images have the best image quality for segmentation, with good contrast and intensity homogeneity, and echocardiogram images are the most challenging, with low contrast and signal-to-noise ratio (SNR). The quality of CMR images is good for segmentation and there are more benefits of using CMR for cardiac function estimation. Unlike echocardiography, CMR does not rely on geometric assumptions for volume and mass quantification and can be applied to both normal and abnormal ventricles, in both of which it has shown to be accurate and reproducible [46]. Comparing to CCT imaging, CMR has no ionizing radiation on subjects and no side effects are known [61], but it is incompatible with ferromagnetic devices such as a pacemaker or an implantable cardioverter defibrillator (ICD).

Myocardial-perfusion MR images are used to assess the perfusion of myocardium by monitoring the first pass of a bolus injection of a contrast agent [57]. An ultrafast

acquisition is required, and the current imaging protocols allow around four slices to be scanned within one heartbeat, with T_1-weighted hybrid turbo GE (TGE) sequences with echoplanar imaging (EPI) readouts. The assessment of myocardial perfusion is important for evaluating patients with known or suspected ischemic heart disease. Compared to the reference catheter-based coronary angiography, perfusion MR is non-invasive and therefore helps to reduce the number of unnecessary invasive angiographic procedures [22]. Another advantage of perfusion MR imaging is that both ventricles and main coronary arteries are visible and therefore the relation between the distribution of myocardial ischemia and coronary artery territories can be identified [57].

Myocardial tagging is used for accurately analysing myocardial deformation [143], and with tagging sequences like complementary spatial modulation of magnetization (C-SPAMM) the whole cardiac cycle can be covered [40]. Normal, shear, and principal strains are quantified from tagging images for analysing myocardial deformation in 2D or 3D [122, 123]. With radio tagging, a combination of SAX slices and LAX slices divides the LV into small cuboidal volumes, and the corner points (or nodes) of the cuboids are the crossing points of the tags and endocardial and epicardial borders. Information from both SAX and LAX images are merged to calculate 3D spatial displacement of the nodes.

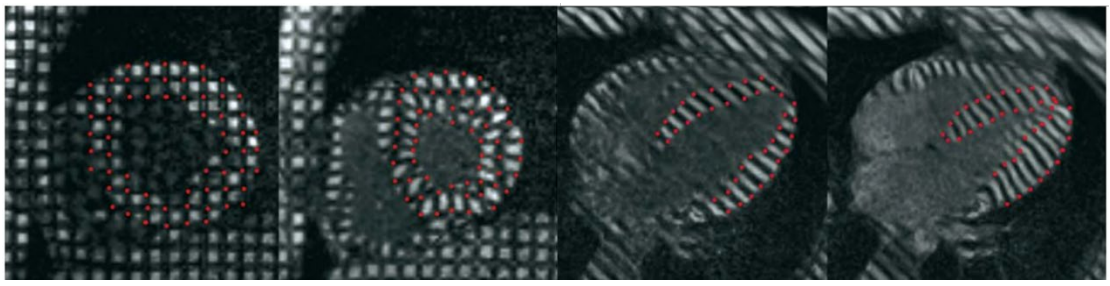


Figure 2. 9 Examples of tagging MR images of SAX and LAX slices in ED and ES frames

reproduced from [22]. The intersections of the tag lines of the SAX slices, and the intersections of the tag lines and the endocardial and epicardial borders are indicated as red points.

Examples of tagging MR SAX and LAX images are shown in Figure 2.9, and by tracking the grid intersection points of the SAX slices and endocardial and epicardial border points along the tags of the LAX slices (the red points shown in Figure 2.9), local myocardial deformation can be analysed. The regional and global deformation of the wall helps to understand how the myocardial fibres contract. Local deformation of myocardium can also be used to assess regional EF by quantifying the contribution of each cuboid to global ejection of the LV [21, 23].

For all MR image acquisitions, the patients are required to hold still in a small chamber with large noises during imaging, which could be difficult for young children or patients with claustrophobia. From an image analysis point of view, there is no spatial consistency preserved between slices through scanning, as cine of slices are acquired at different breath holds [16].

2.2.4 CMR for Mouse Hearts

CMR imaging for mice is more challenging than clinical practice due to the much smaller size of the heart and its short cycle. Fast, high-resolution cine MRI sequences have been specifically developed for in vivo ventricular analysis using a high-field MR system. The sequence in [107] is ECG triggered and respiratory gated, producing artefact-free cine images for accurate estimation of cardiac functional metrics in mice.

However, compared to human CMR images shown in Figure 2.8, the SNR is lower and the contrast between myocardium and surrounding tissues is also lower as shown in Figure 2.10. It is also worth noting, the infarct site indicated by the arrows are extremely thin and such wall thickness makes the segmentation even more challenging.

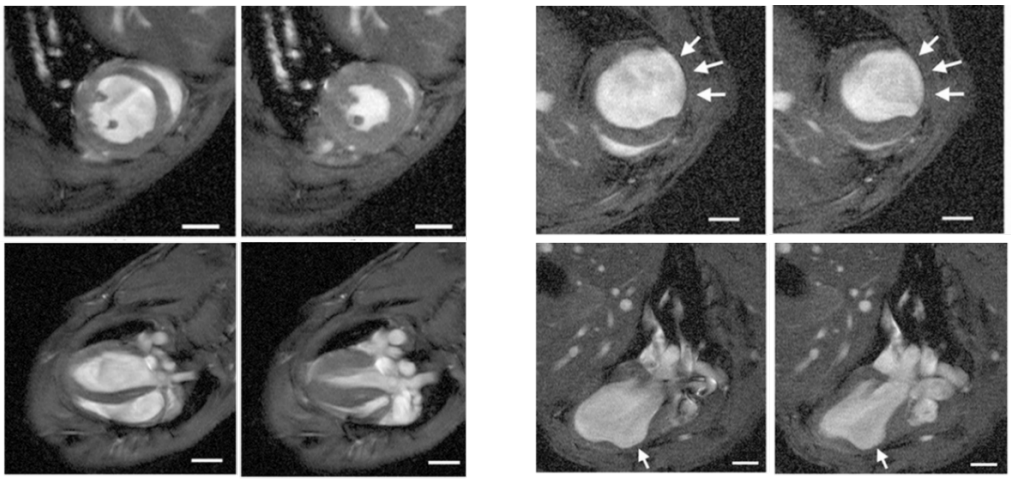


Figure 2. 10 Examples of mouse CMR images for normal and infarct heart reproduced from [107]. Top rows show SAX images in ED (left) and ES frames (right) and the bottom rows show 4-CH LAX images in ED (left) and ES frames (right). Arrows indicate the MI site, which is extremely thin.

2.3 Quantification of Cardiac Function using CMR

The accuracy and reproducibility of cardiac function assessment from CMR has been shown in multiple studies using data from single or multiple centres for human [62, 83, 113] and mouse heart [105, 107]. In this work, we focus on cardiac function assessment of the LV, including estimating global metrics such as volumes and mass from cine MR images; calculating the EF from those volumes; and calculating regional metrics such as wall thickness. For regional quantification, we use the standard 17 segments

established by the American Heart Association (AHA) [8] as illustrated in Figure 2.11.

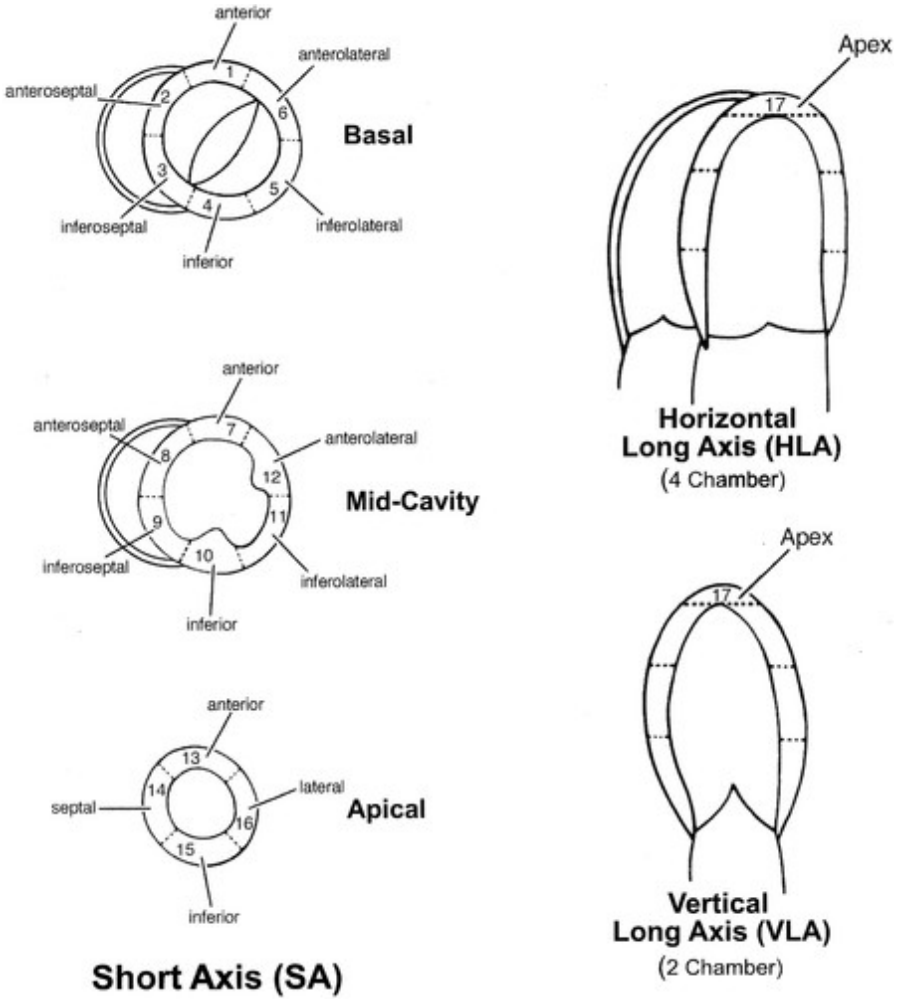


Figure 2. 11 The 17-segment model shown in standard imaging planes reproduced from [8]. Basal segments are labelled as 1 to 6, mid-cavity segments are labelled as 7 to 12, apical segments are labelled as 13 to 16, and the apex segment is labelled as 17.

In this 17-segment model, the heart is first divided into basal, mid-cavity, apical regions and the apical cap along the LAX. This is done by dividing the ventricle into equal thirds after removing the apical cap, which is defined as the area of myocardium beyond the end of the LV cavity. For basal and mid-cavity slices, six segments comprising a 60º arc each are defined, and apical slices are divided into four segments covering 90º each. The name and location of the 17 segments are summarised in the

bull's-eye display shown in Figure 2.12.

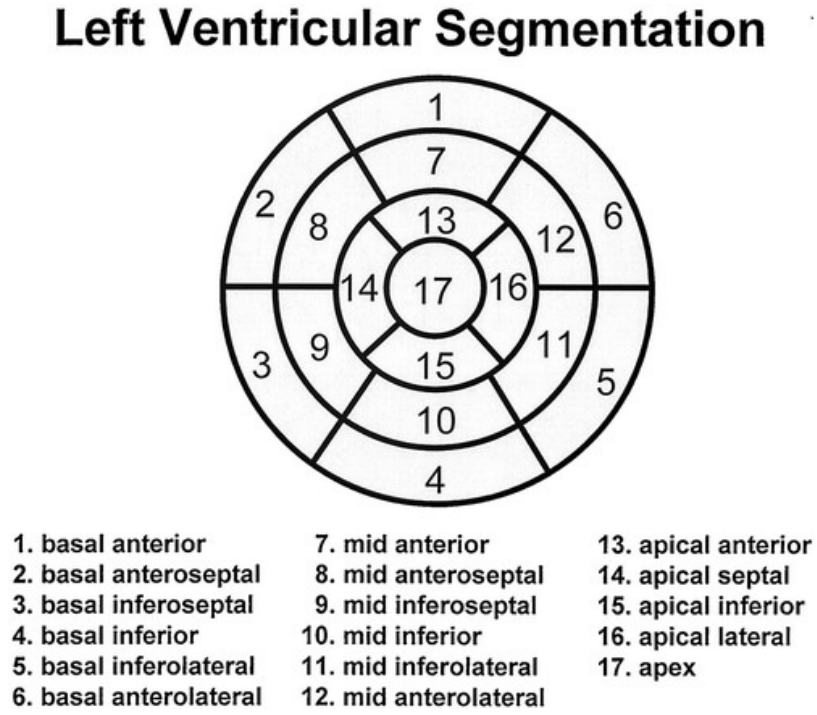


Figure 2. 12 Bull's-eye display of the 17 segments reproduced from [8].

Volumes are commonly calculated from the stack of SAX slices. For this purpose, variations of Simpson's rule (also known as disks summation) are usually applied, using the tracing of the interface between the cavity and the compacted myocardium (also known as endocardium) [70]. In the following chapters, papillary muscle and trabeculation are included in the cavities.

To calculate the myocardial mass, both the outer surface of the myocardium (epicardium) and the endocardium are traced. The volume of the myocardium is calculated as the one enclosed between endocardium and epicardium. This is multiplied by the density of myocardium (1.05 g/ml) to produce an estimated value of mass [70].

2.4 Medical Image Analysis for Cardiac Function Estimation

Segmentation of SAX slice is critical for accurately estimate cardiac function from CMR

images, and medical image segmentation methods can be classified into one of the following categories: thresholding, edge-based method, region-based method, pixel-based classification, deformable and active model, and atlas-based method [19, 99]. In the first half of this section, a brief introduction of each segmentation method will be given by discussing their advantages and limitations in practice. The recent development of deep learning algorithms has a great impact on medical image analysis, and therefore the second half of this section will discuss segmentation neural networks in more details as they are closely related to our methods in the subsequent chapters.

2.4.1 Thresholding

Thresholding using Gray-level value is a simple yet effective segmentation method, and by applying this approach the image is segmented into objects with different labels based on intensity thresholds. A single threshold results in a binary image with label 1 corresponding to the object and label 0 corresponding to the background, however, the images are usually segmented into multiple regions, e.g. LV myocardium and LV cavity for CMR images.

Features of the histogram can be used to select thresholds, such as the minimum value between histograms peaks or the point with maximum curvature [1]. The threshold can also be calculated by maximising or minimising a merit function though optimal thresholding methods, and a good example of non-parametric approach is proposed by Otsu [93], which finds the threshold by maximising the ratio of the between-class variance to the total variance. The shape of the histogram can also be assumed to follow a certain distribution and the problem is then reformulated into estimating the distribution parameters from available data, i.e. parametric

optimal thresholding.

Typical thresholding methods do not take spatial related features from the images into consideration, and therefore they are sensitive to noise and inhomogeneous intensity. These two artifacts can be seen in CMR images and make the threshold selection difficult.

2.4.2 Edge-based Method

Edge-based segmentation methods extract structures based on edges detected from the images using edge operators (such as Sobel, Prewitt, and Canny operators). These edges form a series of boundary candidates, which are then linked to produce accurate boundaries of the structures. Basic boundary detection algorithms include border tracing, graph searching, and dynamic programming.

Border tracing algorithm takes both the edge magnitude and edge orientation calculated from the edge detector to identify the next pixel on the border from the neighbourhood of a given pixel. The algorithm can identify the best candidate from the nearest connected pixels, however, when none of the neighbours satisfies the criteria the algorithm is stopped, making it difficult to deal with complex structures, such as papillary muscles around LV endocardial borders. Graph searching [112] and dynamic programming [17] are more robust algorithms, which aim to identify the optimal path in a graph as the boundary. The difficulty of such approaches is the design of cost functions, which usually include gradient magnitude and direction terms, producing robust and accurate result.

2.4.3 Region-based Method

Region-based segmentation method can be one of the following categories: region growing, region splitting and split-and-merge. Region growing approaches extract image regions with connected pixels satisfised some predefined criteria, and the simplest region growing method requires a starting point (or seed) per region, such as [4] proposed by Adams and Bischof. Region splitting approaches are opposite to region growing. Starting from the whole image, if the homogeneity criteria are not met, the image is split into four sub-images, and the procedure continues until all the sub-images meet the criteria. A key drawback of region splitting is that the final neighbouring sub-images may have identical properties, and split-and-merge approach [52] is proposed to address that issue.

Like thresholding methods, region-based approaches are based on homogeneity property of the images, and therefore also suffer from artifacts of CMR images.

2.4.4 Pixel-based Classification

Traditional pixel-based classification is proposed for multiple-spectral images, e.g. T_1-weighted and T_2-weighted MR images of the same physical space. The intensity features are different depending on the acquisition sequences, and therefore each pixel can be represented as a multi-dimensional feature vector. Multiple features could be obtained from multiple scans naturally, while it could also be calculated from a single scan through feature extraction. Classification and clustering algorithms are then applied to the feature vectors for labelling pixels.

Spatial related information could be incorporated through feature extraction or classification algorithms (such as contextual classifiers [53, 104]) to reduce the noise of the

classified images; however, the spatial information remains local. Deep learning algorithms can be considered as pixel-based classification methods, with feature extraction and classification stages combined within multiple steps of convolutional operations and non-linear operations. Compared to traditional pixel-based classification methods, a great advantage of deep learning algorithms is that they effectively extend the field of view for feature extraction and therefore include large variations of features representing both local and global information for classification.

2.4.5 Deformable and Active Model

It is difficult for image-driven segmentation techniques, including thresholding, edge-based method, region-based method, and classical pixel-based classification, to handle medical images with noise and artifacts, and deformable models have been proposed to produce more robust and accurate segmentation results by incorporating prior information. Deformable models are curves driven by internal and external forces with the image domain. The internal forces are defined by the curve itself to keep it smooth and the external forces are calculated to drive the curve to the desired image features [99].

There are two types of deformable models: parametric deformable models [29, 63] and geometric deformable models [24, 84]. Parametric deformable models are represented in their parametric forms during deformation, and therefore a fast and direct interaction with the model can be implemented. However, it is difficult for parametric deformable models to split or merge during deformation. Geometric deformable models are based on curve evolution and level set method, and the curves are represented as a level set of a higher-dimensional function, which

handles the change of topology very easily.

The extracted boundaries are constrained to be smooth and prior information of the structure can also be incorporated, giving deformable models robustness to image noise and gaps between boundaries. The implementation of deformable models is not based on pixels, and therefore subpixel accuracy can be achieved, which is important for MR images with large pixel spacing. However, these models require manual interactions to place initial contour and are difficult to progress into boundary concavities [128]. To achieve good performance of deformable models, it is also important to choose appropriate parameters [99].

Deformable models have been widely applied for CMR segmentation [39, 82, 124]; however, there are two fundamental issues associated to these methods. The first problem is caused by the weak gradients between myocardium and papillary muscles. Post processing with morphological operations [39] or using circular shape prior [124] can partially solve the problem, but for systolic frames the issue remains. The second drawback of deformable model method comes from the smooth constrains and prior information, which might not cover shape variations caused by diseases [82].

2.4.6 Atlas-based Method

Atlas-guided segmentation uses a single image delineated by an expert in the early days, and the classical atlas-based method reformulated the segmentation task to an image registration task [99]. The atlas image is registered to the novel image, and by applying the resulted mapping, the segmentation labels are transferred to the novel image pixels. A single atlas is then found to be insufficient to capture wide variation of anatomies [34], and therefore several atlases are

used, and probabilistic atlas-based segmentation is first applied [41, 95]. The atlas is built under a single atlas coordinate frame defined through co-registration of all the training images, and the probability of a particular label presenting at a given location is then calculated. A main advantage of probabilistic atlas-based approach is the computation efficiency, as only one registration step is required per novel image. An alternative group of approaches, multi-atlas segmentation methods, is then proposed [50, 66], which does not summarise the atlas in one probabilistic model and instead uses each atlas independently, i.e. a pairwise registration is applied between the novel image and each atlas image. A typical example takes the registration results to transfer the atlas labels to the novel image coordinates, and the most frequent label is selected through majority voting.

The biggest limitation of multi-atlas segmentation methods is the substantial amount of computational cost, and therefore it has not been widely applied in clinical practice. Most of the applications of multi-atlas segmentation have been used for brain MR images, as the registration algorithms achieve good performance, and registration is less effective in other body parts [55].

2.4.7 Deep Learning in Cardiac Function Analysis

Deep learning algorithm for CMR image segmentation can be classified as a pixel-based classification approach. Different from classical pixel-based classification methods, deep learning algorithm typically combines feature extraction and classification in a single neural network, with parameters learnt from a large training dataset, and incorporate both local and global spatial features. Deep learning algorithms

do not require homogeneity assumption and have robust performance with noisy images compared with other image-driven segmentation methods, which is particularly important for segmenting mouse CMR images. Compared to deformable models, deep learning algorithms do not include explicit smoothness terms or strong shape prior, and therefore are more flexible when segmenting pathological cases. Deep learning algorithms usually consume large amount of computation resources in the training stage, but the computational cost in the inferencing stage is extremely small. Compared to multi-atlas segmentation methods, deep learning algorithm has more potential for clinical applications. Deep learning algorithms have great advantages to traditional segmentation methods, and a wide range of deep learning algorithms for segmenting CMR images will be discussed in the following sections.

2.4.7.1 Supervised and Unsupervised Learning

Supervised learning is the most common form of machine learning and is typically used for classification or regression. To train the machine, a large dataset is required with correct labels or reference values assigned for each case, and the internal adjustable parameters (also known as weights) in the machine are optimised with errors between predictions of the machine and the reference [71].

Unsupervised learning is commonly used for clustering or identifying hidden patterns. Compared to supervised learning, unsupervised learning has no requirement of labelling the dataset. The importance of unsupervised learning is overshadowed by the success of supervised learning However, unsupervised learning is critical for discovering the structure of the world

and far more important in the long term [71].

2.4.7.2 Convolutional Neural Networks

Convolutional layers are the basis of CNNs and the output of neurons are calculated with local regions of the input. The computation is a dot product between the weights and a small region of the input for each neuron output, and it largely reduces the number of parameters (the weights) and increases the efficiency of the network implementation.

CNN showed outstanding performance in a natural image classification competition called ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [68], using the network architecture referred to as AlexNet and shown in Figure 2.13. This success started a new era of ANNs applied to image analysis and the number of layers has increased over the years (e.g. VGG net, GoogLeNet and ResNet) [49, 111, 114]. These deeper (i.e., with more sequential layers) networks have been shown to improve performance, and have been fundamental in the analysis of questions such as convolutional kernel sizes and the advantages of skip connections, providing helpful guidance when designing network architectures.

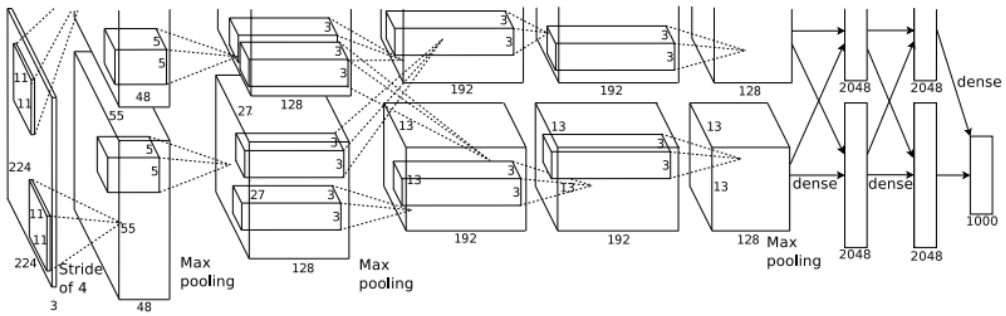


Figure 2. 13 AlexNet architecture reproduced from [68].

Fully convolutional network (FCN) [79] is a type of CNN targeting sematic

segmentation. The FCN makes use of the widely used encoder/decoder structure, even though the decoder of the original FCN (with a one-step expansion) is much shallower than the encoder (with multiple convolutional layers and down-sampling stages) as shown in Figure 2.14.

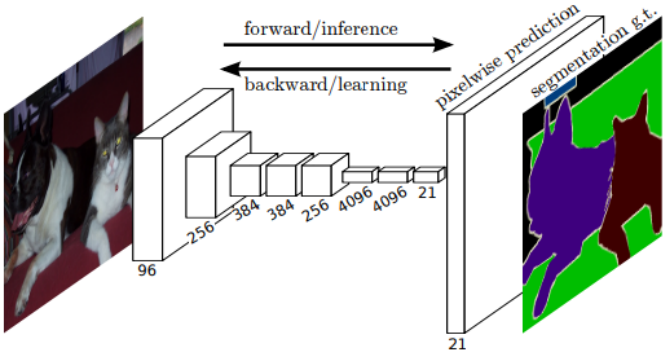


Figure 2. 14 FCN architecture reproduced from [79].

Variations of FCNs have been used for medical image segmentation. Among these, the U-Net [102] is the most successful one. U-Nets have an evenly distributed network structure (as shown in Figure 2.15) with the same depth of encoder and decoder. The layers with the same resolution are connected with skip layers allowing fine image features to pass through the network to achieve more accurate segmentation.

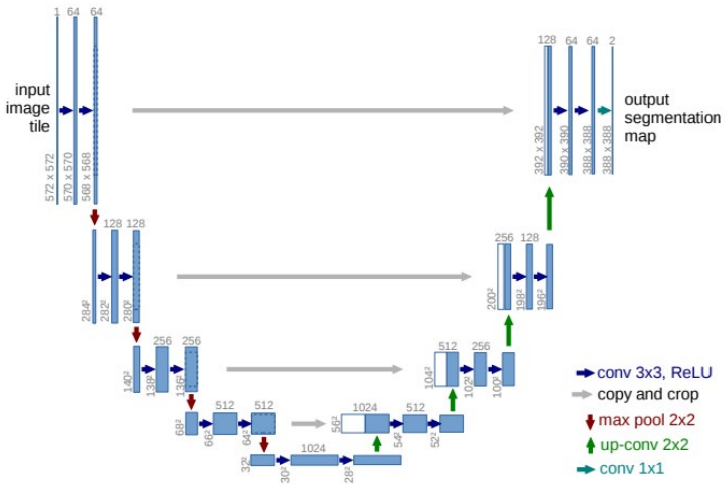


Figure 2. 15 U-Net architecture reproduced form [102].

For CMR image analysis, while the original FCN has also been implemented for

SAX slice image segmentation [117], currently the U-Net is the most widely used network. Recurrent layers have also been combined with a U-Net variation forming a recurrent fully convolutional network (RFCN) [100] to transmit information between SAX slice images during segmentation and thus taking advantages of special dependency of SAX slices. Recurrent layers are commonly used for learning temporal and spatial relationships between instances of sequences; however, due to the vanishing gradient issue training is difficult [97].To overcome this issue, two implementations of the recurrent function are widely used when training recurrent networks, which are long short-term memory (LSTM) [51] and Gated Recurrent Units (GRUs) [27].

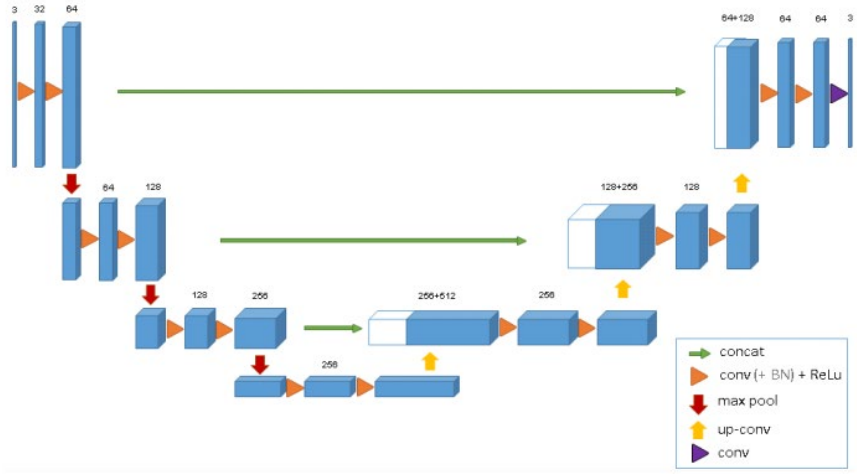


Figure 2. 16 3D U-Net architecture reproduced form [28].

A 3D U-Net was proposed more recently for volumetric segmentation [28] as shown in Figure 2.16, with the scale of the network reduced compared to 2D U-Nets. The reason for reducing the number of down-sampling and up-sampling stages as well as the number of convolutional kernels for each layer is the limited memory size in a GPU for storing features and their derivatives used for updating network parameters with back-propagation. That is also the main reason why four-dimensional (4D)

segmentation networks, which can be used to segment complete cine CMR images as one instance and have the potential of learning cardiac motion, are not available for now.

While segmentation remains the key stage for cardiac function assessment, estimating cardiac function metrics including cavity volumes and myocardial mass and wall thickness directly from images is also possible with CNN [36, 133], and sometimes serves as part of multitask learning network where segmentation remains a critical building block of the network [77, 78, 129].

2.4.7.3 Generative Models

The goal of generative models is to generate instances belonging to a distribution of data usually from a lower dimensional latent space. By successfully achieving this goal the data distribution is also learnt by the generative model, which is fundamental for generation stage. In recent years, deep generative models have shown great success. Among these, generative adversarial networks (GAN) [45] and variational autoencoders (VAE) [35] are the most widely used. Different from deep neural networks for classification or segmentation tasks as shown in previous section, generative models do not require labelling of the dataset, and are therefore unsupervised learning methods. The combination of the two is known as semi-supervised learning.

GAN

GANs [45] were first proposed to sidestep the difficulty raised in maximum likelihood estimation when learning the probability density of the data, by introducing a

discriminative model that learns to distinguish samples between the generative model distribution and the data distribution. The competition between the generative model (also known as generator, *G*) and the discriminative model (also known as discriminator, *D*) improves the performance of both models, improving the fit of the generative model distribution to the data distribution.

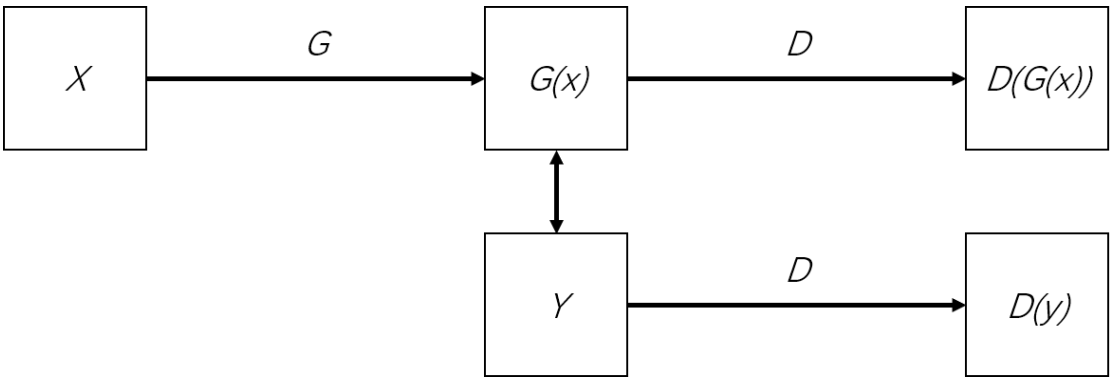


Figure 2. 17 Overview of generalised adversarial network framework. Both generator G and discriminator D are trained to minimize the difference between generated distribution G(x) and original data distribution Y.

By generalising the idea of generative process to a translation from data distribution *X* to data distribution *Y*, as shown in Figure 2.17, the generator $G(x; \theta_g)$ takes sample *x* from data space *X* as input and maps it to the data space *Y*, and the discriminator $D(y; \theta_d)$ takes samples from both generated distribution and original distribution of Y as input and predicts if the input is generated or original, where θ_g and θ_d are the parameters of the generator and discriminator which are usually implemented using deep neural networks.

The general idea of adversarial network framework is that *D(y)* tries to be near 1, and at the same time *D* tries to make *D(G(x))* near 0 and *G* tries to make *D(G(x))* near 1. The overall value function of this two-player minimax game can be expressed as *V(G,*

D):

$$\min_G \max_D V(D,G) = E_{y\sim Y}[\log D(y)] + E_{x\sim X}\left[\log\left(1 - D\left(G(x)\right)\right)\right],$$

which implies the loss function of the *G* is just the negation of loss function of *D*.

In practice, the game is implemented in an iterative approach, alternating between k steps of optimizing D and one step of optimizing G, which results in keeping D at near its optimal solution as long as G changes slowly enough [45]. At the start of the training process, D can easily distinguish generated samples from original data of Y, and therefore $\log\left(1 - D\left(G(x)\right)\right)$ saturates. It is recommended to train *G* to maximize $\log D\left(G(x)\right)$ rather than minimising $\log\left(1 - D\left(G(x)\right)\right)$, and that provides much stronger gradients while the new objective function results in the same fixe point [45].

The Cycle-consistent adversarial network (Cycle-GAN) has proposed for unpaired image-to-image translation [146]. The goal of Cycle-GAN is no longer mapping from latent space to real images, instead the goal becomes mapping from one data distribution to another in both directions. One application of the network is style transfer, and one-directional mapping using CNN has been proposed to generate artistic style images [43].

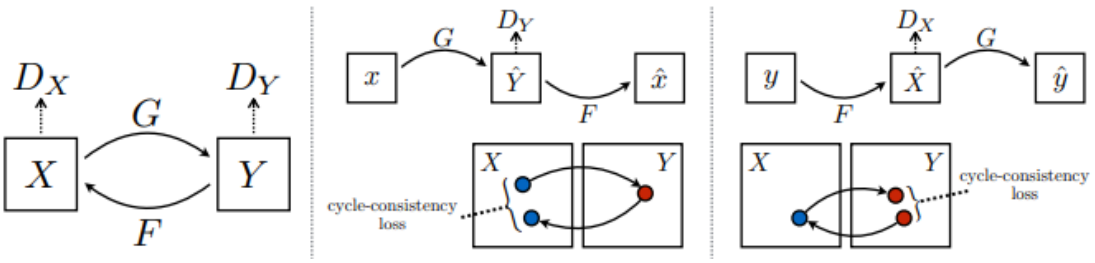


Figure 2. 18 Overview of Cycle-GAN reproduced from [146]. Two generators, tow discriminators and two cycle-consistency are combined to form the Cycle-GAN

mapping between X and Y.

The implementation of Cycle-GAN is shown in Figure 2.18, where two generators G and F map from X to Y and Y to X respectively, and two discriminators D_X and D_Y are used to distinguish between generated and real distributions. The proposed cycle consistency loss:

$$L_{GAN}(G, D_Y, X, Y) = E_{y\sim Y}[\log D(y)] + E_{x\sim X}\left[\log\left(1 - D\big(G(x)\big)\right)\right]$$

is implemented on top of the pair of GANs:

$$L_{cyc}(G, F) = E_{x\sim X}\left[\left\|F\big(G(x)\big) - x\right\|_1\right] += E_{y\sim Y}\left[\left\|G\big(F(y)\big) - y\right\|_1\right]$$

giving the full loss function as:

$$L\big(G, F, D_x, D_y\big) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + L_{cyc}(G, F)$$

for more consistent mapping between the two distributions.

The nature of Cycle-GAN remains unsupervised learning, but when the application of GANs extended to segmentation tasks, the accuracy of the segmentation becomes critical and supervised learning is introduced to the network. Image-to-image translation with conditional adversarial network (cGAN) [58] is therefore implemented for segmentation tasks. From the network point of view, segmentation cGAN is a special GAN which has a segmentation network as the generator and an additional segmentation accuracy loss. From the task point of view, segmentation cGAN is a special segmentation network which is combined with a discriminator introducing an additional adversarial loss. The functionality of such implementation, as we discussed in previous sections, is to increase the similarity of generated and data distributions, in this task meaning the segmentation and the reference distributions. This approach

further reduces the chance for the generator to produce abnormal segmentations, however, when training the network with pathological cases the effect of the discriminator becomes less reliable due to the irregular distribution of the data.

VAE

To conclude the topic of generative models, a VAE based cross-modality segmentation network [94] is discussed here, which takes both CCT and CMR images as input and combines both GANs and VAE. The data used in the example is the MICCAI Multimodality Whole Heart Segmentation Challenge [147] dataset, which consists of 3D CCT and CMR images. The 2D inputs are coronal views reconstructed from the 3D volumes.

The network architecture is shown in Figure 2.19. The objective of this architecture is to solve the issue of training a cross-modality segmentation network when a larger set of annotated data is available for one modality (the source domain) and a much smaller dataset for another modality (the target domain).

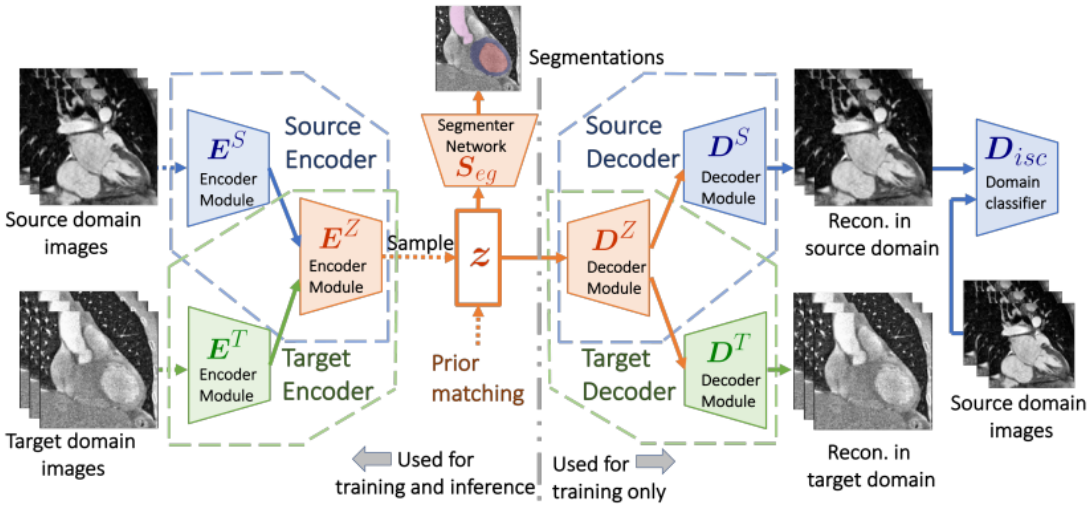


Figure 2. 19 Network architecture overview reproduced from [94].

In this case, the CMR images are taken as source domain and the CCT images

belong to the target domain. The basis of the network is the VAE, which is first trained using the source encoder together with the segmentation network S_{eg} using paired image and labels. The domain classifier D_{isc} (a discriminator) is trained to enhance the similarity between the distributions of the reconstructed source domain and the original source domain. The target encoder and decoder are added to the training, and to prevent overfitting, only their parameters are updated. The domain classifier is also trained to distinguish reconstructed source domain images taking the target domain encoding path from the original source domain images. A cycle-consistency is also implemented, while only the target-to-source-to-target half is applied considering the unbalanced amount of data from two domains. Another task-consistency, which could be considered as a variation of cycle-consistency, is added to the network to optimize the similarity between the reference segmentation distribution and the distribution of segmentations generated from latent variables calculated originating from source image to reconstructed target image and finally through the target encoder [94]. Key elements of generative models appear in the network, and it is shown that a stabilised version can be trained with even one case from the target domain.

2.5 Reconstruction of LV Shape

Reconstruction of LV 3D surfaces meshes has shown promising uses in a wide range of applications including cardiac function analysis, interventional guidance and diagnosis [73, 113, 124] For different modalities, the difficulty of reconstructing LV surfaces also varies and on CCT images for example classical isosurfacing algorithms can be applied,

since the segmented contours have small slice thickness and distance between slices. For cardiac functional analysis, even though the application of the Simpson's rule have shown promising results for accurate and reproducible assessment of cardiac function using manually traced contours as we discussed in previous sections, the accumulated disks suffer from motion artefacts caused by respiration [85, 110] with blocky segments distributed in 3D space. The reconstructed 3D surfaces on the other hand have the advantage of meaningful geometries. The accuracy and robustness of cardiac functional assessment with 3D reconstructions have also been validated for both human and mouse heart [60, 138, 140].

Reconstructing LV surfaces from CMR images is more challenging as the slice thickness and the distance between slices are normally around 20 times larger than for CCT images, and therefore has been normally formulated as a 3D mesh adaptation problem to sparse contours or points [86, 119, 139, 148]. Shape priors are usually incorporated in the fitting process in the form of a regular shape [69, 86, 139] or a statistical template with plausible variations [30], and the geometries are also penalised for low smoothness with an explicit smoothing term during fitting. Both factors regularise the "normality" of the reconstructed surfaces, which could compromise the accuracy of method when handling pathological cases.

2.5.1 Guide-point Modelling

The GPM method [139] has been validated with both human and mouse CMR data [60, 138, 140], where a 3D finite element model of the LV (consisting of 16 elements) was

interactively customized for each of the scans. Visualisations of the model and some guide points are shown in Figure 2.20.

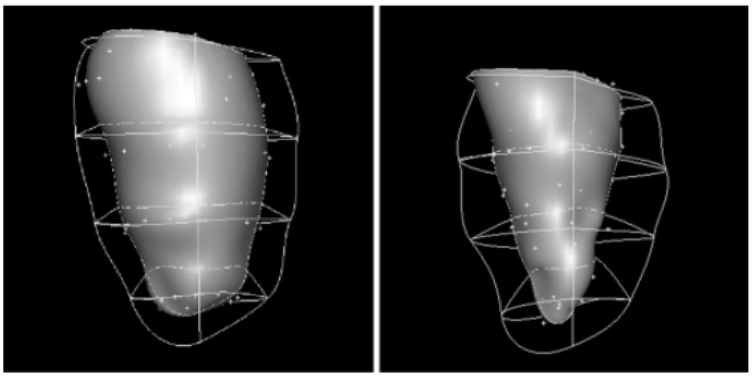


Figure 2. 20 Examples of model and guide points reproduced from [139]. White points are the guide points, and the mesh is the 3D finite element model. The LV cavity is rendered as the grey mesh.

For each element in the 3D finite element model, the geometry is given by a weighted mean of values at nodes:

$$\lambda(\xi_1, \xi_2, \xi_3) = \sum \Psi_n(\xi_1, \xi_2, \xi_3)\lambda^n,$$

where λ^n are the nodal values, Ψ_n are basis functions which give the relative weighting of each nodal value, and (ξ_1, ξ_2, ξ_3) are the element coordinates. The geometry filed λ is defined as the radial coordinate in a prolate spheroidal coordinate system:

$$x = f\,\cosh(\lambda)\cos(\mu)$$
$$y = f\,\sinh(\lambda)\sin(\mu)\cos(\theta)$$
$$z = f\,\sinh(\lambda)\sin(\mu)\sin(\theta)$$

where (λ, μ, θ) are the radial, longitudinal and circumferential coordinates of the prolate spheroidal coordinate system and (x, y, z) are the corresponding rectangular cartesian coordinates. The similarity of the LV to a truncated prolate spheroid makes this coordinate system a natural choice.

The geometry is optimized to manually introduced guide-points and automatically generated data points (from image features and typically given a lower weight) in both SAX and LAX slices by minimizing the error function:

$$E = S(\lambda) + \sum_{g} [\lambda(\xi_g) - \lambda_g]^2,$$

where λ_g are the positions of the guide point data and $\lambda(\xi_g)$ are the model positions at element coordinates ξg corresponding to λ_g. $S(\lambda)$ is a weighted Sobolev norm [115] as the smoothing term.

2.5.2 3D Shape Reconstruction Network

Anatomically Constrained Neural Networks (ACNN) have been proposed to directly produce a high-resolution 3D CMR image of the LV along with the corresponding 3D structure from a standard 2D SAX slice stack [91]. Its baseline network takes a similar format to that of the 3D U-Net, with multiple down-sampling and up-sampling stages and skip connections for the layers with the same resolution, while the outputs are different. The network architecture is shown in Figure 2.21.

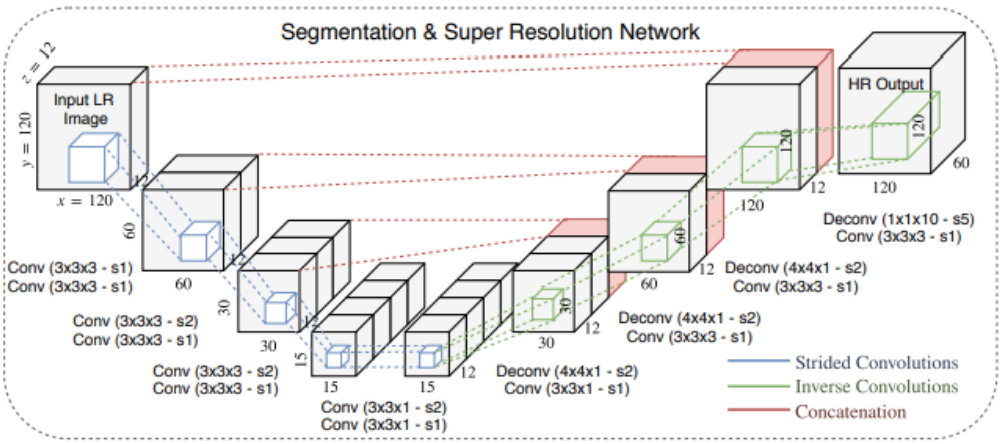


Figure 2. 21 ACNN baseline segmentation and super resolution network architecture

reproduced from [91].

The baseline networks are combined with a convolutional autoencoder, serving as a regularisation model, and the entire model is called T-L network [44] as shown in Figure 2.22. The anatomical meaning is expected to be learnt and used to produce accurate outputs by increasing the similarity between the latent feature space calculated from both image and label map spaces.

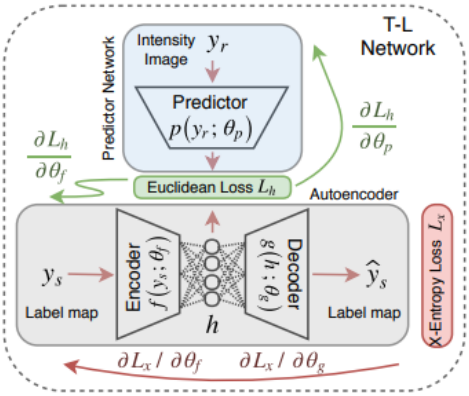


Figure 2. 22 T-L network overview reproduced from [91].

ACNN only takes 2D SAX slices as input, and the spatial information of the slices is not included in the network. Therefore, the output of the network can only take relative voxel spacing values instead of physical voxel spacing values. For example, when the distance between SAX slices is 8 mm for case A and 10 mm for case B while the images are identical, the network outputs are identical, even though the actual shapes of the LV are not the same. For this reason, even though the network is 3D, the method is considered as 2.5D.

3D CNNs [116, 135] have been used for predicting the 3D structure of the LV. In both cited works the feature spaces of input and output volume are the same, and the networks are trained for reconstruction by taking both SAX and LAX images as input.

In [135], CMR images are first segmented, and a variation of 3D U-Net is implemented for reconstruction of the 3D structure. Synthetic data using CT volumes are used for training the 3D network, which has a discriminator added to it forming a GAN. In [116], the 2D images are first transformed into a 3D grid, and a 3D CNN (consisting of two sections connected by an averaging operation of the segmentation of 3D volumes transformed form individual slices) as shown in Figure 2.23 is used to reconstruct 3D LV shape. The novelty of the method is incorporating 3D coordinates into the network, and in order to achieve that both CNN1 and CNN2 take a limited receptive field of 25x25x25 and 21x21x21, giving a patch-wise CNN overall. The 3D coordinates reduce the limitation introduced by the small field of view, provided the position and orientation of the LV are normalised for all cases, as the global information cannot be learnt by the network. The size and shape of the heart will also affect the performance of the network, making it difficult to deal with pathological cases.

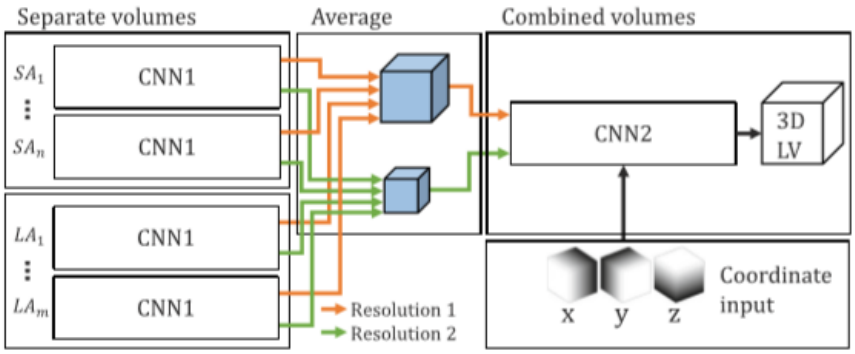


Figure 2. 23 Network overview reproduced from [116].

2.6 Conclusions

Cine CMR images are the current gold standard for anatomical and functional analysis of the heart, and segmentation algorithms are critical for automated quantification of

cardia function. Deep learning algorithms show clear advantages compared to traditional medical image segmentation methods and have been widely applied for both 2D image segmentation and 3D shape reconstruction. However, there has been little work implementing deep learning algorithms for analysing mouse CMR images, and the size of preclinical datasets (which is usually around or less than 20 cases) is partly the reason. The 3D shape reconstruction of LV are still mainly using SAX slices without a flexible and effective approach to incorporate information from LAX slices, and the current fully 3D networks are applied to either synthetic CT dataset with a large field of view, or CMR images in a patch-wised form.

Chapter III. Cardiac Function Estimation of the Left Ventricle

Segmentation is usually applied as the first step in left ventricle (LV) quantification, however, with the recent development of deep learning algorithms, methods with a direct regression also show good results in recent publications. In this chapter, we use our neural network framework to compare the performance of different types of LV quantification methods (i.e., methods for the calculation of relevant cardiac parameters such as cavity volume, myocardial mass and wall thickness) using a publicly available benchmark dataset: the MICCAI 2018 Left Ventricle Full Quantification Challenge dataset, consisting of mid-cavity slices of cardiac cine-MR scans with multiple frames. We propose a multi-mode network that enables us to compare the accuracy obtained with direct regression of the desired metrics, a segmentation-based method and a novel combined approach using segmentation features for metrics evaluation.

We presented part of the work in this chapter as an oral presentation at the Statistical Atlases and Computational Modelling of the Heart (STACOM) workshop of the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2018 conference in Granada, Spain, and it was subsequently published in Lecture Notes in Computer Science [129].

3.1 Introduction

LV quantification has significant clinical importance for identification and diagnosis of cardiac diseases [61] by calculating the cavity volume, LV myocardial mass and regional parameters (such as wall thickness) as described in sections 2.1 and 2.3. Existing LV quantification methods can be classified into one of three types: (1) manual quantification; (2) automated segmentation-based (SB) quantification; and (3) direct regression (DR) [131].

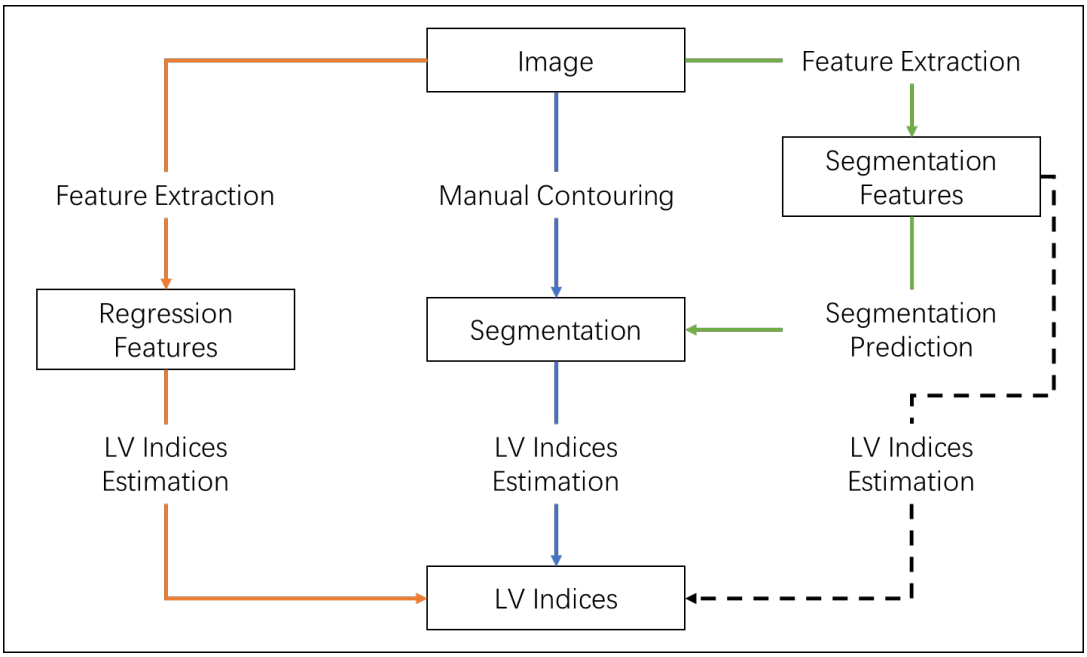


Figure 3. 1 Graphical overview of different types of LV indices quantification methods. The methods on the left (as orange solid lines) and right/centre (as blue and green solid lines) correspond to DR and SB, respectively. The method on the right, including the dashed line on the bottom right corner, represents a novel modification of standard methods in which the LV indices are not directly calculated from the segmentation results, but predicted using an additional neural network.

An overview of these three approaches is shown in Figure 3.1 as solid lines. Quantification of LV anatomy from manual contouring of myocardial borders (manual delineation) is the gold standard in clinical practice [113] and usually takes two steps: manual segmentation and indices calculation, which is the centre line in the Figure 3.1. SB methods follow the same two-step strategy, except with the manual contouring replaced by an automated segmentation; here the focus of the algorithm development is on the segmentation stage. DR methods aim at evaluating LV indices directly from images, without an intermediate segmentation stage, taking advantage of the power of deep neural networks.

In this chapter, we compare methods from two of the above three classes: (2) SB methods and (3) DR methods. One key advantage of evaluating indices from segmentation is that the results can be verified by comparing to manual segmentation quantitatively or to visual inspection qualitatively. DR methods on the other hand only focus on accurately evaluating clinical important metrics and intend to skip the segmentation process, introducing the fundamental limitation that no explanations can be made for failed cases. The recent development of deep learning neural networks has improved the performance of both segmentation algorithms and DR methods. Variations of the fully convolutional neural network (FCN) [79] have shown promising results for segmentation of LV using cine-MR images [91, 100, 117, 118]. Convolutional neural networks (CNNs) have also been applied to directly quantify LV indices and shown steady improvements [131, 132, 133, 134]. Unfortunately, there has been very limited direct comparison between these two types of methods, and the

MICCAI 2018 Left Ventricle Full Quantification Challenge (in conjunction with the STACOM 2018 workshop) provides a great platform for evaluating different methods on the same benchmark dataset. We develop and evaluate our methods on this challenge dataset by proposing a neural network with multiple modes corresponding to different types of methods. In addition to SB and DR, we also propose a new combined approach in the form of multi-task neural network to further investigate the calculation of LV indices using segmentation features.

3.2 Dataset

We used the MICCAI 2018 Left Ventricle Full Quantification Challenge dataset for development and evaluation of our method. There are 145 subjects in the dataset and for each subject 20 frames of one mid-ventricular slice are included to cover the whole cardiac cycle, giving 2900 images in total. Quantifying mid-cavity slices are less challenging than basal and apical slices and therefore the dataset is a good starting point for evaluating DR methods as well as directly comparing with SB methods. The images were collected from three hospitals. The subjects are in a wide range of age and present diverse pathologies, including regional wall motion abnormalities, myocardial hypertrophy, mildly enlarged LV, and atrial septal defect LV dysfunction. The pixel spacings of the MR images range from 0.68 *mm/pixel* to 2.08 *mm/pixel*, with mode of 1.56 *mm/pixel*. For each subject, there is a sequence of 20 frames throughout a cardiac cycle. Before being made available to challenge participants, all cardiac images underwent several pre-processing steps before resulting in approximately aligned

images with dimensions of 80×80 and were manually contoured to obtain the epicardial and endocardial borders, which were double-checked by two experienced cardiac radiologists [130, 131] and were also provided as part of the training dataset. A selection of the first frames from 20 subjects from the dataset are shown in Figure 3.2.

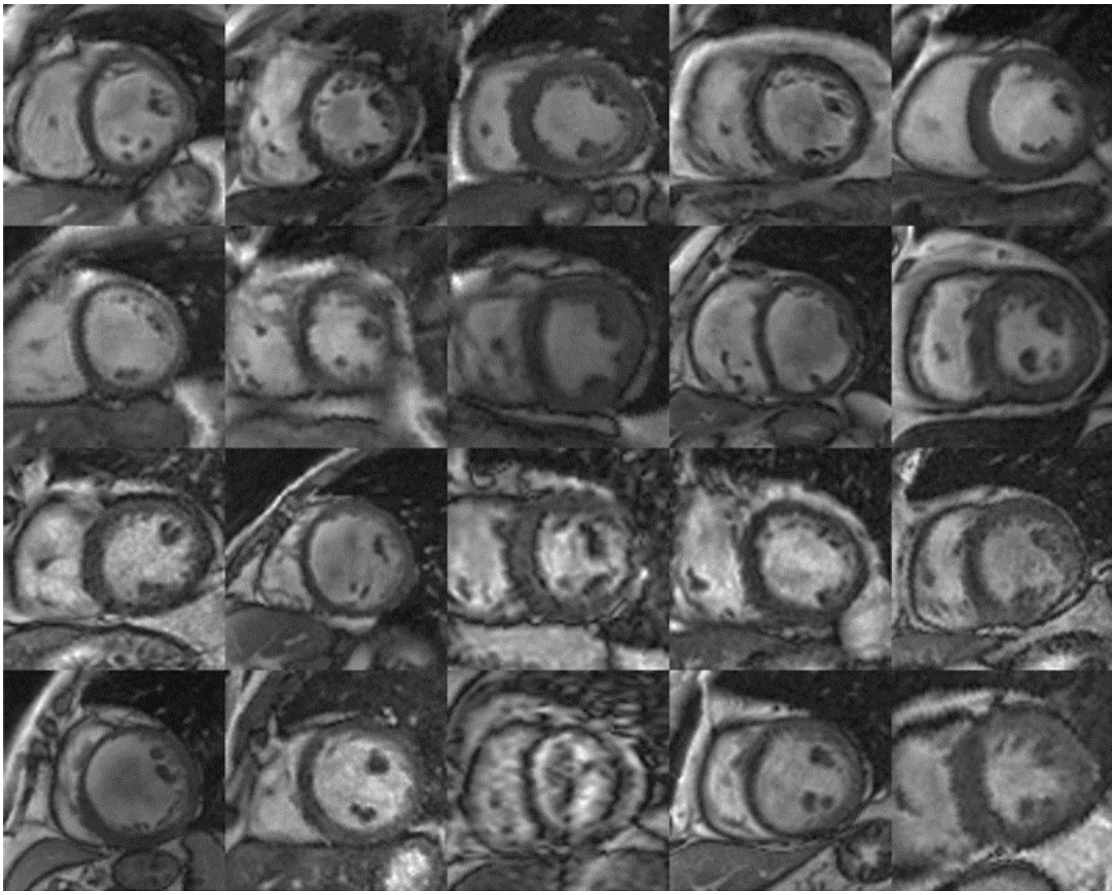


Figure 3. 2 The first frames of 20 subjects from the dataset. The quality of images and anatomies vary widely across the dataset. The hearts are normalised to a similar orientation and size.

Due to the resizing, the actual pixel spacings on the 80×80 images are in a range of [1.13, 2.29] *mm/pixel*, with mean and median values of 1.64 and 1.63 *mm/pixel* respectively. The manual segmentation, to be used for training or as a gold standard for validation, is included in the dataset as two sets of binary masks, representing the region

enclosed by endocardial and epicardial contours delineated manually. For each frame, both global (including normalised areas of the myocardium and LV cavity by the organizer with a factor of 1/6400, where 6400 is the number of pixels per image) and regional indices (three directional LV cavity dimensions and six regional wall thicknesses as shown in Figure 3.3) of the LV and the cardiac phase (diastolic or systolic) [131, 133] are the target estimating metrics, with their reference values (which are also normalised by the organizer with a factor of 1/80, where 80 is the image size) included in the training dataset. Different from other LV indices, cardiac phase is important for indicating the temporal dynamics of the LV, and it has also been shown to provide useful regulations for estimating other LV indices [133]. Therefore, estimating cardiac phase not only provides useful information about cardiac cycle, but also has the potential to further improve the quantification accuracy of other LV indices.

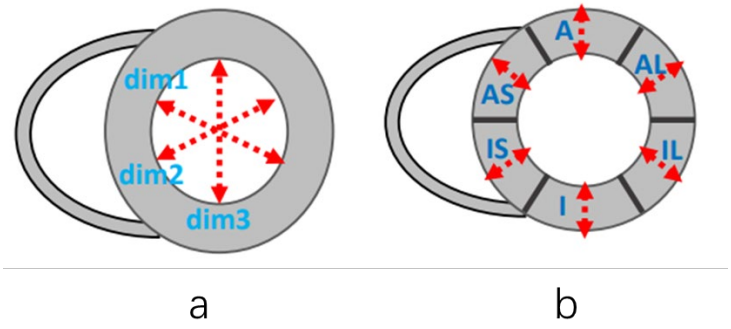


Figure 3. 3 Regional LV indices illustration reproduced from [133]. A, AS, IS, I, IL and AL represent anterior, anteroseptal, inferoseptal, inferior, inferolateral and anterolateral respectively, and the six regional wall thicknesses are estimated for each of the segment. Dimensions 1 – 3 are the directional dimensions of the LV cavity

between AS and IL, IS and AL, and A and I respectively.

3.3 Method

A neural network with three modes is proposed to compare the performance of different LV quantification methods. The three modes are: (1) Reg Mode (DR method); (2) Seg-Reg Mode, regression combined with segmentation-features (DR method); (3) Seg Mode (SB method). The second mode, segmentation-feature regression is a novel method we propose to improve on the original DR method, and is shown as a dashed line in Figure 3.1. The Seg-Reg mode has the form of a multi-task neural network, with a segmentation branch and a regression branch. The concept of the network is strongly regularising the features used for regression with a segmentation loss, and the focus of the network is investigating the impact of such regularisation on the regression task.

3.3.1 Multi-mode Network

One of the key contributions of this chapter is directly comparing DR method to SB method within one network architecture, and to better demonstrate the performance of the DR method the state-of-the-art segmentation network U-Net is integrated in our network architecture as the backbone shared by different modes to ensure an accurate segmentation result. The main network architecture is designed following the same approach as U-Net, including the number of down-sampling and up-sampling stages, the number of layers for each stage, the number of kernels for each layer, and the size of convolutional kernels. We increased the number of convolutional kernels by keeping

the ratio between different stages the same and saw no significant improvements of the network performance, while the GPU memory consumption increased linearly with the number of kernels. Therefore, we keep the same parameter setting as U-Net. For the regression branch, a larger convolutional kernel size is used to increase the field of view, considering the relative smaller number of layers in the branch.

The network architecture we propose is summarized in Figure 3.4. The Feature Extraction Net (FE-Net) consists of two stages: an encoder and a decoder connected with short paths between layers with the same resolutions, using an FCN with skip connections inspired by the U-Net architecture [102]. The encoder transforms the input image into the Middle Features (MDF), while the decoder transforms these into the End Features (EDF), at the original image resolution. This is connected to two output branches (Branch Seg. and Branch Reg.), which perform segmentation and LV indices regression, respectively. The segmentation output layer has sigmoid activation functions and all the other FCN activations are rectified linear units (ReLU) [127]. Two convolutional layers and two batch normalisation (BN) layers [56] form a convolutional block, with the number of kernels for each block specified in Figure 3.4. BN layers are used here for faster convergence [56].

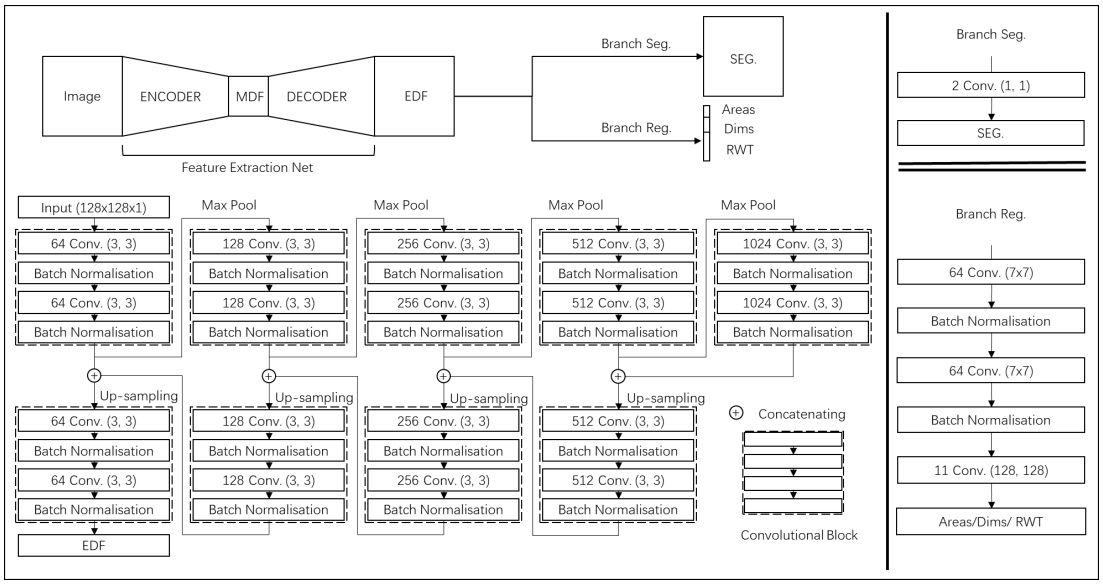


Figure 3. 4 Neural network architecture of the multi-mode network.

Input feature maps of convolutional layers are padded with zeros so that the output feature maps have the same size. Branch Reg. includes two convolutional layers (64 kernels with size 7×7 for each layer) and an additional convolutional layer connecting to the regression output, which has 11 kernels with size 128×128. All the other convolutional layers have kernel size of 3×3. Max-pooling and up-sampling layers have a window size of 2×2 with a stride of 2×2. Branch Seg. has only one convolutional layer with 2 kernels of size 1×1 giving two output layers corresponding to regions enclosed by endocardial and epicardial contours.

The overall loss function is:

$$L_T = \lambda_R L_R + \lambda_S L_S$$

where L_S and L_R are segmentation output and regression output losses respectively. L_R is mean square root error (MSE) and L_S is binary cross-entropy. λ_S and λ_R are the corresponding loss weights, which are set based on the mode: (1) Reg Mode has $\lambda_S = 0$ and $\lambda_R = 1$, focusing the training on the prediction of the final indices without

considering the accuracy of the segmentation; (2) Seg-Reg Mode alternates between $\lambda_S = 1$ and $\lambda_R = 0$ for odd epochs and $\lambda_S = 1$ and $\lambda_R = 1$ for even epochs, combining the accuracy of the intermediate segmentations and the final regressed LV indices; and (3) Seg Mode has $\lambda_S = 1$ and $\lambda_R = 0$, focusing only on the accuracy of the segmentation, from which the LV indices are directly calculated afterwards. In the Seg-Reg Mode, by updating the weights of the FE-Net with segmentation losses at every epoch, and with regression losses only every other epoch, the EDF is heavily regularised by the segmentation task (simulating the segmentation features) and used in the regression task. The implementation of Seg-Reg Mode follows an iterative approach inspired by the GANs [45], in which the generators and discriminators are trained iteratively with different number of steps and different loss functions as described in section 2.4.7.3. Different from the minmax game of GANs, our iterative approach aims to produce EDF that is more biased to the segmentation task.

3.3.2 Calculation of LV indices from Segmentation

For our segmentation method, we normalize the intensity of the input images in three steps: (1) we perform an initial automated segmentation of each slice; (2) from these rough segmentations, we calculate the mean intensity values of the LV cavity and LV myocardial regions (I_{cav} and I_{myo}), and the main parameters of the intensity distribution including population means ($MEAN_{cav}$ and $MEAN_{myo}$) and population standard deviations (STD_{cav} and STD_{myo}) of all the slices used to train and validate the neural network; (3) if I_{cav} is out of the range $MEAN_{cav} \pm STD_{cav}$ or I_{myo} is out of the range

$MEAN_{myo} \pm STD_{myo}$, the intensity values of the slice are divided by a factor of ($I_{cav}/MEAN_{cav} + I_{myo}/MEAN_{myo}$)/2 for normalization. The segmentation outputs are converted to binary masks (by thresholding at 0.5) and cropped to the size of 80×80×2.

The gold standard LV indices provided with the training data were originally calculated from segmentations by the challenge organizers. However, the methods used for this calculation were not provided as part of the challenge information and thus we had to develop our own algorithms for indices calculation. Endocardial and epicardial binary masks are calculated, and for each of them the biggest object is extracted and any potential holes in it are filled. The binary mask of the myocardium is calculated by subtracting the endocardial from the epicardial binary mask. The areas of cavity and myocardium are then normalized by dividing the number of pixels inside the mask over the total number of pixels per image (6400) as required by the challenge organizer for evaluation.

To calculate the LV dimensions and the regional wall thickness, we first sample each of the endocardial and epicardial contours by 360 points located equally in angular coordinate with a one-degree difference between each point. The polar coordinates are defined by setting the zero-degree polar axis parallel to the positive horizontal axis of the image, and the pole *O* located at the centroid of the endocardial border of the binary mask (representing the centre of the LV blood pool). To calculate the location of endocardial and epicardial points for each selected angle, the intensity values of binary masks along the radial direction from *O* are obtained, and the point with the interpolated value of 0.5 is defined as the contour point. For each angle, a middle contour point is

defined as the point with radius of the mean value of the endocardial radius and epicardial radius, and the middle contour point is used for the calculation of the wall thickness. The points are finally converted back to the Cartesian coordinates of the image. An illustration is shown in Figure 3.5.

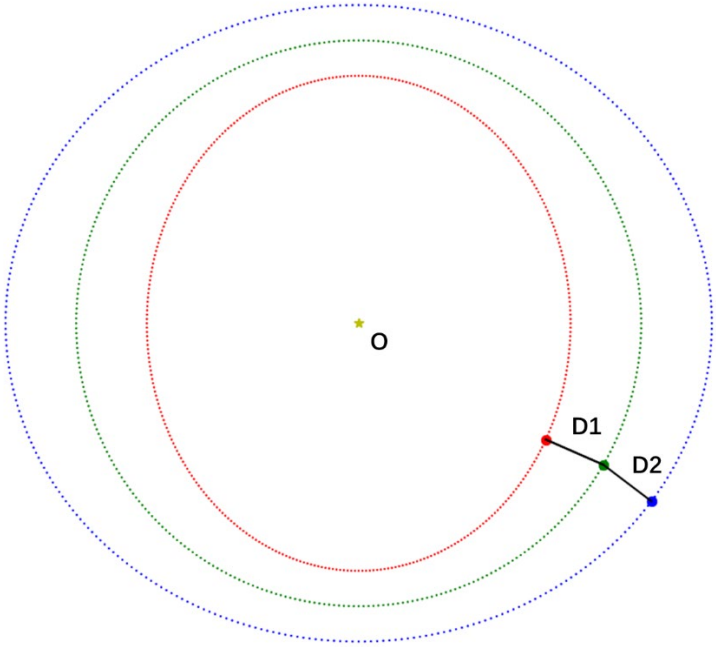


Figure 3. 5 Contour points illustration. The red, green and blue points are endocardial, middle and epicardial contour points. D1 and D2 represent the shortest distance from a middle contour point to the nearest endocardial and epicardial contour points. The point O is the centroid of the endocardial border of the binary mask.

The LV dimensions are calculated in two steps: (1) calculating the distance from the centroid to the endocardial contour points; (2) averaging the values of 60 distances for each region (corresponding to 60 degree sections, i.e. one sixth of the whole myocardium) and adding opposing regions to get the three LV dimensions (Dims) as shown in Figure 3.3. The regional wall thickness (RWT) values are calculated in three steps: (1) for each middle contour point, locating the nearest endocardial and epicardial

contour points; (2) adding the minimum distances from each point to the endocardial and epicardial contours (*D1* and *D2* respectively); (3) averaging the values of 60 points for each region. Both distances are then normalised by the image size (80) as required by the organizer to be compatible with the provided reference values.

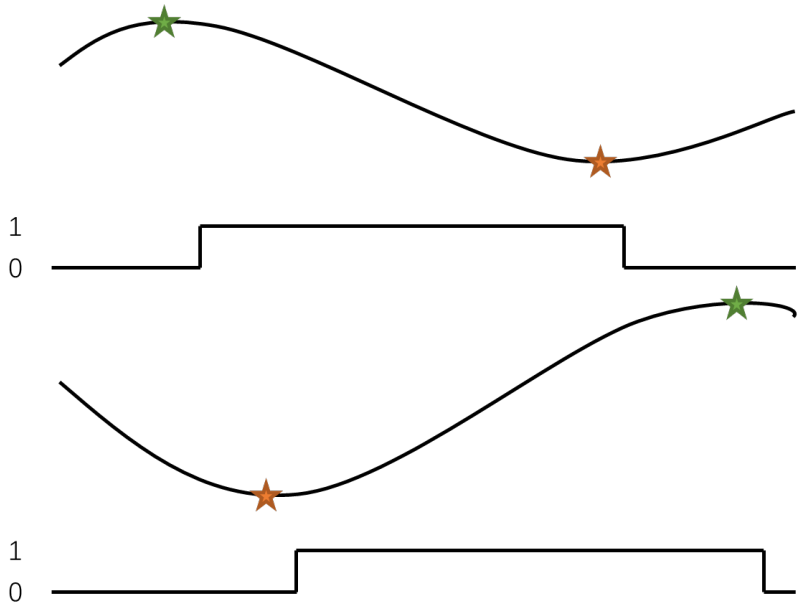


Figure 3. 6 Illustration of phase estimation. The orange and green stars represent the ED and ES frames, and the corresponding frames are labelled 0s and 1s for diastolic and systolic respectively.

Our method uses the cavity segmentation results to estimate phase using a simple procedure as shown in Figure 3.6. The estimated cavity area sequence is first smoothed using a moving average filter with a span of 5 frames. End-diastolic (ED) and end-systolic (ES) frames are then identified as the frames with maximum and minimum cavity areas respectively. The remaining frames are then assigned phase numbers (0 for diastolic and 1 for systolic) sequentially. There are 20 frames for each subject, and there are two possible outcomes of the ED and ES frame numbers: (1) ED frame number is smaller, or (2) ES frame number is smaller. When the ED frame comes first, all frames

from the next frame to the ES frame (inclusive) are labelled as 1, and the other frames are labelled as 0; and when ES frame comes first, all from the next frame to the ED frame (inclusive) are labelled as 0, and the other frames are labelled as 1.

3.3.3 Data Augmentation

To preserve as much information from the cropped images (80×80) as possible, we decided to use a bigger image as the input of the neural networks. We chose 128×128 as a reasonable size providing extra space for data augmentation. We only apply augmentation to the training and validation data, by using a starting point of a background image of size 128×128 filled with zeros, applying random rotation between 0 and 360 degrees, random resizing in the range of ± 10%and random flipping to the 80×80 images and placing the resultant image on a random position of the larger background image while making sure the complete ventricle is included using the binary references (segmentations). The binary references are augmented correspondingly, and the LV indices are also rescaled accordingly. The augmentation process is performed online, i.e. the augmented data is different at each epoch. Comparing with further cropping from the original images, our augmentation method preserves more information from the original images, which should help to prevent over-fitting as the number of trained parameters remains the same. The main drawback of our method is higher memory consumption during training, but by limiting the batch size to 8 we did not experience any issues during training and testing on a single modern GPU.

3.4 Experiments and Results

3.4.1 Implementation

We implemented the neural networks in Keras with Tensorflow as backend, and the networks were trained with Adadelta [142]. We used one NVIDIA GeForce GTX 1080 Ti GPU for all training and testing experiments.

3.4.2 Training Data Experiments

Five-fold cross validation is recommended by the challenge organizer for fair comparison among all the participants. Leave-one-out cross validation is the extreme case of K-fold cross validation and maximises the size of training and validating data, which is beneficial for deep learning algorithms to improve accuracy and prevent overfitting. However, the computational cost is also significantly increased and therefore not usually applied for training CNNs. Reducing the size of training and validating data and increasing the size of testing data sacrifices the network performance and rises the risk of overfitting, while the computational cost is effectively controlled. The validating dataset is not used for parameter tuning but used to prevent overfitting during the training process and a larger proportion of validating data tends to produce more general networks, and the cost is a smaller size of training data which usually leads to a performance drop of the network. For all the training data experiments we performed five-fold cross validation (29 subjects on each fold), and randomly assigned 100 subjects as training data and 16 subjects as validating data from the

remaining dataset.

Table 3. 1 Average results of compared methods. Values are shown as mean error ± standard deviation, followed by the correlation coefficient shown in brackets.

Dataset	Method	Avg. Areas (mm^2)	Avg. Dims (mm^2)	Avg. RWT (mm^2)	Phase Error (%)
Training and Validating Data	Indice-Net [132]	200 ± 130 (0.90)	NA	1.4 ± 0.7 (0.76)	NA
	FullLVNet [134]	190 ± 130 (0.94)	2.7 ± 1.6 (0.92)	1.4 ± 0.7 (0.76)	10
	DMTRL [131]	180 ± 120 (0.95)	2.5 ± 1.6 (0.93)	1.1 ± 0.7 (0.77)	8.2
	Reg Mode	160 ± 160 (0.91)	2.1 ± 1.8 (0.92)	1.5 ± 1.3 (0.80)	9.4
	Seg-Reg Mode	140 ± 130 (0.93)	1.9 ± 1.6 (0.94)	1.4 ± 1.2 (0.82)	8.6
	Seg Mode	**120 ± 110** (0.95)	**1.3 ± 1.1** (0.97)	**1.0 ± 0.9** (0.91)	**7.8**
	Seg Ref	7 ± 6 (1.0)	0.19 ± 0.16 (1.0)	0.17 ± 0.13 (1.0)	6.3

The results are shown in Table 3.1 compared with the state-of-the-art methods applied on the same datasets [131, 132, 134], with the values shown as Mean Error ± Standard Deviation, followed by the correlation coefficient shown in brackets. In our initial experiments, the SB method performed best and thus we trained two additional sets of networks and ensembled the segmentation results by averaging the predictions,

yielding an overall improvement of around 3%.

Since the method used for indices evaluation from segmentation was not available to us as part of the challenge information, to calculate the indices from contours we developed our quantification method as detailed in Section 3.3.2. We evaluated this method by applying it to the manual segmentation references, comparing the indices thus calculated to the indices provided by the challenge organizers. This showed a small discrepancy between our calculations and those provided in the challenge. This discrepancy is shown in Table 3.1 as Seg Ref and is likely to have affected the results of our Seg Mode method negatively, albeit to a small extent. The Reg Mode achieves comparable results to the state-of-the-art methods, as expected due to their common use of end-to-end training using indices. When combined with segmentation in our novel method, the Seg-Reg Mode produces better results for all indices. Indice-Net uses image reconstruction features from an auto-encoder and has two convolutional layers for regression [132]. Comparing Indice-Net and our Seg-Reg Mode, we see a clear advantage in using segmentation features for area evaluation and a slight improvement on RWT evaluation. Our best performing DR method (Seg-Reg Mode) shows comparable results for RWT and Phase estimation with the best existing method (DMTRL), and more accurate results in area and dimension indices calculation. Our SB method out-performs DR methods for all indices and achieved better results comparing to the state-of-the-art methods.

3.4.3 Analysis of Segmentation Results

On top of the better accuracy of the SB method when predicting clinically important indices, another key advantage of it is the ability to use segmented binary masks to better understand the predictions. We further investigated our SB method both qualitatively and quantitatively.

For qualitative analysis, we performed a visual inspection of the segmentation results and 18 randomly chosen examples are shown in Figure 3.7, where endocardial and epicardial contours are represented by red and blue contours. We also show the reference binary masks of the same cases in Figure 3.8 for pairwise comparison.

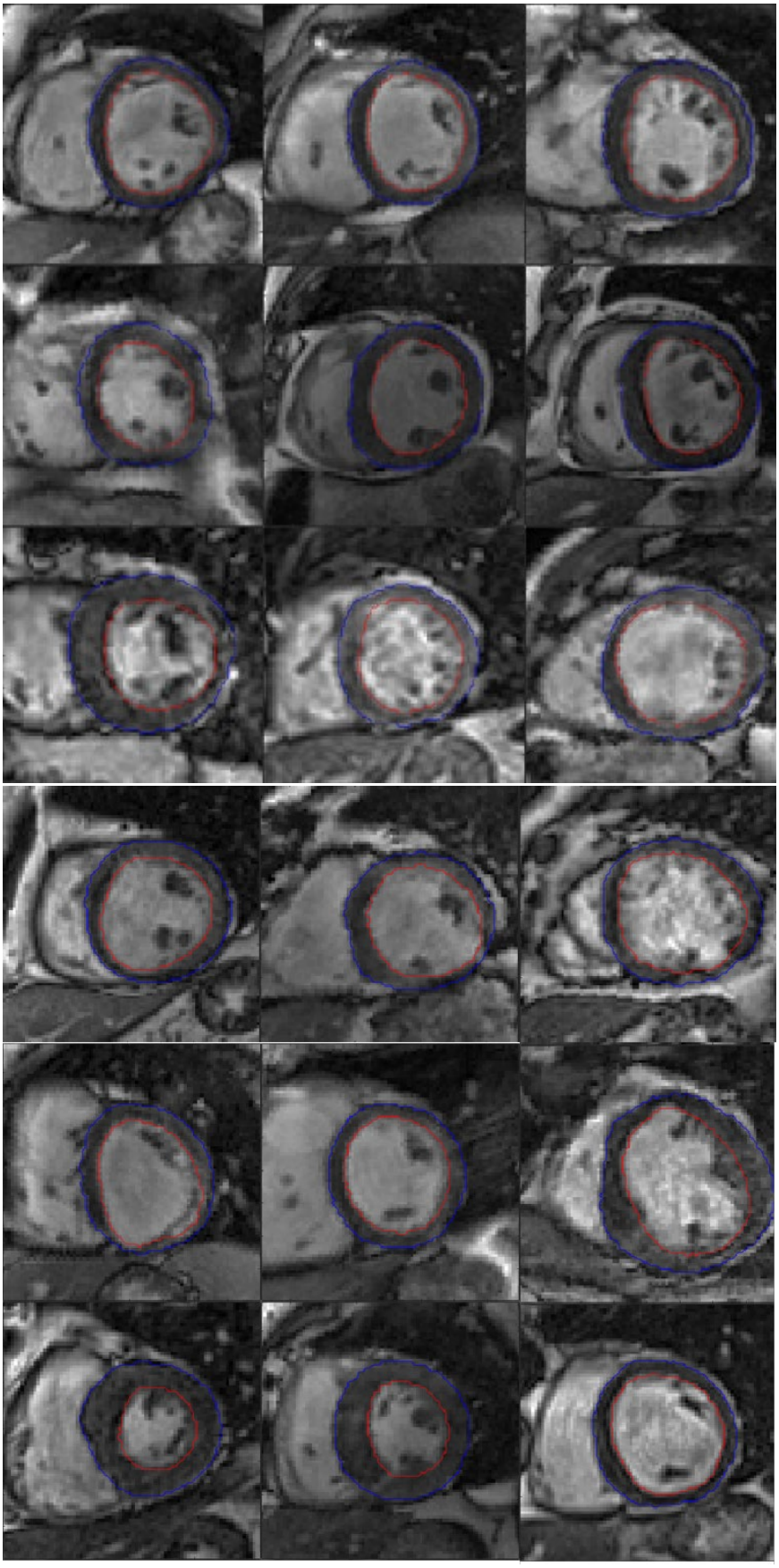


Figure 3. 7 Training data segmentation results for visual inspection. The red contours

are the endocardial contours and the blue contours are the epicardial contours.

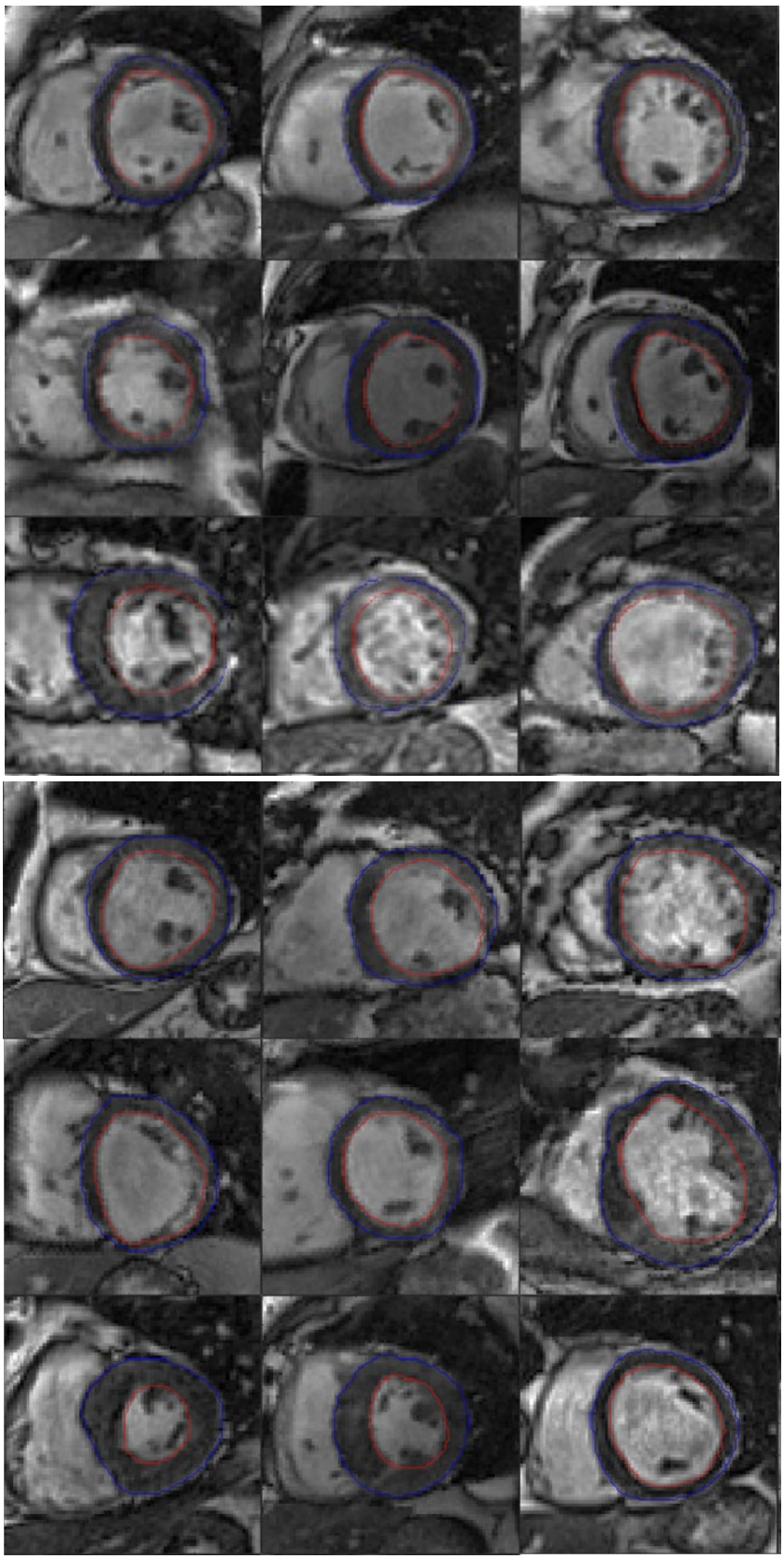


Figure 3. 8 Training data reference contours. The red contours are the endocardial

contours and the blue contours are the epicardial contours. The segmentation results in Figure 3.7 are similar to the reference contours.

Despite the wide variety of shapes and sizes of the LV myocardium, we can see that our SB method produces good quality segmentation results, which show great similarity to the reference binary masks.

For quantitative analysis, we calculate Dice similarity metric between our segmentation results (Y') and the reference binary masks (Y):

$$Dice = \frac{2 \times Y \cap Y'}{|Y| + |Y'|}$$

The Dice metric has a value range between 0 and 1, with the extremes indicating no spatial overlap and complete overlap respectively.

We calculated Dice metrics for endocardial, epicardial and myocardial binary masks. The results (mean and standard deviation) were: 0.96 ± 0.02, 0.97 ± 0.01 and 0.91 ± 0.04 respectively. We show box plots of the Dice scores in Figure 3.9. For Dice values of the endocardial region, the lower adjacent, 25th percentile, median, 75th percentile, higher adjacent values are: 0.93, 0.96, 0.97, 0.98, 0.99 respectively, where the lower adjacent and upper adjacent are 1.5 times the interquartile range away from the top and bottom of the box respectively. For Dice values of the epicardial region, the lower adjacent, 25th percentile, median, 75th percentile, higher adjacent values are: 0.95, 0.97, 0.98, 0.98, 0.99 respectively. For Dice values of the myocardial region, the lower adjacent, 25th percentile, median, 75th percentile, higher adjacent values are: 0.83, 0.89, 0.91, 0.93, 0.97 respectively. The number of outliers for endo/epicardial and myocardial Dice values are: 138, 115 and 168 respectively, corresponding to 4.8%, 4.0% and 5.8%

out of the population (2900 images in total).

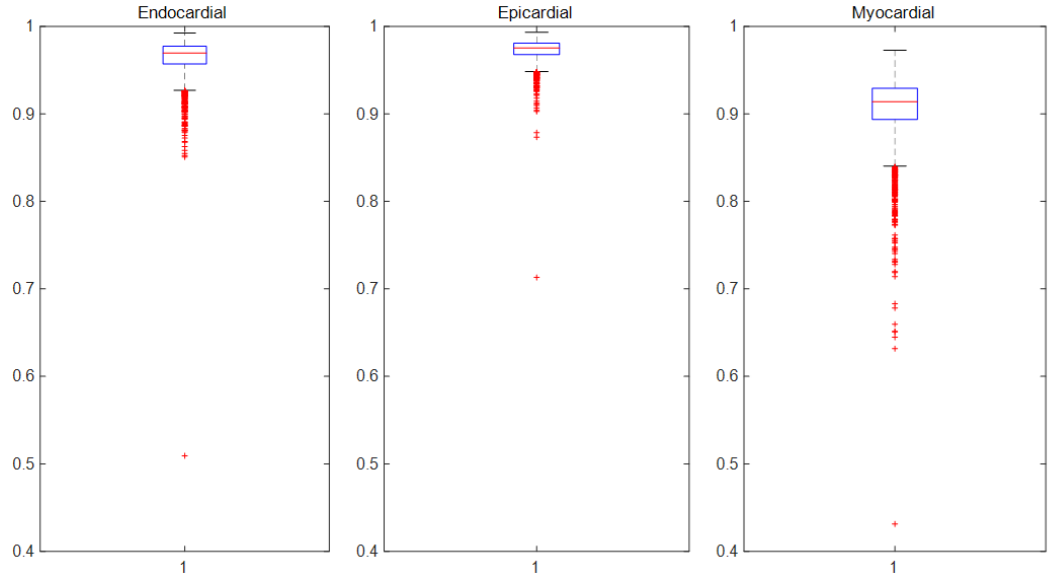


Figure 3. 9 Box plots of dice metric values for areas enclosed by endocardial and epicardial contours and myocardial area.

In one of the cases the Dice metric values for all three regions are substantially smaller than in all others. The MR image, the predicted segmentation results and the reference segmentations are shown in Figure 3.10 to better understand the reasons behind this worst case.

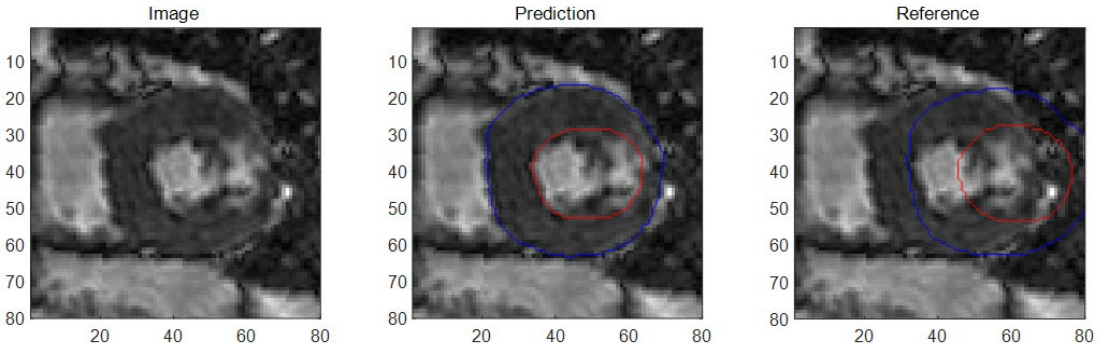


Figure 3. 10 Worst performing case of our segmentation method. The reference contours are completely off.

As clearly visible in Figure 3.10, the reference binary masks are misaligned with the images, while the predicted binary masks seem to be closer to the actual contours.

Because the reference is completely off, the Dice coefficient value is small, which is usually interpreted as bad performance of the segmentation method but does not reflect the quality of segmentation result for this case. Some other cases might also be affected by the manual segmentation inaccuracies, which would reduce the overall Dice scores.

3.5 Conclusions

In this chapter, we directly compare two different approaches for evaluating LV cardiac function: the SB and DR methods, under the same neural network architecture proposed for this comparison. From Table 3.1, we can see that our SB method produces the best results compared to both the state-of-the-art and our DR methods. We also develop a novel DR method using segmentation features, which shows better performance compared to our DR method. We joined the MICCAI 2018 Left Ventricle Full Quantification Challenge and got the Third Place Award with our SB methods (details of the methods and the testing results will be discussed in Section 4.1).

We also evaluated the segmentation results of our SB method and showed another key advantage of the SB method compared to using metrics to validate algorithms. Using the Dice similarity metric, in our experiment, to evaluate the performance of our segmentation neural network indicates one case is obviously worse than the others. However, by visualising the segmentation results, we have confidence on our SB method as we spot that the low Dice value is due to the reference binary masks being misplaced. DR methods predict clinically important metrics and the only way to validate the method is calculating the difference between predicted values and reference

values, while SB methods offer alternative approaches to evaluate the results.

The dataset we have used is the MICCAI 2018 Left Ventricle Full Quantification Challenge dataset, which has images collected from three hospitals with subjects in a wide range of age and in presence of diverse pathologies. However, the dataset is limited to mid-cavity slice CMR images, which is reasonable for a preliminary experiment but does not cover basal and apical short-axis (SAX) slices or long-axis (LAX) slices. The reference metrics are calculated from the reference segmentation provided, but the evaluation method is unknown, therefore introducing additional error into our pipeline which could be avoidable in clinical practice.

Limitations of our work should also be noticed. The iterative implementation of the Seg-Reg Mode involves frequent changes of the loss function, which may cause interruptions during the optimization process, even with both segmentation and regression losses aiming to improve the estimation accuracy. Alternative loss functions (e.g. with fixed values of λ_S and λ_R) and different combinations of λ_S and λ_R should be investigated in the future for Seg-Reg Mode. Although the comparison between Seg-Reg Mode and Reg Mode shows the advantage of including the segmentation loss for DR method, the segmentation result of Seg-Reg Mode could also be investigated to better understand the performance of the network. We experimented with different number of convolutional kernels and decided to use the same parameter setting as the original U-Net, but the network architecture may be further optimized to this particular task by exploring different number of pooling stages, different number of layers per stage, different number of convolutional kernels, different kernel sizes, different

activation functions and other related parameters. For quantitative analysis of the segmentation result, global similarity metrics (such as Dice score) are used and demonstrate the advantage of SB method for evaluating LV indices. More complete analysis could be carried out by including Hausdorff distance, which highlights local differences between predicted and reference contours that cannot be reflected by global similarity metrics.

With the rapid development of deep learning algorithms, direct regression of LV indices begins to show promising results, and the end-to-end nature of such neural networks provides the exact information that clinician requires and any manual adjustments of the estimated indices (without the need for manual correction of the segmentation results) can be directly back-propagated to the network parameter for further development. However, with the state-of-the-art deep learning algorithms, SB methods are still more accurate than DR methods, which is more important for medical image analysis, and therefore the clinical utility of DR methods is still limited. By considering the advantages of SB methods compared to DR methods, in the following of the thesis, we focus on developing segmentation algorithms and cardiac function quantification methods based on segmentation.

Chapter IV. Segmentation Neural Network Development

In the previous chapter, we evaluated LV cardiac function using a direct regression method, a segmentation-based method and a multi-task approach which predicts clinical indices using segmentation features. We concluded that the segmentation-based method achieves the best performance. In this chapter, we further develop the segmentation neural network by implementing temporal dynamic modelling (TDM), comparing widely used network architectures (U-Net and VGG variations) and designing a network architecture targeted for better segmentation accuracy of mouse cardiac cine-MR images with both short-axis (SAX) and long-axis (LAX) slices.

We presented part of the work in this chapter as oral presentations at the Statistical Atlases and Computational Modelling of the Heart (STACOM) workshop of the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2018 conference in Granada, Spain, and the Statistical Atlases and Computational Modelling of the Heart (STACOM) workshop of the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2019 conference in Shenzhen, China. The corresponding papers were published in Lecture Notes in Computer Science [2, 129].

4.1 TDM for Cardiac Cine-MR

4.1.1 Introduction

TDM has been applied for both segmentation-based (SB) and direct-regression (DR) methods [100, 131, 133], with researchers claiming that it improved the results. The underlying theory is that for cine cardiac MR (CMR) images, each slice contains a sequence of frames usually covering a full cardiac cycle and TDM allows information propagating in time to assist segmentation or metrics evaluation. In this section, we introduce TDM using LSTM units [51] for our SB method (Seg-Mode described in section 3.3) and aim to evaluate the improvement we achieve with this modification.

We joined the MICCAI 2018 Left Ventricle Full Quantification Challenge with ensembled results using both the SB method and our modified SB-TDM method. The details of the methods are described in section 4.1.2 and the results are discussed in section 4.1.3.

4.1.2 Dataset and Method

We use the same dataset described in section 3.2 (training data of MICCAI 2018 Left Ventricle Full Quantification Challenge dataset) to develop and validate our SB-TDM method. When LSTM is used, we change the middle convolutional block (with the two 1024 Conv. (3, 3) layers shown in Figure 3.4) to a convolutional LSTM layer with the same size (1024 kernels of size 3×3) and maintain the size of bottleneck feature space. In this way we modify the multi-mode network proposed in section 3.3.1 to include an

additional mode: Seg-TDM Mode, as shown in Figure 4.1. For the Seg-TDM Mode, the FE-Net becomes a recurrent FCN (RFCN), due to the additional LSTM layer applied to the Middle Features (MDF), in a similar way to [100]. Data augmentation is applied consistently for the complete sequence, unlike in the other modes where data augmentation is applied to individual images. The loss function and metrics evaluation methods remain the same as in the Seg Mode introduced in Chapter 3.

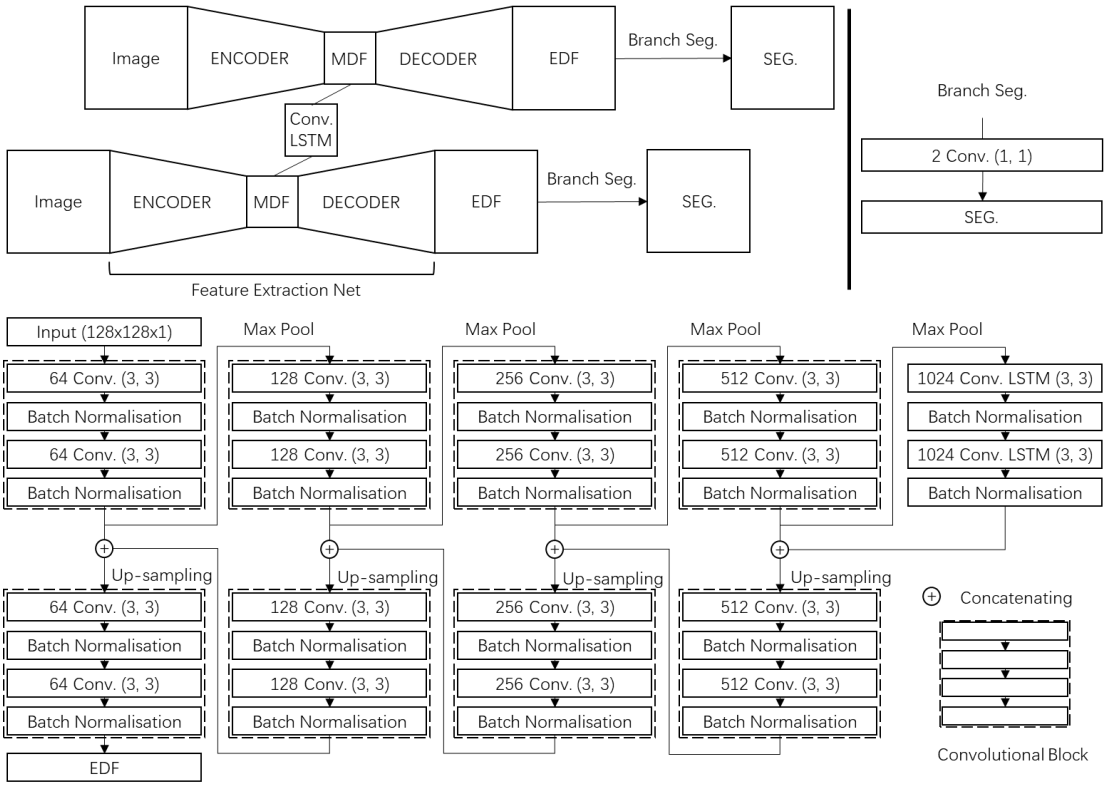


Figure 4. 1 Neural network architecture of Seg-TDM Mode. Multiple frames of the images are segmented simultaneously with our Seg-TDM Mode.

4.1.3 Experiments and Results

We performed five-fold cross validation (as described in section 3.4.2) for our Seg-TDM Mode experiments, and the results are shown in Table 4.1. There are three DR methods (Indice-Net, FullLVNet and DMTRL [131, 132, 134]) developed and

evaluated with the same dataset and published before the challenge, and we included these results in the table as the state-of-the-art methods.

We also included the best performing methods of the challenge [48, 65, 75], in order to compare to our method. The participants of the challenge are ranked based on the testing results, and we included all the methods awarded the first, second and third places.

Table 4. 1 Average results of compared methods. Values are shown as mean error ± standard deviation, followed by the correlation coefficient shown in brackets.

Dataset	Method	Avg. Areas (mm^2)	Avg. Dims (mm)	Avg. RWT (mm)	Phase Error (%)
Training and Validating Data	Indice-Net [132]	203 ± 133 (0.903)	NA	1.44 ± 0.71 (0.758)	NA
	FullLVNet [134]	190 ± 128 (0.937)	2.68 ± 1.64 (0.917)	1.41 ± 0.72 (0.761)	10.4
	DMTRL [132]	180 ± 118 (0.945)	2.51 ± 1.58 (0.925)	1.39 ± 0.68 (0.768)	8.2
	Seg Mode	120 ± 105 (0.951)	1.25 ± 1.06 (0.972)	1.04 ± 0.85 (0.908)	**7.8**
	Seg-TDM Mode	**116 ± 103** (0.952)	**1.23 ± 1.11** (0.972)	**1.03 ± 0.86** (0.908)	8.2
Testing	Seg/Seg-TDM	341 ± 178 (0.953)	2.67 ± 1.70 (0.979)	2.28 ± 1.29 (0.829)	10.3

Data	[75]	301 ± 165 (NA)	2.15 ± 1.61 (NA)	2.03 ± 1.23 (NA)	9.5
	[65]	301 ± 183 (NA)	2.33 ± 1.76 (NA)	2.15 ± 1.36 (NA)	10
	[48]	309 ± 170 (NA)	2.85 ± 3.45 (NA)	2.31 ± 1.48 (NA)	12.3

Comparing to Seg Mode, the mean absolute error (MAE) values of Areas, Dims, and RWT favour Seg-TDM Mode, but when we calculate the p-values (0.05, 0.71, 1.1e − 6 respectively) only RWT values show a significant (though small) difference between the two methods. Seg Mode produces a more accurate phase prediction.

Therefore, we ensembled the segmentations from all SB networks (15 Seg Mode and 15 Seg-TDM Mode) and calculated the LV indices using the method described in section 3.3.2.

4.1.3.1 Analysis of TDM

From the quantitative results we see very small improvements after introducing TDM, by comparing Seg Mode and Seg-TDM Mode. We then evaluate the results of Seg-TDM Mode qualitatively by visualising, in a three-dimensional representation, the evolution of the predicted two-dimensional contours over time, for both reference binary segmentation and Seg-TDM Mode. The 2D contours are stacked, with time providing the third coordinate (the vertical axis, ascending from top to bottom), as shown in Figures 4.2 and 4.3. We show three pairs of endocardial and epicardial surfaces in Figures 4.2 and Figure 4.3.

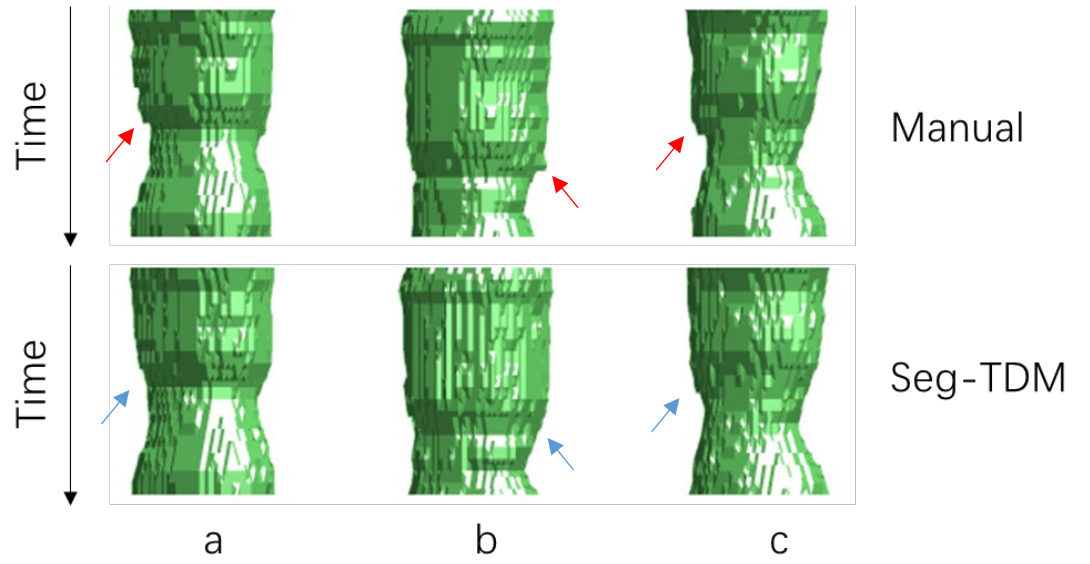


Figure 4. 2 Visualisation of endocardial surfaces for three representative cases (a), (b) and (c) are the iso-surfaces formed of three endocardial contours in time, generated from manual binary segmentations (top row) and the predicted binary segmentation using the Seg-TDM Model (bottom row).The red arrows indicate sudden changes of manual segmentations in time, and the blue arrows point at corresponding places in Seg-TDM Mode results.

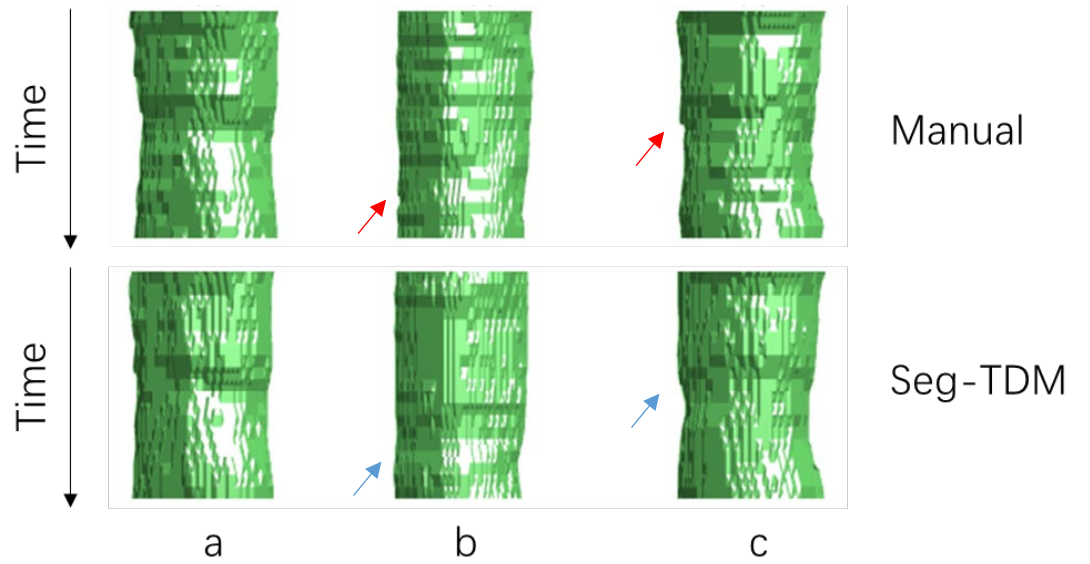


Figure 4. 3 Visualisation of epicardial surfaces for three representative cases (a), (b) and (c) are the iso-surfaces formed of three epicardial contours in time, generated from

manual binary segmentations (top row) and the predicted binary segmentation using the Seg-TDM Model (bottom row). The red arrows indicate sudden changes of manual segmentations in time, and the blue arrows point at corresponding places in Seg-TDM Mode results.

We can see that there are sudden changes in the reference surfaces from one frame to the next one (indicated by the red arrows), and this is caused by fundamental limitations of the manual segmentation procedure used here, in which only one frame is segmented at a time. We can see that the Seg-TDM networks learn a more general model and produce temporally smoother results (indicated by the blue arrows) compared to the reference surfaces.

Endocardial contours have larger deformations in time, and therefore the sudden changes are more obvious than epicardial contours. In Figure 4.4 to Figure 4.6, we show three adjacent frames, with both cine MR images and segmentation results, around the sudden changes of the endocardial contours in time for the three cases shown in Figure 4.2 and Figure 4.3, in order to compare manual segmentation and the segmentation produced by Seg-TDM Mode. Red arrows are used to indicate places with sudden changes of the manual segmented endocardial contours, and blue arrows are pointing at the corresponding places of the Seg-TDM results.

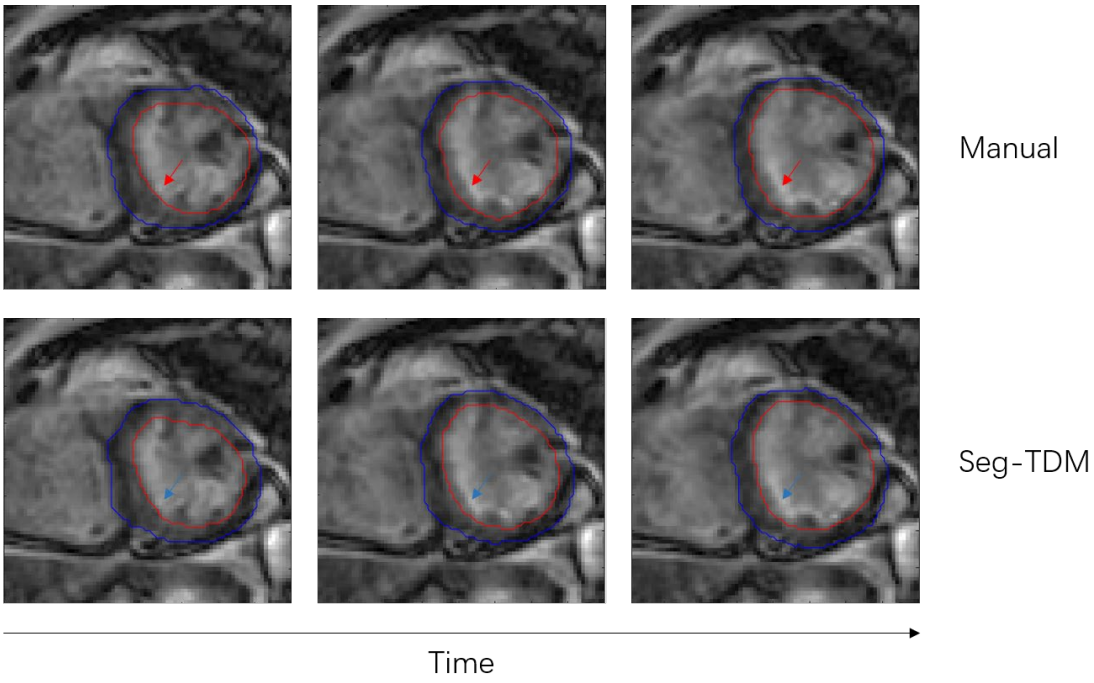


Figure 4. 4 Manual segmentation and Seg-TDM segmentation comparison for case (a). The red contours are endocardial contours and the dark blue contours are epicardial contours. The red and blue arrows indicate the places with sudden changes of manual endocardial contours and the corresponding places for Seg-TDM results.

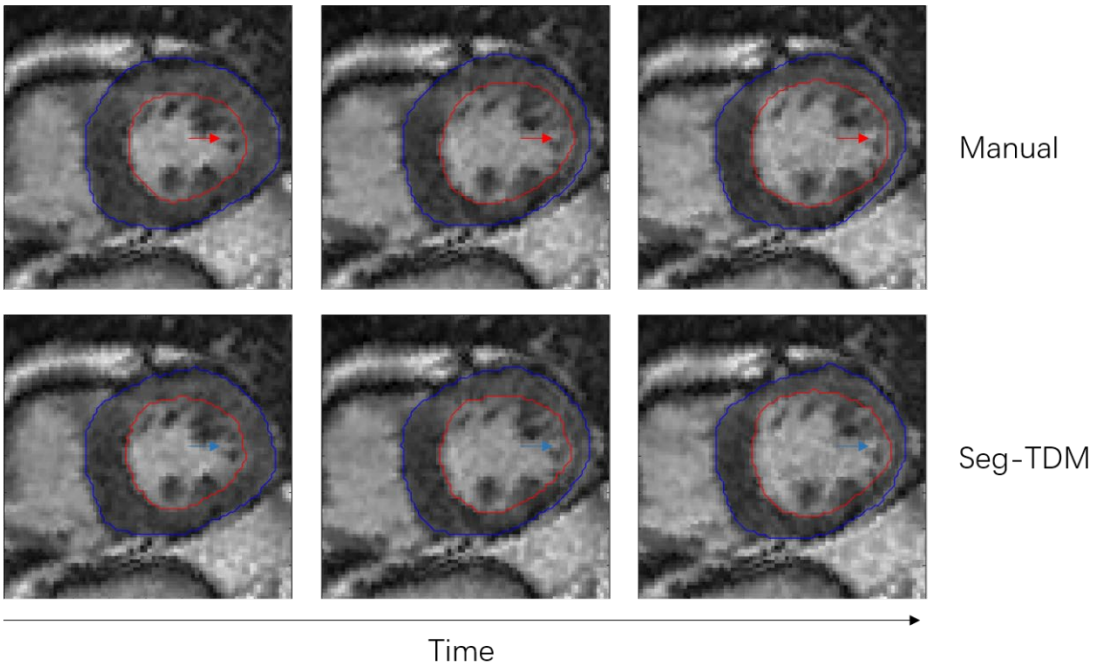


Figure 4. 5 Manual segmentation and Seg-TDM segmentation comparison for case (b). The red contours are endocardial contours and the dark blue contours are epicardial contours. The red and blue arrows indicate the places with sudden changes of manual endocardial contours and the corresponding places for Seg-TDM results.

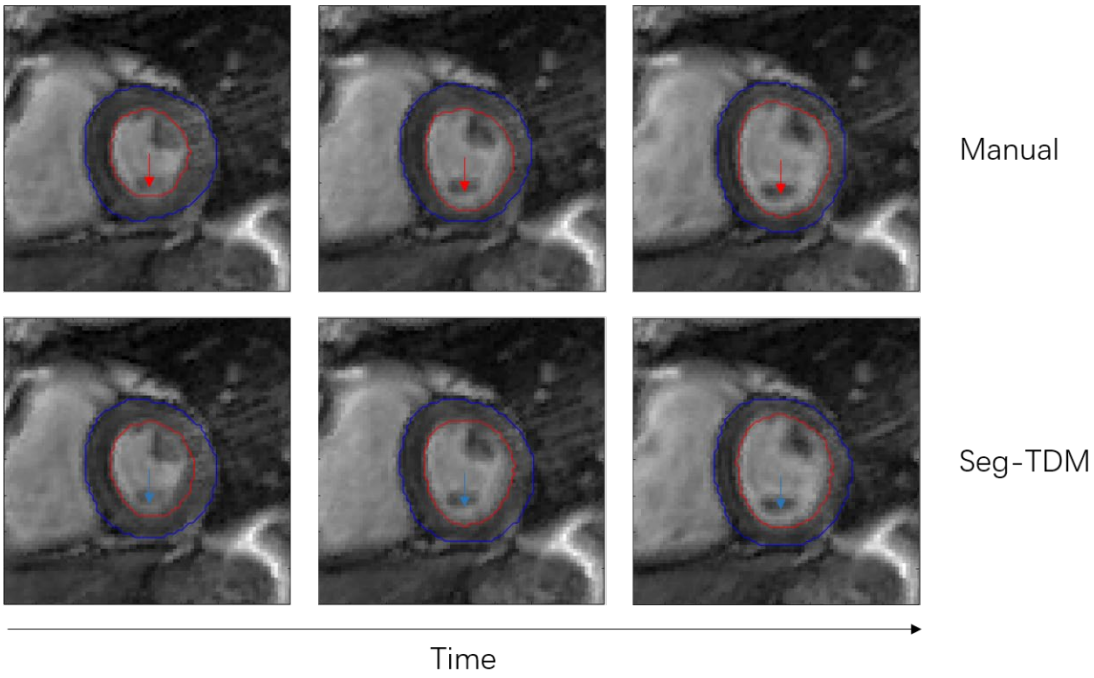


Figure 4. 6 Manual segmentation and Seg-TDM segmentation comparison for case (c). The red contours are endocardial contours and the dark blue contours are epicardial contours. The red and blue arrows indicate the places with sudden changes of manual endocardial contours and the corresponding places for Seg-TDM results.

Because unlike Seg-TDM Mode, Seg Mode segments each frame independently, and the quantitative analysis compares with the manual reference, Seg Mode is likely to produce more similar results to the reference in those cases skewing the quantitative validation values. By increasing the LSTM convolution kernel sizes and numbers or the number of recurrent layers, more complete patterns could be learnt by the network; however, that will raise the chance of over-fitting or gradient vanishing and increase the GPU memory consumption at the same time.

4.1.3.2 Testing Data Experiments

As described in section 3.4.2, for each of the five-fold cross validation experiments we

trained three randomly initiated networks from scratch, and in the testing stage we ensembled the segmentations from all SB networks (15 Seg Mode and 15 Seg-TDM Mode) and calculated the LV indices using the method described in section 3.3.2. The indices are then normalised following the procedure required by the challenge organisers (as described in section 3.3.2) before submission.

Surprisingly, from Table 4.1 we can see that compared to our cross-validation results on training data, the testing data results are 2 to 3 times larger in MAE values, while the correlation coefficients are similar or even slightly better for areas and dimensions. The combination of large correlation and increased error suggests a constant bias between our result and the reference values. We performed a visual inspection of the segmentation results and we show the first frame of all testing sequences in Figure 4.7 and Figure 4.8, where endocardial and epicardial contours are represented by red and blue contours. Since the reference binary masks are not available for the testing data, we cannot compare the segmentation results and the reference to evaluate our SB methods qualitatively by showing the similarity between them. However, by comparing the segmentation results with the training data (as shown in Figure 3.7), we can see that the quality of the segmentation of both datasets are similar to each other suggesting our segmentation methods also have a robust performance on the testing data.

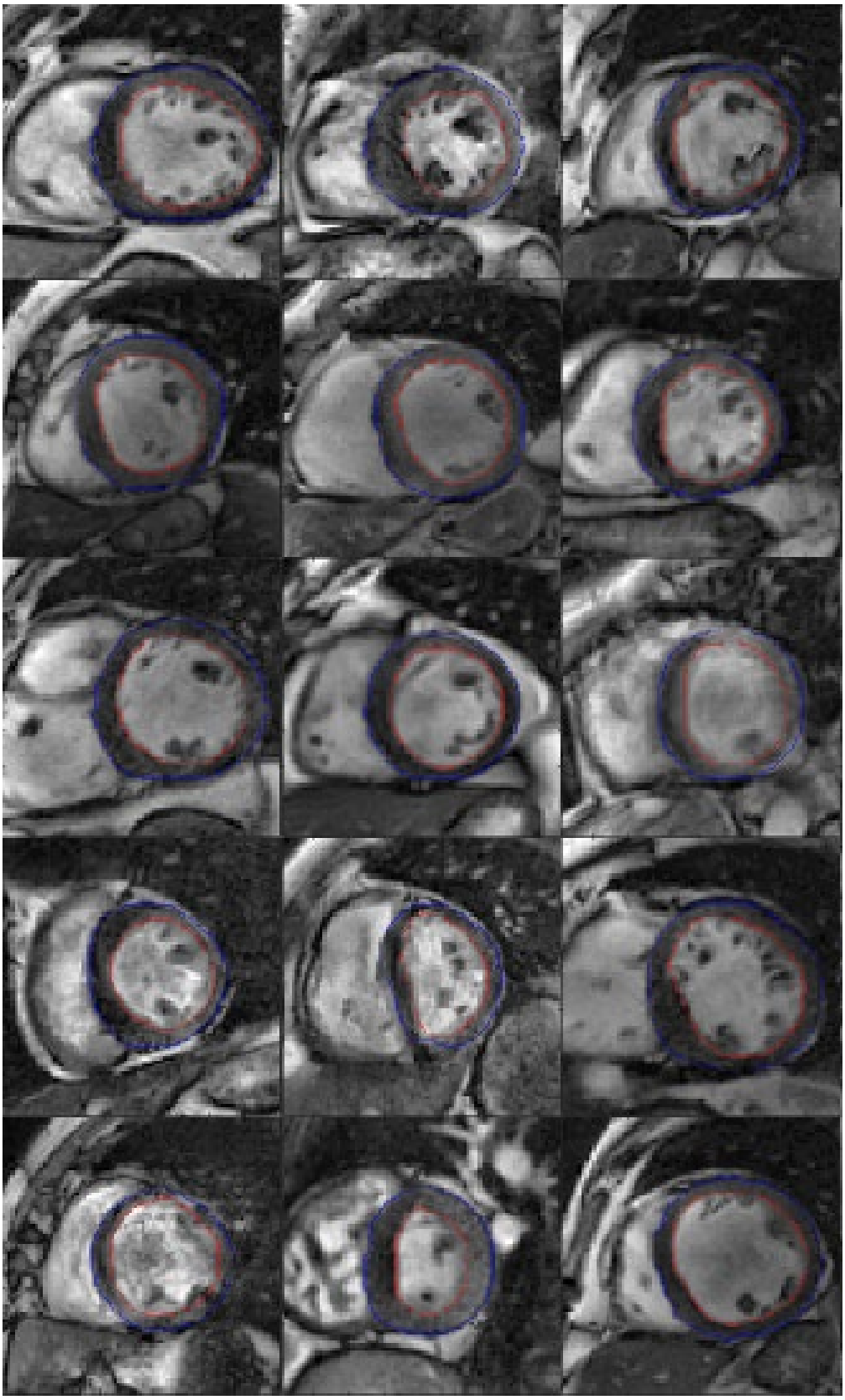


Figure 4. 7 Testing data (1 - 15) segmentation result visual inspection. The red contours

are the endocardial contours and the blue contours are the epicardial contours.

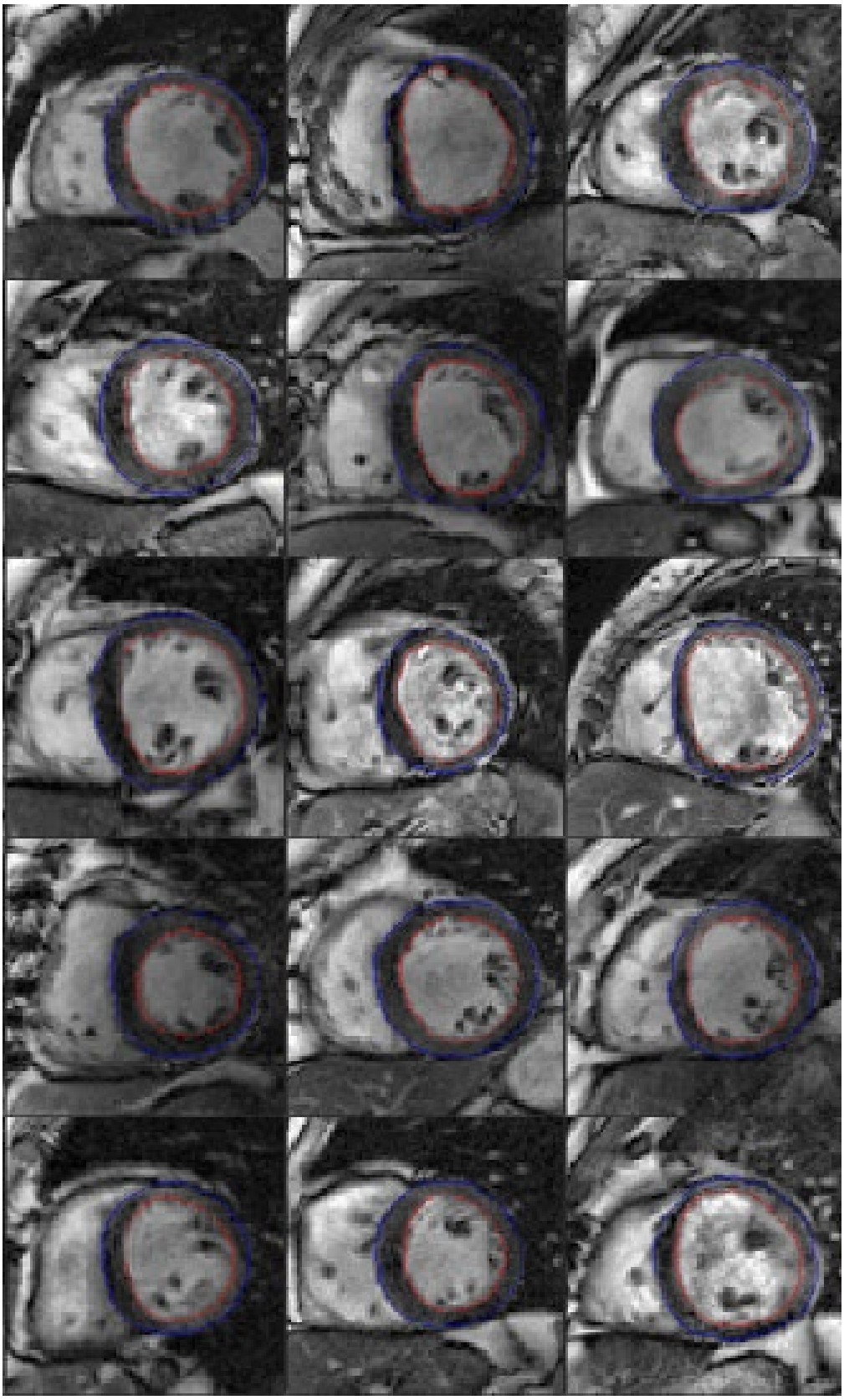


Figure 4. 8 Testing data (16 - 30) segmentation result visual inspection. The red

contours are the endocardial contours and the blue contours are the epicardial contours.

We could not appreciate the differences between our cross validation and the final testing results or locate the source of the bias. We see similar performance drop in testing data results from other challenge participants [48, 65, 75] as shown in Table 4.1, and suspect that it might have been introduced due to potential underlying differences between the references of training and testing datasets.

4.2 Comparison of Segmentation Networks

4.2.1 Introduction

Both U-Net [102] and VGG networks [111] have pairwise down-sampling and up-sampling paths with skip connections between them allowing multi-level features passing through. The key differences between the two networks are the depths of down-sampling and up-sampling paths; U-Net has equally distributed down-sampling and up-sampling paths while the VGG network has a deeper down-sampling path and a much shallower up-sampling path. The overall depth of the VGG network is shorter than that of the U-Net, and therefore VGG requires a smaller number of parameters, reducing the chance of overfitting during training. In this section, we compare the two networks with comparable designs of the architectures.

4.2.2 Dataset and Method

We developed and evaluated our pipeline using the MICCAI 2019 Left Ventricle Full Quantification Challenge dataset, which consists of 56 training subjects and 30 testing

subjects. For each subject in the training dataset, a single mid-cavity SAX CMR sequence consisting of 20 frames was provided together with a set of clinically significant LV indices including regional wall thicknesses, cavity dimensions, cavity and myocardium areas and cardiac phase for each frame. Endocardial and epicardial segmentation binary masks were also made available as reference, and pixel-spacing values of the images were also given for metrics evaluation. For subjects in the testing dataset, only CMR image sequences and pixel-spacing values were provided.

Compared to MICCAI 2018 Left Ventricle Full Quantification Challenge, which had 145 training subjects and 30 testing subjects, the size of the training dataset was reduced by 61.4% while the testing dataset remained the same size [129]. The smaller dataset makes it more difficult for network training, and the risk of overfitting is increased. The new dataset is also more challenging because the pre-processing stages are removed, and thus the LV is no longer centred or oriented in a consistent way as in the previous challenge images shown in Figure 3.2. A comparison between the two datasets is shown in Figure 4.9.

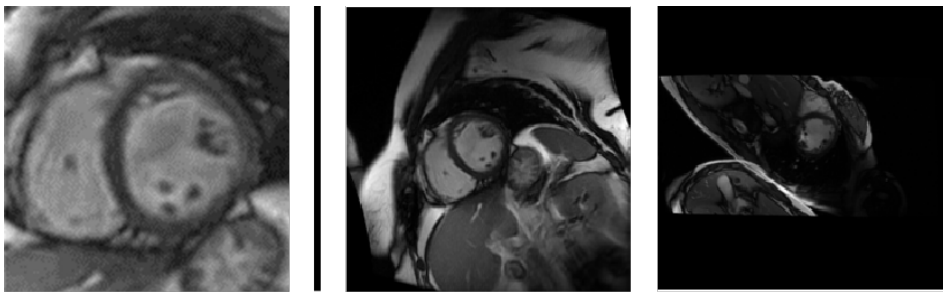


Figure 4. 9 Comparison between challenge datasets. (a) is an example from the MICCAI 2018 Left Ventricle Full Quantification Challenge dataset, which is pre-processed so that the heart is centred with the same orientation. (b) and (c) are

examples from the MICCAI 2019 Left Ventricle Full Quantification Challenge dataset, which have different sizes and orientations of the heart.

In Figure 4.10, we show 20 randomly chosen examples of the new training dataset showing the variation of the images without pre-processing. We propose a complete pipeline for quantifying the LV from cardiac MR (CMR) images [2], and in this section we focus on the comparison between two network architectures which take pre-processed images as input.

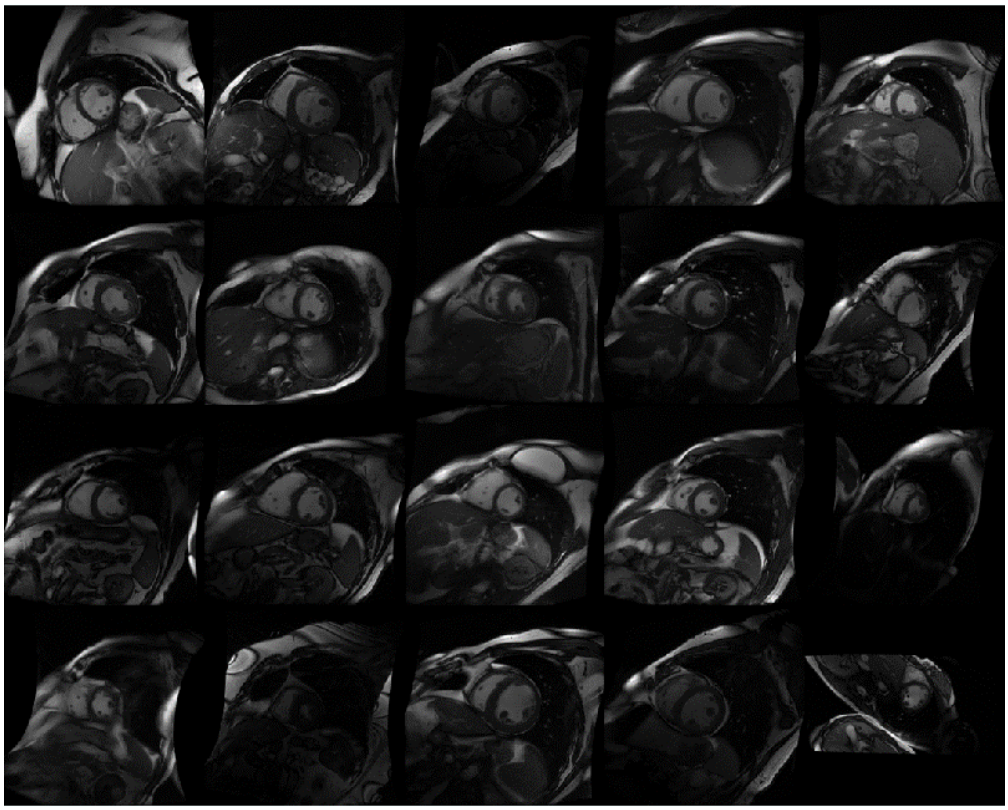


Figure 4. 10 Randomly chosen 20 subjects from the dataset; the first frames are displayed. The position and orientation of the LV varies within the dataset.

A variation of U-Net has been implemented in our multi-mode network described in section 3.3.1, and the network architecture is designed following the same parameter setting as the original U-Net. The segmentation results are accurate and robust, and

therefore we used the same network architecture to compare to our variation of VGG-16 network. We set the number of convolutional kernels for each down-sampling stage of our VGG network the same as our U-Net, and that results a big difference in the number of parameters between the two networks. To compensate for such difference we add one convolutional layer for each of the last two down-sampling stages in our VGG network.

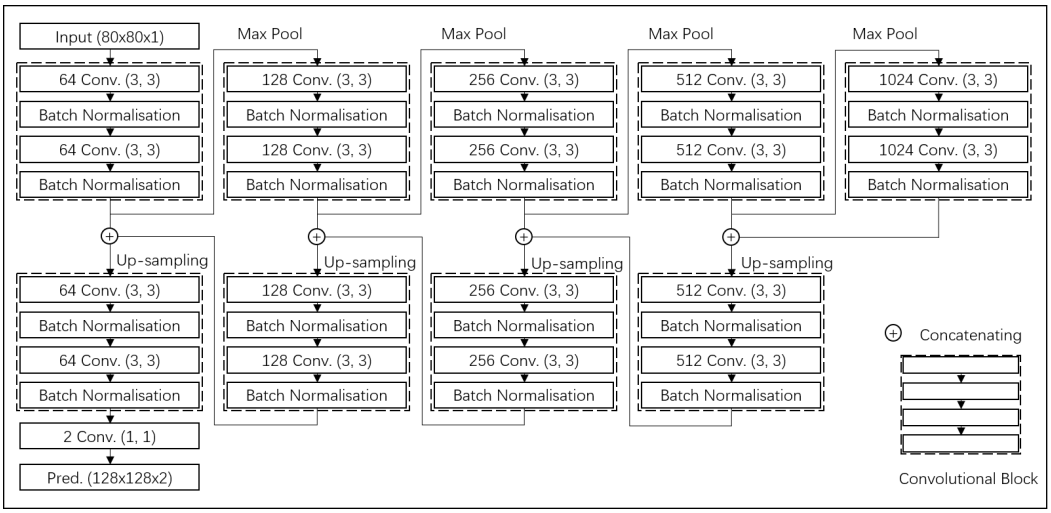


Figure 4. 11 Neural network architecture of the U-Net variation.

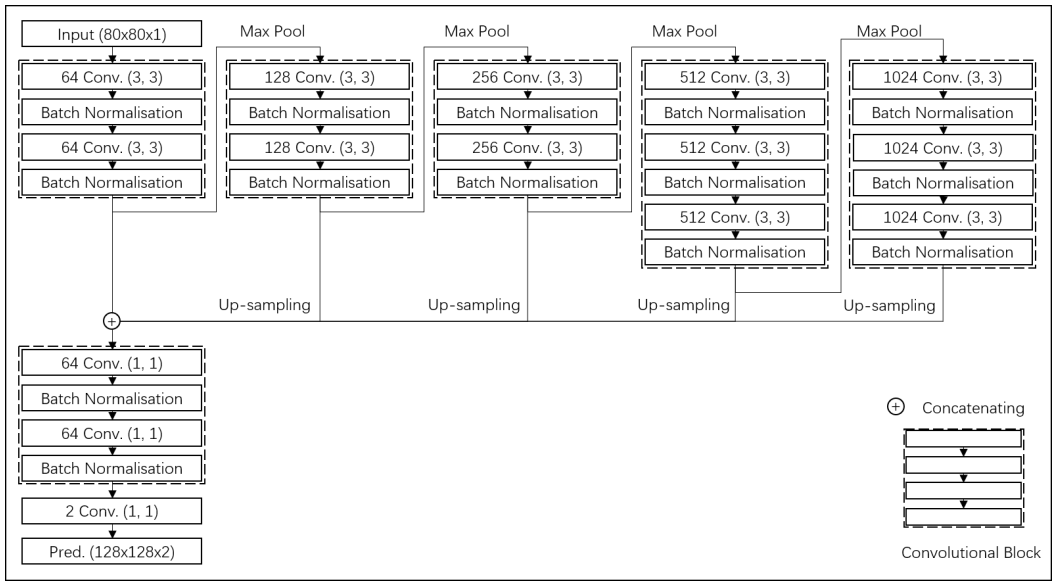


Figure 4. 12 Neural network architecture of the VGG-16 variation.

Details of our network architectures are shown in Figure 4.11 and Figure 4.12. For

both networks, there are four 2×2 max-pooling stages with the stride of 2×2, and the number of filters is 64, 128, 256, 512 and 1024 for each stage accordingly. The size of all kernels is 3×3 and the activation functions are ReLU [127] for all layers other than the output layer, which is a sigmoid. Our U-Net variation has a set of step-by-step up-sampling stages and our VGG variation concatenates up-samples from each scale in one step, which is the key difference between them.

We compare the performance of networks with different augmentation methods. The standard method is based on random flipping, rotations (in the range from 0 to 360 degrees) and translations (within ± 11 *mm* in horizontal and vertical axis of the image). The SMOD+ method is an augmentation approach based on statistical models of deformation in order to produce anatomically meaningful deformations within the space of plausible shapes. The development and evaluation of SMOD+ is not part of the contributions of this work, and this section is focused more on comparing the performance of U-Net and VGG network.

4.2.3 Experiments and Results

We performed 3-fold cross-validation experiments on the training dataset, with the size of each fold being 18, 19 and 19. The subjects were randomly assigned to one-fold, and for each cross-validation experiment we used 4 subjects as validation set and kept the rest as training set. We implemented the neural networks in Keras with Tensorflow as backend, with cross-entropy as the loss function and Adadelta [142] as optimizer. By keeping the pre-processing stages, data augmentation strategies (four data

augmentation approaches are implemented in our experiment), post-processing stages and metric evaluation algorithms the same [2], we were able to compare the segmentation performance of the network and the LV quantification performance in equivalent conditions. The results of these tests are shown in Table 4.1.

Table 4. 2 Average results of compared methods. Values are shown as mean error ± standard deviation. The best results of overall metrics are shown in bold. For details about the A, B, C and D methods, refer to main text.

	U-Net A	U-Net B	U-Net C	U-Net D	VGG A	VGG B	VGG C	VGG D	Test
Dice Endo	0.950 ±0.026	0.949 ±0.033	**0.953** ±0.023	0.950 ±0.026	0.938 ±0.045	0.941 ±0.031	0.943 ±0.032	0.941 ±0.031	NA
Dice Epi	0.966 ±0.014	0.965 ±0.019	**0.967** ±0.014	0.965 ±0.013	0.957 ±0.026	0.958 ±0.020	0.959 ±0.021	0.959 ±0.020	NA
A1 (mm^2)	102 ±87	107 ±111	92 ±83	100 ±83	101 ±121	95 ±77	90 ±79	101 ±84	184
A2 (mm^2)	132 ±105	142 ±105	121 ±98	135 ±105	140 ±121	155 ±131	126 ±110	140 ±112	525
Areas (mm^2)	117 ±97	125 ±109	**106** ±91	118 ±96	121 ±122	125 ±111	108 ±97	120 ±101	355
Dim1 (mm^2)	1.59 ±1.53	1.73 ±2.30	1.46 ±1.28	1.58 ±1.45	1.40 ±1.78	1.40 ±1.16	1.32 ±1.08	1.51 ±1.23	2.59

Dim2 (mm^2)	1.70 ±1.39	1.81 ± 2.18	1.53 ± 1.31	1.60 ± 1.38	2.17 ± 3.01	1.72 ± 1.48	1.89 ± 1.65	1.96 ±1.77	2.33
Dim3 (mm^2)	1.65 ± 1.32	1.63 ± 1.59	1.56 ± 1.28	1.64 ± 1.26	2.01 ± 2.54	1.76 ± 1.55	1.83 ± 1.62	1.81 ±1.48	2.40
Dims (mm^2)	1.65 ±1.42	1.72 ±2.05	**1.52** ±1.29	1.60 ±1.36	1.86 ±2.52	1.63 ±1.42	1.68 ±1.49	1.76 ±1.52	2.44
RWT1 (mm)	1.01 ±1.03	0.98 ±0.89	0.85 ±0.68	0.89 ±0.68	0.89 ±0.75	1.01 ±0.89	0.91 ±0.80	0.91 ±0.84	2.40
RWT2 (mm)	1.23 ±0.85	1.23 ±0.92	1.05 ±0.78	1.18 ±0.86	1.19 ±0.90	1.19 ±0.94	1.15 ±0.84	1.22 ±0.87	2.39
RWT3 (mm)	1.27 ±0.97	1.22 ±0.95	1.10 ±0.87	1.26 ±1.03	1.21 ±0.97	1.21 ±0.97	1.15 ±0.92	1.26 ±1.00	2.20
RWT4 (mm)	1.20 ±0.90	1.27 ±0.97	1.21 ±1.02	1.23 ±1.00	1.22 ±0.99	1.16 ±0.93	1.13 ±0.91	1.29 ±1.10	1.91
RWT5 (mm)	0.91 ±0.75	1.02 ±0.79	1.00 ±0.74	0.97 ±0.78	1.10 ±1.30	0.93 ±0.78	0.99 ±0.89	1.14 ±0.97	1.98
RWT6 (mm)	0.91 ±0.76	0.85 ±0.72	0.84 ±0.66	0.88 ±0.67	1.08 ±1.31	1.02 ±0.78	0.99 ±0.92	0.93 ±0.74	2.21
RWT (mm)	1.09 ±0.9	1.10 ±0.89	**1.01** ±0.81	1.07 ±0.86	1.12 ±1.06	1.09 ±0.89	1.06 ±0.89	1.12 ±0.94	2.18
Phase (%)	7.9	8.3	8.0	8.1	8.2	**7.8**	8.4	8.2	9.5

The four data augmentation strategies are the standard method (A) and SMOD+ approach [3] with σ = 1, 2, 3 (B, C, D respectively) specified in [2].

Comparing the two networks with the same strategy, U-Net performs better than the VGG network in general. For the best performing strategy (C), U-Net outperforms VGG network for both Dice similarity metrics and LV quantification metrics, suggesting that removing multiple stages of up-sampling has a negative effect on the segmentation task, even though the VGG network has a deeper down-sampling stage, and the network architecture of U-Net has a more balanced parameter distribution.

4.3 Segmentation of Mouse CMR

4.3.1 Introduction

In previous sections, we develop and evaluate our segmentation method using human cine CMR images. For mouse CMR, imaging is more challenging due to the much smaller size of the heart, resulting in difficulties for the development of segmentation neural networks, such as the lower contrast and signal-to-noise ratio (SNR) of the images as shown in Figure 4.13 to Figure 4.15.

Furthermore, methods for automated segmentation of CMR images have been largely focused on the SAX slices. This is because for cardiac functional analysis, a stack of SAX slices is usually used as the gold standard with manually delineated endocardial and epicardial contours [54, 103]. Mid-cavity slices cover most of the cavity and are the most important for volume or mass estimation, while the accuracy of segmentation results is typically higher than in apical and basal slices. LAX slice

images assess the cardiac chambers from a different angle, and they are critical when evaluating LV volume and myocardial mass in three-dimensional (3D) models for both human and mouse hearts [86, 119, 137]. However, the segmentation approach of the LAX images usually adapts the same network structure for SAX slice segmentation. In this section, we further develop our method using mouse cine CMR images, for both SAX and LAX slices.

4.3.2 Dataset and Method

The dataset consists of SAX and LAX cine CMR images of 100 mice, with both Sham and Infarct cases, acquired using a 9.4T (400 *MHz*) MR system (Varian Inc., Palo Alto, CA) comprising a horizontal magnet (bore size 210 *mm*), a VNMRS Direct Drive™ console and a shielded gradient system (1000 mT/m, rise time 130 μs).

For each subject there is a stack of SAX slices (with the number of slices between 7 and 13) and 2 LAX slices: a four-chamber (4-CH) view and a two-chamber (2-CH) view, forming a total number of 22167 images with all time frames. We standardised our dataset by resizing all images to 256 × 256 with image resolution of 0.1 × 0.1 mm^2 (if necessary when the original image size was 512 × 512 with image resolution of 0.05 × 0.05 mm^2). For each cardiac cycle, the number of frames varies between 10 and 36, containing both the end-diastolic (ED) frame and end-systolic (ES) frame while the whole cycle might not be covered. We also normalised the intensity range of the images to 0 and 255. Examples of stacks of SAX images are shown in Figure 4.13 and Figure 4.14 for ED and ES frame respectively. Examples of LAX images are shown in Figure

4.15. The low contrast to noise ratio and artefacts presented in the image making the process of mouse CMR images more challenging compared to clinical CMR images as shown in Figure 4.7, Figure 4.8 and Figure 4.9.

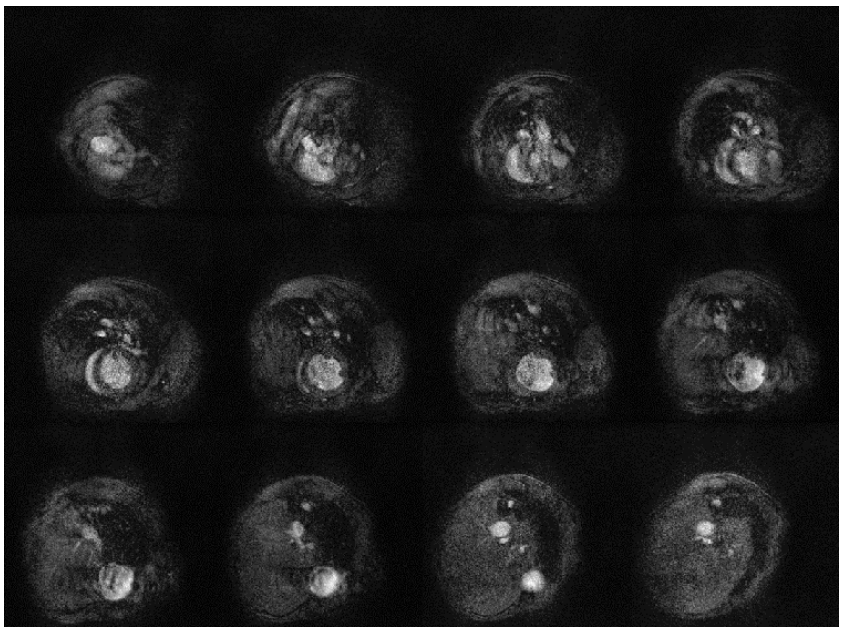


Figure 4. 13 SAX slice images of the ED frame. The images are from basal to apical slices. The images are noisy compared to clinical CMR images.

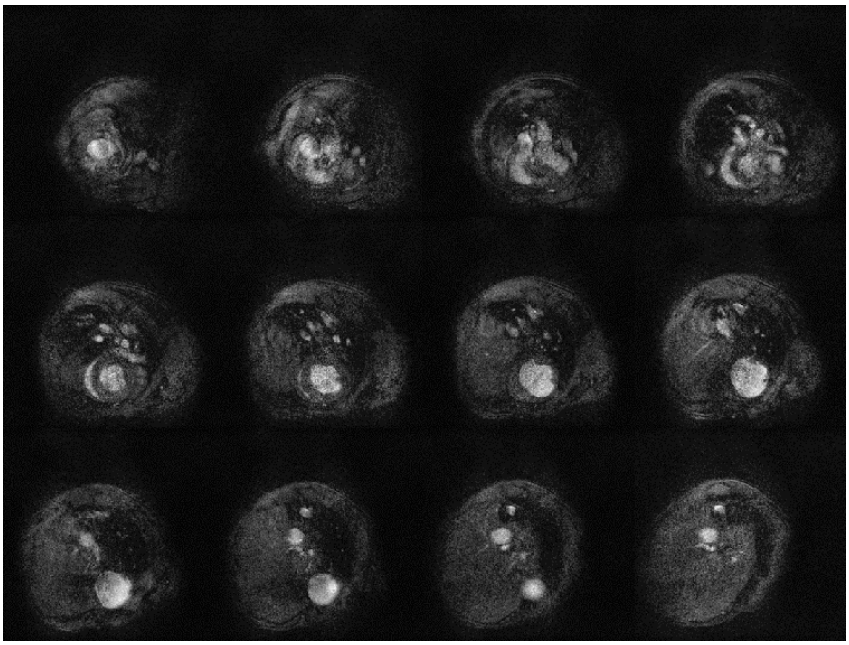


Figure 4. 14 SAX slice images of the ES frame. The images are from basal to apical

slices. The images are noisy compared to clinical CMR images.

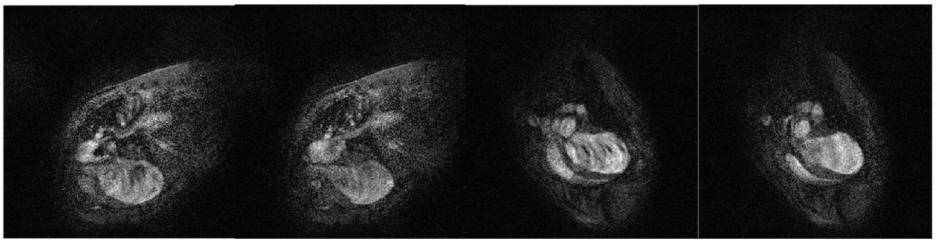


Figure 4. 15 LAX slice images of both ED and ES frames. The images from left to right are: 2-CH ED, 2-CH ES, 4-CH ED and 4-CH ES. The images are noisy compared to clinical CMR images.

The size of the input image of the segmentation neural network is 128 × 128, which is large enough to cover the LV for all images. Augmentation is applied for both training and validating data, starting by placing the centre of the LV (calculated from the intersecting points between slices) in the middle of the input image padded with zeros if the LV is too close to the sides of the image, followed by random rotation (0 to 360 *degrees*) and rigid transformation along rows and columns by ± 10 *pixels*.

Unlike the challenge datasets, which use manually traced contours as the gold-standard in the study, here we use the CIM software package (with guide-point modelling or GPM technique [129] introduced in section 2.4.6) for the analysis of LV function, which takes manually introduced guide-points to optimize the LV geometry iteratively and produces a pair of endocardial and epicardial surfaces. The CIM model is therefore considered as the reference and used for preparing the output binary masks by first calculating the physical positions of pixels for all image planes using their meta data and assigning them the corresponding class according to the spatial relationship regarding to the CIM model. Pixels outside the epicardial surface, pixels between

endocardial and epicardial surfaces and pixels within endocardial surfaces are labelled as background, LV myocardium and LV cavity correspondingly, forming an output layer with three channels in the same order. Examples of reference segmentation generated from three CIM models are shown in Figure 4.16 to Figure 4.25, with basal, mid-cavity and apical SAX slices and 2-CH and 4-CH LAX slices for both ED and ES frames. The first two columns on the left are examples Sham hearts and the last column is an example of an infarct heart.

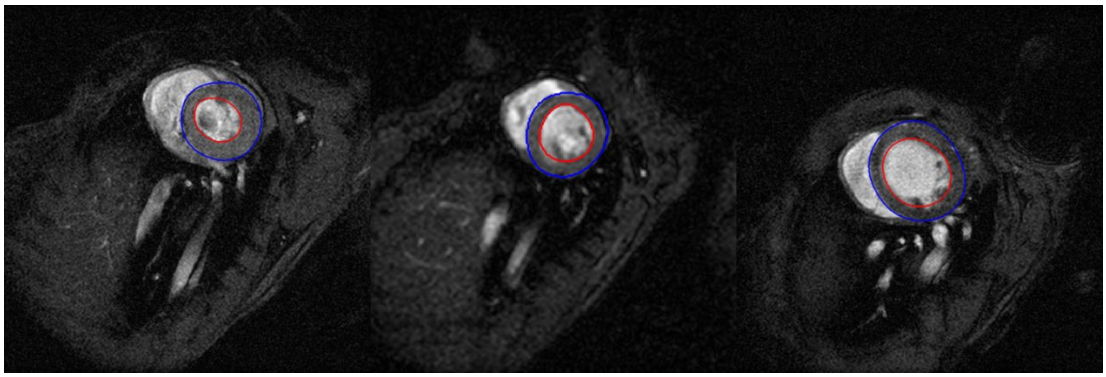


Figure 4. 16 Examples of LV reference contours: basal slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

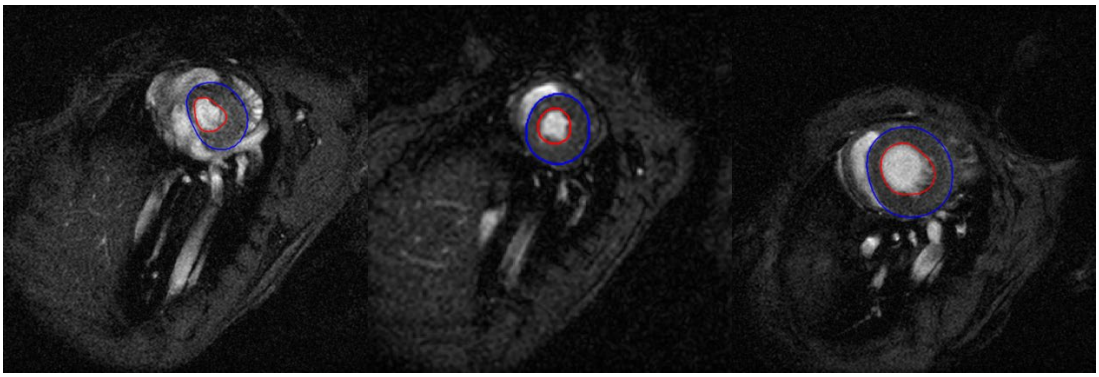


Figure 4. 17 Examples of LV reference contours: basal slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

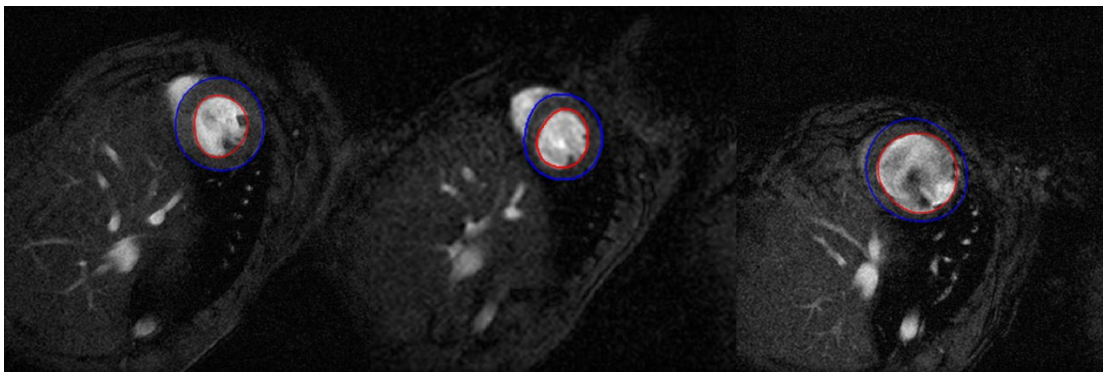


Figure 4. 18 Examples of LV reference contours: mid-cavity slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

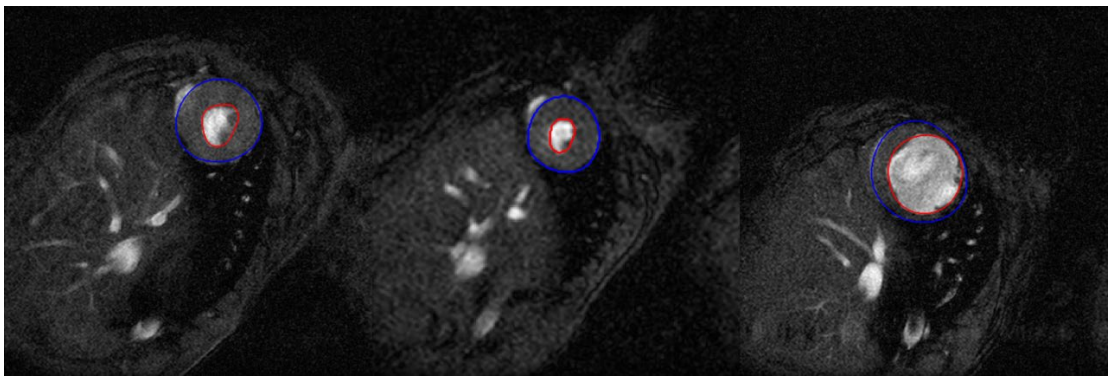


Figure 4. 19 Examples of LV reference contours: mid-cavity slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

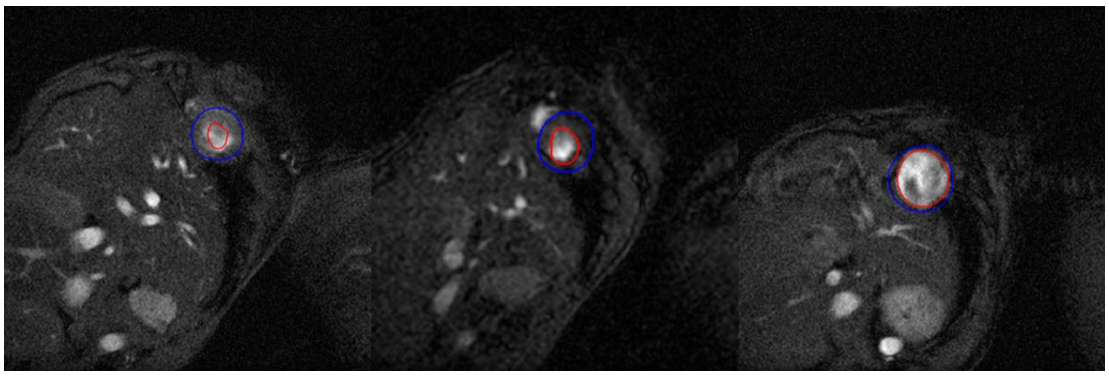


Figure 4. 20 Examples of LV reference contours: apical slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

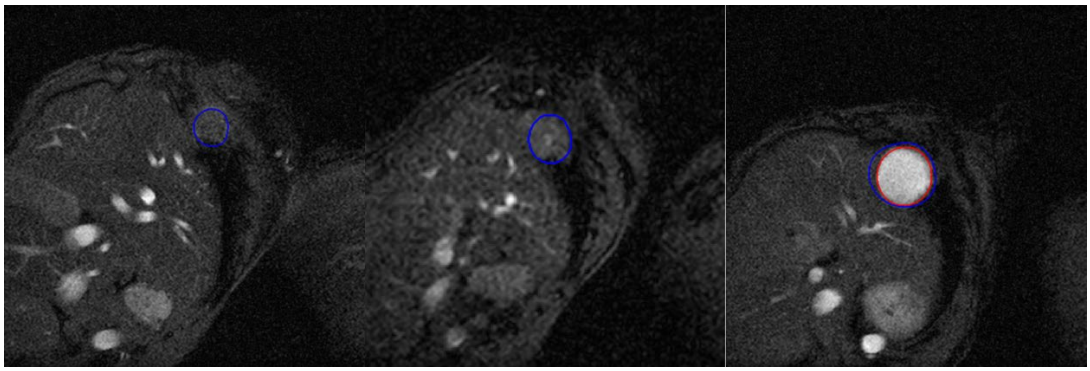


Figure 4. 21 Examples of LV reference contours: apical slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

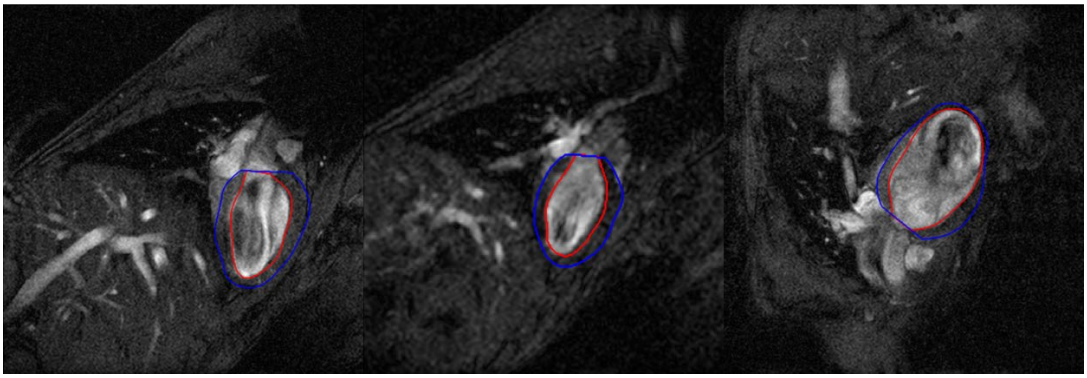


Figure 4. 22 Examples of LV reference contours: 2-CH slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

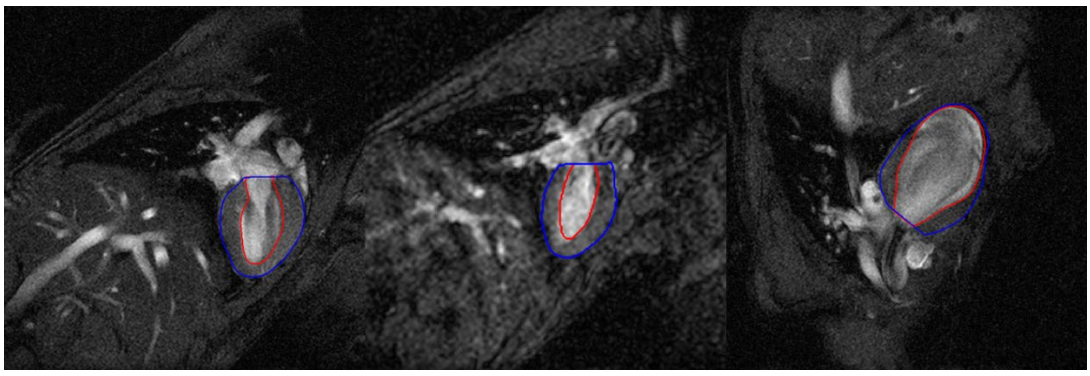


Figure 4. 23 Examples of LV reference contours: 2-CH slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

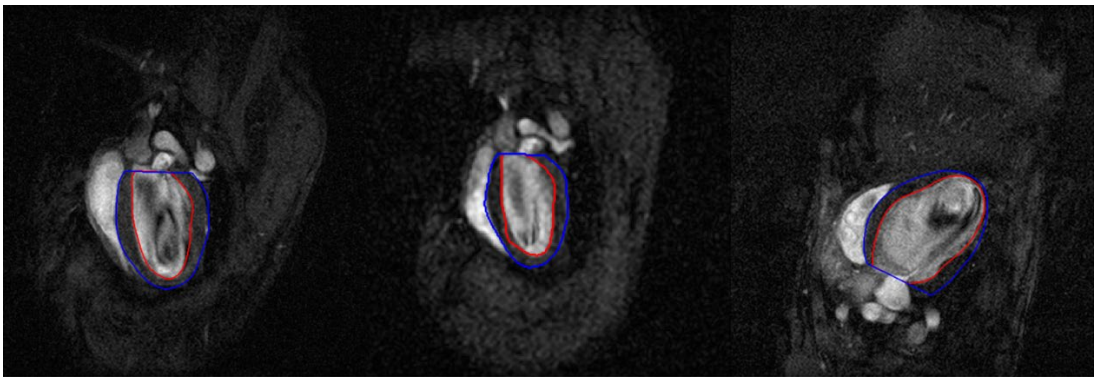


Figure 4. 24 Examples of LV reference contours: 4-CH slice, ED frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

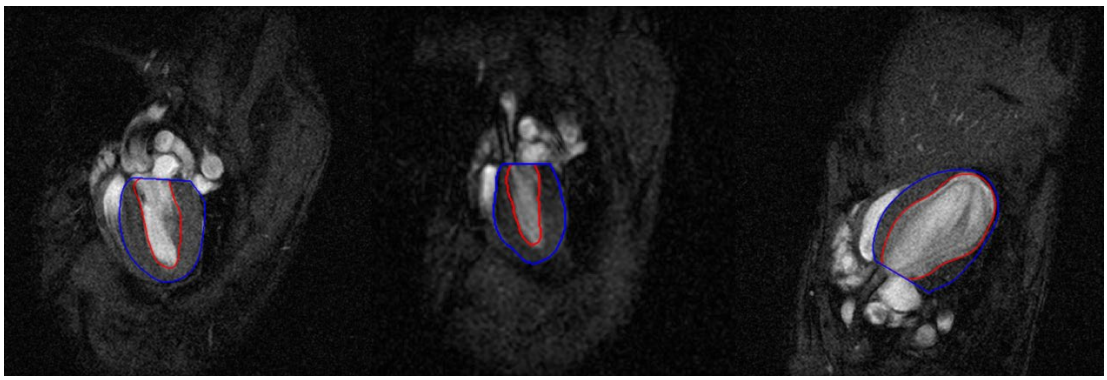


Figure 4. 25 Examples of LV reference contours: 4-CH slice, ES frame. The red contours are the endocardial contours and the blue contours are the epicardial contours.

From the examples of Sham images, good contractions can be visualised from both the myocardial wall thickening and the shortening of the LV, which can be observed from both LAX slices and SAX slices (especially the basal and apical slices) by comparing the ED and ES frames. For the infarct heart example, both the thickening of the myocardium and the shortening of the LV are not obvious. It should also be noticed that the myocardial wall becomes very thin around the apex and remains the same for both ED and ES frame, which can be visualised in Figure 4.15 to Figure 4.25 where the blood pool regions disappear in the ES frame for the Sham hearts due to contraction while the contours are almost the same for the infarct heart. The geometry of the heart

varies a lot across the database, which further increases the difficulty of the segmentation task in addition to the image quality.

To tackle this more challenging task, we made additional modifications to the network architecture (shown in Figure 4.26) inspired by the 3D U-Net in [28], by doubling the number of filters of the second convolutional layer for each down-sampling stage before max-pooling, to avoid representational bottlenecks within the network. The input layer has a value of N equal to 1, and the output layer has three convolutional filters giving three layers of the prediction representing the background, myocardium and blood pool of the LV.

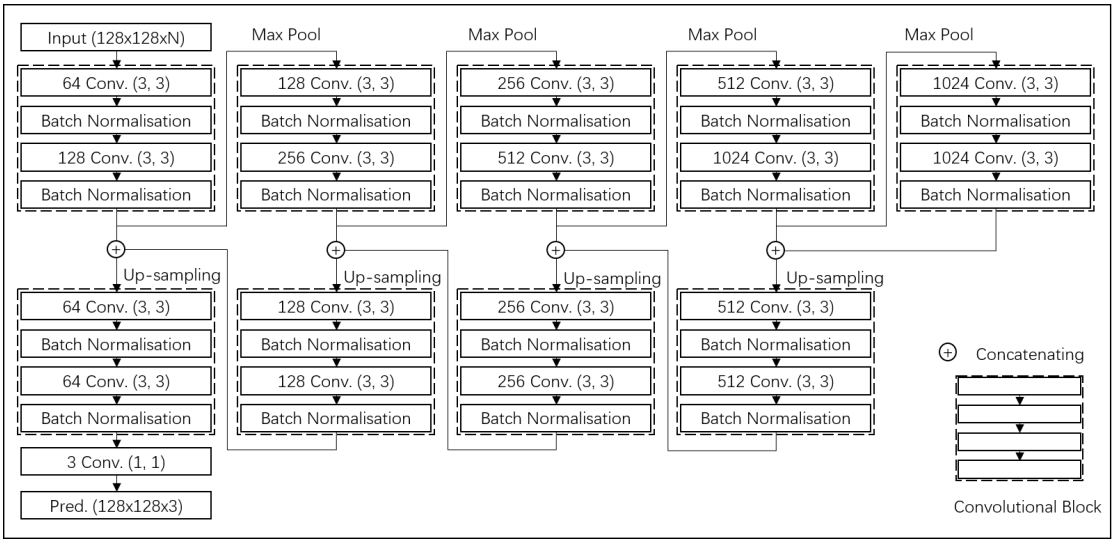


Figure 4. 26 Segmentation neural network architecture.

4.3.3 Experiments and Results

We performed a 5-fold cross-validation experiment with 20 test cases for each fold, dividing the remaining 80 cases into training (70 cases) and validation (10 cases). We implement the neural networks in Keras with Tensorflow as backend, with Dice similarity as the loss function and Adadelta as optimizer.

To obtain a comparison between the performance of segmentation neural network for human and mouse CMR images, we first evaluate the mid-cavity slice using the Dice similarity metric. The Dice metrics for endocardial and myocardial regions have the mean and standard deviation values of: 0.94 ± 0.07 and 0.89 ± 0.07 respectively. The results for human and mouse CMR images are similar for mid-cavity slices, while the standard deviation values of human CMR data are larger.

We then further evaluate all SAX slice images (with LV cavity or myocardium presented in the slice according to the reference label masks) by considering each image independently (19918 images in total). The Dice metrics for endocardial and myocardial regions have mean and standard deviation values of: 0.88 ± 0.21 and 0.81 ± 0.22 respectively. There is a clear drop of Dice values after including basal and apical slices, as the size of the regions are smaller, and those images are more challenging to segment compared to mid-cavity slices.

Segmentation results are usually validated on individual slices, and in order to estimate cardiac function of the LV, a full stack of the SAX slices must be summarised for evaluation. Considering each frame independently for Dice metric calculation (2444 frames in total), but accumulating all slices from each frame in a single value, results in Dice metrics for endocardial and myocardial regions with mean and standard deviation values of: 0.91 ± 0.05 and 0.85 ± 0.06 respectively. The summation of 2D SAX slices has its limitations because the slice thickness is approximately ten times larger than in-plane pixel spacing. However, the segmentation of the SAX stack is a good approximation of the 3D representation for functional analysis of the LV and is

usually employed as gold standard. Therefore, we also calculated the mean and standard deviation of the absolute percentage error between the estimated and reference values of the cavity volume and myocardial mass, obtaining values of: 8.1 ± 9.1 % and 7.5 ± 7.6 % respectively.

We also evaluated the segmentation results of the LAX slices, and the Dice metrics for endocardial and myocardial regions have the mean and standard deviation values of: 0.92 ± 0.10 and 0.80 ± 0.12 respectively. The positions of LAX slices are intended to be placed around the centre of the LV along the LAX, which is similar to the mid-cavity SAX slice except for the respective orientation of the slices, which are approximately perpendicular to each other. The Dice score for the LV cavity is comparable to the mid-cavity slices; however, for segmentation of the myocardium the difference between the mean values is obvious. From the results, there is the motivation of further improving the network performance for LAX slices segmentation, especially on the myocardial segmentation, and in the next section we further develop our method by feeding back SAX segmentation results into LAX segmentation to improve the accuracy and robustness of the method.

4.3.4 Segmentation of LAX slices of mouse hearts

To improve LAX slice segmentation, we modified the segmentation by concatenating a second layer (named as projected segmentation layer), constructed using the segmentation of the SAX stack, to the input LAX slice. SAX segmentation shows good performance (see previous section), and can thus be used to guide LAX segmentation

which is more challenging. The projected segmentation layer is calculated by assigning the segmented label of the SAX slices (background, myocardium and blood pool, labelled as 0, 1 and 2 respectively) into the pixels corresponding to the intersection with the LAX slices. Since the SAX slices are placed with a large interslice separation, the majority of the pixels in the LAX slices are not assigned any SAX labels and are thus assigned to background (labelled as 0). To match the intensity values, the range of label values of the projected segmentation layers are then normalised to 0 and 255, and to fit the enhanced input image, the input layer parameter N shown in Figure 4.26 is changed to 2. Paired projected segmentation layers of LAX slice images shown in Figure 4.15 are shown in Figure 4.27.

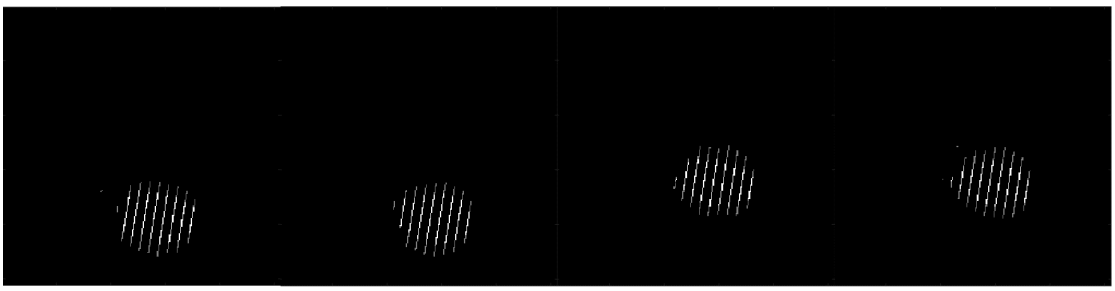


Figure 4. 27 Projected segmentation layers. Black, grey, and white pixels are corresponding to background, myocardium and blood pool respectively.

We performed the same 5-fold cross-validation experiment, and the segmentation results were evaluated using Dice scores. The mean and standard deviation values for endocardial and myocardial regions were improved to: 0.94 ± 0.07 and 0.84 ± 0.08 respectively. The mean values improve significantly, especially for myocardial regions, with the mean value of the Dice score for LAX endocardial segmentation matching the mid-cavity SAX slices. The standard deviation values also improve for both regions, suggesting the method becomes more robust with the enhanced input layer.

4.4 Conclusions

In this chapter, we summarise the development of the segmentation network. Following the previous chapter, we introduce TDM into our segmentation network using an LSTM layer, and evaluate the improvement using the LV quantification metrics. Compared to Seg Mode, there is a slight improvement after introducing TDM. We then plotted the surfaces formed by endocardial and epicardial contours contracting in time and show that the surfaces constructed using reference binary masks have worse temporal smoothness, which is presented in the surfaces formed with TDM results. Even though TDM shows limited benefits for the 2018 LV quantification challenge dataset when validated quantitatively, the main reason behind that is the reference binary masks are segmented frame by frame individually, showing the importance to have a better reference than manual contours in 2D. Considering the similar performance of both SB methods, we ensemble all 30 networks for testing data, and the segmentation results show a robust performance of our method by visual inspection. However, the big difference of the MAE values between training and testing datasets of the challenge cannot be fully explained, but we suspect that one of the main reasons is the underlying differences between the manual reference of the two datasets. Five-fold cross validation with a good proportion of validating data has been applied to our experiments of the training dataset of the challenge in order to prevent the networks from overfitting to the proportions of training data, and the results of the experiments are promising. When we apply our networks to the testing dataset of the challenge, there is an obvious performance drop. This is supported by the fact that several other challenge participants

experienced the same issue, and such differences between training and testing datasets should be prevented by the challenge organizers.

We then investigated the distribution of parameters of the network by implementing two widely used segmentation networks with comparable parameter settings. U-Net has evenly distributed down-sampling and up-sampling paths with multiple stages connected by skip layers, and the VGG-16 network has a deeper down-sampling path but an extremely short up-sample path with only one step. The number of parameters of our U-Net is larger than our VGG-16 network, and therefore the U-Net has the potential of better performance with a big dataset, while the risk of overfitting is also higher when the size of the dataset is constrained. The overall performance of the U-Net is better than that of the VGG-16 network when validated using the 2019 LV quantification challenge dataset, which is much smaller compared to large databases such as the UK Biobank, and the U-Net is a better option when the size of the dataset is comparable or larger than the challenge dataset. Therefore, we keep a variation of the U-Net for the following chapters.

We further developed our U-Net by referring to previously published versions of the 3D U-Net and validated the network with mouse cine CMR images. Even though the image quality is not as good as that of human CMR images, the network performance for mid-cavity slices is comparable to our segmentation networks developed using the LV quantification challenge datasets. The segmentation results for apical and basal SAX slices are worse than those of mid-cavity slices, as the Dice similarity score drops when all SAX slices are evaluated. However, since the LV cavity

and myocardial volumes in apical and basal slices are smaller than mid-cavity slices, the mid-cavity slices have more impact for cardiac functional analysis of the LV. We then integrate the SAX stack for each frame to estimate LV volume and myocardial mass, and we see good similarity between our segmentation results and the references binary masks.

For LAX slice segmentation, implementing the same neural network as for the SAX slices is the most straightforward and widely applied approach. However, for each case there are usually two LAX slices and around ten SAX slices, and by considering the amount of training data the accuracy of the segmentation network is expected to be higher for mid-cavity SAX slices than LAX slices, which have similar orientation and image quality as shown in previous sections. In our experiments of mouse CMR images, the performance of the network on mid-cavity SAX slices was better than LAX segmentation network, especially for the prediction of epicardial contours. We propose to utilise information from SAX slices for LAX segmentation and see clear improvements after introducing an additional layer to the input image constructed using SAX segmentation results, which also suggests for mouse CMR images the SAX and LAX are aligned well in space.

We finalise the development of our segmentation network after evaluating TDM and a variety of network architectures on both human and mouse CMR images. From these results, we conclude that a variation of U-Net provides the best overall performance and is used for 2D segmentation in the following chapters.

Chapter V. A Neural Network for Reconstruction of 3D Shapes from Sparse Contours

In previous chapters, we presented our segmentation method in two dimensions (2D). However, estimating cardiac function from integrating 2D short-axis (SAX) slices is limited due to inter-slice distance, which affects the accuracy of the three-dimensional (3D) reconstruction, particularly at the areas close to the apex and base where the curvature is higher. To fully integrate information of the complete dataset requires the development of a method for reconstruction of 3D ventricular surfaces from 2D cardiac MR (CMR) images, which is challenging due to the sparsity of the input data and the presence of interslice misalignment. We have developed and validated our method using a statistical shape model on both reconstructing 3D shapes from spatially consistent and misaligned input data, and datasets from the DETERMINE database to further investigate the effect of introducing LAX slices as well as the implementation of an adversarial loss component for the reconstruction neural network.

We presented part of the work in this chapter as an oral presentation at the Functional Imaging and Modelling of the Heart (FIMH) 2019 conference in Bordeaux, France, and it was published in Lecture Notes in Computer Science [130].

5.1 Introduction

The generation of anatomically accurate 3D surface meshes has shown promising uses in a wide range of applications including cardiac function analysis, interventional guidance and diagnosis [73, 113, 121]. Personalization of cardiac surfaces in 3D is also the first step required for computational simulations of cardiac electromechanics using the finite element method [9, 31, 141]. CMR imaging provides accurate shape information of the heart non-invasively [113]. A standard clinical CMR study includes a stack of SAX slices, covering at least from the left/right ventricular (LV/RV) apex to the base, plus at least two LAX views: horizontal long-axis (HLA, also known as four chamber view or 4-CH) and vertical long-axis (VLA, also known as two chamber view or 2-CH) [8].

Classical isosurfacing algorithms cannot be directly used due to the sparsity of the input data and because of the presence of motion artefacts (misalignment between slices caused by multiple breath holding and possible body movement during acquisition) [120], which make the task of reconstructing the 3D structure from CMR data particularly challenging. Examples of different levels of misalignment are shown in Figure 5.1.

Reconstruction of 3D surfaces from CMR data is normally formulated as a 3D mesh adaptation problem to sparse contours or points [86, 119, 139, 147], and solutions often incorporate shape priors during that process. The form of such prior could be a regular shape [69, 86, 139] or a statistical template with plausible variations [30]. Deviations from regular and smooth geometries are commonly penalized in a fitting

process. These methods usually include an explicit smoothing term during fitting to further regularize the shape in addition to the use of shape priors. Both regularizers can introduce bias, compromising the accuracy of the reconstructed surfaces when handling pathological hearts [6, 145].

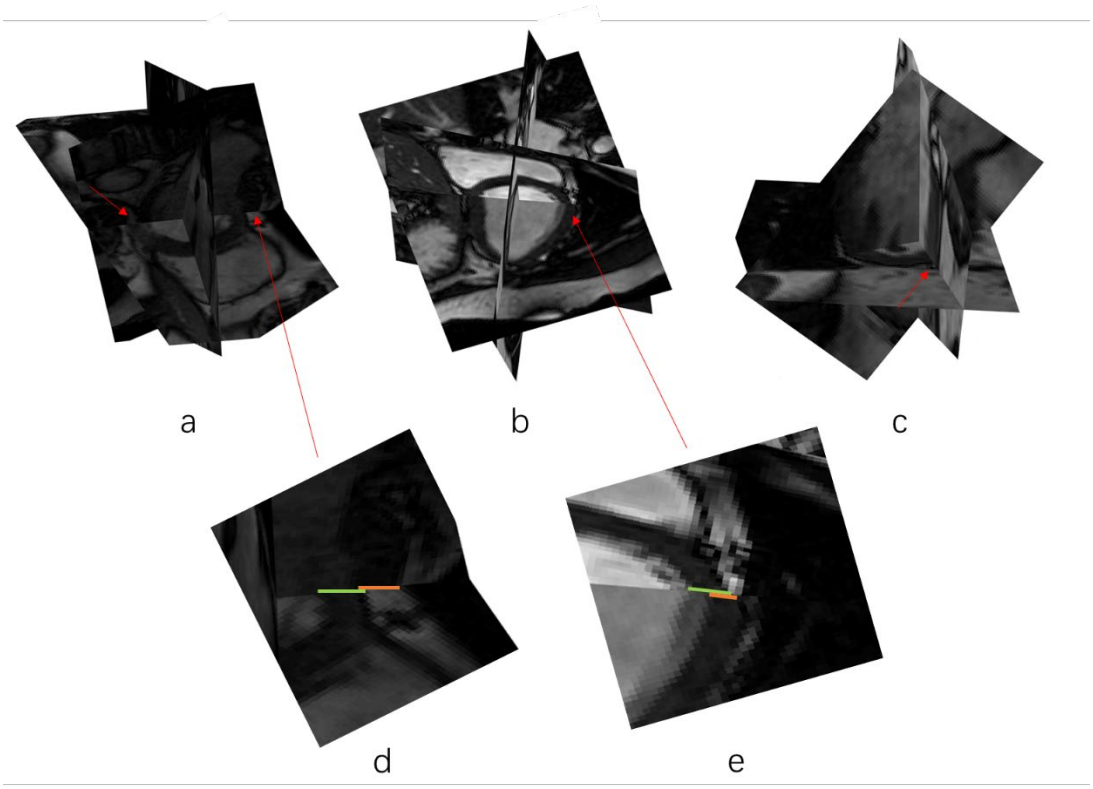


Figure 5. 1 Motion artefact examples with different level of misalignment. The red arrows indicate the positions of the misalignment, and from (a) to (c) the misalignment is less obvious. More details are shown in (d) and (e), where the green and orange lines indicate the cross-sections of the myocardium for each intersecting slices.

In order to generate anatomically meaningful surfaces while also preserving the wide variability of the shape, reconstruction methods should incorporate a wide collection of plausible shape priors to refer to when fitting the input data. Deep learning

algorithms have shown their capacity for storing a large number of accurate mappings between paired data [144], without the need for an explicit smoothing term, thus allowing the appearance of sharp edges and corners during reconstruction if necessary. However, for Convolutional Neural Networks (CNNs), input data are required to be highly structured (usually on a regular grid in 2D or 3D), and therefore most state-of-the-art deep learning methods for generating LV and RV 3D meshes using CMR data only take the SAX stack as input with the assumption that all SAX images are parallel [37, 91]. This assumption limits the possibility of utilizing the LAX slices, which has been shown to improve the reconstruction of ventricular meshes [119], as well as the possibility of incorporating out-of-plane geometric transformations for better correction of motion artefacts due to respiration [85, 110].

In this chapter, we propose to consider the sparse 3D information from contours in a volumetric form, and therefore tackle the problem similarly to volumetric image inpainting [18]. We then transform the problem of mesh fitting from sparse input into a 3D volumetric mapping problem followed by isosurfacing from dense volumetric data. The sparse volumetric input can take contours at any location and orientation within the volume, and therefore our method can incorporate both SAX and LAX slices and allows out-of-plane motion artefact correction. Our method also has the capacity of reconstructing 3D meshes from misaligned input contours without a specific misalignment correction step.

We implement a variation of 3D U-Net [28] as our baseline method and use a bi-ventricular statistical shape model to develop and validate the method as a proof of

theory. We then use clinical datasets to further develop and validate our method by introducing an adversarial loss to our baseline method and evaluating the importance of LAX slices. To explicitly demonstrate the advantage of utilising intersecting slices, we also perform experiments to compare the performance of our baseline method and the adversarial loss modification for both spatially consistent and misaligned input sparse input volumes.

5.2 Dataset

We first use the statistical shape model published by Bai et al [30], in which the authors registered 1093 segmented hearts to a template space using rigid registration followed by the application of principal component analysis (PCA) to the surface meshes. The model is formed by labelled images with labels for background, LV myocardium and LV/RV cavities. We downloaded from the publicly available dataset (http://wp.doc.ic.ac.uk/wbai/data), the mean shape model, the first 100 principal components (PCs) and the corresponding eigenvalues or variances. We then use these to generate 120 different shapes, limiting the variations to six standard deviations for any of the PCs.

To generate output references, we placed these shapes into a 3D volume with the size of 128 × 128 × 128 voxels and voxel size of 2 × 2 × 2 mm^3, by aligning the centre of the smallest sphere enclosing the corresponding contours and the centre of the 3D volume, and labelled voxels enclosed by surfaces accordingly. To simulate input sparse volumes in conditions similar to real clinical acquisition, we used real image planes

from clinical datasets (consisting of both a SAX stack and two LAX slices), and set the thickness of the reconstructed slices to 1 *mm*. Examples of reference bi-ventricular shapes from the statistical shape model are illustrated in Figure 5.2.

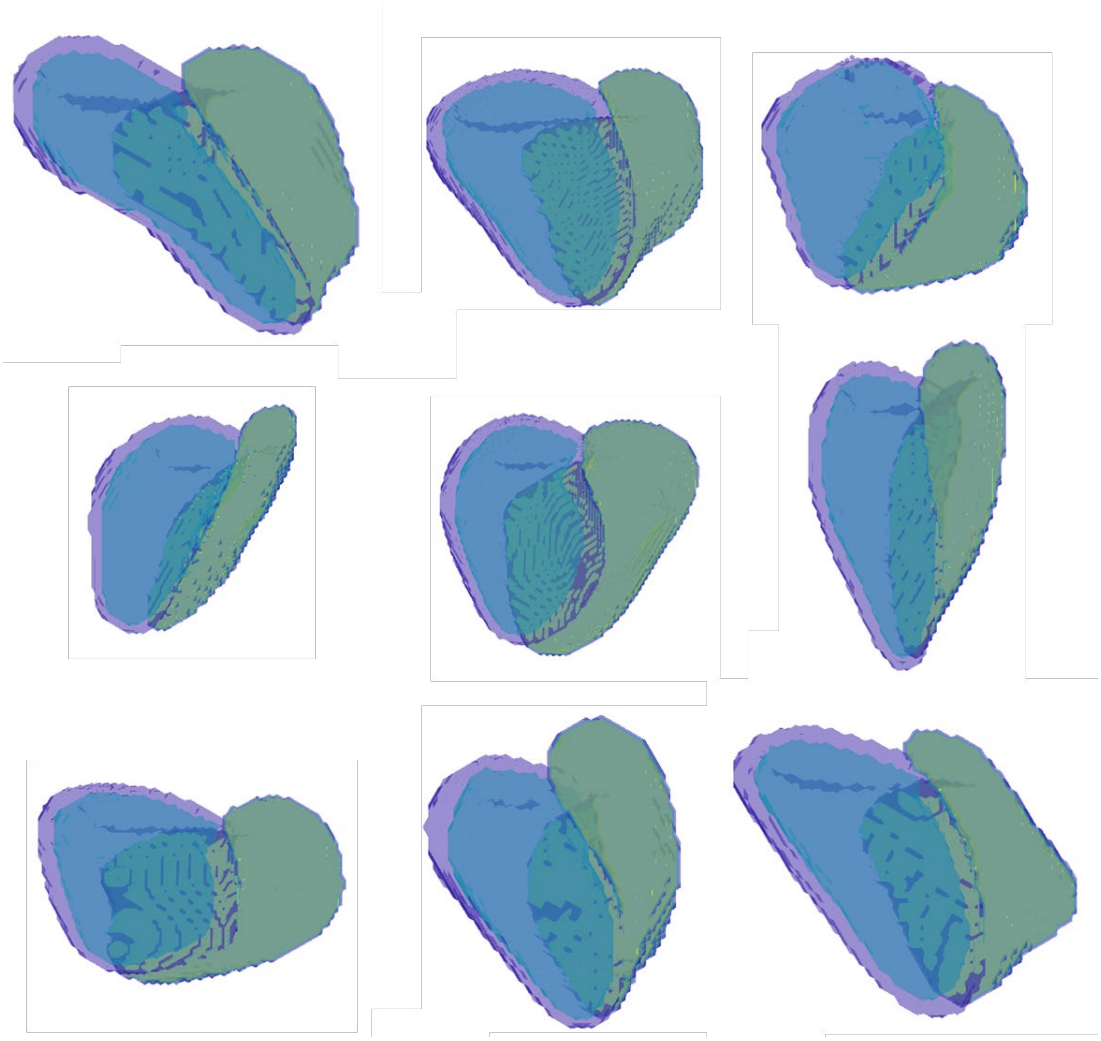


Figure 5. 2 Examples of reference LV shapes from the statistical shape model. The purple, blue and green surfaces represent LV epicardial, LV endocardial and RV endocardial surfaces. A great variation of plausible shapes is covered within the generated datasets.

To generate spatially consistent input datasets, we first aligned one set of image planes to each of the shapes, and the voxels located within 0.5 *mm* (half of the slice thickness) away from the planes were then given their original labels. Voxels located

more than 0.5 *mm* away from the plane were assigned a special label that we refer to as Unknown. To mimic the misalignment caused by motion artefacts, we first fixed image planes; then, before assigning labels for each plane we applied random rotations no larger than 10º and random translations no more than 4 *mm* to the reference shape, which are sufficient for simulating motion artefacts caused by respiration as suggested by [85, 110]. For intersecting voxels with multiple labels assigned to them, we assign the label of Unknown to them at the end of the process.

We then randomly selected 200 cases from the DETERMINE (Defibrillators to Reduce Risk by Magnetic Resonance Imaging Evaluation) dataset [60] consisting of patients with clinical evidence of myocardial infarction, from the Cardiac Atlas Project database [27], to further investigate our method. There is some variability in the dataset, but the typical study includes retrospectively gated SSFP cine imaging (typical thickness ≤ 10 *mm*, gap ≤ 2 *mm*, TR 30-50 *ms*, TE 1.6 *ms*, flip angle 60º, FOV 360 *mm*, image matrix 128×256, 25 frames, depending on heart rate and breath-hold duration). Enough SAX slices are acquired to cover the whole heart, and all the cases we selected had LAX slices in 4-CH and 2-CH views.

For each case, a finite element model of the LV is fitted to the cine MR images using the CIM software package (with guide-point modelling or GPM technique [137] introduced in section 2.4.6) interactively by experts. The endocardial and epicardial surfaces of the CIM model are considered as reference and used for input/output volume preparation. For output reference volume, we generated a regular grid with the size of 128×128×128 and voxel size of PS×PS×PS mm^3 (where PS is the pixel spacing from

corresponding DICOM headers, with an mean value of 1.5 *mm*) by aligning the centre of the epicardial surface of the finite element model of a single frame to the centre of the volume. Examples of generated LV shapes from the DETERMINE data are illustrated in Figure 5.3.



Figure 5. 3 Examples of generated LV shapes from the DETERMINE data. The purple and golden surfaces represent the epicardial and endocardial surfaces of the LV. The size and shape of the LV and the thickness of myocardium varies within the database.

For the preparation of sparse input volumes, we chose the distance between voxel

and image plane to be 1 *mm* to validate robustness of the performance of our network on a variety of slice thickness. The same protocol is used to simulate the motion artefacts by introducing random motion to the reference shape.

5.3 Method

5.3.1 Method Overview

Our 3D reconstruction method consists of two steps: 1. generating dense 3D volumetric predictions of LV myocardium, LV cavities and RV cavity (for statistical shape model only) using our 3D U-Net as shown in Figure 5.4; 2. generating 3D meshes from the predictions using isosurfacing. U-Net is well known as a segmentation network, which maps from images to segmentations; however, it has also been successfully implemented for image-to-image translation (including segmentation-to-image mapping or sketch-to-photo mapping) [58]. We reformulate our 3D reconstruction process in a similar way to volumetric inpainting, and variations of U-Net have been applied to image inpainting tasks with promising results [64, 76, 96]. The sparse input volume of our network is prepared using the statistical shape model or the finite element model.

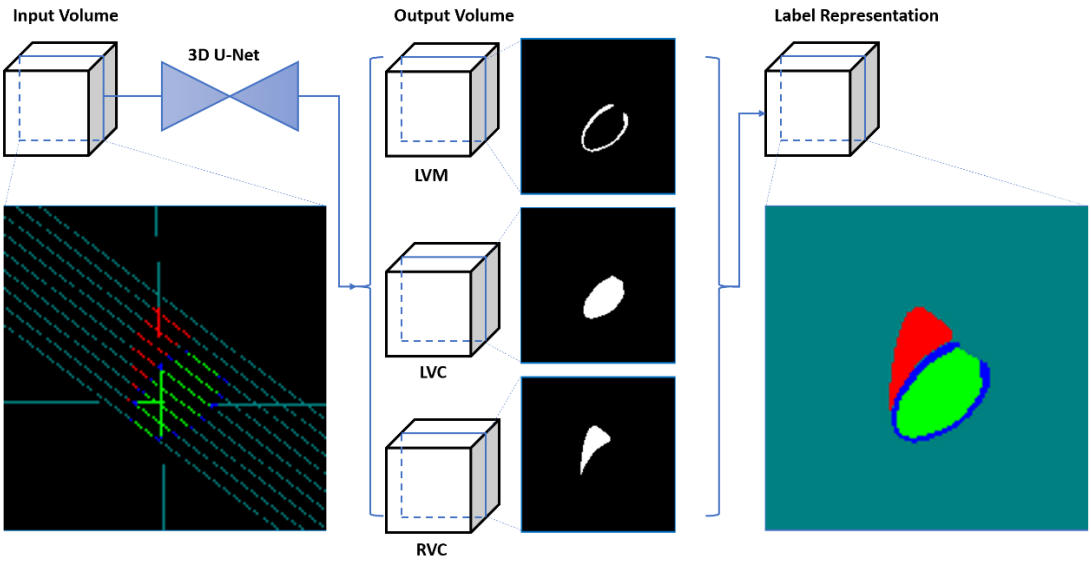


Figure 5. 4 Method overview demonstrated using the statistical shape model. The illustration shows an individual slice, but all the datasets and operations are in 3D. The same architecture is implemented as both the baseline networks and reconstruction networks (the generator, which is trained simultaneously with adversarial loss function), and therefore the testing phases of them are identical.

We use an example of the statistical shape model to illustrate the reconstruction process of our 3D U-Net. The cross-sectional views corresponding to a specific slice (marked with dotted lines on the volumes) are shown in Figure 5.4, where the slice does not correspond to any of the CMR slices. In the input volume, only voxels corresponding to the centre of CMR slices are assigned specific labels (2: red, 3: green, 4: blue and 1: cyan for RV cavity, LV cavity, LV myocardium and background, respectively), while all others are assigned the label ”Unknown” (0). For each of the 3D output volumes, a value between 0 and 1 is assigned to each voxel representing the likelihood of belonging to the respective region. These 3D volumes are then combined to produce the label representation of the reconstructed volume, in which an average of

94.5% of the voxels are transformed from the original Unknown label to one of the defined classes.

For the DETERMINE dataset, given that the manual reference only includes LV endocardial and epicardial surfaces, we make modifications to our method. The neural network architecture remains the same, while only three output volumes are produced representing LV cavity, LV myocardium and background or RV cavity for DETERMINE dataset or statistical shape model respectively. We also use a different labelling strategy (the maximum distance from the labelled voxels away from the image plane is larger) and therefore change the sparsity of the input volume, resulting in the percentage of voxels with Unknown label dropping to around 90%.

5.3.2 Baseline Neural Network

We adapt a variation of the 3D U-Net as the baseline method for volumetric mapping. The schematic diagram of the reconstruction neural network architecture is shown in Figure 5.5.

The network consists of an encoder and a decoder, with skip connections between feature maps with the same resolution. The encoder starts from the input layer with the size of 128×128×128×1 and has three max-pooling stages where each pooling layer has a window size of 2×2×2 and a stride of 2×2×2, giving feature maps with four different resolutions. The decoder then upsamples the feature maps back to the original resolution also in three stages using upsampling layers with factors of 2×2×2. The heavy duty of the computation within the network is carried out by the convolutional blocks

shown in Figure 5.5, with each of them having two convolutional layers with kernel size of 3×3×3 and two batch normalisation (BN) layers.

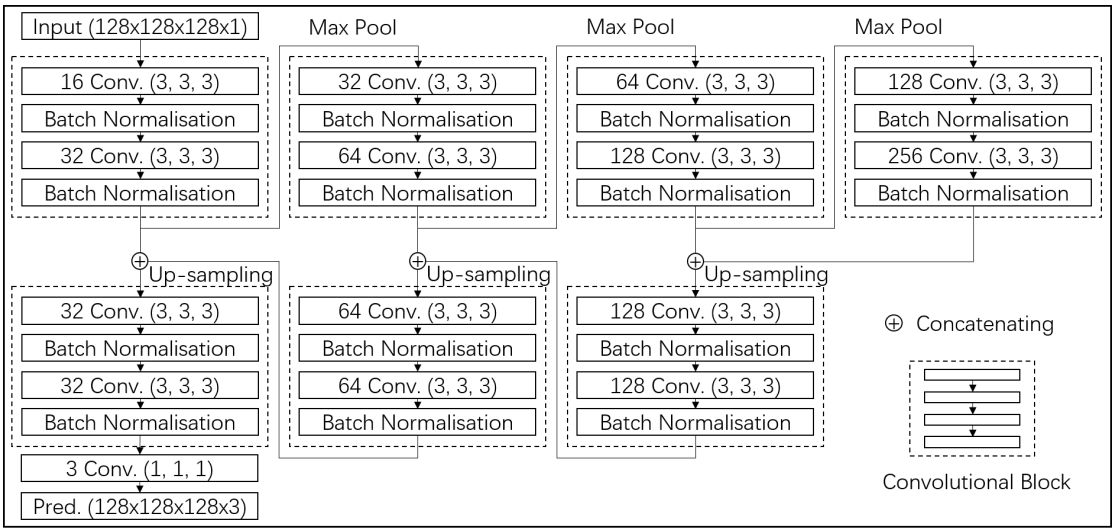


Figure 5. 5 Schematic representation of reconstruction neural network architecture.

Other than the output layer, rectified linear units (ReLU) are used throughout the network. The output layer has different types of activation functions for experiments with the statistical shape model and DETERMINE dataset due to the difference between the output volumes. For the statistical shape model, no background volume is included in the output and we use sigmoid activation function for the output layer as softmax is not beneficial. For the DETERMINE dataset, all voxels are assigned to a designed label as the background is included, and softmax is used as the activation function and helps the classification for each voxel. Including background layer in the output volume improves the accuracy and robustness when converting the output volumes to label maps, and therefore we begin to use this implementation in our experiments with the DETERMINE dataset as well as in the chapter. The output layer has 3 convolutional kernels with the size of 1×1×1 giving the prediction in the form of 128×128×128×3 volume representing LV myocardium, LV cavity and RV cavity (for the statistical shape

model) or background (for the DETERMINE dataset). The Dice coefficient loss implemented in [87] is used during the training of the network to deal with class-imbalance between foreground and background voxels, and the loss function L_g is defined as:

$$L_g = 1 - \frac{2\sum_i^N p_i r_i}{\sum_i^N p_i + \sum_i^N r_i}$$

where the sums run over the $N = 128^3$ voxels of the predicted binary segmentation volume $p_i \in P$ and the ground truth binary volume $r_i \in R$. The value of L_g varies between 0 and 1 when the similarity between P and R changes, and the value drops when P becomes more like R.

5.3.3 Adversarial Loss

We further develop our reconstruction neural network using the DETERMINE dataset by introducing an adversarial loss function L_d through a discriminator network:

$$L_d = \sum_i^M (rc_i - pc_i)^2 / M$$

where the sums run over the $M = 16^3$ voxels of the predicted class volume $pc_i \in PC$ (as shown in Figure 5.6) and the reference class volume $rc \in RC$. The reference class volume RC is filled with zeros for predicted binary segmentation volume P and filled with ones for reference binary volume Rt. The value of L_d also varies between 0 and 1.

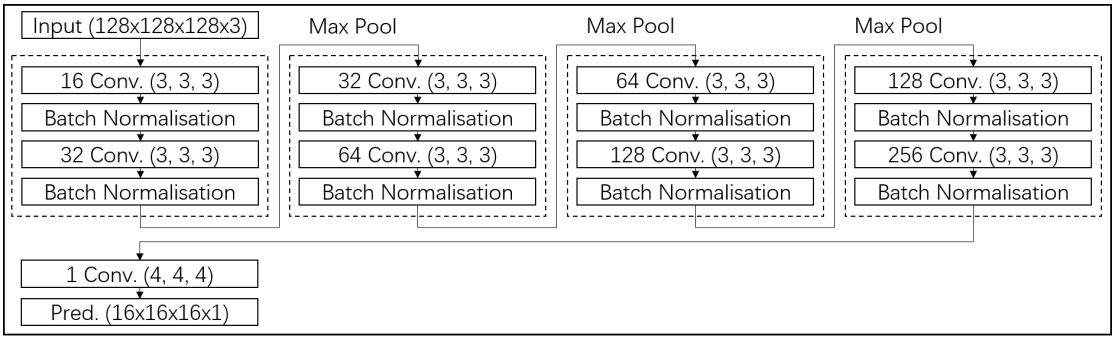


Figure 5. 6 Schematic representation of discriminator neural network architecture.

The network has the same encoder of the reconstruction network, followed by 1 convolutional kernel with the size of 4×4×4 giving the prediction in the form of a 16×16×16×1 volume representing the classification result. Rectified linear units (ReLU) are used throughout the network except for the output layer where the sigmoid activation function is used. Even though the discriminator network is designed to distinguish predicted from reference binary volumes, acting as a classifier, the loss function L_d is used during training and the actual classification result is not used for both training and testing.

Two networks are implemented as a variation of generative adversarial neural network (GAN) [45] with the reconstruction network considered as the generator, and to simultaneously train the reconstruction neural network we minimize the loss function L:

$$L = \alpha \times L_g + L_d$$

where we set the value of α as 1000 to keep the losses within a similar magnitude and achieve the convergence of the network during training.

5.3.4 Isosurfacing

The prediction of the network is converted into three 128 × 128 × 128 binary volumes and for each 3D volume the largest object was selected as the expected structure, used to form the isosurfaces at 0.5 using marching cubes [80].

5.4 Experiment and Results

5.4.1 Implementation

We implemented the neural networks in Keras with Tensorflow as backend. The networks were trained with Adadelta [142]. We used one NVIDIA GeForce GTX 1080 Ti GPU for all training and testing experiments.

5.4.2 Statistical Shape Model

We performed two 4-fold cross-validation experiments with spatially consistent data and misaligned data. For both experiments, the same number of bi-ventricular shapes were generated as described in section 5.2, and they were randomly grouped into 4 sets of training data (80 shapes), validating data (10 shapes) and testing data (30 shapes) also in the same way, with the 4 testing datasets covering all generated shapes. During the training phase of each cross-validation experiment, the network was randomly initiated, and the network parameters were updated through backpropagation using the spatially consistent training data. The optimization was early stopped by evaluating the loss function L_g on the validating data to prevent over-fitting. Once the training of the network for the spatially consistent cases is terminated, the learned parameters are used

to initiate the network for reconstructing misaligned cases and fine-tuned using the misaligned training data.

An overview of the data preparation and network training progress is summarised in Figure 5.7. The label maps generated from the statistical shape model are used to prepare the input volumes (for both spatially consistent and misaligned experiments) and the output binary volumes, and the spatially consistent network is then trained first and used to initiate the parameters of the network that is fine-tuned with misaligned data.

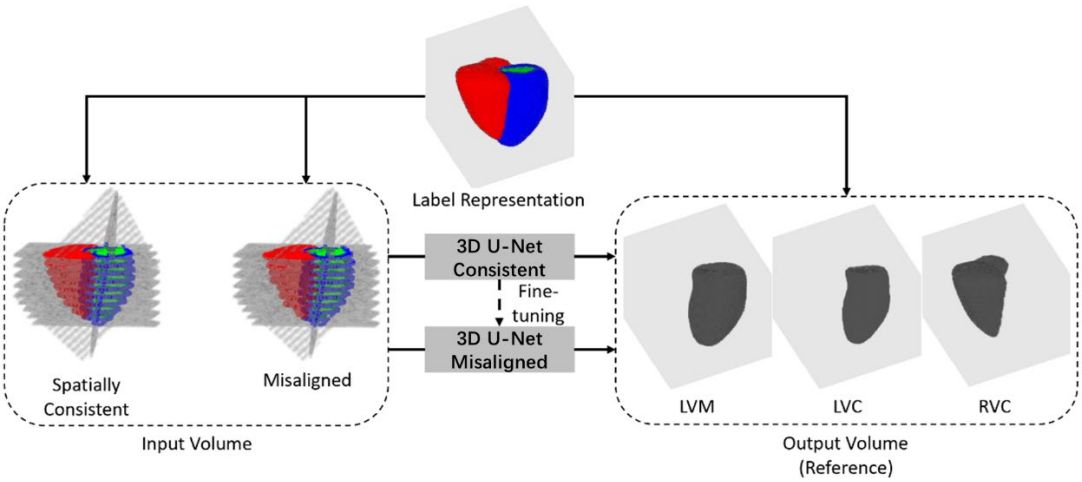


Figure 5. 7 Data preparation and training progress overview. Red, green, blue, and light grey voxels are RV cavity, LV cavity, LV myocardium, and background. The dark grey voxels are foreground in binary volumes.

The testing data is only used for the evaluation of the method, and the predictions from the network for the testing data are then used to generate 3D meshes as described in section 5.3.4. In both experiments, Dice coefficient and Hausdorff Distance (HD) are the metrics used for quantitative analysis of the method, and the results are shown in Table 5.1.

Table 5. 1 Mean and standard deviation of Dice score and HD of the statistical shape model experiments

	Dice			HD (*mm*)		
	LVM	LVC	RVC	LVM	LVC	RVC
Consistent	0.94±0.01	0.98±0.01	0.98±0.01	4.56±1.66	3.17±1.07	3.87±2.28
Misaligned	0.83±0.01	0.95±0.01	0.93±0.01	6.03±5.21	4.94±3.61	5.91±2.55

It should be noted that, for the experiment with misaligned contours we register the reconstructed shape to the reference shape with a rigid transformation before calculating the metrics. An example of reconstructed and reference surfaces of LV myocardium is shown in Figure 5.8, illustrating the difference before and after the rigid transformation.

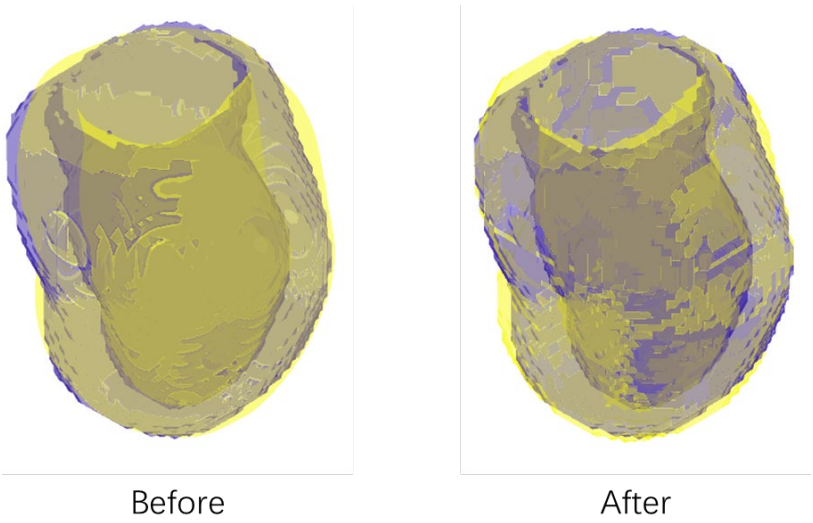


Figure 5. 8 Illustration of the effect of rigid transformation. The purple surfaces represent the reconstructed LV myocardium and the yellow surfaces represent the reference LV myocardium. The reconstructed surface aligns better with the reference surface after the rigid transformation.

For spatially consistent input data, the reconstructed LV/RV cavities shape matched

the reference shape accurately with mean Dice coefficient values of 0.98. The LV myocardium has a more complicated shape with larger surface area, and therefore the Dice score is more sensitive to small changes, while the mean value is still 0.94. The voxel size is $2 \times 2 \times 2$ mm^3, and the mean values of HD are around 2 voxels. As the input data has only around 5.5% of voxels with known labels, the results suggest that our 3D U-Net stored a collection of plausible mappings between sparse input data and corresponding ventricular shapes during training and is able to utilize that information during the reconstruction of the testing cases.

The experiments on misaligned input data also achieve good accuracy, with above 0.93 mean Dice coefficient for LV/RV cavities and 0.83 for LV myocardium. As expected, comparing to the experiment on spatially consistent input data, the accuracy of the experiment on misaligned slices is lower. This is unavoidable given the additional challenges involved in reconstructing the 3D shapes from misaligned input data, including discrepancies between slices, which results in a much wider range of mapping from a set of input volumes to the same plausible output shape and vice versa. It is important to highlight the limitations of the validation method itself. The misalignment of all slices, combined, will result in an overall rigid transformation of the underlying three-dimensional shape. This rigidly transformed shape is the most likely shape to be reconstructed by our network. If we compare this reconstruction with the original shape, the accuracy will be significantly affected, even if the shapes are exactly equal. We attempted to compensate for this unwanted effect by performing a rigid registration, but this still leaves residual errors that are partly responsible for the decrease of accuracy

shown in Table 5.1.

5.4.3 DETERMINE Dataset

We performed 2-fold cross-validation for all experiments with the DETERMINE dataset. The 200 cases are randomly grouped into 2 sets of training data (90 cases), validating data (10 cases) and testing data (100 cases) also in the same way, with the 2 testing datasets covering all the cases. Comparing to the 4-fold cross-validation of the statistical shape model experiments, we halved the number of experiments for the DETERMINE dataset (as the computational cost, especially for training, was also halved) in order to demonstrate the effect of including/excluding the LAX slices as well as the adversarial loss. Therefore, for each fold we perform a set of eight experiments formulated with all possible combinations from: spatially consistent or misaligned, with or without adversarial loss, and with or without LAX slices.

The same training protocol for the statistical shape models was first applied to the DETERMINE dataset to train the baseline networks, which were then used to initiate the parameters of the reconstruction networks trained with adversarial loss simultaneously. The parameters of the discriminator and generator were updated iteratively, and each time one step of optimization was performed for each network by minimizing L_d and L (where $\alpha = 1000$ for L) respectively. The optimization processes for all training networks were early stopped by evaluating the loss function L_g on the validating data to prevent over-fitting. An example of our U-Net (which is also the generator of our GAN) validating loss in training progress is shown in Figure 5.9.

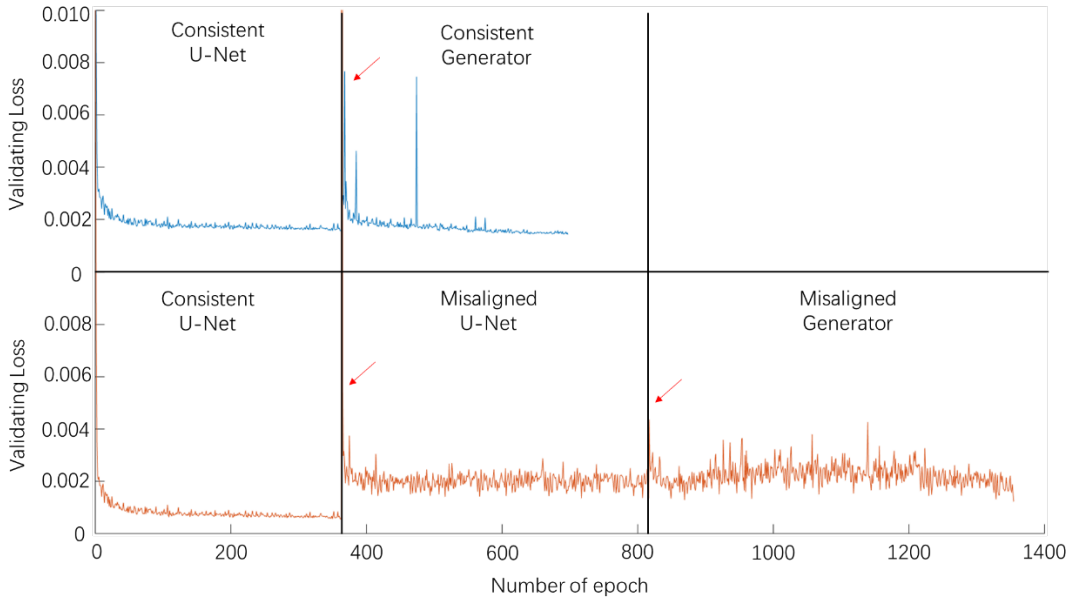


Figure 5. 9 Validating loss of U-Net (or generator of GAN) in training progress. The experiment starts with training the U-Net with spatially consistent input data, which is the same in both the blue and orange lines. The second stage of the blue line illustrates the training progress after adding the adversarial loss to the consistent U-Net. The second stage of the orange line illustrates the training progress of taking the consistent U-Net parameters to initiate the training of the U-Net with misaligned input data. The final stage of the orange line illustrates the training progress after adding the adversarial loss to the misaligned U-Net. The red arrows show a sudden increase of the loss function when the network is fine-tuned with a different type of dataset.

In the same way as in the statistical shape model experiments, we used Dice coefficient and HD for quantitative analysis of our method on the DETERMINE dataset, and a rigid transformation was also applied to the reconstructed surfaces from misaligned input volumes. One volume for each testing case was randomly generated to calculate the metrics to avoid bias caused by different number of frames from one case to another. The results are shown in Table 5.2.

Table 5. 2 Mean and standard deviation of Dice score and HD of DETERMINE dataset experiments

	Dice		HD (*mm*)	
	LVM	LVC	LVM	LVC
Consistent SAX	0.95±0.08	0.97±0.05	13.4±8.6	5.7±4.4
Consistent SAX Adversarial	0.96±0.07	0.97±0.06	12.9±8.9	5.3±5.0
Consistent S/LAX	0.96±0.05	0.98±0.04	11.8±6.2	4.3±3.9
Consistent S/LAX Adversarial	0.98±0.03	0.99±0.02	10.6±5.8	3.6±2.7
Misaligned SAX	0.81±0.10	0.90±0.08	16.0±8.2	8.5±6.4
Misaligned SAX Adversarial	0.84±0.08	0.92±0.07	15.1±8.2	8.0±6.3
Misaligned S/LAX	0.84±0.09	0.93±0.07	14.5±6.9	6.5±4.2
Misaligned S/LAX Adversarial	0.88±0.06	0.93±0.06	14.3±6.2	6.0±4.2

In order to evaluate the performance of the network after including LAX slices, adversarial loss or both of them, we apply paired t-tests with a significance level of 5% on all corresponding pairs of experiments using both Dice score and HD values on both LV myocardium and cavity. For experiments with/without LAX slices (e.g. Consistent SAX and Consistent S/LAX), most pairs are significantly different for both metrics on both anatomies, except the Misaligned SAX Adversarial and Misaligned S/LAX Adversarial pair, which is not significantly different for HD on LV myocardium with p-values of 0.13. For experiments with/without adversarial loss (e.g. Consistent SAX and Consistent SAX Adversarial), most pairs are significantly different for both metrics on both anatomies, except the Consistent SAX and Consistent SAX Adversarial pair, which is not significantly different for both Dice score and HD on LV cavity with p-

values of 0.28 and 0.16 respectively, and the Misaligned S/LAX and Misaligned S/LAX Adversarial pair, which is not significantly different for HD on LV myocardium with p-values of 0.76. For experiments with/without both LAX slices and adversarial loss (e.g. Consistent SAX and Consistent S/LAX Adversarial), all pairs are significantly different for both metrics on both anatomies. For the Consistent SAX and Consistent SAX Adversarial pair, the slices are well aligned in space and the structure of LV cavity is relatively simple, and both factors simplify the task resulting in no significant difference after introducing the adversarial loss. T-tests on HD values of LV myocardium for both Misaligned SAX Adversarial and Misaligned S/LAX Adversarial pair and the Misaligned S/LAX and Misaligned S/LAX Adversarial pair, on the contrary, reflect the most challenging task of all experiments, which has misaligned slices and the more complicated structure (LV myocardium). Different from Dice score, which shows the global similarity between two structures and has significant differences for both LV myocardium and cavity, HD represents a localised difference and we make further analysis to understand the reasons behind.

We first compare the results with the statistical shape model results, using the corresponding rows: Consistent S/LAX and Misaligned S/LAX. For Consistent S/LAX experiments, the statistical shape model Dice scores of LV myocardium and cavity are 0.94 ± 0.01 and 0.98 ± 0.01, while for the DETERMINE data results the corresponding values are 0.96 ± 0.05 and 0.98 ± 0.04. The Dice scores are comparable to each other, and for HD of LV cavity the results are also similar with the values of 3.17 ± 1.07 *mm* and 4.34 ± 3.92 *mm* for the statistical shape model and the DETERMINE data

respectively. However, for HD of LV myocardium, we see a bigger difference between the statistical shape model results and the DETERMINE data results, which are 11.8 ± 6.15 *mm* and 6.03 ± 5.21 *mm* respectively. For Misaligned S/LAX experiments, the Dice scores of LV myocardium and cavity (0.83 ± 0.01 and 0.95 ± 0.01 for the statistical shape model results, and 0.84 ± 0.09 and 0.93 ± 0.07 for the DETERMINE data results) and the HD values of LV cavity (4.94 ± 3.61 *mm* for the statistical shape model results, and 6.54 ± 4.23 *mm* for the DETERMINE data results) are also comparable between two sets of experiments, and the difference between the HD values of LV myocardium (6.03 ± 5.21 *mm* for the statistical shape model results, and 14.45 ± 6.91 *mm* for the DETERMINE data results) is also comparatively larger.

The overall performance of the network is comparable for both the statistical shape model and the DETERMINE data experiments, despite the DETERMINE dataset consisting of pathological cases, showing that our method has consistent performance even when the shape of LV differs from a healthy heart. The HD value of LV myocardium is around two times larger than the result of the statistical shape model, and by considering the Dice score and HD values of LV cavity, the large HD value is likely caused by epicardial surfaces. Also, by considering the similar Dice score values for LV myocardium, the large HD value reflects a localised difference between the reference surfaces and reconstructed surfaces, and an example is shown in Figure 5.11.

We expected the performance of our method to improve after introducing the LAX slices. However, due to the motion artefacts, for misaligned slices there are discrepancies between them, which might have a negative effect on the results. By

comparing corresponding SAX rows with S/LAX rows, we can see clear improvements for both spatially consistent and misaligned input data. The improvements for misaligned cases are even larger than for the consistent cases, which shows that LAX slices have a more positive effect to the reconstruction process by serving as good references for aligning SAX stacks, and therefore it is important to incorporate information from LAX slices into the segmentation and reconstruction process for better accuracy.

Our implementation of the adversarial loss follows the training progress of the baseline network, which has ended up in a local minimum after the optimization. By comparing the corresponding rows with and without Adversarial (loss), we can see an overall improvement after introducing the adversarial loss, which guides the optimization process to an equivalent or better local minimum. For LV myocardium, the improvement is clearer than for LV cavity, which has a more regular shape compared to the myocardium. This suggests that our adversarial implementation is not necessary when the baseline method produces reconstructed surfaces close to real cases and additional regularisation would not improve the result much more, as no sufficient gradient is provided for the reconstruction network to learn. For both spatially consistent and misaligned experiments, reconstruction networks with adversarial loss and utilising both SAX and LAX slice information have the best performance.

To investigate the large HD compared to the statistical shape model and locate the source of the difference between the reference and prediction surfaces, we divide the LV into basal, mid-cavity and apical regions along the LAX equally in thickness based

on the reference surfaces. We then calculate HD for each section respectively. The results are shown in Table 5.3.

Table 5. 3 Mean and standard deviation of regional HD of DETERMINE dataset experiments

	Base		Mid-cavity		Apex	
	LVM	LVC	LVM	LVC	LVM	LVC
Consistent SAX	13.3 ±8.6	5.5 ±4.4	2.4 ±7.2	1.9 ±3.4	2.6 ±3.5	2.3 ±2.7
Consistent SAX Adversarial	12.8 ±8.9	4.9 ±4.2	2.4 ±7.3	1.9 ±3.4	2.6 ±3.8	2.2 ±2.7
Consistent S/LAX	11.8 ±6.1	4.2 ±3.9	1.9 ±3.1	1.9 ±3.5	2.2 ±2.3	2.1 ±2.7
Consistent S/LAX Adversarial	10.5 ±5.9	3.5 ±2.2	1.8 ±2.4	1.6 ±1.6	1.9 ±1.8	1.8 ±1.4
Misaligned SAX	15.8 ±8.1	8.9 ±6.0	6.0 ±5.8	8.3 ±6.6	5.8 ±2.5	6.3 ±5.4
Misaligned SAX Adversarial	15.0 ±8.2	8.6 ±6.0	5.4 ±6.3	7.2 ±6.3	5.7 ±5.2	6.0 ±3.7
Misaligned S/LAX	14.4 ±6.8	7.0 ±4.6	5.0 ±4.7	6.4 ±5.6	4.8 ±2.3	5.0 ±4.2
Misaligned S/LAX Adversarial	14.3 ±6.2	6.3 ±4.3	4.2 ±3.2	5.3 ±4.8	4.2 ±2.8	4.4 ±3.5

From the results, we can see that the source of the large HD comes from the basal region. For mid-cavity and apical regions, the results are comparable with the statistical shape experiments. In these regions, both SAX and LAX slices are contributing in the reconstruction process, however, as the valve is open in the basal region and only LAX are available for reconstruction, the uncertainty of the reconstructed shape increases

resulting in high HD values. The basal HD values are slightly different for whole ventricular evaluation because the LV is divided into three sections for HD calculation.

The uncertainty around the base is mainly determined by the arrangement of imaging slices, as only two LAX slices partially cover the most basal regions, which will affect both the semi-automated references and the network predicted results, giving a lack a consistency locally. However, for LV functional analysis, by considering the global and reginal Dice scores and HD values for both the statistical shape model and the DETERMINE data experiments, the reconstructed shape can be used reasonably.

5.4.4 Visual Inspection of Reconstructed Surfaces

We also analyse our results qualitatively by visual inspections.

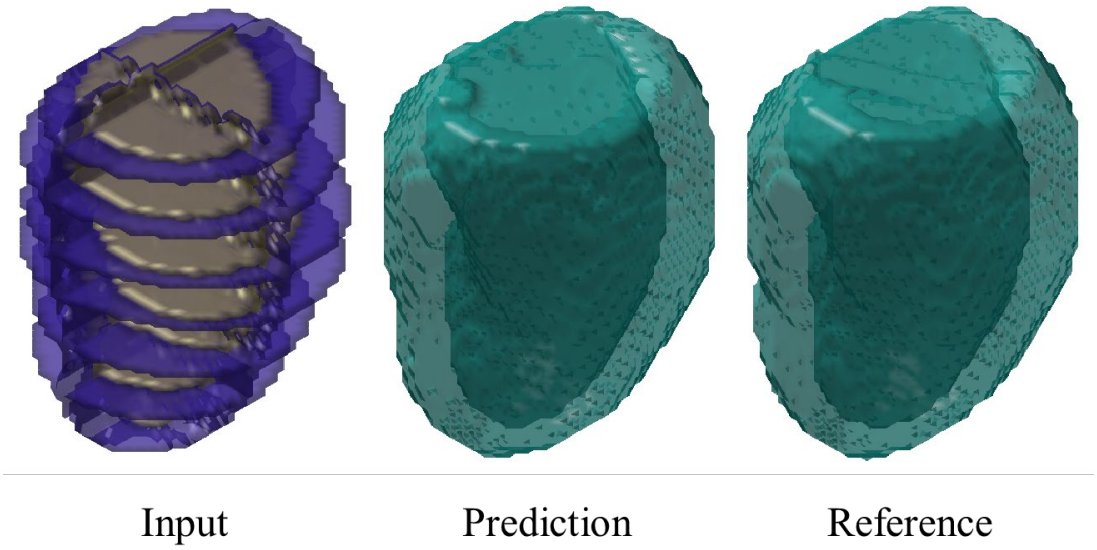


Figure 5. 10 Visualisation of the reconstructed LV surfaces for a spatially consistent input volume. The golden and purple regions represent LV cavity and LV myocardium in the input voulume. The predicted and reference surfaces are in green.

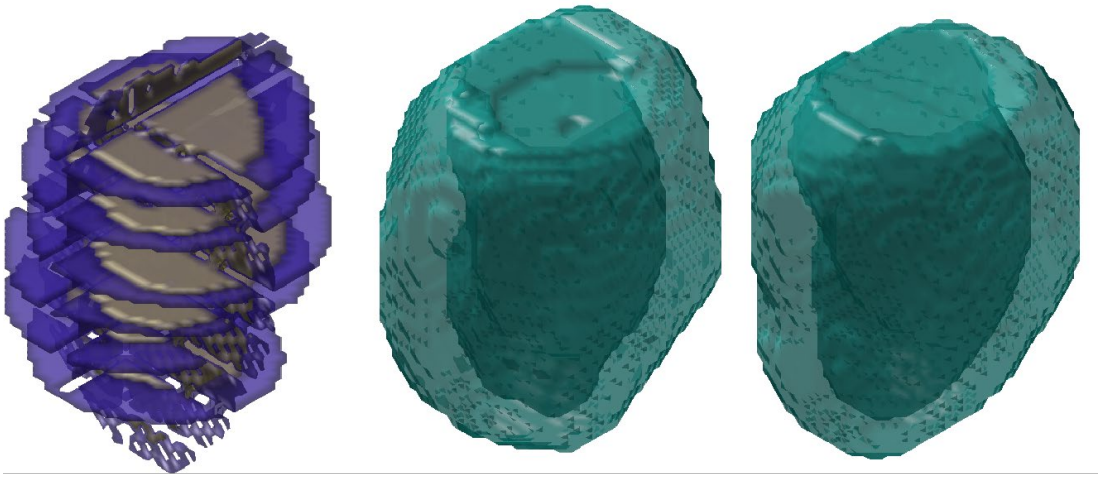


Figure 5. 11 Visualisation of the reconstructed LV surfaces for a misaligned input volume. The golden and purple regions represent LV cavity and LV myocardium in the input voulume. The predicted and reference surfaces are in green.

In Figure 5.10 and Figure 5.11 we show a representative example (with Dice scores and HD values close to the mean values of the experiments) with reference surfaces and reconstructed surfaces using our best performing networks from DETERMINE dataset experiments, with the corresponding input volume representations, where the golden region is the labelled cavity voxels and the purple region is the labelled myocardial voxels.

For the spatially consistent case, the reconstructed surfaces look very similar to the reference surface with distinguishable features for each of them in the basal region. The second row shows the misaligned case, and from the input volume representation we can see the artefact is very severe with one of the LAX slices shifted towards the apex and leaving the other LAX slice for the reconstruction of the most basal part alone. The overall reconstructed surfaces show good accuracy, and the apical region is not heavily

affected by the misaligned LAX slice. The wall thickness in the mid-cavity region, however, is thicker compared to the reference surfaces, caused by two SAX slices moving towards opposite directions with no LAX to define the borders. The reconstructed basal myocardium has a thin layer covering the cavity, which explains the relatively high HD values in the basal region.

5.5 Conclusions

In this chapter, we develop and evaluate our 3D reconstruction method using a statistical shape model (built from healthy hearts) and the DETERMINE dataset (consisting of infarct hearts) for both spatially consistent and misaligned CMR slices. Our method achieved comparable results with state-of-the-art methods [88, 91], and has the key advantage of utilising intersecting slices, which allows us to introduce LAX slice information for the reconstruction of LV.

The baseline method is first developed and validated using bi-ventricular shapes generated from the statistical shape model. The generated shapes cover a wide range of healthy hearts by allowing the variation of PC values to six standard deviations, and the reconstructed surfaces have good similarity to the reference surfaces for both Dice coefficient and HD evaluations. We repeat our baseline method on the DETERMINE dataset, and the mean value of HD for LV myocardium becomes more than two times larger, while the other metrics remains similar despite the surfaces being from pathological hearts. By dividing the reconstructed LV surfaces into basal, mid-cavity and apical sections, we located the source of high HD value is in the basal region, and

this is confirmed with visual inspection of the reconstructed surfaces.

The difficulty of reconstructing the basal region is the lack of relevant information from CMR images, and both the GPM method and our networks suffer from this issue. For the GPM method, due to the small number of available guide points used to regularize the base, the initial shape and the smoothing term become more important when producing the basal region reconstruction, which could result in a larger variation and inconsistency. Our networks are trained to learn the same pattern from the GPM method reference shapes, and achieve good accuracy for middle and apical regions while the difference is relatively larger for the basal region, which also suggests that the reconstructed base given by GPM method is quite arbitrary. We have showed the difference between the two methods, however, limited by the ground truth, we cannot conclude quantitatively which method works better. From the overall performance, this localised discrepancy has little effect on functional analysis of the LV.

As mentioned in previous chapters, LAX slices are acquired regularly in clinical practice, however, mostly used to estimate the location of the LV rather than for analysing it. Including LAX slices into 3D reconstruction has potential important benefits, and we perform experiments to illustrate that its overall effect on our method is positive, particularly for misaligned input volumes and the reconstruction of the basal region. For spatially consistent input volumes, the benefit is slightly smaller as the SAX stack is aligned very well and the improvements are focused on the most basal and apical regions. However, as we have seen in previous chapters, the 2D segmentation results for basal and apical slices are comparatively worse than the mid-cavity slices,

and the importance of including LAX slices for 3D reconstruction is more obvious when input volumes are created from segmentation of CMR images.

Finally, we implemented adversarial loss to our baseline method, to investigate the effect of indirect regularisation using a distribution of prior shapes. This improves the overall accuracy of the reconstructed surfaces, and similar to the inclusion of LAX slices, the most obvious improvements come from misaligned input volumes and basal region. The cost of training the reconstruction network simultaneously with adversarial loss is an increase of training time, and in practice the cost can be justified by the improvements as the networks are most likely to be trained only once before used for prediction.

We develop our reconstruction method with simulated input volumes from reference surfaces, and in the following chapter we combine our segmentation and reconstruction methods to create an end-to-end pipeline which takes CMR images as input and produces a 3D reconstruction of the LV with its functional analysis as output.

Chapter VI. LV Function Analysis of the Mouse Heart

In previous chapters, we developed our two-dimensional (2D) segmentation neural network for cardiac cine MR images and our three-dimensional (3D) neural network to reconstruct shapes of left/right ventricles (LV/RV) from sparse volumetric input. In this chapter, we develop a fully automated pipeline for the assessment of mouse LV using cardiac MR (CMR) images by integrating our segmentation and reconstruction methods and utilising both short-axis (SAX) and long-axis (LAX) slices. We developed and validated our pipeline using two datasets collected from different imaging centres, consisting of both pathological and control subjects. Similar to our experiment with the DETERMINE database, a guide-point modelling (GPM) based method was used to produce manually verified LV shapes by experienced operators, and we used the LV shapes to prepare training data for both our segmentation and reconstruction networks.

6.1 Introduction

CMR is the current gold standard imaging modality for cardiac function analysis of human heart, and the state-of-the-art deep learning algorithms have been widely applied to automate the image analysis of clinical CMR images. Mouse CMR images have also been verified for accurate estimation of cardiac function [107]; however, the image analysis stage is still largely relying on manual contouring and disks summation approach of the SAX slices. Semi-automated segmentation methods have been successfully applied to mouse CMR images to improve the efficiency of image analysis [138, 140], and the results are verified against manual segmentations. Fully automated LV segmentation and quantification methods using mouse CMR images, on the other hand, are still missing in the literature. A main limitation of preclinical studies is the small number of datasets [105, 107, 149], which largely increases the difficulty when implementing deep learning algorithms. In this chapter, to our best knowledge, we build the first fully automated segmentation and 3D reconstruction pipeline for analysing mouse CMR images by implementing deep neural networks, and the object of our work is the assessment of our pipeline for accurately quantifying LV function of mouse heart.

In chapter 4, we modified our segmentation method for mouse CMR images. We first adjusted the network architecture, inspired by the state-of-the-art 3D U-Net [28]. To improve the segmentation accuracy of the LAX slices, we introduced a second input layer, constructed using the segmentation results of the SAX slices, into our segmentation network for the LAX images. In this way, the network can use the SAX segmentation results to improve LAX segmentation accuracy. The overall segmentation

strategy for mouse CMR images thus becomes a three-step method, as illustrated in Figure 6.1 to Figure 6.3.

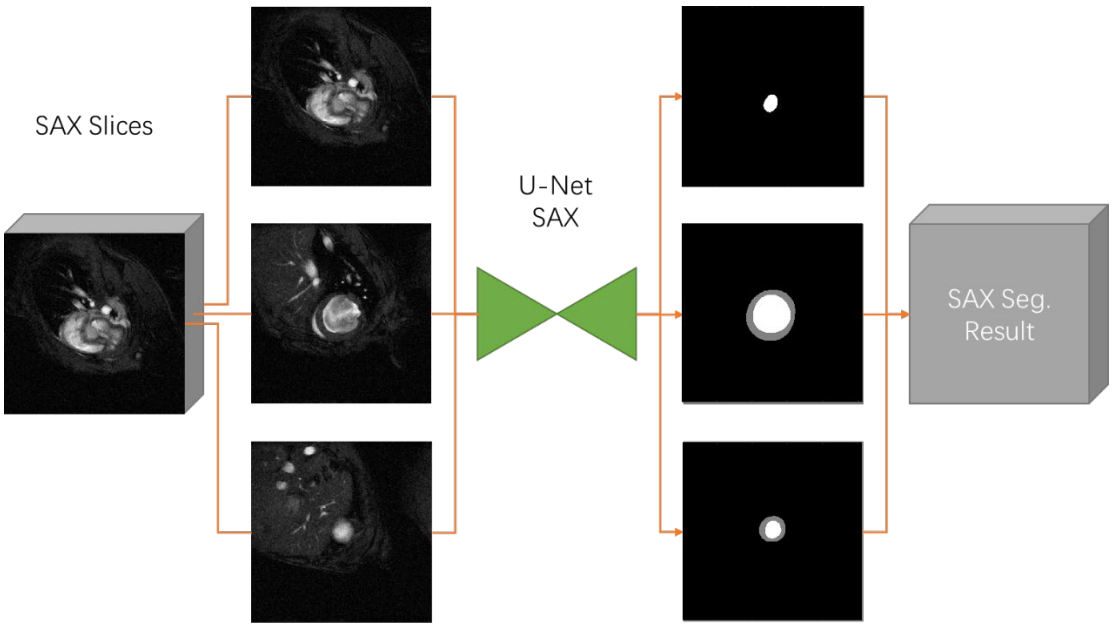


Figure 6. 1 Overview of the mouse CMR image segmentation method. We first segment the SAX slices using a variation of U-Net. The three slices represent the basal, mid-cavity and apical slices. A typical SAX stack usually consists of 10 to 15 slices.

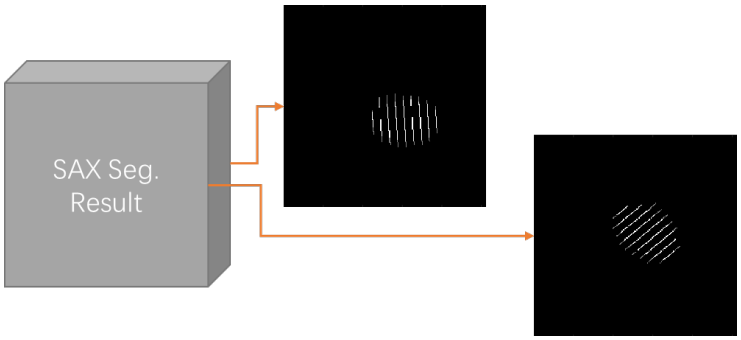


Figure 6. 2 In the second step, we project the segmentation of the SAX slices onto the LAX slice planes (usually two slices for each case: a 2-CH view and a 4-CH view.) This projected segmentation layer assists the segmentation of the LAX slices in the next step.

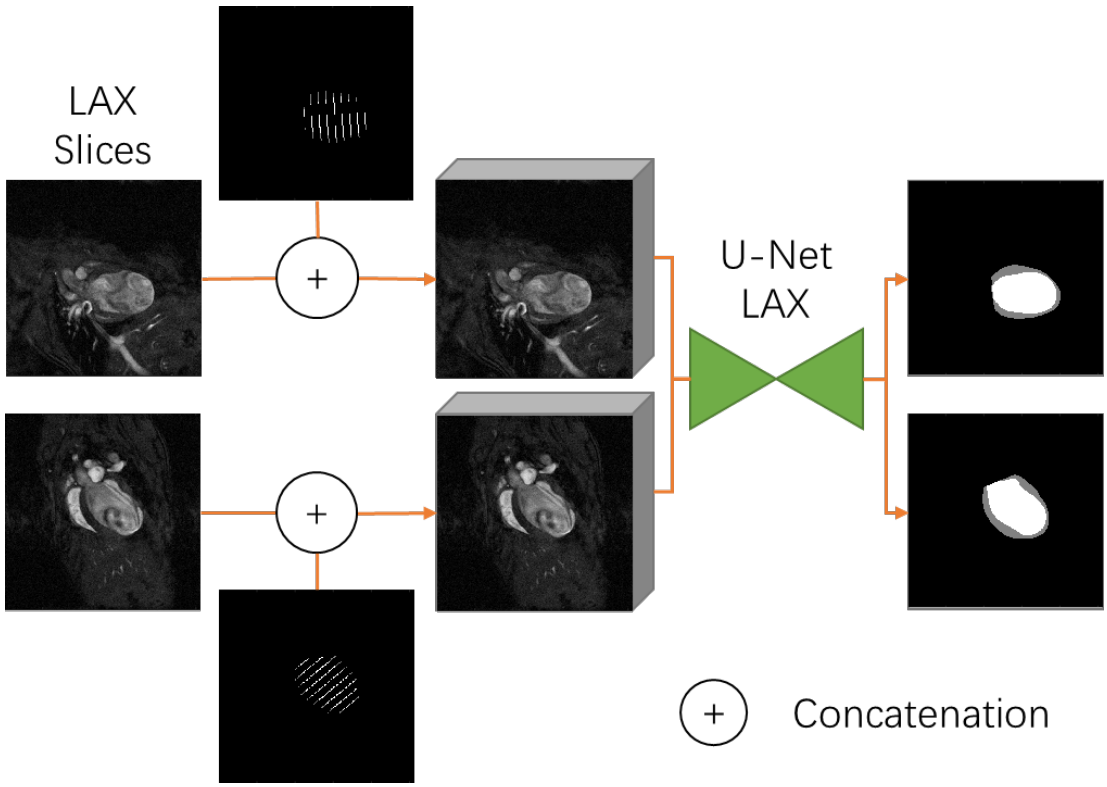


Figure 6. 3 Finally, we concatenate the image and projected segmentation layers and use our modified U-Net to segment the LAX slices. This completes the segmentation method.

In chapter 5, we used a bi-ventricular statistical shape model to develop and validate our neural network for reconstructing 3D ventricular shapes from sparse contours. We then focused on the LV and used the LV shapes extracted from the DETERMINE database, consisting of patients with clinical evidence of myocardial infarction, to validate our method on clinical datasets. We also demonstrated the benefits of including the LAX slices and the adversarial loss, particularly for the experiments with misaligned slices that are commonly seen in human CMR images.

For mouse CMR images, due to the high beating rate of the heart (around 500 *beats/min* as discussed in section 2.1.2) and ECG gated imaging (as discussed in section 2.2.4), we do not see obvious misalignments of the slices, as shown in Figure 6.15. Also,

by considering our LAX segmentation strategy, the potential small discrepancies between the SAX and LAX slices are already partially corrected (this could bring the drawback that the LAX segmentation becomes biased to the SAX segmentation results, but the improved LAX segmentation accuracy suggests that such bias is beneficial and the slices are indeed spatially consistent.) Therefore, we believe the mouse CMR segmentation results can be treated as largely free of motion artefacts. In the previous chapter, we showed the improvement in accuracy in LV shape reconstruction resulting from combining SAX and LAX slices (see Table 5.2), and thus we decide to use our 3D U-Net as reconstruction network to limit the computational costs of our experiments and achieve good reconstruction accuracy at the same time.

For developing and validating our reconstruction network, we prepared 200 cases for both the statistical model and the DETERMINE data experiments. Mouse CMR experiments typically involve smaller samples, and at this point in the development we only had access to 100 subjects as described in section 4.3.2 (which we refer to as the Oxford cohort). In order to have a comparable range of LV shape variation in the training dataset, we expanded our database by introducing a second 100 cases (which we refer to as the Tubingen cohort) from a different imaging centre. The same experienced operator prepared the manually verified reference CIM models as our gold standard for training both networks.

In this chapter, we integrate our segmentation method and shape reconstruction network forming a complete pipeline for analysing the mouse LV function from CMR images. We increase the size of database providing more training data for both our

segmentation and reconstruction networks. We include the LAX slices, which have shown to improve the accuracy of the reconstructed LV shapes especially around the base and implemented a SAX segmentation projection layer to improve the LAX segmentation network performance. For the reconstruction network, we choose our 3D U-Net (without the adversarial component), assuming the slices of mouse CMR images to be spatially consistent as discussed above.

6.2 Dataset

Our dataset consists of a total number of 200 CMR images from two cohorts, including 34 infarct cases and 166 normal cases. Different from the Oxford cohort, the CMR imaging measurements of the Tubingen cohort were performed on a 7.0 T Bruker Pharmascan 70/16, equipped with a B-GA 9S *HP* (760 *mT/m*) gradient system, using a 72 *mm* volume transmit-only and a cryogenic 4 element 1H array receive-only coil and the IntraGate™ self-gating tool (Bruker, Ettlingen, Germany). The measurements were based on the Intragate FLASH method (repetition time = 6.2 *ms*, echo time = 1.3 *ms*, flip angle = 10 *degrees*, field of view = 22 × 22 mm^2, slice thickness = 1.0 *mm*, matrix = 128 × 128 and oversampling = 100).

The imaging plane was localized using scout images showing the 2-CH and 4-CH views of the heart (while only the 4-CH view was saved and available in the dataset), followed by acquisition in SAX slices, orthogonal to the septum in both scouts. Multiple contiguous SAX slices consisting of around 10 slices were acquired for complete coverage of the left and right ventricles. The total number of slices (including SAX and

LAX slices) is 897. To cover the cardiac cycle 10 frames were saved for each slice, giving a total number of 8970 images (which is less than half the amount of the Oxford cohort with 22167 images in total). One example of the Tubingen cohort datasets is shown in Figure 6.4 and Figure 6.5 with SAX slices and 4-CH view LAX for ED and ES frames respectively.

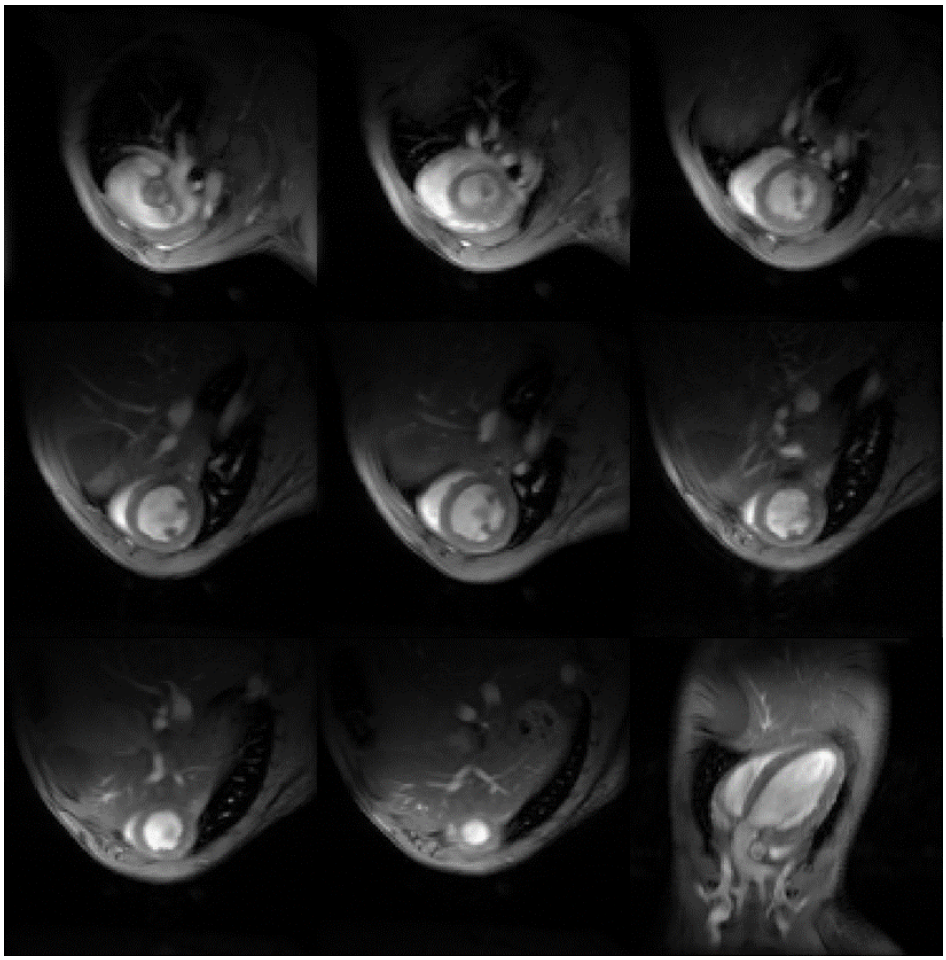


Figure 6. 4 SAX slices (base to apex) and 4-CH LAX slice images of the ED frame. Compared to Figure 4.13 to Figure 4.15, the number of SAX slice is smaller, and the contrast is better overall, while segmenting epicardial contours remains challenging.

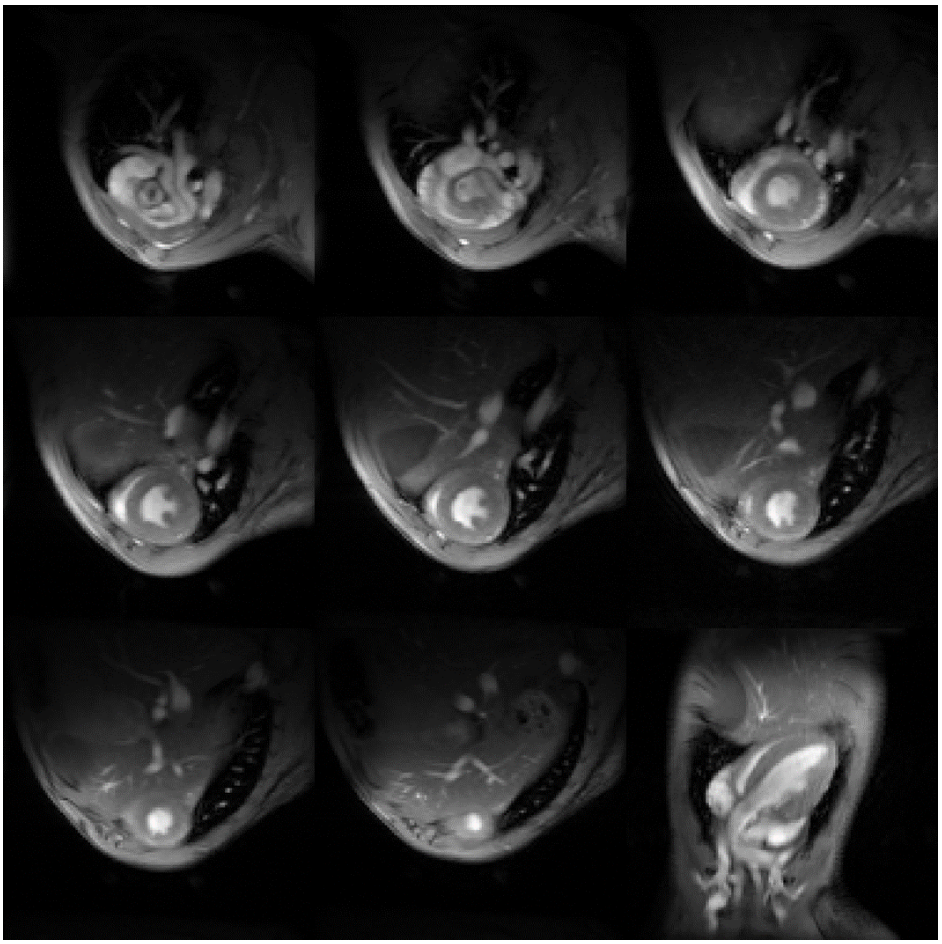


Figure 6. 5 SAX slices (base to apex) and 4-CH LAX slice images of the ES frame. Compared to Figure 4.13 to Figure 4.15, the number of SAX slice is smaller, and the contrast is better overall, while segmenting epicardial contours remains challenging.

The contrast of the images is better compared to the Oxford cohort. However, due to the limited number of SAX slices, the most apical slice, where the blood pool of the LV disappears in the ES frame due to contraction as shown in Figure 4.13, is not included, making the LAX slice more important for reconstructing the apex.

6.2.1 GPM Analysis

In order to preserve the consistency of the database, the same experienced operator performed LV analysis using the CIM software package (which applies the GPM

method introduced in section 2.4.5). The GPM method has been validated using both human and mice cine CMR data and we have introduced applications of the method on both types of data in previous chapters, where a 3D mathematical model of the LV was interactively customized for each of the MR scans. For the generation of gold standard volumes, the geometry was optimized by manually introduced guide-points in an iterative process. At the end of the process, a pair of endocardial and epicardial LV surfaces is saved as a CIM model (a surface model), which we use as the reference shape of the LV.

6.2.2 Data Preparation for 2D Image Segmentation

The size of each original image was 128 × 128 and the field of view is 22 × 22 mm^2, giving the pixel size of 0.17 × 0.17 mm^2, which is larger than the resampled image resolution of 0.10 × 0.10 mm^2 that we used for the Oxford cohort. To match the image resolution, we resized the images to 256 × 256, giving a pixel size of 0.086 × 0.086 mm^2, which was close to 0.10 × 0.10 mm^2 and introduced a slightly wider range of LV sizes, serving as an additional level of augmentation. We also used the CIM model to prepare the output binary masks for background, LV myocardium and LV cavity (which are concatenated as three layers of the output in the same order). We first aligned the image planes with the LV shape in 3D space using image metadata, and by calculating the relative position of each pixel a corresponding class was assigned to it accordingly. Examples of endocardial and epicardial contours of the Tubingen cohort are shown in Figure 6.6 to Figure 6.12.

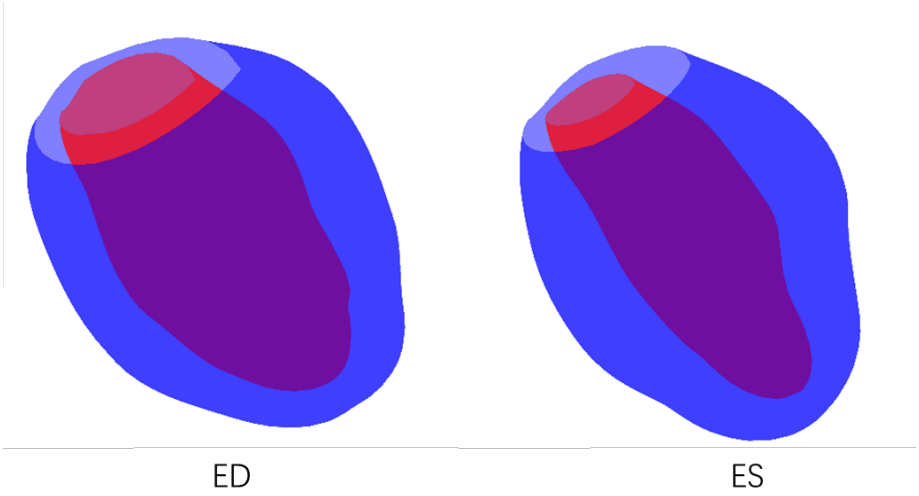


Figure 6. 6 Examples of LV reference contours (a) CIM model surfaces from which the contours were calculated, ED and ES frames. The red and blue surfaces are endocardial and epicardial surfaces, which are corresponding to the red (endocardial) and blue (epicardial) contours in Figure 6.7 to Figure 6.12.

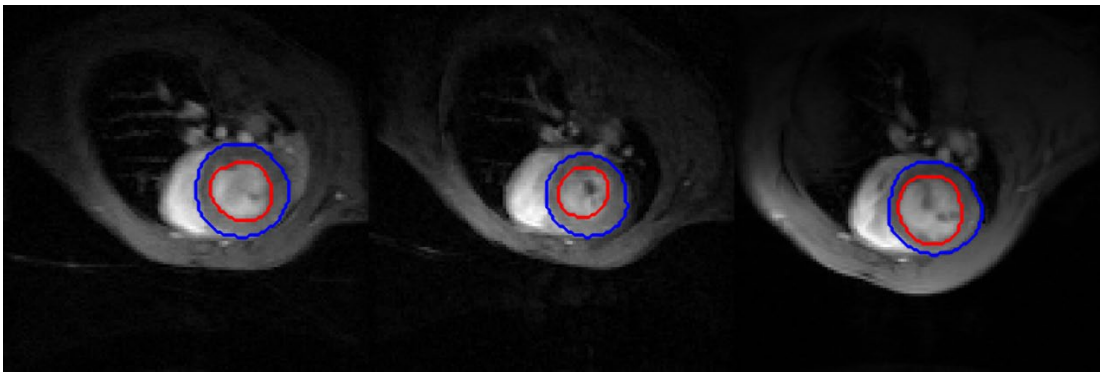


Figure 6. 7 Examples of LV reference contours (b): basal slice, ED frame.

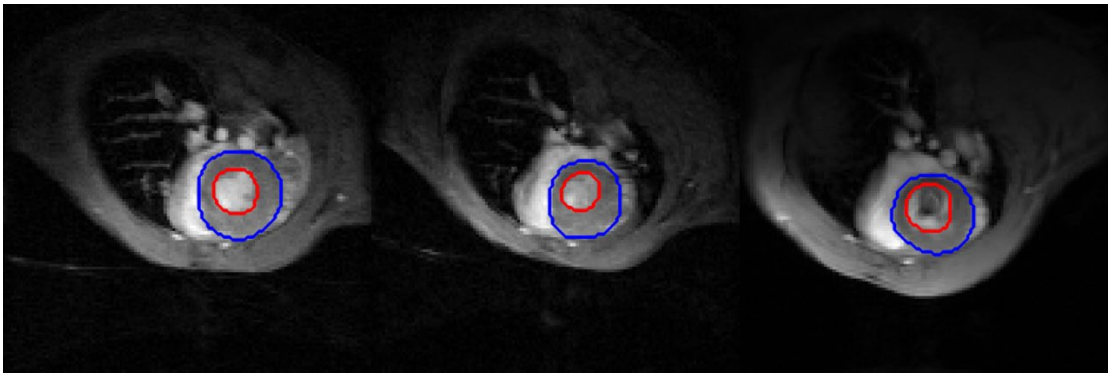


Figure 6. 8 Examples of LV reference contours (c): basal slice, ES frame.

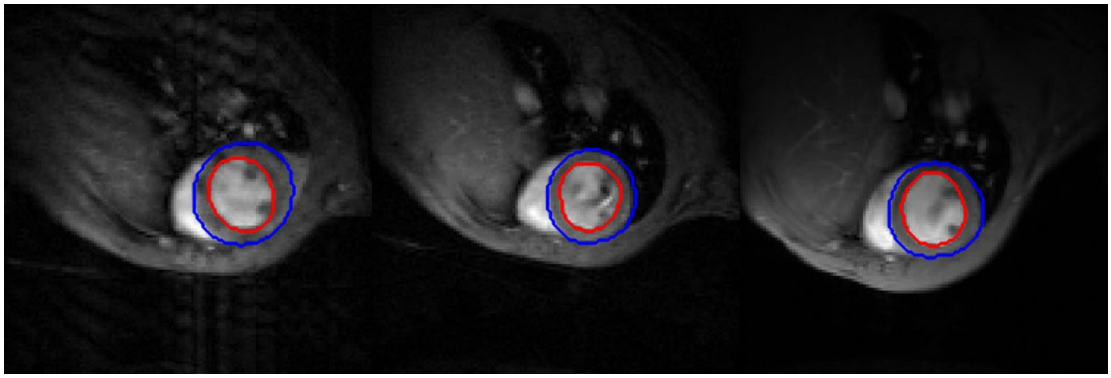


Figure 6. 9 Examples of LV reference contours (d): mid-cavity slice, ED frame.

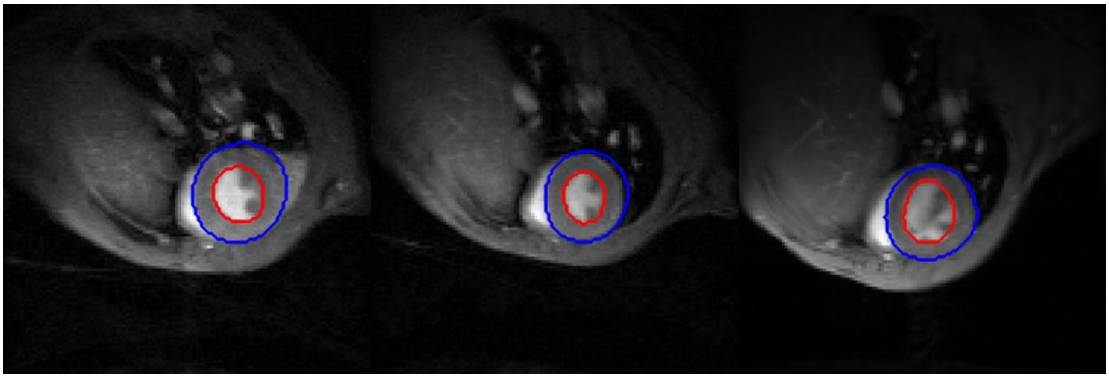


Figure 6. 10 Examples of LV reference contours (e): mid-cavity slice, ES frame.



Figure 6. 11 Examples of LV reference contours (f): apical slice, ED frame.

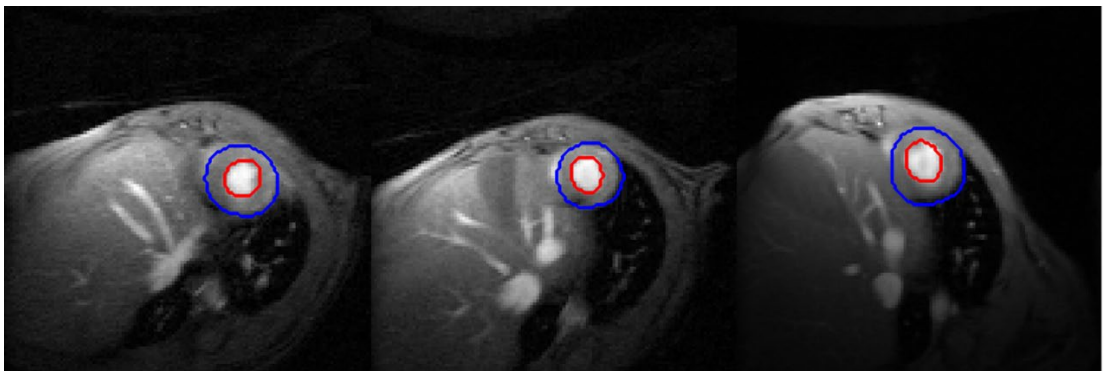


Figure 6. 12 Examples of LV reference contours (g): apical slice, ES frame.

To prepare the input layer images, we applied the same augmentation strategy to both training and validating data as for the Oxford cohort, including placing the centre of the LV in the middle of the image, 360 degrees random rotation, rigid translation along rows and columns by ± 10 *pixels* and zero padding in areas not covered by the original images.

6.2.3 Data Preparation for 3D Shape Reconstruction

To generate output reference volumes, we placed the centre of the LV shape in the centre of a regular grid with the size of 128 × 128 × 128 voxels and voxel size of 0.12 × 0.12 × 0.12 mm^3. The voxel size was chosen so that the volume was big enough to enclose all LV shapes as well as keeping a reasonable proportion of the volume (around 50% considering the statistical shape model and the DETERMINE data experiments) occupied by LV. Examples of the LV shapes from both datasets are shown in Figure 6.13 and Figure 6.14.

The shapes of the mouse LV are similar to human LV shapes as can be confirmed by comparing these to Figure 5.3. The relative size the LV cavity is smaller compared to human LV (i.e. relatively thicker myocardium), suggesting a powerful contracting mechanism, and for infarct mouse hearts we can see that the region of thin myocardium after remodelling is relatively larger and located around the apex. The orientations of the LV shapes are generally arranged to have the base at the top and the apex at the bottom, while the mitral valves could be pointing along the paper, into the paper or out of the paper as shown in the first row of Figure 6.13.

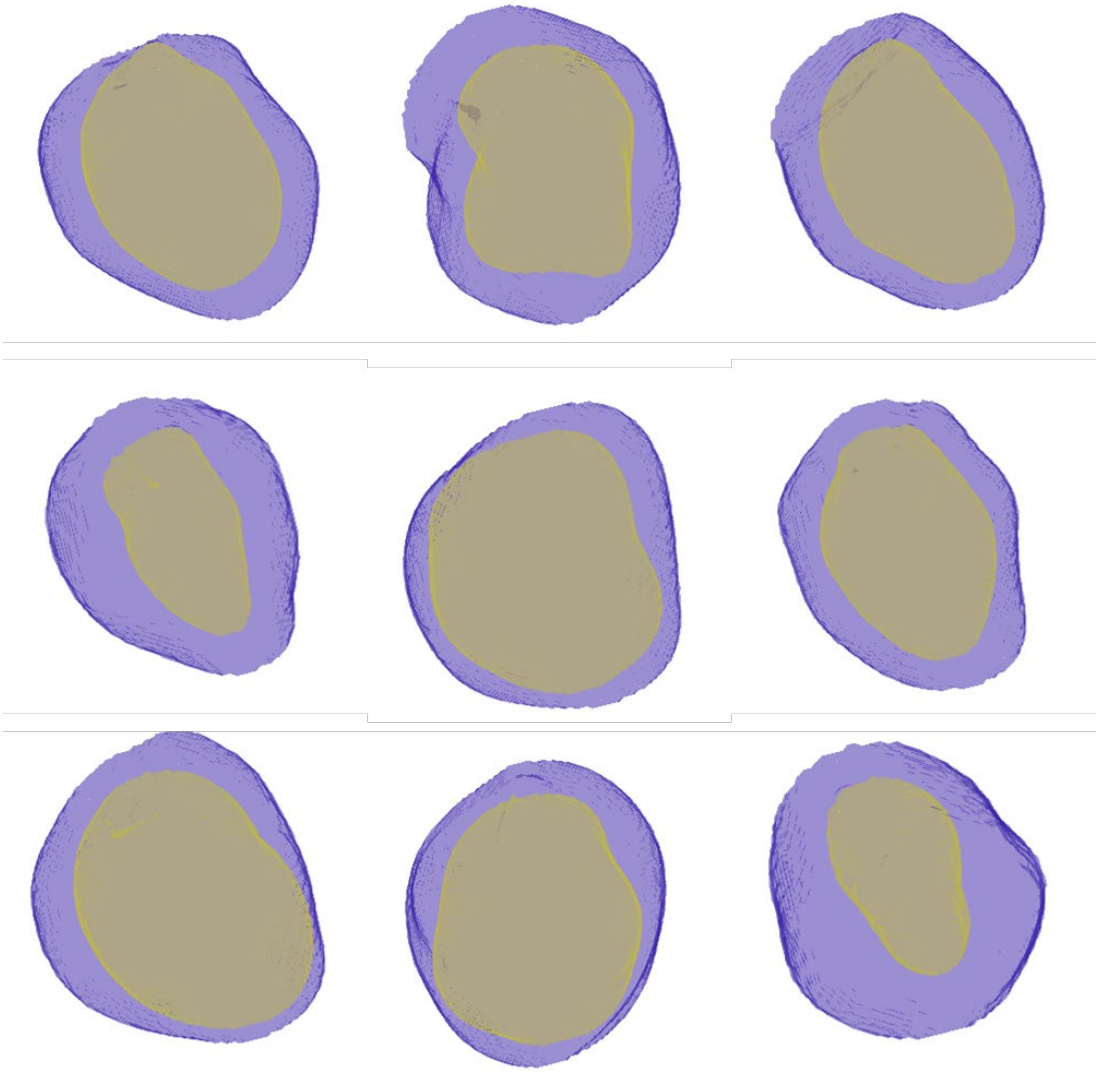


Figure 6. 13 Examples of reference LV shapes from the Oxford cohort. The purple and golden surfaces represent the epicardial and endocardial surfaces of the LV. The size and shape of the LV and the thickness of myocardium varies within the database.

The average thickness of the myocardial wall is larger in the Tubingen cohort as there are no infarct cases included, while the shapes of the LV are still quite different from each other.

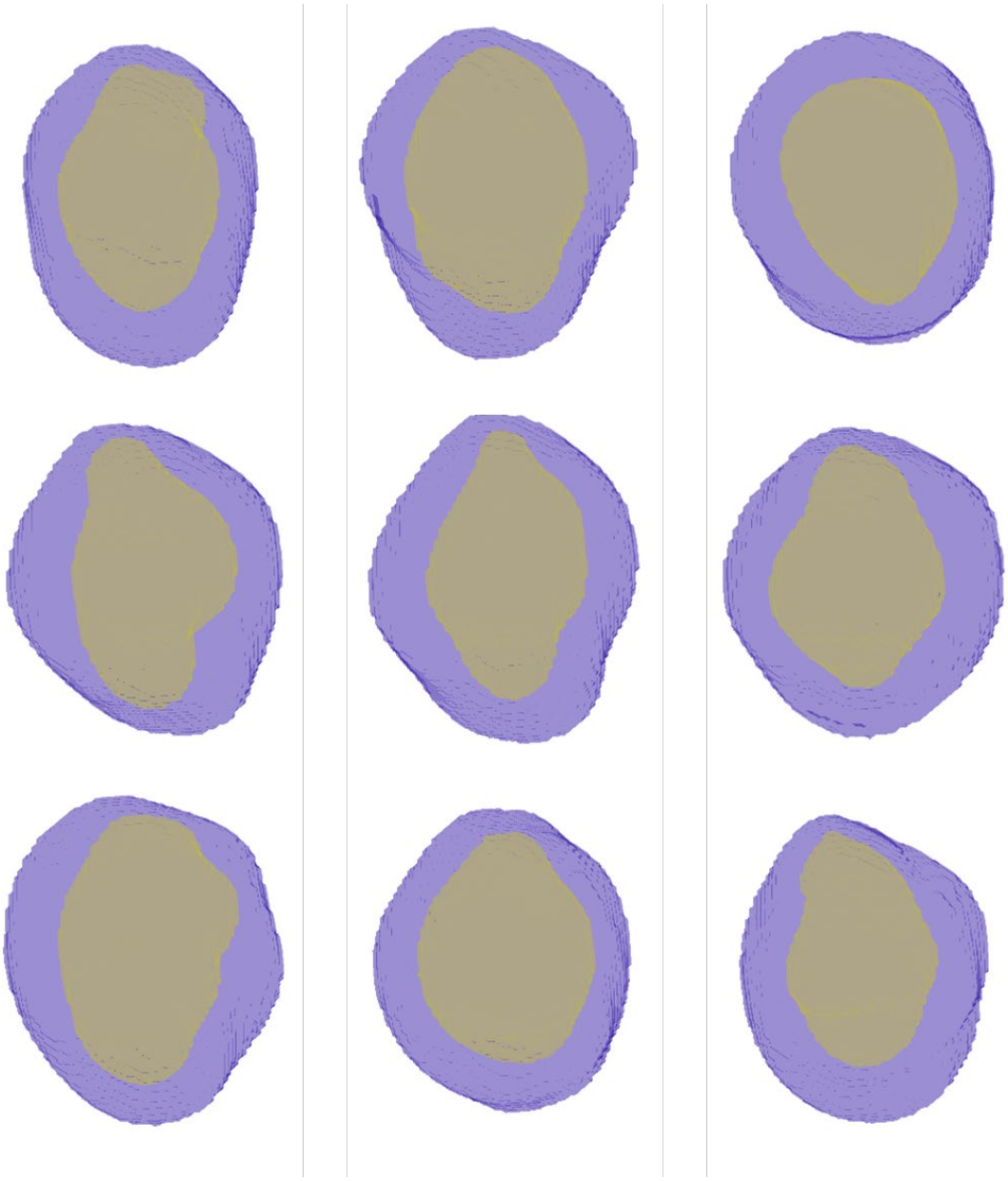


Figure 6. 14 Examples of reference LV shapes from the Tubingen cohort. The purple and golden surfaces represent the epicardial and endocardial surfaces of the LV. The size and shape of the LV and the thickness of myocardium varies within the database.

To prepare the input sparse volume, we first aligned the segmentation of the image planes with the 3D volume, and for each pixel on the 2D images, we assigned the corresponding label to the nearest voxel in the 3D volume by calculating the distances between the centre of the pixels and voxels, as long as the minimum distance was

smaller or equal to 0.05 *mm*. This produced a slice thickness equivalent to 0.1 *mm*. The ratio of slice thickness / voxel size was in the same range as for the statistical shape model and the DETERMINE data experiments.

As there were only 4-CH view LAX slices for the Tubingen cohort, the reconstruction task became more challenging, since the network must be able to cope with sparse input volumes including either one or two LAX slices in order to utilise all the available images. To train a network that was more adaptive to the LAX arrangement, for the Oxford cohort we prepared three different input volumes for each case, including a sparse volume with only a 2-CH view LAX slice, a sparse volume with only a 4-CH view LAX slice, and a sparse volume with both 2-CH and 4-CH view LAX slices.For reconstructing LV shapes both types of sparse volume with single LAX slice are similar to each other (except for the orientation, which also becomes indistinguishable after augmentation.) Examples of input sparse volumes from both datasets are shown in Figure 6.15.

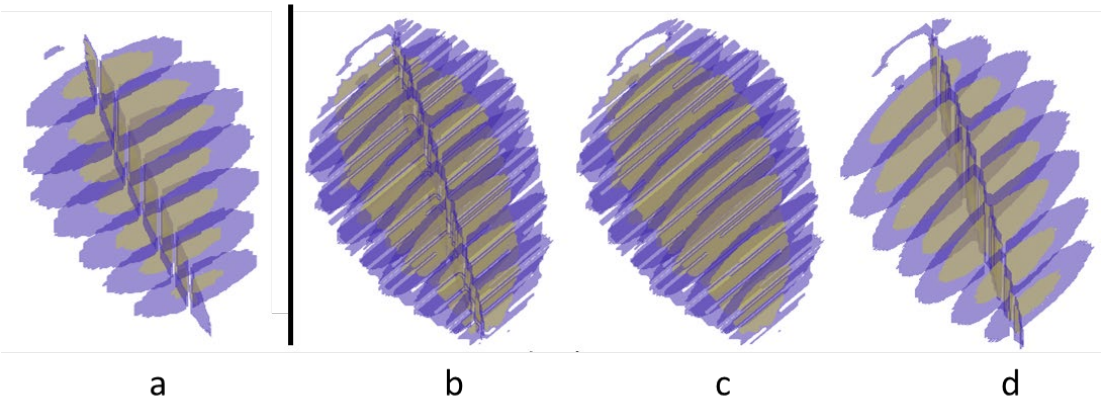


Figure 6. 15 Examples of input sparse volumes. For the Oxford cohort, three combinations (both 2-CH and 4-CH views, 2-CH view only and 4-CH view only) of LAX slices are shown from (b) to (d). For the Tubingen cohort, only one LAX slice is

available as shown in (a). The colours are the same as Figure 6.13 and Figure 6.14.

From the coherence of the slices in Figure 6.15, it is apparent that the assumption of spatial consistency between SAX and LAX slices (also inducted from section 4.3.5) can be made without much loss of accuracy.

We applied augmentation to the training and validation datasets (paired input and output volumes prepared using segmentation results and reference LV shape respectively), including random rotation (between 0 and 360 *degrees*) along the LAX, random tilting (between ±10 *degrees*) along the two perpendicular axis to the LAX, and random translation (up to ± 5 voxels) in all three axes.

6.3 Method

6.3.1 Pipeline Overview

A summary of our LV anatomy and functional analysis pipeline is illustrated in Figure 6.16.

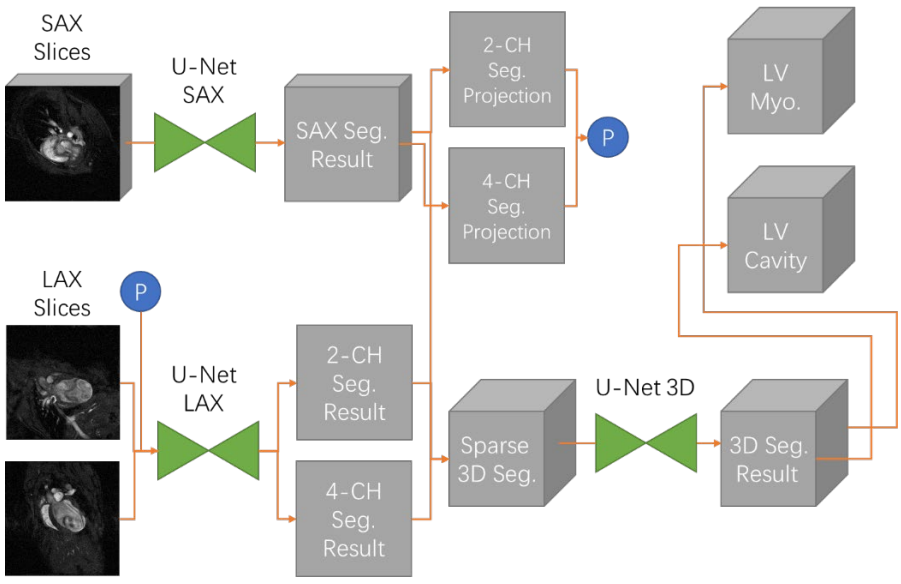


Figure 6. 16 Pipeline for LV anatomy and functional analysis of mouse heart. The SAX

slice images are segmented with our variation of U-Net, and the results are used to produce the projections of SAX segmentation on the LAX image plane(s). The projection layer(s) are then combined with the corresponding LAX image(s) and passed through our LAX segmentation U-Net giving the LAX segmentation. The segmentation results from all slices are used to produce the sparse input volume for 3D shape reconstruction, and the prediction from our variation of 3D U-Net is finally converted to LV cavity and myocardium for anatomical and functional analysis. The point P represents the segmentation projection layer for LAX segmentation.

Compared to an end-to-end method, which takes the 2D images as input and 3D predictions as output, we broke the pipeline into three steps with simplified targets and developed algorithms. The three steps are linked with two relatively complicated (compared to convolutions) operations: generating the segmentation projection layer for LAX slice segmentation and generating sparse input volumes for 3D shape reconstruction. The pipeline is fully automatic and at the same time capable of producing intermediate results including 2D segmentation results.

6.3.2 Segmentation Neural Network

The architecture of the networks we used to segment 2D images are shown in Figure 6.17. For the SAX segmentation network, the input layer consists of only one layer with image intensity values (i.e. N = 1). For the LAX segmentation network, the input layer consists of two layers with image intensity values and projected SAX segmentation (i.e. N = 2). Different from the original U-Net convolutional block pattern in the encoding

path, which has the same number of filters within each block, we doubled the size of the second convolutional layer inspired by the 3D U-Net [28], which was published later based on the U-Net [102], to prevent the bottleneck effect of down-sampling.

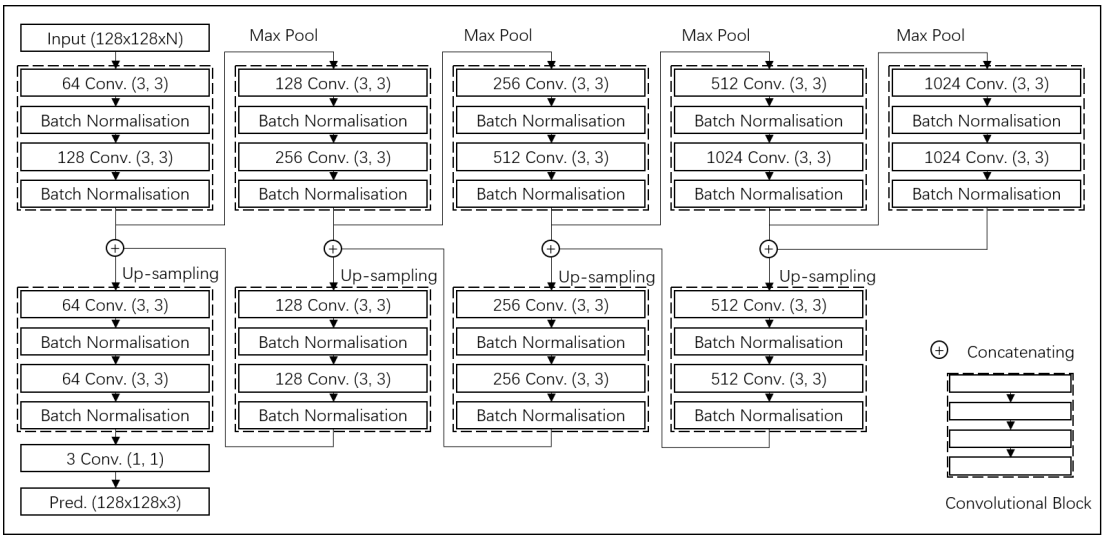


Figure 6. 17 Illustration of the segmentation network architecture.

6.3.3 Reconstruction Neural Network

By assuming the segmentation of 2D images are spatially consistent, based on both visual inspection (see Figure 6.15) as well as the fact that SAX slice segmentation results improved the accuracy of the LAX slice segmentation, we used our 3D U-Net for reconstructing the LV shape using mouse CMR images, leaving out the adversarial component that we used before, since it would increase the complexity of the method and its main use was to correct for misalignment. The network architecture is shown in Figure 6.18.

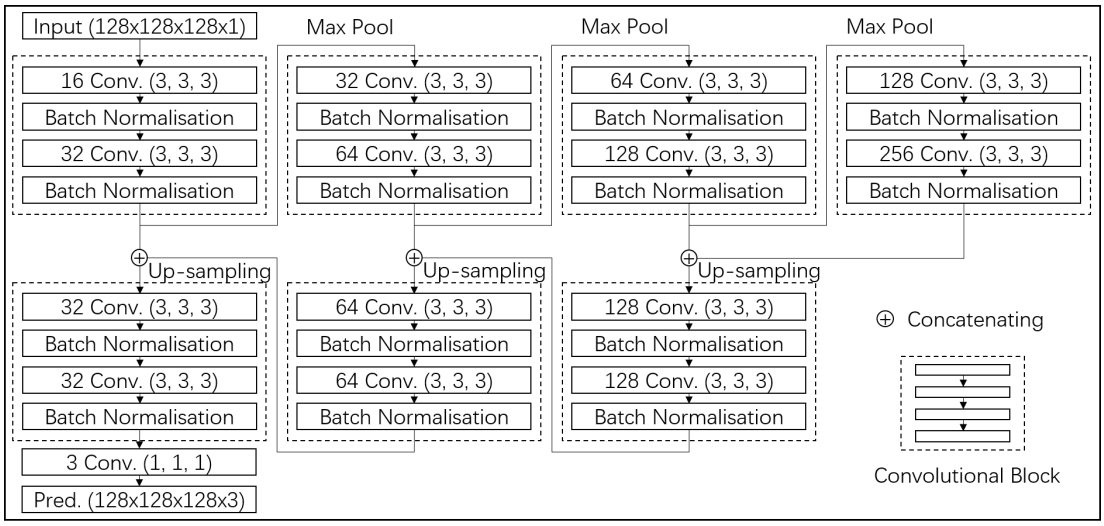


Figure 6. 18 Illustration of architecture of neural network for LV shape reconstruction.

The number of down-sampling stages was limited to three and the number of convolutional filters was comparatively smaller than in our 3D U-Net, as the features calculated within the network are four-dimensional (4D) with a linear expansion of the batch-size for each sub-epoch, giving a large memory consumption, and the scale of the network is limited by the computational hardware. The number of parameters within the network, on the contrary, are smaller than in our 2D U-Net due to the limited number of filters.

6.3.4 Cardiac Function Estimation

To analyse the LV function, we focused on evaluating the cavity volume and myocardial mass. For each frame, we calculated the cavity and myocardial volumes by integrating the voxels of corresponding binary volumes calculated from the predictions of the 3D U-Net. The volume of myocardium was then converted to myocardial mass (LVM) by multiplying the myocardial volume times the myocardial density of 1.05 *g/ml*. We then used the volume sequence of the cardiac cycle to identify the end-diastolic (ED) and

end-systolic (ES) frames (which were defined as the frames with maximum and minimum cavity volumes respectively), and the values of ED cavity volume (EDV) and ES cavity volume (ESV) were used to calculate the stroke volume (SV) and ejection fraction (EF) as described in section 2.1.1:

$$SV = EDV - ESV$$

$$EF = 100\% \times SV/EDV.$$

In previous chapters, our development of segmentation and reconstruction methods focused on the similarity between the predicted structure and the reference structure. In order to produce meaningful information for LV function analysis, here we include metrics evaluation at the end of our pipeline as the target of the task.

6.4 Experiments and Results

We performed a 5-fold cross-validation experiment with 40 cases (out of the total 200) in the test set for each fold. Each time, for the remaining 160 cases, 140 cases were randomly chosen as training data and 20 cases as validation data. All training, validation and testing cases are mixed with an equal number of cases from both datasets.

6.4.1 Implementation

We implemented the neural networks in Keras with Tensorflow as backend. The networks were trained with Adadelta [142]. We used one NVIDIA GeForce GTX 1080 Ti GPU for all training and testing experiments.

6.4.2 Intermediate Evaluation

To investigate the performance of the intermediate stages of our pipeline, including the segmentation of 2D images and reconstruction of 3D structure, we calculated the Dice similarity metrics between our predicted results and the reference binary masks (in 2D or 3D) prepared from the manually verified LV shapes. The average Dice scores for SAX slice segmentation results of LV cavity and myocardium were 0.87 and 0.79 respectively, and the Dice scores for LAX slice segmentation results of LV cavity and myocardium were 0.94 and 0.84 respectively. The performance of the network remained at a similar level compared to our experiment with the Oxford cohort as discussed in section 4.3, where the corresponding Dice score values were 0.88 and 0.81 for SAX slices, and 0.94 and 0.84 for LAX slices.

The average Dice scores between the predicted and reference 3D binary volumes of LV cavity and myocardium were 0.89 and 0.82 correspondingly. We assumed the sparse input volume to be spatially consistent; however, compared to our experiment with the DETERMINE data of consistent SAX/LAX slices, which had the corresponding Dice scores of 0.98 and 0.96, there was an important difference. There were mainly three sources that contributed to the difference: the error accumulated from 2D segmentation, the variation of the 3D U-Net performance between the distributions of the LV shape for human and mouse hearts, and the assumption of spatial consistence between slices for some cases.

For the DETERMINE data experiments, the sparse input volume and the reference volume were aligned in space. However, for our pipeline in this chapter, the

segmentation results and the reference volume were not perfectly aligned, and even though the segmentation results are spatially consistent, the generated LV shape is not perfectly aligned to the reference volume. The LV shapes of the mouse heart were more complicated with a larger variation, as shown in Figure 5.3, Figure 6.13 and Figure 6.14, which could lead to a difference in the network performance. Finally, when visually evaluating cases with low Dice scores, we found cases with 2D segmentations not well aligned (usually caused by inaccurate 2D segmentation), and thus the assumption of spatially consistent input sparse volume would not hold for those cases. In these cases, the overall condition of the input volumes from segmentation results became more similar to the misaligned experiments of the DETERMINE data, and the Dice scores agreed better to those values (0.92 and 0.84) as well. An example is shown in Figure 6.19, where one mid-apical slice and one mid-basal slice were poorly segmented, and the LAX segmentation result was also corrupted, ending up with misalignments between the slices.

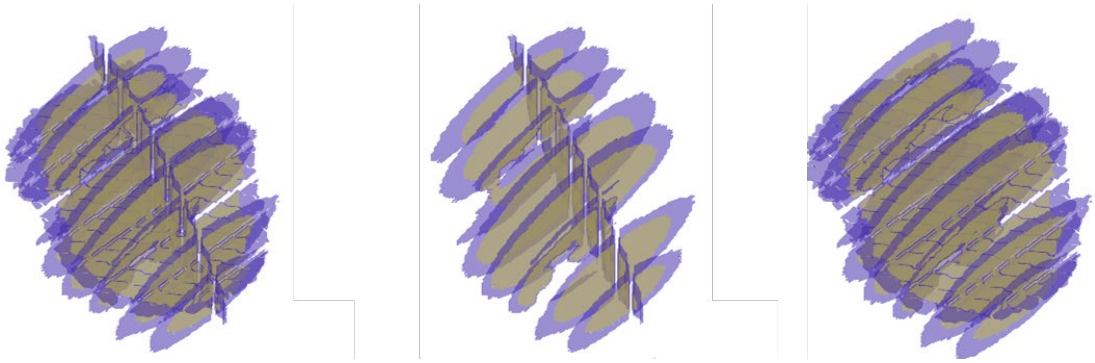


Figure 6. 19 Illustration of misalignment of sparse volume with suboptimal segmentation results. All combinations of SAX and LAX are shown, with both 2-CH and 4-CH views, 2-CH view, or 4-CH view. The colours are the same as Figure 6.13 and

Figure 6.14.

6.4.2 LV Functional Metrics

The segmentation and reconstruction errors introduced in the intermediate stages could have different effects for quantitative analysis of the LV function. We calculated the absolute errors between metrics evaluated from our pipeline and the reference analysis, and the absolute errors were 7.6 ± 9.7 *mg* in LVM, 5.5 ± 4.8 μl in EDV, 4.5 ± 5.4 μl in ESV, 4.6 ± 3.6 μl in SV, and 5.6 ± 4.5 *%* in EF. The mean values and the standard deviations of reference LVM, EDV, ESV, SV and EF are 97 ± 28 *mg*, 70 ± 34 μl, 38 ± 35 μl, 32 ± 7 μl, and 52 ± 16 *%* respectively, showing a large variation within the dataset.

We plotted both Bland-Altman plots of predicted metrics and the reference metrics and the correlation between predicted metrics and the reference metrics in Figure 6.20 to Figure 6.24. The orange line in the correlation plots represents y = x, and for metrics directly estimated from the pipeline (i.e. EDV, ESV and LVM) the data points align well around the lines, showing strong linear relation between the predicted and reference metrics, with very few outliers for LV with large cavity volumes and myocardial mass. The data points in SV and EF correlation plots spread out more compared to EDV, ESV and LVM, suggesting a lower correlation between the predicted and reference metrics. We also calculated the correlation coefficient for LVM, EDV, ESV, SV and EF, which were 0.92, 0.98, 0.98, 0.81 and 0.93 respectively, with all p-values smaller than 10^{-45}.

The three red lines in the Bland-Altman plots were plotted using the mean and

standard deviation (std) values of the difference between metrics, where from top to bottom, the three lines represent mean – 1.96 × std, mean, and mean + 1.96 × std. Data points lying outside the top and bottom lines were considered as outliers, and for all the metrics only a small number of outliers could be seen in the Bland-Altman plots, suggesting the robustness of our pipeline.

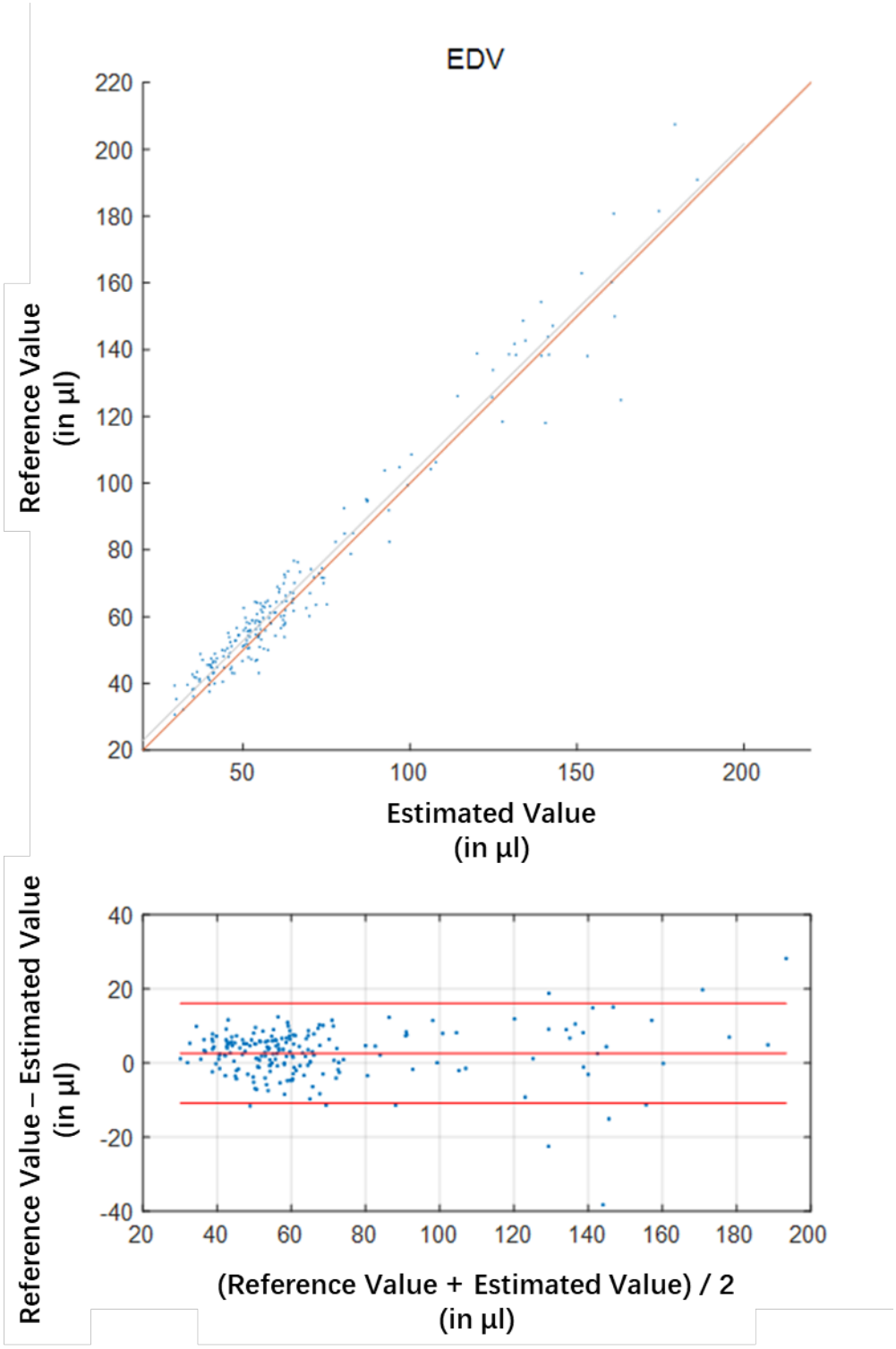


Figure 6. 20 Bland-Altman plot and correlation plot for LV functional metrics: EDV.

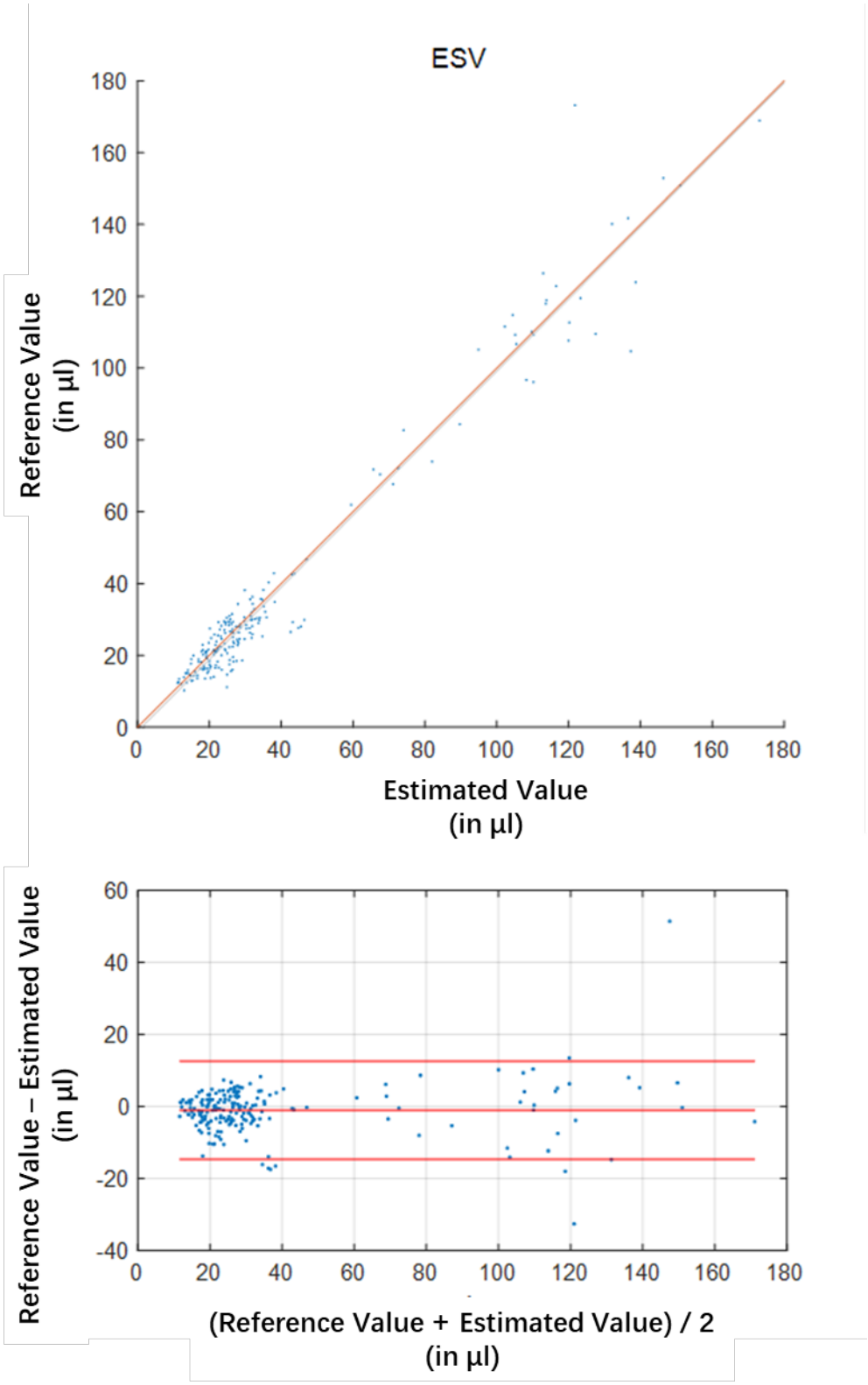


Figure 6. 21 Bland-Altman plot and correlation plot for LV functional metrics: ESV.

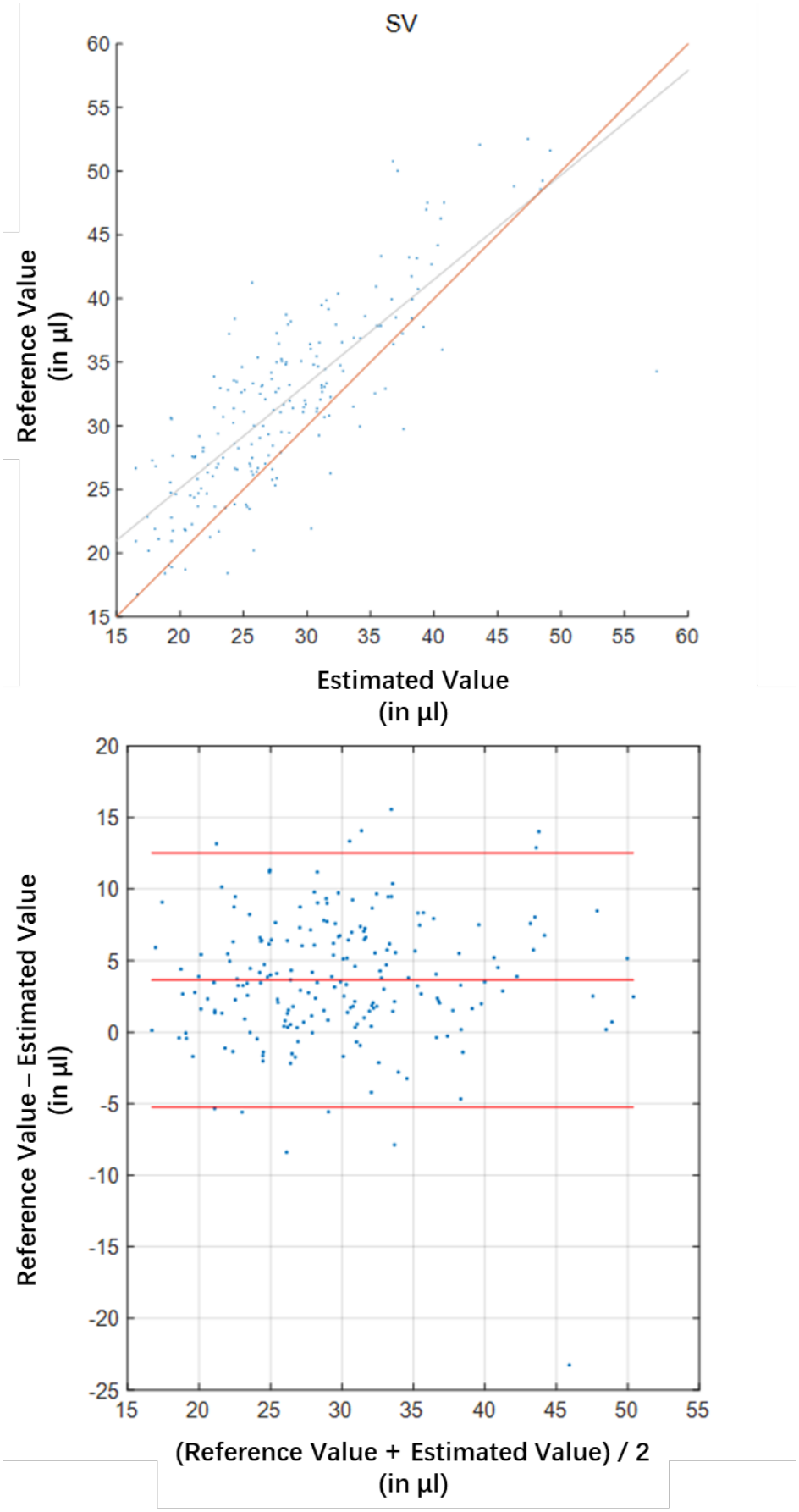


Figure 6. 22 Bland-Altman plot and correlation plot for LV functional metrics: SV.

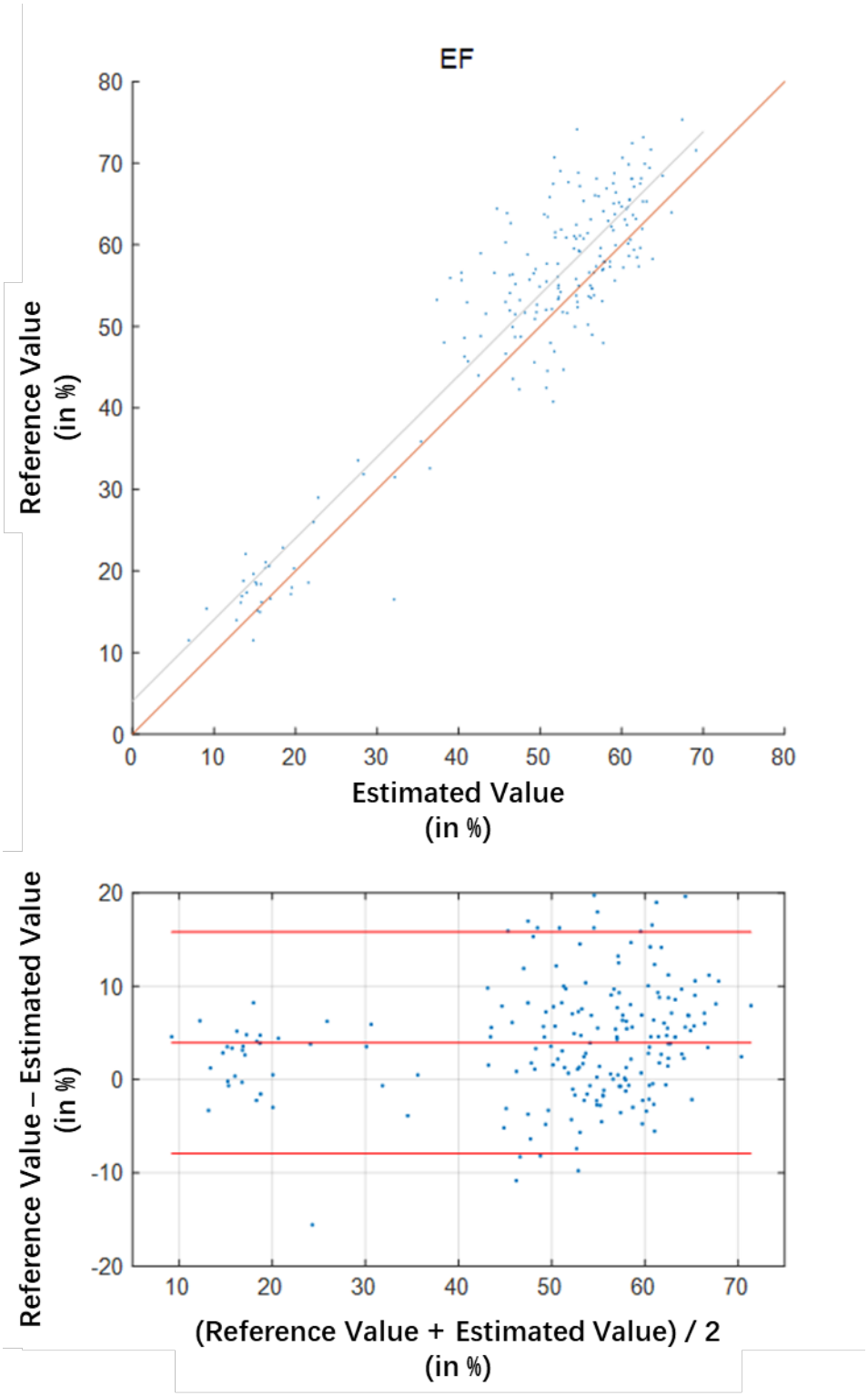


Figure 6. 23 Bland-Altman plot and correlation plot for LV functional metrics: EF.

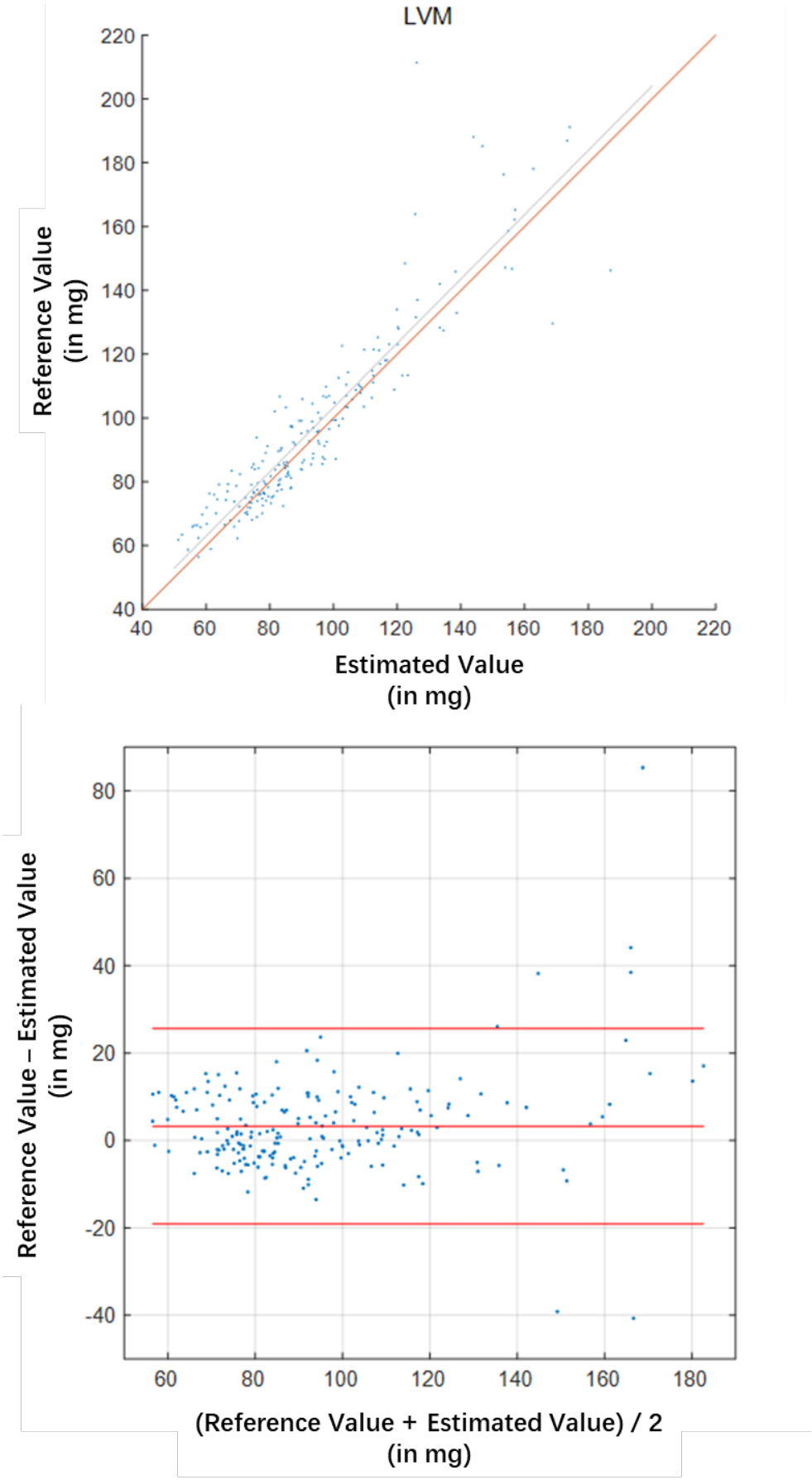


Figure 6. 24 Bland-Altman plot and correlation plot for LV functional metrics: LVM.

We could also see biases in the Bland-Altman plots. For LV volumes, there was a positive bias (+2.6 μl) between the reference EDV values and the estimated EDV values and a negative bias (-1.0 μl) between the reference ESV values and the estimated ESV values, as shown in Figure 6.20 and Figure 6.21, meaning that our pipeline underestimated EDV and overestimated ESV on average. SV was calculated using EDV and ESV and therefore had a larger positive bias (+3.6 μl) between the reference SV values and the estimated SV values, as shown in Figure 6.22. Similarly, EF had a positive bias (+3.9 %) between the reference EF values and the estimated EF values, as shown in Figure 6.23. For LV mass, there was also a positive bias (+3.2 *mg*) between the reference LVM values and the estimated LVM values, as shown in Figure 6.24, meaning that our pipeline underestimated LVM (i.e. predicted thinner myocardium) on average.

EDV, ESV and LVM were directly evaluated from the network output, and the corresponding Bland-Altman plots suggested that the accuracy of our pipeline reduced when the heart became larger. The distribution of the data points showed that the majority of our dataset were smaller hearts, and by considering the data-driven nature of our pipeline, the negative effect of variation in size was clearly shown in the results and was not removed by our augmentation strategy.

EF is fundamental for assessment of cardiac function, and from the Bland-Altman plot we can clearly see a separation between the infarct hearts (with relatively small EF values) and the normal hearts. From the distribution of the data points we can see that there were more normal hearts, however, the infarct hearts had a better overall accuracy

and were not affected by the class-imbalance situation.

We also compared the percentage absolute errors to inter- and intra-observer variability of manual analysis [105, 107] to investigate the accuracy of our method. The comparison is shown in Table 6.1.

Table 6. 1 Percentage absolute error compared to intra- and inter-observer variability.

	[107]		[105]		Absolute Error / %
	Intraobserver / %	Interobserver / %	Intraobserver / %	Interobserver / %	
LVM	NA	NA	NA	NA	8 ± 7
Diastolic mass	7 ± 3	11 ± 2	3 ± 2	5 ± 3	NA
Systolic mass	4 ± 2	3 ± 3	4 ± 3	3 ± 3	NA
EDV	3 ± 1	5 ± 3	1 ± 1	3 ± 3	8 ± 6
ESV	3 ± 4	9 ± 4	3 ± 1	6 ± 5	14 ± 12
SV	6 ± 3	2 ± 4	3 ± 2	4 ± 4	16 ± 12
EF	3 ± 2	3 ± 2	2 ± 1	3 ± 2	12 ± 10

For LV mass evaluation, diastolic and systolic masses were considered independently in [105, 107], and in our method we calculated the mean value of masses of all frames as LVM. The percentage absolute error of our method was similar to the intra- and inter-observer variability values. For LV volume evaluation, the percentage absolute errors of our method were comparatively larger. A key difference was that the

intra- and inter-observer variability compared between integrated values from SAX slices, where the papillary muscles were considered as part of myocardium in [105, 107] (for the purpose of in-vivo and ex-vivo comparison in the study). The variability would increase in theory when papillary muscles were included in the blood pool as in our method, especially for ES frame. For SV and EF, because the errors accumulated from EDV and ESV, because our dataset included infarct hearts, which had extremely small SV and EF values, the overall percentage errors were further increased compared to normal hearts, as in [105, 107].

We also compared our average errors to [138], which investigated the accuracy of manual CIM analysis on mouse hearts that we used to produce our reference results. The average error between CIM analysis and integration of SAX slices for LVM, EDV, ESV, SV and EF were 2.3 ± 5.8 mg, 1.7 ± 3.2 μl, 2.3 ± 3.1 μl, -0.6 ± 3.3 μl, and -2.7 ± 4.3 % respectively. The average error between CIM analysis and our pipeline for LVM, EDV, ESV, SV and EF were 3.2 ± 11.4 mg, 2.6 ± 6.8 μl, -1.0 ± 7.0 μl, 3.6 ± 4.5 μl, and 3.9 ± 6.1 % respectively. There were similar levels of bias in both comparisons, and the standard deviations were larger in our experiments, which could be caused also by the large variability of our cohort.

6.5 Conclusions

In this chapter, we integrated our segmentation and 3D reconstruction networks for LV shape reconstruction, forming a fully automated pipeline for LV shape reconstruction and functional analysis using mouse CMR images. The performance of our pipeline

was promising, with an accuracy similar to intra- and inter-observer variability and manual GPM analysis. The database from two imaging centres largely increased the variability of both the image quality and the LV shape, and we also included infarct cases which made the task even more challenging. We also used a total number of 200 cases, which was much larger compared to previous pre-clinical studies for mouse heart.

We developed our segmentation network for mouse heart in chapter 4, and we used a variation of U-Net for SAX slices. For LAX slices, we introduced a segmentation projection layer to the U-Net, forming a two-layer input image, which improved the overall segmentation accuracy of the LAX slices. We kept the same segmentation network in our pipeline and achieved expected accuracy. However, the errors were accumulated from SAX segmentation to LAX segmentation, and finally affected the reconstruction accuracy. Therefore, for future work, improving segmentation accuracy is critical, such as introducing quality control and additional regularisations to improve the worst cases.

We developed our LV shape reconstruction network in chapter 5, and we decided to balance the performance and computational cost and used our variation of 3D U-Net which takes sparse input volume and output volumes representing LV cavity and myocardium, by assuming the segmentation results were spatially consistent, which was supported by the special characteristics of mouse CMR imaging and our LAX segmentation network approach. The differences between our experiments in chapter 5 and our pipeline included: initial segmentation errors, the difference between populations of human and mouse LV shapes, and some datasets with misalignment

between slices even with SAX-assisted segmentation of LAX. The performance of our reconstruction network in the pipeline was eventually similar to the experiments using the DETERMINE data with misaligned input volumes. For future work, including the adversarial loss that we evaluated in chapter 5 has the potential of further improving the accuracy of the reconstructed LV shapes when the limitation of computational power is relaxed.

We finally evaluated our method for LV functional analysis by calculating EDV, ESV, SV, EF and LVM. We compared the differences between our predicted values and our reference values (from CIM analysis) to both intra- and inter-observer variability of standard manual analysis and the differences between CIM analysis and standard manual analysis. The comparison showed that our method was reasonably accurate, and there was still room for improvements. The Bland-Altman plots showed that the size of the heart had obvious impact on the performance of the network, which was expected to be removed by our augmentation strategy. To resolve this issue, two alternative approaches could be investigated: 1. instead of generating random variations from each case, the augmentation strategy would be targeting on producing evenly distributed size of the heart in the training dataset; 2. instead of sampling all images to the same pixel/voxel sizes, the size of the hearts appeared on the input layer/volume would be normalised to simplify the task of our neural network. Even though there were only a small number of outliers, evaluating the worst cases was important and beneficial for further development. We did not see obvious impact from the class-imbalance issue between infarct and normal hearts, but the negative bias of our LVM prediction could

be caused by including infarct cases. The myocardium was very thin for infarct mouse heart, and since the detection of epicardial boarders was more challenging in both cases (from Oxford cohort), the network could learn to segment the myocardium wall with an intermediate thickness. Balancing the infarct and normal hearts in the training dataset could not resolve this potential issue, and further evaluation of reginal myocardial wall thickness was required to fully investigate it.

The current overall pipeline included three neural networks, joined by more complicated and time-consuming operations (compared to convolution operations), such as generating sparse volume input for reconstruction network from 2D segmentation results. In future work, it is of interest to replace these by neural networks to achieve a faster and end-to-end trainable pipeline with multiple tasks combined together.

A third cohort of dataset (acquired from Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds) was expected to be included in our study. Due to COVID-19 related restrictions, experiments had to be stopped and the third cohort could not be acquired and analysed before the submission of the thesis.

Chapter VII. Conclusions and Future Work

In this chapter, I summarise the thesis and discuss potential future work.

7.1 Summary

Cardiac magnetic resonance (CMR) has unique advantages for anatomical and functional analysis of the heart. Accurate manual segmentation of two-dimensional (2D) short-axis (SAX) slices is critical when applying this imaging technique to the quantification of cardiac anatomy and function, but it suffers from limitations such as long processing time and intra- and inter-observer variability. This is the motivation for developing automated segmentation methods, which has remained as an open question for decades. Deep neural networks exploit the advantages of data expansion and computational hardware development and are showing excellent results in image analysis applications including segmentation of CMR images. The outstanding performance of the deep learning algorithms also opens new possibilities for analysing CMR images. In this thesis, we explored the applications of deep neural networks for cardiac function assessment of human and mouse hearts using CMR images. Key contributions included: investigating the importance of segmentation in deep neural networks for evaluating clinically important metrics; developing segmentation neural networks for 2D CMR images, especially for long-axis (LAX) slices; developing fully three-dimensional (3D) neural networks for reconstructing ventricular shapes from sparse contours; and integrating and validating our pipeline using mouse CMR images.

In chapter 2, we introduced the background information related to cardiac physiology, cardiac imaging and functional analysis, and deep learning algorithms implemented in the thesis. We first explained the structure and function of the heart and compared the human heart and the mouse heart in details. We then compared CMR

imaging with other techniques and discussed its advantages and limitations. Finally, for deep learning algorithms, we introduced the multilayer perceptron (MLP), autoencoders (AE), convolutional neural networks (CNN), generative adversarial networks (GAN) and variational autoencoders (VAE), which are related to our work either conceptually or in practice. We also introduced the guide-point modelling (GPM) method, which we have used as the key algorithm for generating left ventricle (LV) shapes that we used to develop our 3D networks.

In chapter 3, we developed a single deep learning framework to investigate the importance of segmentation for evaluating LV function metrics, including volumes, dimensions and myocardial wall thickness. Our segmentation-based method showed better performance compared to our direct regression method and our mixture of the two methods, and the results were comparable to the stare-of-the-art methods on the same benchmark dataset. Another key advantage of the segmentation-based method was that the segmentation results could be used to analyse the source of errors, which the direct regression method could not offer. We developed our framework based on the U-Net [102], which is a thoroughly validated segmentation network but might not be optimal for a direct regression approach. However, by considering the advantages and drawbacks of both types of methods, we decided to use the segmentation-based method in subsequent work in the thesis for better accuracy and lower risks.

In chapter 4, we investigated the impact of introducing temporal dynamic modelling (TDM) using an LSTM layer [51]. The improvement after implementing TDM in our segmentation network was significant but small. Another limitation of

using manually traced contours as reference for training neural networks was explored, in that as each frame was usually segmented independently, no temporal smoothness was guaranteed within the reference segmentation. Therefore, the benefit of implementing TDM was underestimated in our quantitative validation. We also compared popular segmentation networks and decided to use U-Net given its better performance. We then developed a SAX-dependent approach for segmenting LAX mouse CMR images. This improved accuracy compared to our baseline U-Net. The motivation of introducing SAX information was to simplify the task of segmenting LAX slices, as the number of SAX slices is normally ten times larger than the number of LAX slices and the segmentation performance of SAX slices was therefore better than the LAX slices. The improved segmentation accuracy also suggested that for mouse CMR images, SAX and LAX slices were well aligned in space.

In chapter 5, we developed our deep neural network for the reconstruction of 3D ventricular shapes. Our method takes sparse contours from 2D slices and combines them into a sparse volumetric representation, which is then used as the input for our reconstruction network. The network is a variation of the 3D U-Net [28], which was also designed for segmentation but generally used for volumetric mapping, and it predicts a corresponding dense output volume from which the ventricular shapes are extracted. Apart from the reconstruction accuracy, another motivation for us to use neural networks for ventricular shape reconstruction is that the output shapes are more variable compared to mesh fitting approaches, in which a mean shape or an atlas is normally used to initiate the fitting. Based on our 3D U-Net, we investigated the impact

of introducing LAX slices and an adversarial loss for reconstructing from both spatially consistent and misaligned 2D contours, and we showed that the best performance was achieved by integrating SAX and LAX slices and including the adversarial loss for additional regularisation for both cases.

In chapter 6, we integrate our segmentation method and reconstruction method into a fully automated pipeline for assessment of LV function of mouse heart. The scale of pre-clinical studies using mouse CMR images is normally in a range of 10 to 20 subjects. We implemented our neural network-based method on 200 CMR images, collected from two imaging centres, and the performance of the networks was similar to previous chapters. For reconstructing LV shapes, we decided to use our U-Net model for balancing the performance and the computational cost, as the mouse CMR images were assumed to be well aligned in space. We finally evaluated volume and mass metrics and compared our results to intra- and inter-observer variability of standard manual analysis and the difference between CIM analysis and standard manual analysis. Our pipeline achieved good accuracy, and a better performance was expected after introducing the adversarial loss in chapter 5. We noted the segmentation error weakened the assumption of spatial consistency among 2D segmentation results, and thus the impact of the adversarial loss would be more obvious than expected. We also noted that the size variation was related to the accuracy and investigating normalisation or augmentation approaches could potentially further improve the performance of our method.

7.2 Future Work

In this thesis, we have developed a fully automated pipeline for anatomical and functional analysis of the LV. We investigated the importance of segmentation, the implementation of TDM, the use of LAX slices, and most importantly the development of 3D neural network for ventricular shape reconstruction. We finally validated our method using mouse CMR images and achieved reasonable accuracy. However, there are still investigations required to fully understand the performance of the method to further improve it without fundamental changes of our pipeline. The networks within our pipeline are trained independently; however, the lower stream networks are dependent on the upper stream networks and operations. Therefore, it is also worth investigating fundamental changes of our method, including fusing all networks into one end-to-end trainable network. Similar to the development of our method, future applications of our method could also be classified into short-term and long-term work. We summarise the directions of our future work as short-term or long-term for both algorithm development and application of our method.

7.2.1 Short-term Future Work

Segmentation of 2D CMR images is the fundamental stage of our method and improving segmentation accuracy remains critical for future work. From the overall performance of our pipeline, we saw worse performance for hearts with larger sizes, which were outnumbered by smaller hearts. We kept the variation of the size to preserve the generality of our networks and implemented an augmentation strategy with zooming

to prevent overfitting to the most common sizes. However, the generality of the network had a negative impact on the accuracy of the network, which was more important for our method. Therefore, to improve the segmentation accuracy without modifying the network, the task of the network could be simplified by normalising the size of the heart. Similar to our segmentation network, the reconstruction network was also designed to allow variation of sizes and therefore a potential source of accuracy reduction. Normalising the size of the heart could be worth investigating for better overall performance. If the computational cost limitation is relaxed, implementing the adversarial loss could also improve the reconstruction accuracy as we discussed in chapter 5.

Short-term applications include implementing our pipeline for analysing human CMR images. Our reconstruction network was developed and validated using sparse volumetric inputs generated from the reference shapes, and therefore the impact of manual segmentation errors was not investigated. In order to verify our pipeline for clinical datasets, additional experiments with the DETERMINE dataset could be performed to investigate the performance of the reconstruction network suffering from both segmentation and misalignment errors. The image quality would be comparatively better; however, the misalignment between slices would be combined to the segmentation errors, making the overall task as challenging as analysing mouse CMR images. We also plan to integrate our pipeline for analysing mouse CMR images into the workflow of Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds. We applied and validated our method using the Oxford and Tubingen cohorts,

and directly applying the trained networks to the upcoming Leeds cohort will not provide the optimal results, as the imaging protocols are different. Including all three cohorts in the training, validating, and testing datasets to retrain the networks is the most straightforward short-term plan.

7.2.2 Long-term Future Work

Our current pipeline has three networks linked with intermediate operations. These operations are not compatible with neural networks, and therefore each network must be trained independently, losing the advantage of backpropagating errors from later stages to parameters within previous stages. Replacing these intermediate stages with neural networks would allow us to build an end-to-end trainable network. However, this involves spatial transformation between 2D slices or from 2D slices to 3D, and the number of slices also varies between CMR images making the modification more challenging. Spatial transformation within 2D images has been shown in [118] using a spatial transformer network [59], and to achieve spatial transformation in 3D it is necessary to utilise metadata as the conditions of the network. To cope with variation of slices, multiple networks could be trained with a fixed slice arrangement for each of them, similar to our approach to deal with LAX slices in chapter 6.

For preclinical studies, new imaging protocols will be more frequently applied compared to clinical studies, and the image qualities are expected to be different from our current database. Accumulating a reasonable size of training data is the most straightforward approach; however, pre-clinical studies often have smaller dataset

comparing to clinical databases. Therefore, investigation of transferring our existing networks to new datasets, such as fine-tuning or more advanced domain adaptation techniques, is important in our long-term future work.

References

[1] Aboutanos, G.B., Nikanne, J., Watkins, N. and Dawan, B.M., 1999. Model creation and deformation for the automatic segmentation of the brain in MR images. *IEEE Transactions on Biomedical Engineering*, 46(11), pp.1346-1356.

[2] Acero, J.C., Xu, H., Zacur, E., Schneider, J.E., Lamata, P., Bueno-Orovio, A. and Grau, V., 2019, October. Left Ventricle Quantification with Cardiac MRI: Deep Learning Meets Statistical Models of Deformation. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 384-394). Springer, Cham.

[3] Acero, J.C., Zacur, E., Xu, H., Ariga, R., Bueno-Orovio, A., Lamata, P. and Grau, V., 2019, June. 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, Cham.

[4] Adams, R. and Bischof, L., 1994. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), pp.641-647.

[5] Afshin, M., Ayed, I.B., Islam, A., Goela, A., Peters, T.M. and Li, S., 2012, October. Global assessment of cardiac function using image statistics in MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 535-543).

Springer, Berlin, Heidelberg.

[6] Alba, X., Pereanez, M., Hoogendoorn, C., Swift, A.J., Wild, J.M., Frangi, A.F. and Lekadir, K., 2015. An algorithm for the segmentation of highly abnormal hearts using a generic statistical shape model. *IEEE Transactions on Medical Imaging,* 35(3), pp.845-859.

[7] Allison, J.D., Flickinger, F.W., Wright, J.C., Falls III, D.G., Prisant, L.M., VonDohlen, T.W. and Frank, M.J., 1993. Measurement of left ventricular mass in hypertrophic cardiomyopathy using MRI: comparison with echocardiography. *Magnetic Resonance Imaging*, 11(3), pp.329-334.

[8] American Heart Association Writing Group on Myocardial Segmentation and Registration for Cardiac Imaging:, Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T. and Verani, M.S., 2002. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation*, 105(4), pp.539-542.

[9] Arevalo, H.J., Vadakkumpadan, F., Guallar, E., Jebb, A., Malamas, P., Wu, K.C. and Trayanova, N.A., 2016. Arrhythmia risk stratification of patients after myocardial infarction using personalized heart models. *Nature Communications*, 7(1), pp.1-8.

[10] Arjovsky, M., Chintala, S. and Bottou, L., 2017. Wasserstein GAN. *arXiv:*1701.07875.

[11] Hedén, B., Edenbrandt, L., Haisty Jr, W.K. and Pahlm, O., 1994. Artificial neural networks for the electrocardiographic diagnosis of healed myocardial infarction. *The American Journal of Cardiology*, 74(1), pp.5-8.

[12] Badano, L.P., Boccalini, F., Muraru, D., Dal Bianco, L., Peluso, D., Bellu, R., Zoppellaro, G. and Iliceto, S., 2012. Current clinical applications of transthoracic three-dimensional echocardiography. *Journal of Cardiovascular Ultrasound*, 20(1), p.1.

[13] Badano, L.P., Ginghina, C., Easaw, J., Muraru, D., Grillo, M.T., Lancellotti, P., Pinamonti, B., Coghlan, G., Marra, M.P., Popescu, B.A. and De Vita, S., 2010. Right ventricle in pulmonary arterial hypertension: haemodynamics, structural changes, imaging, and proposal of a study protocol aimed to assess remodelling and treatment effects. *European Journal of Echocardiography*, 11(1), pp.27-37.

[14] Bai, W., Shi, W., de Marvao, A., Dawes, T.J., O'Regan, D.P., Cook, S.A. and Rueckert, D., 2015. A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of

healthy subjects and an analysis of shape and motion. *Medical Image Analysis*, 26(1), pp.133-145.

[15] Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A.M., Aung, N., Lukaschuk, E., Sanghvi, M.M. and Zemrak, F., 2018. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1), p.65.

[16] Bellenger, N.G., Francis, J.M., Davies, C.L., Coats, A.J. and Pennell, D.J., 2000. Establishment and performance of a magnetic resonance cardiac function clinic. *Journal of Cardiovascular Magnetic Resonance*, 2(1), pp.15-22.

[17] Bellman, R., 1966. Dynamic programming. *Science*, 153(3731), pp.34-37.

[18] Bertalmio, M., Sapiro, G., Caselles, V. and Ballester, C., 2000, July. Image inpainting. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 417-424).

[19] Beutel, J., Kundel, H.L. and Van Metter, R.L., 2000. *Handbook of Medical Imaging* (Vol. 1). Spie Press.

[20] Blank, S., Chen, V., Hamilton, N., Salerno, T.A. and Ianuzzo, C.D., 1989. Biochemical

characteristics of mammalian myocardia. *Journal of Molecular and Cellular Cardiology*, 21(4), pp.367-373.

[21] Bogaert, J., Bosmans, H., Maes, A., Suetens, P., Marchal, G. and Rademakers, F.E., 2000. Remote myocardial dysfunction after acute anterior myocardial infarction: impact of left ventricular shape on regional function: a magnetic resonance myocardial tagging study. *Journal of the American College of Cardiology*, 35(6), pp.1525-1534.

[22] Bogaert, J., Dymarkowski, S., Taylor, A.M. and Muthurangu, V. eds., 2012. *Clinical Cardiac MRI*. Springer Science & Business Media.

[23] Bogaert, J., Maes, A., Van de Werf, F., Bosmans, H., Herregods, M.C., Nuyts, J., Desmet, W., Mortelmans, L., Marchal, G. and Rademakers, F.E., 1999. Functional recovery of subepicardial myocardial tissue in transmural myocardial infarction after successful reperfusion: an important contribution to the improvement of regional and global left ventricular function. *Circulation*, 99(1), pp.36-43.

[24] Caselles, V., Catté, F., Coll, T. and Dibos, F., 1993. A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1), pp.1-31.

[25] Cathers, I., 1995. Neural network assisted cardiac auscultation. *Artificial Intelligence in Medicine*, 7(1), pp.53-66.

[26] Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W. and Rueckert, D., 2020. Deep Learning for Cardiac Image Segmentation: A Review. *Frontiers in Cardiovascular Medicine*, 7, p.25.

[27] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv*:1406.1078.

[28] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O., 2016, October. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 424-432). Springer, Cham.

[29] Cohen, L.D., 1991. On active contour models and balloons. *CVGIP: Image Understanding*, 53(2), pp.211-218.

[30] de Marvao, A., Dawes, T.J., Shi, W., Minas, C., Keenan, N.G., Diamond, T., Durighel, G., Montana, G., Rueckert, D., Cook, S.A. and O'Regan, D.P., 2014. Population-based studies of myocardial hypertrophy: high resolution cardiovascular magnetic resonance atlases improve statistical power. *Journal of Cardiovascular Magnetic Resonance*, 16(1), pp.1-10.

[31] Deng, D., Zhang, J. and Xia, L., 2010. Three-dimensional mesh generation for human heart model. In *Life System Modeling and Intelligent Computing* (pp. 157-162). Springer, Berlin, Heidelberg.

[32] Desai, K.H., Sato, R., Schauble, E., Barsh, G.S., Kobilka, B.K. and Bernstein, D., 1997. Cardiovascular indexes in the mouse at rest and with exercise: new tools to study models of cardiac disease. *American Journal of Physiology-Heart and Circulatory Physiology*, 272(2), pp.H1053-H1061.

[33] Desjardins, B. and Kazerooni, E.A., 2004. ECG-gated cardiac CT. *American Journal of Roentgenology*, 182(4), pp.993-1010.

[34] Doan, N.T., de Xivry, J.O. and Macq, B., 2010, March. Effect of inter-subject variation on the accuracy of atlas-based segmentation applied to human brain structures. In *Medical Imaging 2010: Image Processing* (Vol. 7623, p. 76231S). International Society for Optics and Photonics.

[35] Doersch, C., 2016. Tutorial on variational autoencoders. *arXiv:*1606.05908.

[36] Doevendans, P.A., J. Daemen, M., de Muinck, E.D. and Smits, J.F., 1998. Cardiovascular phenotyping in mice. *Cardiovascular Research*, 39(1), pp.34-49.

[37] Duan, J., Bello, G., Schlemper, J., Bai, W., Dawes, T.J., Biffi, C., de Marvao, A., Doumoud, G., O'Regan, D.P. and Rueckert, D., 2019. Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. *IEEE Transactions on Medical Imaging*, 38(9), pp.2151-2164.

[38] Ejection Fraction Heart Failure Measurement, *American Heart Association*. https://www.heart.org/en/health-topics/heart-failure/diagnosing-heart-failure/ejection-fraction-heart-failure-measurement. (accessed 8 October 2020).

[39] El Berbari, R., Bloch, I., Redheuil, A., Angelini, E., Mousseaux, E., Frouin, F. and Herment, A., 2007, August. An automated myocardial segmentation in cardiac MRI. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4508-4511). IEEE.

[40] Fischer, S.E., McKinnon, G.C., Maier, S.E. and Boesiger, P., 1993. Improved myocardial tagging contrast. *Magnetic Resonance in Medicine*, 30(2), pp.191-200.

[41] Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S. and Montillo, A., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), pp.341-355.

[42] Fonseca, C.G., Backhaus, M., Bluemke, D.A., Britten, R.D., Chung, J.D., Cowan, B.R., Dinov, I.D., Finn, J.P., Hunter, P.J., Kadish, A.H. and Lee, D.C., 2011. The Cardiac Atlas Project—an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*, 27(16), pp.2288-2295.

[43] Gatys, L.A., Ecker, A.S. and Bethge, M., 2015. A neural algorithm of artistic style. *arXiv:*1508.06576.

[44] Girdhar, R., Fouhey, D.F., Rodriguez, M. and Gupta, A., 2016, October. Learning a predictable and generative vector representation for objects. In *European Conference on Computer Vision* (pp. 484-499). Springer, Cham.

[45] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems* (pp. 2672-2680).

[46] Grothues, F., Smith, G.C., Moon, J.C., Bellenger, N.G., Collins, P., Klein, H.U. and Pennell, D.J., 2002. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. *The American Journal of Cardiology*, 90(1), pp.29-34.

[47] Grupp, I.L., Subramaniam, A.R.U.N., Hewett, T.E., Robbins, J.E.F.F.R.E.Y. and Grupp, G.U.N.T.E.R., 1993. Comparison of normal, hypodynamic, and hyperdynamic mouse hearts using isolated work-performing heart preparations. *American Journal of Physiology-Heart and Circulatory Physiology*, 265(4), pp.H1401-H1410.

[48] Guo, F., Ng, M. and Wright, G., 2018, September. Cardiac MRI left ventricle segmentation and quantification: a framework combining U-Net and continuous max-flow. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 450-458). Springer, Cham.

[49] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778).

[50] Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D. and Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1), pp.115-126.

[51] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9(8), pp.1735-1780.

[52] Horowitz, S.L. and Pavlidis, T., 1976. Picture segmentation by a tree traversal

algorithm. *Journal of the ACM (JACM)*, 23(2), pp.368-388.

[53] Hummel, R.A. and Zucker, S.W., 1983. On the foundations of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (3), pp.267-287.

[54] Iaizzo, P.A. ed., 2009. *Handbook of Cardiac Anatomy, Physiology, and Devices*. Springer Science & Business Media.

[55] Iglesias, J.E. and Sabuncu, M.R., 2015. Multi-atlas segmentation of biomedical images: a survey. *Medical Image Analysis*, 24(1), pp.205-219.

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

[57] Ishida, N., Sakuma, H., Motoyasu, M., Okinaka, T., Isaka, N., Nakano, T. and Takeda, K., 2003. Noninfarcted myocardium: correlation between dynamic first-pass contrast-enhanced myocardial MR imaging and quantitative coronary angiography. *Radiology*, 229(1), pp.209-216.

[58] Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition (pp. 1125-1134).

[59] Jaderberg, M., Simonyan, K. and Zisserman, A., 2015. Spatial transformer networks. In *Advances in Neural Information Processing Systems* (pp. 2017-2025).

[60] Kadish, A.H., Bello, D., Finn, J.P., Bonow, R.O., Schaechter, A., Subacius, H., Albert, C., Daubert, J.P., Fonseca, C.G. and Goldberger, J.J., 2009. Rationale and design for the defibrillators to reduce risk by magnetic resonance imaging evaluation (DETERMINE) trial. *Journal of Cardiovascular Electrophysiology*, 20(9), pp.982-987.

[61] Karamitsos, T.D., Francis, J.M., Myerson, S., Selvanayagam, J.B. and Neubauer, S., 2009. The role of cardiovascular magnetic resonance imaging in heart failure. *Journal of the American College of Cardiology*, 54(15), pp.1407-1424.

[62] Karamitsos, T.D., Hudsmith, L.E., Selvanayagam, J.B., Neubauer, S. and Francis, J.M., 2007. Operator induced variability in left ventricular measurements with cardiovascular magnetic resonance is improved after training. *Journal of Cardiovascular Magnetic Resonance*, 9(5), pp.777-783.

[63] Kass, M., Witkin, A. and Terzopoulos, D., 1988. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), pp.321-331.

[64] Kaur, A., Raj, A., Jayanthi, N. and Indu, S., 2020, July. Inpainting of Irregular Holes in a Manuscript using UNet and Partial Convolution. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 778-784). IEEE.

[65] Kerfoot, E., Clough, J., Oksuz, I., Lee, J., King, A.P. and Schnabel, J.A., 2018, September. Left-ventricle quantification using residual U-Net. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 371-380). Springer, Cham.

[66] Klein, A., Mensh, B., Ghosh, S., Tourville, J. and Hirsch, J., 2005. Mindboggle: automated brain labeling with multiple atlases. *BMC Medical Imaging*, 5(1), pp.1-14.

[67] Kramer, K., van Acker, S.A., Voss, H.P., Grimbergen, J.A., van der Vijgh, W.J. and Bast, A., 1993. Use of telemetry to record electrocardiogram and heart rate in freely moving mice. *Journal of Pharmacological and Toxicological Methods*, 30(4), pp.209-215.

[68] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).

[69] Lamata, P., Niederer, S., Nordsletten, D., Barber, D.C., Roy, I., Hose, D.R. and Smith, N., 2011. An accurate, fast and robust method to generate patient-specific cubic Hermite meshes. *Medical Image Analysis*, 15(6), pp.801-813.

[70] Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T. and Lancellotti, P., 2015. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *European Heart Journal-Cardiovascular Imaging*, 16(3), pp.233-271.

[71] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, 521(7553), pp.436-444.

[72] Lee, H., Battle, A., Raina, R. and Ng, A.Y., 2007. Efficient sparse coding algorithms. In *Advances in Neural Information Processing Systems* (pp. 801-808).

[73] Lehmann, H., Kneser, R., Neizel, M., Peters, J., Ecabert, O., Kühl, H., Kelm, M. and Weese, J., 2009, June. Integrating viability information into a cardiac model for interventional guidance. In *International Conference on Functional Imaging and Modeling of the Heart* (pp. 312-320). Springer, Berlin, Heidelberg.

[74] Levick, J.R., 2013. *An Introduction to Cardiovascular Physiology*. Butterworth-Heinemann.

[75] Li, J. and Hu, Z., 2018, September. Left ventricle full quantification using deep layer aggregation based multitask relationship learning. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 381-388). Springer, Cham.

[76] Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A. and Catanzaro, B., 2018. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 85-100).

[77] Liu, J., Li, X., Ren, H. and Li, Q., 2018, September. Multi-estimator full left ventricle quantification through ensemble learning. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 459-465). Springer, Cham.

[78] Liu, L., Ma, J., Wang, J. and Xiao, J., 2018, September. Automated full quantification of left ventricle with deep neural networks. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 412-420). Springer, Cham.

[79] Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431-3440).

[80] Lorensen, W.E. and Cline, H.E., 1987. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4), pp.163-169.

[81] Lygate, C.A., Fischer, A., Sebag-Montefiore, L., Wallis, J., ten Hove, M. and Neubauer, S., 2007. The creatine kinase energy transport system in the failing mouse heart. *Journal of Molecular and Cellular Cardiology*, 42(6), pp.1129-1136.

[82] Lynch, M., Ghita, O. and Whelan, P.F., 2008. Segmentation of the left ventricle of the heart in 3-D+ t MRI data using an optimized nonrigid temporal model. *IEEE Transactions on Medical Imaging*, 27(2), pp.195-203.

[83] Mahrholdt, H., Wagner, A., Holly, T.A., Elliott, M.D., Bonow, R.O., Kim, R.J. and Judd, R.M., 2002. Reproducibility of chronic infarct size measurement by contrast-enhanced magnetic resonance imaging. *Circulation*, 106(18), pp.2322-2327.

[84] Malladi, R., Sethian, J.A. and Vemuri, B.C., 1995. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(2), pp.158-175.

[85] McLeish, K., Hill, D.L., Atkinson, D., Blackall, J.M. and Razavi, R., 2002. A study of the motion and deformation of the heart due to respiration. *IEEE Transactions on Medical Imaging*, 21(9), pp.1142-1150.

[86] Medrano-Gracia, P., Cowan, B.R., Bluemke, D.A., Finn, J.P., Lima, J.A., Suinesiaputra, A. and Young, A.A., 2013, June. Large scale left ventricular shape atlas using automated model fitting to contours. In *International Conference on Functional Imaging and Modeling of the Heart* (pp. 433-441). Springer, Berlin, Heidelberg.

[87] Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565-571). IEEE.

[88] Mor-Avi, V., Sugeng, L., Weinert, L., MacEneaney, P., Caiani, E.G., Koch, R., Salgo, I.S. and Lang, R.M., 2004. Fast measurement of left ventricular mass with real-time three-dimensional echocardiography: comparison with magnetic resonance imaging. *Circulation*, 110(13), pp.1814-1818.

[89] Ng, A., 2011. Sparse autoencoder. *CS294A Lecture Notes*, 72(2011), pp.1-19.

[90] Nikitin, N.P., Constantin, C., Loh, P.H., Ghosh, J., Lukaschuk, E.I., Bennett, A., Hurren, S., Alamgir, F., Clark, A.L. and Cleland, J.G., 2006. New generation 3-dimensional echocardiography for left ventricular volumetric and functional measurements: comparison with cardiac magnetic resonance. *European Journal of Echocardiography*, 7(5), pp.365-372.

[91] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., De Marvao, A., Dawes, T., O'Regan, D.P. and Kainz, B., 2017. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2), pp.384-395.

[92] Olshausen, B.A. and Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), pp.607-609.

[93] Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp.62-66.

[94] Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J. and Rueckert, D., 2019, October. Data efficient unsupervised domain adaptation for cross-modality image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 669-677). Springer, Cham.

[95] Park, H., Bland, P.H. and Meyer, C.R., 2003. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Transactions on Medical Imaging*, 22(4), pp.483-492.

[96] Park, H.S., Chung, Y.E., Lee, S.M., Kim, H.P. and Seo, J.K., 2017. Sinogram-consistency learning in CT for metal artifact reduction. *arXiv:*1708.00607, 2(3).

[97] Pascanu, R., Mikolov, T. and Bengio, Y., 2013, February. On the difficulty of training recurrent neural networks. In *International Conference on Machine Learning* (pp. 1310-1318).

[98] Petitjean, C. and Dacher, J.N., 2011. A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis*, 15(2), pp.169-184.

[99] Pham, D.L., Xu, C. and Prince, J.L., 2000. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1), pp.315-337.

[100] Poudel, R.P., Lamata, P. and Montana, G., 2016. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In *Reconstruction, Segmentation, and Analysis of Medical Images* (pp. 83-94). Springer, Cham.

[101] Rockman, H.A., Ross, R.S., Harris, A.N., Knowlton, K.U., Steinhelper, M.E., Field, L.J., Ross, J. and Chien, K.R., 1991. Segregation of atrial-specific and inducible expression of an atrial natriuretic factor transgene in an in vivo murine model of cardiac hypertrophy. In *Proceedings of the National Academy of Sciences*, 88(18), pp.8277-8281.

[102] Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234-241). Springer, Cham.

[103] Rosen, S.E., Bauch, L., Fisher, E.A., Vorchheimer, D., Budd, J.H. and Goldman, M.E., 1993. Current status of echocardiographic automated border detection. *Cardio*, 12, pp.44-54.

[104] Rosenfeld, A., Hummel, R.A. and Zucker, S.W., 1976. Scene labeling by relaxation operations. *IEEE Transactions on Systems, Man, and Cybernetics*, (6), pp.420-433.

[105] Ruff, J., Wiesmann, F., Hiller, K.H., Voll, S., von Kienlin, M., Bauer, W.R., Rommel, E., Neubauer, S. and Haase, A., 1998. Magnetic resonance microimaging for noninvasive quantification of myocardial function and mass in the mouse. *Magnetic Resonance in Medicine*, 40(1), pp.43-48.

[106] Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1985. *Learning Internal Representations by Error Propagation* (No. ICS-8506). California Univ San Diego La Jolla Inst for Cognitive Science.

[107] Schneider, J.E., Cassidy, P.J., Lygate, C., Tyler, D.J., Wiesmann, F., Grieve, S.M., Hulbert, K., Clarke, K. and Neubauer, S., 2003. Fast, high-resolution in vivo cine magnetic resonance imaging in normal and failing mouse hearts on a vertical 11.7 T system. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 18(6), pp.691-701.

[108] Schneider, J.E., Wiesmann, F., Lygate, C.A. and Neubauer, S., 2006. How to perform an accurate assessment of cardiac function in mice using high-resolution magnetic resonance imaging. *Journal of Cardiovascular Magnetic Resonance*, 8(5), pp.693-701.

[109] Shah, B.N., 2013. Echocardiography in the era of multimodality cardiovascular imaging. *BioMed Research International, 2013*.

[110] Shechter, G., Ozturk, C., Resar, J.R. and McVeigh, E.R., 2004. Respiratory motion of the heart from free breathing coronary angiograms. *IEEE Transactions on Medical Imaging*, 23(8), pp.1046-1056.

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

[112] Sonka, M., Winniford, M.D. and Collins, S.M., 1995. Robust simultaneous detection of coronary borders in complex images. *IEEE Transactions on Medical Imaging*, 14(1), pp.151-161.

[113] Suinesiaputra, A., Ablin, P., Alba, X., Alessandrini, M., Allen, J., Bai, W., Cimen, S., Claes, P., Cowan, B.R., D'hooge, J. and Duchateau, N., 2017. Statistical shape modeling of the left ventricle: myocardial infarct classification challenge. *IEEE Journal of Biomedical and Health Informatics*, 22(2), pp.503-515.

[114] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9).

[115] Terzopoulos, D., 1986. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4), pp.413-424.

[116] Tilborghs, S., Dresselears, T., Claus, P., Bogaert, J. and Maes, F., 2019, October. 3D Left Ventricular Segmentation from 2D Cardiac MR Images Using Spatial Context. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 90-99). Springer, Cham.

[117] Tran, P.V., 2016. A fully convolutional neural network for cardiac segmentation in short-axis MRI. *arXiv:*1604.00494.

[118] Vigneault, D.M., Xie, W., Ho, C.Y., Bluemke, D.A. and Noble, J.A., 2018. Ω-net (omega-net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Medical Image Analysis*, 48, pp.95-106.

[119] Villard, B., Grau, V. and Zacur, E., 2018. Surface mesh reconstruction from cardiac MRI contours. *Journal of Imaging*, 4(1), p.16.

[120] Villard, B., Zacur, E., Dall'Armellina, E. and Grau, V., 2016, October. Correction of slice misalignment in multi-breath-hold cardiac MRI scans. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 30-38). Springer, Cham.

[121] Vukicevic, M., Mosadegh, B., Min, J.K. and Little, S.H., 2017. Cardiac 3D printing and its future directions. *JACC: Cardiovascular Imaging*, 10(2), pp.171-184.

[122] Waldman, L.K., Fung, Y.C. and Covell, J.W., 1985. Transmural myocardial deformation in the canine left ventricle. Normal in vivo three-dimensional finite strains. *Circulation Research*, 57(1), pp.152-163.

[123] Waldman, L.K., Nosan, D., Villarreal, F. and Covell, J.W., 1988. Relation between transmural deformation and local myofiber direction in canine left ventricle. *Circulation Research*, 63(3), pp.550-562.

[124] Wang, Y. and Jia, Y., 2006, August. Segmentation of the left ventricle from cardiac MR images based on degenerated minimal surface diffusion and shape priors. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 4, pp. 671-674). IEEE.

[125] Webb, S., Brown, N.A. and Anderson, R.H., 1996. The structure of the mouse heart in late fetal stages. *Anatomy and Embryology*, 194(1), pp.37-47.

[126] Winchester, D.E., Wymer, D.C., Shifrin, R.Y., Kraft, S.M. and Hill, J.A., 2010, April. Responsible use of computed tomography in the evaluation of coronary artery disease and chest pain. In *Mayo Clinic Proceedings* (Vol. 85, No. 4, pp. 358-364). Elsevier.

[127] Xu, B., Wang, N., Chen, T. and Li, M., 2015. Empirical evaluation of rectified activations in convolutional network. *arXiv*:1505.00853.

[128] Xu, C. and Prince, J.L., 1998. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3), pp.359-369.

[129] Xu, H., Schneider, J.E. and Grau, V., 2018, September. Calculation of anatomical and functional metrics using deep learning in cardiac MRI: comparison between direct and segmentation-based estimation. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 402-411). Springer, Cham.

[130] Xu, H., Zacur, E., Schneider, J.E. and Grau, V., 2019, June. Ventricle surface reconstruction from Cardiac MR slices using deep learning. In *International Conference on Functional Imaging and Modeling of the Heart* (pp. 342-351). Springer, Cham.

[131] Xue, W., Brahm, G., Pandey, S., Leung, S. and Li, S., 2018. Full left ventricle quantification via deep multitask relationships learning. *Medical Image Analysis*, 43, pp.54-65.

[132] Xue, W., Islam, A., Bhaduri, M. and Li, S., 2017. Direct multitype cardiac indices estimation via joint representation and regression learning. *IEEE Transactions on Medical Imaging*, 36(10), pp.2057-2067.

[133] Xue, W., Lum, A., Mercado, A., Landis, M., Warrington, J. and Li, S., 2017, September. Full quantification of left ventricle via deep multitask learning network respecting intra- and inter-task relatedness. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 276-284). Springer, Cham.

[134] Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S. and Li, S., 2017, June. Direct estimation of regional wall thicknesses via residual recurrent neural network. In *International Conference on Information Processing in Medical Imaging* (pp. 505-516). Springer, Cham.

[135] Yang, D., Liu, B., Axel, L. and Metaxas, D., 2018, September. 3D LV Probabilistic Segmentation in Cardiac MRI Using Generative Adversarial Network. In *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 181-190). Springer, Cham.

[136] Yang, T.F., Devine, B. and Macfarlane, P.W., 1994. Use of artificial neural networks within deterministic logic for the computer ECG diagnosis of inferior myocardial infarction. *Journal of Electrocardiology*, 27, pp.188-193.

[137] Young, A.A., Barnes, H., Davison, D., Neubauer, S. and Schneider, J.E., 2009. Fast left ventricular mass and volume assessment in mice with three-dimensional guide-point modeling. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 30(3), pp.514-520.

[138] Young, A.A., Cowan, B.R., Thrupp, S.F., Hedley, W.J. and Dell'Italia, L.J., 2000. Left ventricular mass and volume: fast calculation with guide-point modeling on MR images. *Radiology*, 216(2), pp.597-602.

[139] Young, A.A., Cowan, B.R., Thrupp, S.F., Hedley, W.J. and Dell'Italia, L.J., 2000. Left ventricular mass and volume: fast calculation with guide-point modeling on MR images. *Radiology*, 216(2), pp.597-602.

[140] Young, A.A., Medway, D.J., Lygate, C.A., Neubauer, S. and Schneider, J.E., 2011. Accelerating global left-ventricular function assessment in mice using reduced slice acquisition and three-dimensional guide-point modelling. *Journal of Cardiovascular Magnetic Resonance*, 13(1), p.49.

[141] Zacur, E., Minchole, A., Villard, B., Carapella, V., Ariga, R., Rodriguez, B. and Grau, V., 2017. MRI-based heart and torso personalization for computer modeling and simulation of cardiac electrophysiology. In *Imaging for Patient-Customized Simulations and Systems for Point-of-Care Ultrasound* (pp. 61-70). Springer, Cham.

[142] Zeiler, M.D., 2012. Adadelta: an adaptive learning rate method. *arXiv*:1212.5701.

[143] Zerhouni, E.A., Parish, D.M., Rogers, W.J., Yang, A. and Shapiro, E.P., 1988. Human heart: tagging with MR imaging--a method for noninvasive assessment of myocardial motion. *Radiology*, 169(1), pp.59-63.

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

[145] Zhang, X., Cowan, B.R., Bluemke, D.A., Finn, J.P., Fonseca, C.G., Kadish, A.H., Lee, D.C., Lima, J.A., Suinesiaputra, A., Young, A.A. and Medrano-Gracia, P., 2014. Atlas-based quantification of cardiac remodeling due to myocardial infarction. *PloS One*, 9(10), p.e110243.

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

[147] Zhuang, X. and Shen, J., 2016. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Medical Image Analysis*, 31, pp.77-87.

[148] Zou, M., Holloway, M., Carr, N. and Ju, T., 2015. Topology-constrained surface reconstruction from cross-sections. *ACM Transactions on Graphics (TOG)*, 34(4), pp.1-10.

[149] Zuo, Z., Subgang, A., Abaei, A., Rottbauer, W., Stiller, D., Ma, G. and Rasche, V., 2017. Assessment of longitudinal reproducibility of mice LV function parameters at 11.7 T derived from self-gated CINE MRI. *BioMed Research International*, *2017*.