FULLY AUTOMATED SEGMENTATION OF THE LEFT VENTRICLE IN CINE CARDIAC MAGNETIC RESONANCE IMAGING

TAN LI KUO

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **TAN LI KUO**

Matric No: **KHA140033**

Name of Degree: **DOCTOR OF PHILOSOPHY**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

FULLY AUTOMATED SEGMENTATION OF THE LEFT VENTRICLE IN CINE CARDIAC MAGNETIC RESONANCE IMAGING

Field of Study: **BIOMEDICAL ENGINEERING**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;

(2) This Work is original;

(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;

(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

FULLY AUTOMATED SEGMENTATION OF THE LEFT VENTRICLE IN CINE CARDIAC MAGNETIC RESONANCE IMAGING

ABSTRACT

Cardiovascular diseases (CVD) are the primary cause of death globally, accounting for approximately 31% of all deaths worldwide. Cardiac magnetic resonance imaging (MRI) is the reference standard for the medical assessment of cardiac volumes and regional functions due to its accuracy and reproducibility. Most standard cardiac MRI protocols begin with assessing the left ventricle (LV) structure and functions due to the LVs' role in supplying most of the body with oxygenated blood. In standard clinical practice, quantification of LV function is performed via manual delineation of the LV myocardium within the MR images, for the end-diastole (ED) and end-systole (ES) cardiac phases. This enables the evaluation of standard diagnostic clinical measurements such as LV ED and ES blood volumes, ejection fraction, and LV mass. Despite delineating only two cardiac phases, such manual tracing can take up to 20 minutes by a radiologist. Full delineation across all cardiac phases would enable useful quantification of motion parameters to identify regional LV dysfunction. However, the excessive effort required for manual full delineation makes it impractical for clinical adoption. In this thesis, two fully automatic algorithms for cardiac MRI were presented: the first for localization of the LV blood pool – a sub-problem for enabling subsequent automatic segmentation; and the second for segmentation of the LV with full coverage from base to apex across all cardiac phases. The novel use of neural network regression for image segmentation was introduced, whereby multiple independent networks were designed and trained for the inference of LV landmarks, LV centrepoints, and myocardial contours, respectively. A large range of data sources was utilized for training and validation, including both in-house and publicly available databases, representing a heterogeneous mix of scanner types, imaging protocols, and parameters. Tested against the public 2011 Left Ventricle

Segmentation Challenge (LVSC) database, a final Jaccard index result of 0.77 ± 0.11 was obtained for segmentation accuracy. This represents the best published LVSC performance to date for a fully automated algorithm. Tested against the public 2016 Kaggle Second Annual Data Science Bowl challenge, a final result of +7.2 ± 13.0 mL and −19.8 ± 18.8 mL was obtained for clinical blood volume measurement accuracy in the ES and ED phases, respectively. This performance is comparable to published inter-reader variability values for multiple independent expert readers. The execution speed is approximately 12 s per case. In conclusion, two algorithms were developed and tested leading to fully automatic segmentation of LV in cardiac cine MRI. These were validated against a diverse set of publicly available and in-house cardiac cine MRI data. The strong performance overall is suggestive of practical clinical utility.

Keywords: cardiac MRI, LV localization, LV segmentation, automated segmentation, neural network regression

SEGMENTASI AUTOMATIK PADA VENTRICLE KIRI UNTUK PENGIMEJAN RESONANS MAGNETIK CINE JANTUNG

ABSTRAK

Penyakit kardiovaskular (CVD) adalah punca kematian utama di seluruh dunia, merangkumi kira-kira 31% daripada semua kematian di seluruh dunia. Pengimejan resonans magnetik (MRI) jantung adalah sistem rujukan klinikal untuk menilai isipadu dan fungsi jantung, disebabkan oleh ketepatan dan kebolehulangannya. Kebanyakan protokol MR jantung standard bermula dengan menilai struktur dan fungsi ventrikel kiri (LV) kerana peranan LV dalam membekalkan seluruh badan dengan darah oksigen. Dalam amalan klinikal yang standard, kuantifikasi fungsi LV dilakukan melalui lukisan manual otot LV dalam imej MR, untuk fasa jantung di akhir-diastol (ED) dan akhir-sistol (ES). Ini membolehkan penilaian ukuran klinikal diagnostik piawai seperti isi padu darah LV ED dan ES, fraksi ejeksi, dan jisim LV. Walaupun hanya dua fasa jantung dianalisakan, lukisan manual sedemikian boleh mengambil masa selama 20 minit oleh pakar radiologi. Kuntur otot LV pada semua fasa akan membolehkan kuantifikasi yang berguna dilakukan seperti pergerakan jantung untuk mengenal pasti fungsi tempatan yang luar biasa di LV. Walau bagaimanapun, usaha berlebihan yang diperlukan untuk melukis secara manual pada semua fasa menjadikannya tidak praktikal untuk diterima pakai di klinik. Dalam tesis ini, saya membentangkan dua algoritma automatik untuk cine MRI jantung: yang pertama untuk penyetempatan kawasan darah LV – suatu applikasi permulaan untuk membolehkan segmentasi automatik berikutnya; dan kedua untuk segmentasi otot LV dengan liputan penuh dari bawah ke atas, bagi semua fasa jantung. Saya memperkenalkan teknik novel bernama regresi rangkaian neural untuk segmen imej, di mana pelbagai rangkaian direkabentuk dan dilatih untuk menyimpulkan landasan, titik pusat, dan kontur LV. Saya menggunakan pelbagai sumber data, termasuk pangkalan data dalaman dan awam, untuk latihan dan pengesahan teknik ini. Sumber data ini mengandungi imej-imej yang diperolehi daripada pelbagai jenis pengimbas, protokol pengimejan, dan parameter. Apabila diuji pada pangkalan data awam, iaitu 2011 Left Ventricle Segmentation Challenge (LVSC), saya memperolehi keputusan akhir indeks Jaccard 0.77 ± 0.11 sebagai ketepatan segmentasi. Ini mewakili prestasi LVSC yang terbaik yang pernah diterbitkan untuk algoritma automatik sepenuhnya. Apabila diuji terhadap pangkalan data awam 2016 Kaggle Second Annual Data Science Bowl Challenge, saya memperolehi keputusan akhir ketepatan pengukuran isi padu darah LV klinikal sebanyak +7.2 ± 13.0 mL pada fasa ES dan -19.8 ± 18.8 mL pada fasa ED. Prestasi ini adalah setanding dengan nilai variabiliti antara pakar-pakar klinikal. Kelajuan pelaksanaan adalah lebih kurang 12 s bagi setiap kes. Kesimpulannya, saya memperkenal dan menguji dua algoritma yang membawa kepada segmentasi LV secara automatik untuk MRI jantung. Teknik-teknik ini telah disahkan dengan menggunakan pelbagai pangkalan data MRI jantung dalaman dan awam. Prestasi keseluruhannya yang tepat dan cepat mencadangkan utiliti klinikal secara praktikal.

Kata kunci: MRI jantung, penyetempatan LV, segmentasi LV, segmentasi automatik, regresi rangkaian neural

ACKNOWLEDGEMENTS

First and foremost, I thank God for all He has done, and for bestowing me the perseverance to complete this thesis.

I respectfully and sincerely express my gratitude to my main supervisors, Dr. Liew Yih Miin and Prof. Dr. Robert A McLaughlin for their supervision and guidance. This PhD would not have been possible without both of your support. I also thank my co-supervisor, Dr. Einly Lim for her backing and advice. I have also benefited from the collaboration of a fellow student, Mr. Yong Yan Ling, the members of the Asian Cardiac Engineering laboratory research team, as well as clinicians from the University of Malaya, Prof. Dr. Yang Faridah Abdul Aziz and Prof. Dr. Chee Kok Han. I thank them for all their help.

Last but not least, I wish to highlight my family, in particular my beloved wife, Dr. Lim Jen Nee Jones, my wonderful parents, Prof. Dr. Tan Chong Tin and Ms. Irene Yek Siew Hong, and my dear children, Ms. Tan Yue Xin and Mr. Tan Shang Jie. I thank them for keeping me in their prayers, supporting and sustaining me throughout. This thesis is personally dedicated to all of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ANN	Artificial neural network
CDF	Cumulative distribution function
CN	Convolutional layer
CNN	Convolutional neural network
CNR	Convolutional network regression
CPL	Centrepoint localization
CRPS	Continuous ranked probability score
CT	Computed tomography
CTR	Centrepoint
CVD	Cardiovascular disease
DI	Dice index
DICOM	Digital imaging and communications in medicine
ED	End-diastole
EDV	End-diastolic volume
EF	Ejection fraction
ELU	Exponential linear units
ES	End-systole
ESV	End-systolic volume
FC	Fully connected layer
FCON	Fully connected neural network
FOV	Field-of-view
FT-H1	First harmonic Fourier transform
IVOCT	Intravascular optical coherence tomography
IVUS	Intravascular ultrasound
GPU	Graphics processing unit
JI	Jaccard index
LA	Long axis

LA2C	Long axis two-chamber
LA4C	Long axis four-chamber
LM	Left ventricle landmarks
LV	Left ventricle
LVOT	Left ventricular outflow tract
LVSC	STACOM 2011 Left Ventricle Segmentation Challenge
MICCAI	Medical Image Computing and Computer Assisted Intervention
MB	Myocardial borders
MHD	Modified Hausdorff distance
MR	Magnetic resonance
MRI	Magnetic resonance imaging
NPV	Negative predictive values
OCT	Optical coherence tomography
PPP	Pre- and post-processing
PPV	Positive predictive values
ReLU	Rectified linear unit
ROI	Region of interest
RV	Right ventricle
SA	Short axis
SD	Standard deviation
SSFP	Steady-state free precession
STACOM	Statistical Atlases and Computational Modelling of the Heart
STAPLE	Simultaneous truth and performance level estimate

CHAPTER 1: GENERAL INTRODUCTION

1.1 Motivation

Cardiovascular diseases (CVDs) are the most common cause of death globally; it is estimated that 31% of all global deaths in 2012 were due to CVDs (Low, Lee, & Samy, 2014; Mendis, 2014). These include congenital heart disease, where birth defects affect the normal operation of the heart; coronary heart disease, where the blood supply to the heart muscle is occluded; and strokes, where the blood supply to the brain is occluded.

Cardiac magnetic resonance imaging (MRI) is currently considered the gold standard for the assessment of various aspects of CVDs (Abdul Aziz et al., 2013). Quantification of key parameters from cardiac MRI is now recommended as a standard diagnostic procedure by cardiovascular expert groups (Schulz-Menger et al., 2013). Of the standardized protocols for CVD imaging, 10 out of 13 require the quantification of LV function (Kramer, Barkhausen, Flamm, Kim, & Nagel, 2013). These quantifications typically require the delineation of LV myocardial borders, enabling clinically diagnostic measurements such as LV blood volume and cardiac ejection fraction. Clinically, most physicians restrict delineation to only the end-diastole (ED) and end-systole (ES) cardiac phases, which can require up to 20 minutes to delineate manually (Petitjean & Dacher, 2011). Complete delineation across the entire cardiac cycle would be desirable, but modern 20+ cardiac acquisition framerates make this far too tedious and time consuming to be performed manually.

Computer aided semi- and fully-automated techniques for the segmentation of LV myocardium are valuable, both for the reduction in human labour as well as the reduction of inter-observer variability. There have been numerous published approaches tackling this task, ranging from semi-automated single phase, single slice segmentation of the LV inner wall (endocardium), to fully-automated full cycle, base-to-apex segmentation of the full LV myocardium (Petitjean & Dacher, 2011; Tavakoli & Amini, 2013). Due to the far more challenging nature of a fully automated approach (i.e. no human interaction), there are correspondingly fewer published works addressing this task. At time of writing, the LV quantification tools used in this author's healthcare institution for routine clinical practice, still require significant manual input for their operation.

1.2 Project scope & objectives

The ultimate goal of this thesis is the development and implementation of a fully automated algorithm for the segmentation of LV myocardium in cardiac magnetic resonance cine images. For this thesis, computerised image segmentation techniques are broadly classify into two families: (1) *expert knowledge-based* techniques, which are defined here as techniques largely derived from human hand-crafted features or algorithms, and (2) *data-driven* techniques, which are defined as techniques based on elementary operations with minimal human enforced explicit constraints or assumptions, where the performance is almost completely driven by the provided training data.

Expert knowledge-based techniques was first utilized to develop a method for fully automatic localization of the LV blood pool, as presented in Chapter 3. LV localization is a sub-problem enabling subsequent automatic full myocardial segmentation, as indicated in numerous published LV segmentation approaches (Hu, Liu, Gao, & Huang, 2013; Nambakhsh et al., 2013; Yin Wu et al., 2015) which require the initial manual localization of the LV blood pool. The system was developed utilizing a chain of image processing functions, each targeting a logical subset of the problem employing techniques such as motion analysis and shape morphology. There are a limited number of published approaches (Jolly, 2008; X. Lin, Cowan, & Young, 2006) focused on LV localization. It was shown that this technique can provide more reliable performance in the presence of motion and scanning artifacts.

Next, data-driven techniques were utilized to develop a method for fully automatic segmentation of the LV myocardium from base-to-tip, across all cardiac phases, presented in Chapter 4 and Chapter 5. The specific technique used was convolutional neural network regression. This was approached in two stages, first by developing a semi-automated LV segmentation system targeting solely short axis (SA) images (Chapter 4). This required manual input to identify LV slice coverage. The algorithm was then extended to handle long axis (LA) images (Chapter 5), making the segmentation system fully automatic. Existing published approaches (Baumgartner, Koch, Pollefeys, & Konukoglu, 2018; Tran, 2016) for LV segmentation based on neural networks all utilize the specific method of *per-pixel classification*. A different approach was adopted, by parameterising the location of the LV as a distance from a central point using neural network regression. This provided an improvement in performance compared to existing methods.

In summary, the objectives of this project are two-fold:

(1) To develop and validate a fully automated algorithm for localization of the LV blood pool.

(2) To develop and validate a fully automated algorithm for segmentation of the LV myocardium from base to apex, for all cardiac phases.

1.3 Thesis organization

Chapter 2 provides background on human heart anatomy, as well as on standard clinical cardiac MR acquisition protocols and standard clinical LV quantification practices. Expert knowledge-based LV segmentation techniques are briefly reviewed and contrasted to LV segmentation techniques utilizing the neural network data-driven approach.

Chapter 3 (Tan, Liew, et al., 2018) presents an expert knowledge-based method for fully automatic localization of the LV blood pool, a sub-problem for enabling subsequent automatic segmentation. Chapter 4 (Tan, Liew, Lim, & McLaughlin, 2017) presents a neural network regression-based method for semi-automatic base-to-apex segmentation of the LV myocardium in SA images across all cardiac phases. Chapter 5 (Tan, McLaughlin, Lim, Abdul Aziz, & Liew, 2018) builds on the neural network regression technique, extending it to LA images and making the LV segmentation algorithm fully automatic. Chapter 6 (Yong, Tan, McLaughlin, Chee, & Liew, 2017) describes the extension of the neural network regression technique to a separate clinical task – automatic segmentation of vessel lumen wall in optical coherence tomography (OCT). This extension demonstrates the generalizability of the underlying technique.

Finally, Chapter 7 concludes this thesis, and provides some suggestions on future work.

1.4 Research contribution

The chapters in this thesis are primarily derived from four published full-length journal articles: three first-authored, and one co-authored. The specific contributions of this author to each journal article (and by extension, the respective thesis chapters), are stated below.

Chapter 3: **Tan, L. K.**, Liew, Y. M., Lim, E., Abdul Aziz, Y. F., Chee, K. H., & McLaughlin, R. A. (2018). Automatic localization of the left ventricular blood pool centroid in short axis cardiac cine MR images. *Medical & Biological Engineering & Computing*, (in press). (doi:10.1007/s11517-017-1750-7)

> TLK (this author) was the principal author of this article. TLK collected and prepared the imaging data, devised and implemented the image processing

code, conducted the experiments, led the data analysis, and led the writing of the manuscript, which was edited and reviewed by all co-authors.

Chapter 4: **Tan, L. K.**, Liew, Y. M., Lim, E., & McLaughlin, R. A. (2017). Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Medical Image Analysis*, 39, 78–86.

TLK was the principal author of this article. TLK collected and prepared the imaging data, devised and implemented the neural network architecture, conducted the experiments, led the data analysis, and led the writing of the manuscript, which was edited and reviewed by all co-authors.

Chapter 5: **Tan, L. K.**, McLaughlin, R. A., Lim, E., Abdul Aziz, Y. F., & Liew, Y. M. (2018). Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression. Journal of Magnetic Resonance Imaging, (in press). (doi:10.1002/jmri.25932)

TLK was the principal author of this article. TLK collected and prepared the imaging data, devised and implemented the neural network architecture, conducted the experiments, led the data analysis, and led the writing of the manuscript, which was edited and reviewed by all co-authors.

Chapter 6: Yong, Y. L., **Tan, L. K.**, McLaughlin, R. A., Chee, K. H., & Liew, Y. M. (2017). Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography. *Journal of Biomedical Optics*, 22(12), 126005.

YYL and LYM were the principal authors of this article. YYL and LYM collected and prepared the imaging data, led the data analysis, and led the writing of the manuscript, which was edited and reviewed by all co-authors. TLK (this author) devised and implemented the neural network architecture.

CHAPTER 2: LITERATURE REVIEW

2.1 Preface

This chapter provides background to the problem domain of left ventricular (LV) segmentation, and reviews the published literature on automated methods tackling this task. Section 2.2 provides a brief background on human heart anatomy; Section 2.3 introduces typical cardiac magnetic resonance (MR) acquisition protocols and clinical quantification practices; Section 2.4 provides a brief review of published LV automated segmentation approaches utilizing *expert knowledge* techniques; Finally, Section 2.5 introduces the specific data-driven technique of convolutional neural networks, also known as *deep learning*.

2.2 Human heart anatomy

The human heart is an approximately fist-sized muscular organ in the body located between the lungs, and is responsible for generating the primary pumping force for the circulation of blood throughout the body. There are two main loops of blood circulation in the body: the first, smaller loop (pulmonary circulation) sends deoxygenated blood to the lungs where it is oxygenated, and the second, larger loop (systemic circulation) sends the oxygenated blood throughout the rest of the body. The heart consists of four chambers, divided into left/right and upper/lower; the upper/lower chambers are known as the *atrium* and *ventricle*, respectively. In brief, the atrium chambers receive blood supply from the lungs (left atrium) or from the body (right atrium) through the veins, and the ventricles send blood supply out to the lungs (right ventricle) or to the body (left ventricle) through the arteries. The right atrium and ventricle serve the pulmonary circulation loop (from the rest of the body to the lungs), while the left atrium and ventricle serve the systemic circulation loop (from the lungs to the rest of the body) (Figure 2.1). In both cases, the pumping force is primarily generated by the *myocardium* (heart muscle) surrounding the ventricles (Walsh, Fang, Fuster, & O'Rourke, 2012a).

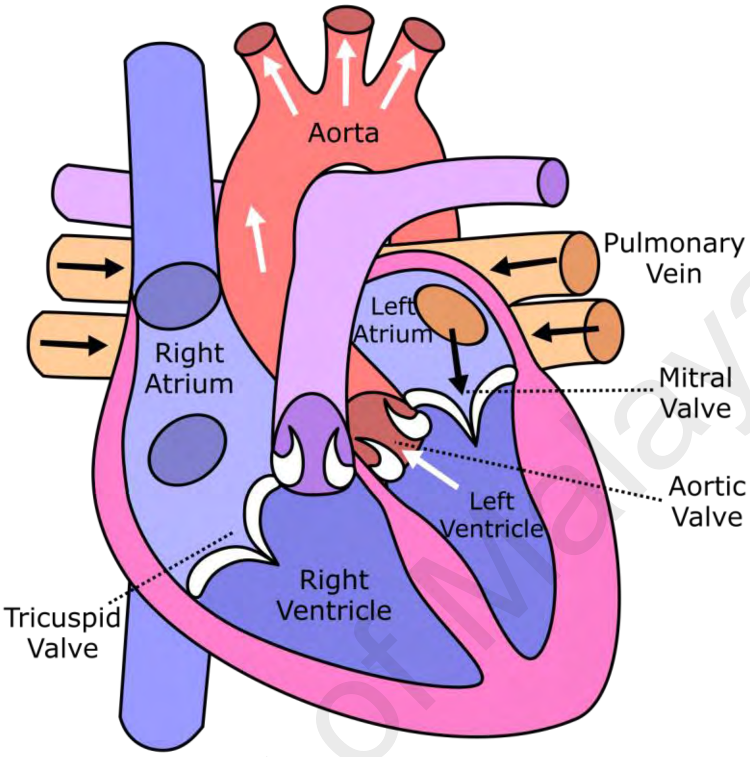


Figure 2.1: Illustration of the human heart, focusing on the systemic circulation loop. Blood flow follows the sequence of pulmonary veins (from the lungs) → left atrium → mitral valve → left ventricle → aortic valve → aorta (to rest of the body). Adapted from ("Heart," 2017)

The systemic circulation loop supplies the entire body with oxygenated blood excepting the lungs, thus the left heart needs to generate a substantially larger pumping force compared to the right heart. Anatomically this manifests as the substantially thicker myocardium wall of the LV compared to its right counterpart. This relative importance is also reflected in clinical practice, whereby most heart disease imaging protocols are founded on or begin with a structural and functional assessment of the LV (Kramer et al., 2013).

In a complete cardiac cycle, the heart periodically contracts and relaxes the myocardium surrounding the atrium and ventricles. This contraction is the primary force generator for the blood pump, and a typical young adult heart might beat at 70 cycles per minute (Hall, 2015). Focusing on the LV, the time period where the myocardium contracts is known as *systole* – this is where ventricular pressure peaks to expel the blood through the aorta to supply the rest of the body. The time where the myocardium then relaxes is known as *diastole* – this is where ventricular pressure drops to a minimum, allowing the ventricle to be refilled with blood for the next cycle. End-diastole (ED) and end-systole (ES) are two specific time points often referenced in LV quantification. The temporal state of the cardiac cycle is controlled by the electrical activity within the myocardium, however in practice, for imaging use, ED and ES are usually defined as the time points where the LV has maximum and minimum blood volumes, respectively (Schulz-Menger et al., 2013). This is also taken to correspond to the point of maximum dilation and maximum contraction of the LV, respectively.

The orientation of the heart is typically defined by the wall (*septum*) that divides the left and right sides; this approximately defines the longitudinal or long-axis (LA) plane (Figure 2.2). Viewed in this light, the LV coverage can be localized by two landmarks: the tip where the heart tapers off (*apex*), and the approximate midpoint of the mitral valve (*base*). Orthogonal to the LA plane is the transverse or short-axis (SA) plane; this defines a plane roughly parallel with the orientation of the mitral valve. Finally, the four-chamber orientation defines a plane that cuts through all four chambers of the heart, and is approximately orthogonal to the LA (along the septum wall) and SA planes (Walsh et al., 2012a). It should be noted that these three planes (SA, LA, four-chamber) do not correspond to the standard anatomic planes of the body (coronal, sagittal, transverse); the heart is oriented obliquely within the chest.

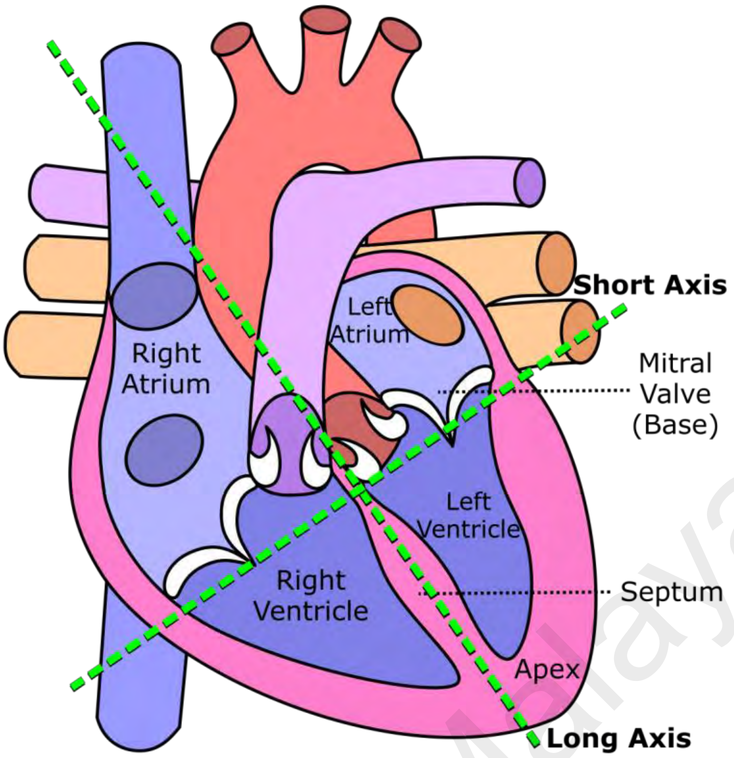


Figure 2.2: Illustration of the orientation and landmarks of the human heart. The LA plane is approximately defined by the septum wall dividing the left and right sides. The SA plane defines a plane approximately parallel with the orientation of the mitral valve. The LV coverage is commonly localized through the apical (apex) and basal (base) points. Adapted from ("Heart," 2017)

2.3 Magnetic resonance acquisition and quantification of the LV

Magnetic resonance is an imaging modality that utilizes magnetic fields and radiofrequency pulses to measure and image tissue properties. Compared to computed tomography (CT) scans, MR does not use ionizing radiation, offers better soft tissue contrast, and is able to natively image at arbitrary oblique planes. Compared to ultrasound scans, MR has a larger field-of-view, significantly better spatial resolution and image quality, and has little to no dependence on operator skill. For these and related reasons, MR imaging is currently considered the gold standard for diagnostic cardiac imaging (Walsh, Fang, Fuster, & O'Rourke, 2012b).

MR is a relatively complex imaging modality. By adjusting key acquisition parameters, different acquisition protocols or *pulse sequences* can be designed, which measure different tissue characteristics (Bitar et al., 2006). These pulse sequences are usually categorized based on their principle method of triggering the measured return signal, the most common being the *spin echo* and *gradient echo* sequence families. Steady-state free precession (SSFP) is at present the most common clinical pulse sequence used for standard 2D+time multi-slice cine cardiac MR acquisitions of the heart. It is a gradient echo technique, and provides strong contrast between the blood (high intensity) and myocardium (low intensity) (Figure 2.3).

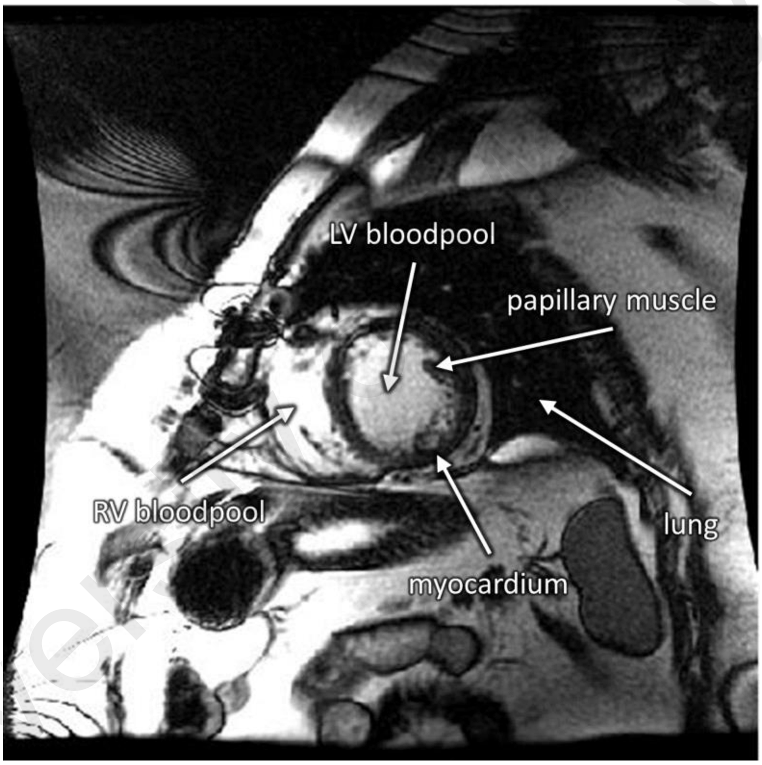


Figure 2.3: Sample SSFP acquisition of the LV in the SA plane. Good contrast is seen between the blood pool (high intensity) and myocardium (low intensity). Papillary muscles can also been seen; these connect the mitral valve cusps to the main body of the heart.

Clinical SSFP acquisitions are still largely 2D with slice thickness around 8−10 mm, and 3D volumes are acquired slice-by-slice. Clinical quantification is usually performed against images captured in the SA plane, as the contraction motion of the LV is best visualized there. Multi-slice SA acquisitions are performed for coverage of the LV from apex tip to base (mid of mitral valve); typically this would encompass around 10 slice locations (Kramer et al., 2013). To capture the cardiac motion, electrocardiography gated acquisitions are used to image individual time points evenly sampled through a full cardiac cycle; typically this would encompass around 20 separate cardiac phases. Thus, a representative SA SSFP cine acquisition with 10 slice locations and 20 cardiac phases would result in 200 individual 2D images (Kramer et al., 2013).

In-plane spatial resolution for clinical protocols commonly varies around 1−2 mm/pixel. Given typical clinical protocols utilizing 8−10 mm slice thickness, it can be seen that multi-slice SA volumes are significantly under-sampled along the long axis, and thus not useful for any form of multi-planar reconstruction. To address this, most clinical guidelines recommend additional supporting acquisitions at various useful LV-focused orientations. Typically these would include the 2-chamber vertical LA view, which cuts through the LV and left atrium through the centre of the mitral valve, the 4-chamber LA view, which cuts through all four chambers of the heart through the mitral and tricuspid valves, and less frequently the LV outflow tract view, which cuts through the LV and the centres of the mitral and aortic valves (Figure 2.4) (Kramer et al., 2013).

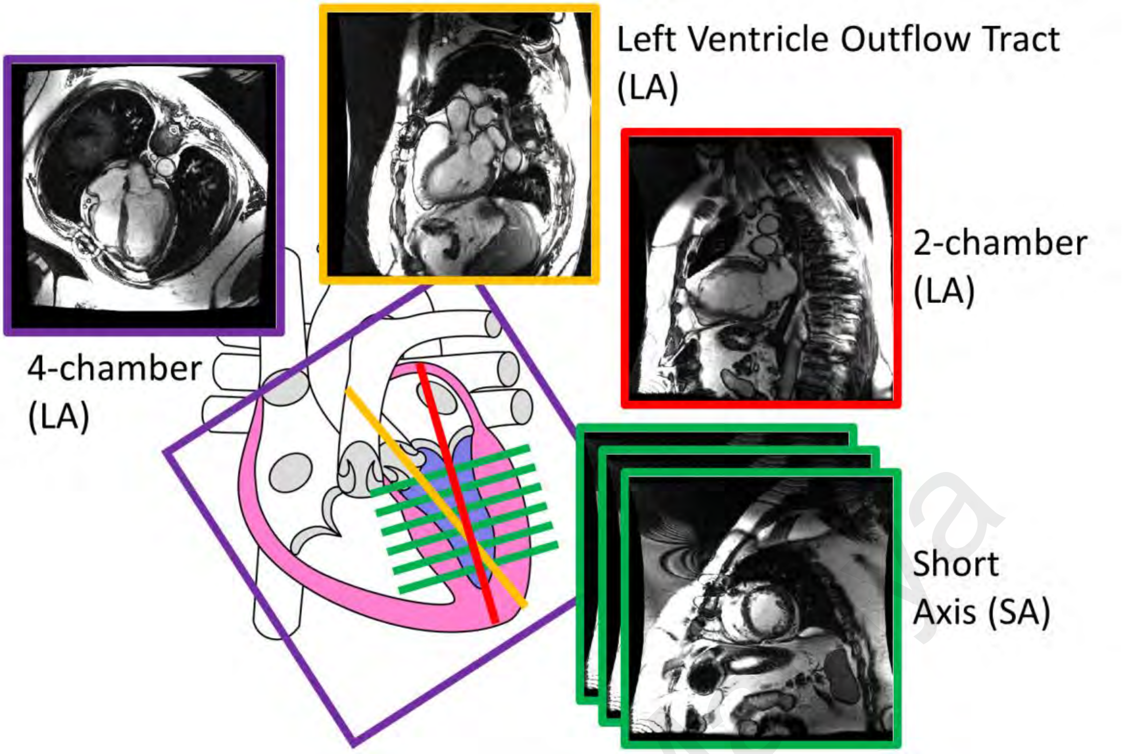


Figure 2.4: Common cardiac MR acquisition views for structural and functional assessment of the LV. (Red) **Vertical LA 2-chamber view aligned through the apex and centre of the mitral valve.** (Green) **Multi-slice SA view, typically aligned perpendicular to the 2-channel view.** (Purple) **4-chamber LA view aligned through all four chambers of the heart through the mitral and tricuspid valves.** (Yellow) **LV outflow tract view aligned through the LV and the centres of the mitral and aortic valves.**

Standard guidelines for clinical assessment call for the quantification of multiple parameters including LV ED and ES volumes, LV ejection fraction, LV stroke volume, cardiac output, and LV mass. These parameters are quantified via the delineation of the myocardium walls across the entire stack of SA multi-slice images (Schulz-Menger et al., 2013). In the SA plane, the LV myocardium resembles a connected low intensity ring, defined by the *endocardial* (the inner wall neighbouring the LV blood pool) and *epicardial* contours (the outer wall neighbouring the RV and lungs). Certain clinical parameters only require delineation of the endocardial contours (e.g. LV ED and ES volumes and LV ejection fraction), whereas others require full myocardium delineation (e.g. LV mass). Papillary muscles are technically part of the myocardium, but it is a common clinical practice to exclude them during quantification (Schulz-Menger et al., 2013) (Figure 2.5).

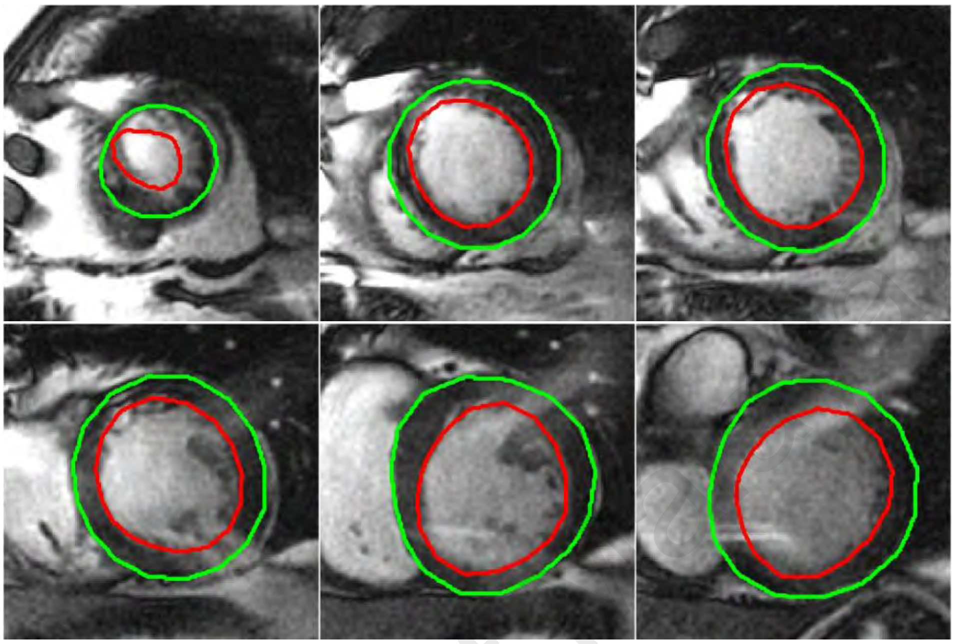


Figure 2.5: Sample delineation of the LV myocardium for quantification of clinical parameters. Red contour indicates endocardial wall, green contour indicates epicardial wall. Images from left-to-right, top-to-bottom, reflect selected slice positions in apex-to-base order.

2.4 LV segmentation algorithms

Given the importance of LV quantification, there has been significant research focused on automated segmentation of LV myocardium, particularly for MR cine images in the SA plane. This section focuses on published approaches utilizing *expert knowledge* segmentation techniques, which are defined here as techniques largely derived from human hand-crafted features or algorithms. A common characteristic of these techniques is a set of assumptions derived from human expert knowledge. In the next section (Section 2.5), an alternate neural network data-focused approach with no expert assumptions required will be discussed.

In this section, existing expert knowledge-based LV segmentation techniques are categorized into three general categories: (1) purely image- or pixel-driven methods such as intensity thresholding or distribution modelling utilizing blood pool to myocardium contrast; (2) methods incorporating statistical or geometric shape methods to model the LV such as truncated ellipsoids; and (3) anatomical atlas-based registration. Frangi, Niessen, & Viergever (2001), Petitjean & Dacher (2011), Tavakoli & Amini (2013), Zhuang (2013), as well as Peng et al. (2016) have provided comprehensive reviews of state-of-the-art techniques.

2.4.1 Image- or pixel-driven methods

Image- or pixel-driven methods encompass basic image processing techniques such as intensity thresholding or binning, as well as more complex texture classification methods or edge and regional energy techniques such as active contours (also known as *snakes*). A selection of representative papers are listed below.

Nachtomy et al. (1998) utilized minimum error thresholding to categorize pixel intensities into three classes: lung, myocardium, and blood. The threshold was applied against a dynamically expanding region-of-interest centred on an initial seed point, utilizing the assumption that the image histogram would change from unimodal (blood only) to bi-modal (blood and myocardium) to tri-modal (blood, myocardium, and lung) as the region-of-interest expanded. The algorithm was semi-automatic – requiring manual localization of the LV centrepoint. It was evaluated against the data of 20 subjects.

Lynch, Ghita, & Whelan (2006) used a modified *k*-means clustering algorithm to iteratively cluster and merge pixel intensities into connected regions. The LV blood pool was then identified via multiple morphological criteria including shape (circularity) and continuity between slices. This was then used to identify the similar intensity right ventricle (RV) blood pool, and to identify the myocardium (septum) separating the LV and RV. Finally, the identified septum was extrapolated via a cubic spline fit to surround the LV blood pool, to represent the full myocardium. The algorithm was fully automatic, and it was evaluated against the data of 25 subjects.

Üzümcü, van der Geest, Swingen, Reiber, & Lelieveldt (2006) used dynamic programming to optimize the myocardial contours in 2D+time volumes. Dynamic programming is a general optimization technique; in this instance it was used to solve for the optimal connected path given a cost function (image gradient) and constraints (spatial and temporal shifts between neighbours). The algorithm was semi-automatic – it required an initial manual contour to be performed. This was used to generate 32 equally spaced 2D+time region of interests sampled uniformly around the contour, which was then used as the search space for the optimal connected path. The algorithm was evaluated against the data of 20 subjects.

Jolly (2006) developed a flexible system for LV segmentation in both MR and CT images, for both single frame and multi-frame (propagation across time) applications. The system started with LV localization: an LV cross sectional intensity profile was generated from training data and a Hough-based voting procedure was used to localize its position in evaluated images. For LV segmentation, a Gaussian mixture model was used to categorize pixel intensities to air, muscle, and blood. This rough segmentation was then used as the starting point for an active contour fit against the image gradient including temporal propagation of the curve. The system was fully automatic, and it was evaluated against the data of 29 subjects.

2.4.2 Statistical geometric models

In comparison to the image- or pixel-driven methods described in Section 2.4.1, the methods described in this section incorporate stronger assumptions about shape. In the case of LV segmentation, this might be reflected in the circular or elliptical shape of the

LV in 2D SA images, or the truncated cone shape of the LV in LA or 3D SA volumes. The representative papers listed here include techniques based on geometric parameters as well as learned statistical objects derived from training data.

Pluempitiwiriyawej, Moura, Wu, & Ho (2005) utilized a modified active contour-based technique, incorporating strong shape priors in the energy function in addition to the standard edge and smoothness terms. Their function operated on 2D SA images, and the shape prior was modelled as a five-parameter ellipse function. The algorithm was semi-automatic – requiring an initial circle bounding the LV to be defined. It was evaluated against the data of 48 subjects, though only at the mid-level LV.

O'Brien, Ghita, & Whelan, (2011) collated a training set of SA images with a 32 point contour defined for each endo- and epicardial contour. After rigid registration of the training set, principal component analysis was applied to quantize the training shape parameters to 98% variation coverage. This derived statistical parametric shape was then incorporated to an active contour model for the LV segmentation. The authors also utilized endo-epicardium dependencies by building a statistical model deriving an additional weighted function for the epicardium size, given a prior evaluated endocardial contour. The algorithm was semi-automatic – requiring manual localization of the LV centrepoint. It was evaluated against the data of 33 subjects.

Assen et al. (2006) built a training set of SA and LA images, including LA planes captured at regularly spaced radial angles. From this a point cloud of LV landmarks was defined, and principal component analysis was applied to quantize the shape parameters. When matching the statistical shape to the target volume, the authors transformed the image intensity values to categorical tissue classes via fuzzy C-means clustering. This allowed the algorithm to be generalized to multimodality use, the only assumption being that air, myocardium, and blood pool have successively higher intensity values. The algorithm was semi-automatic – requiring manual delineation at the basal and apex slices. It was evaluated against the data of 20 subjects.

Zhang, Wahle, Johnson, Scholz, & Sonka (2010) first built a 4D template model of the LV and RV, then utilized a combination of 4D and 3D active shape and active appearance models to perform segmentation. The 4D template was generated as sixteen 3D point clouds in Euclidean space. For the appearance model, a reduced resolution texture was sampled from the template surfaces of the LV and RV. The algorithm was semi-automatic, segmentation was performed in two steps: a manually assisted first pass fit the 4D template model to the 4D volume data. The second pass refined the segmentation based on 3D fitting at each individual phase. The algorithm was evaluated against the data of 25 subjects.

2.4.3 Anatomical atlas-based registration

The methods described in Section 2.4.2 are often based on a sparse point cloud template of the LV; the derived shape or atlas coverage is usually restricted to the myocardium surface. In contrast, anatomical atlas-based registration is usually based on a dense 3D or 4D voxel volume derived from training data, which includes surrounding non-myocardium tissue like blood and lung. Segmentation via anatomical atlas-based registration is thus comparatively more data-driven. Nevertheless, strong human-enforced assumptions and constraints are present, e.g. atlas creation is usually performed via an arithmetic mean function across co-registered training data. The registration algorithm also carries strong assumptions and constraints via the choice of transformation used. E.g., rigid (translation + rotation + scale), affine (rigid + shear), or more complicated geometric transforms such as meshed warp fields (Vercauteren, Pennec, Perchant, & Ayache, 2009).

Zhuang et al. (2008) utilized a high resolution single individual volume for their anatomical atlas. They initialized the registration with separate masked similarity transforms for the LV and RV, respectively, resolving overlap conflicts via a distance weighted interpolation. Following that, a non-rigid fluid registration was performed to determine the final transformation. This was a gridded diffeomorphic deformation field based off a viscous fluid model, i.e. the transformation was invertible. The algorithm was fully automatic, and it was evaluated against the data of eight subjects.

Zhuang, Rhode, Razavi, Hawkes, & Ourselin (2010) was an evolution on the work of Zhuang et al. (2008). The authors built an anatomic atlas from the combined, averaged scans of ten healthy volunteers. The registration was initialized with separate masked affine transforms for all four chambers of the heart plus the major veins and arteries, resolving overlap conflicts via a distance weighted interpolation. Following that, a non-rigid adaptive control point free-form deformation registration was performed to determine the final transformation. This was a gridded diffeomorphic deformation field, designed to identify key control points around which the search space was focused, for the reduction of computation time. The algorithm was fully automatic, and it was evaluated against the data of 37 subjects.

Rikxoort et al. (2010) acquired 15 cardiac scans from a lung cancer screening trial, for atlas construction. However, instead of combining them into a single composite volume, the authors maintained them as separate volumes in a multi-atlas reference library. During segmentation, a fast first pass was performed, utilizing affine registration to match the target volume to each individual volume in the multi-atlas reference. The difference image was then calculated for each registration and used to determine which individual atlas best matched the target, and subsequently a full non-rigid B-spline registration was performed for the final registration and segmentation. The algorithm was fully automatic, and it was evaluated against the data of 29 subjects, though for cardiac CT rather than MR.

Table 2.1: Expert knowledge-based approaches for LV segmentation

Publication	Description	Manual input	No. of subjects
Image- or pixel-driven methods			
Nachtomy et al. (1998)	Pixel intensity categorization via minimum error thresholding	LV centroid	20
Lynch et al. (2006)	Pixel intensity categorization via *k*-means clustering	None	25
Üzümcü et al. (2006)	Contour modelled as optimal connected path using dynamic programming	Initial contour	20
Jolly (2006)	Pixel intensity categorization via Gaussian mixture model	None	29
Lu et al. (2013)	Shape metric + intensity thresholding + region-growing	Select mid-slice	133
Eslami et al. (2013)	Random walks guided by database of subjects	Seed LV and background	104
Albà (2014)	Graph cut with shape and interslice smoothness constraints	None	35
Statistical geometric models			
Pluempitiwiriyawej et al. (2005)	2D active contour using five-parameter ellipse function	Initial contour	48
Assen et al. (2006)	Principal component analysis on LV point cloud + fuzzy C-means clustering	Initial contour	20
Zhang et al. (2010)	3D+4D active shape and appearance model	Adjust first pass	25
O’Brien et al. (2011)	Principal component analysis + active contour model	LV centroid	33
Wu et al. (2013)	Circular active contour + modified edge gradient vector convolution	None	171
Woo et al. (2013)	Level sets with coupled endo- & epicardium shape constraints	LV centroid and radius	15
Queirós et al. (2014)	B-spline explicit active surface + optical flow motion tracking	Select apex & base	45
Anatomical atlas-based registration			
Zhuang et al. (2008)	Masked rigid + non-rigid registration	None	8
Zhuang et al. (2010)	Individual chamber masks + non-rigid registration	None	37
Rikxoort et al. (2010)	(CT) Multi-atlas library + non-rigid registration	None	29
Bai et al. (2015)	Multi-atlas library + augmented feature vector including neighbouring appearance	None	83

2.5 Neural networks and deep learning

2.5.1 Introduction

Artificial neural networks (ANNs) are a family of mathematical functions with numerous recent successes in tackling artificial intelligence problems, including image processing and recognition (LeCun, Bengio, & Hinton, 2015). Originating in the 1960s, ANNs showed early promise, but limitations in hardware computational power largely led to disappointing performance in complex tasks. ANNs experienced an extended period of low interest from the 1980s onwards, but have seen a strong resurgence in recent years, largely due to improvements in computing hardware and the availability of large quantities of training data (Goodfellow, Bengio, & Courville, 2016).

The expert knowledge image processing techniques described in Section 2.4 are strongly dependent on hand-crafted features or algorithms. They tend to be compact in their representation; though a technique such as active contours may be complex in description, its execution is only dependent on a small number of tuned parameters. In contrast, ANNs are almost completely data-driven. The algorithms used in ANNs tend to be structured as a network of elementary operations, and the underlying training data and learning process provides the bulk of the network performance. Post-training, the "learned" parameters typically number in the millions.

ANNs can be understood as a chain of linear operations interspersed with various nonlinear *activation* functions. Each group in the chain is more commonly known as a *layer*, which consists of a matrix of weights, W, and a vector of biases, b. For each individual layer the input vector is multiplied and summed against W and b, respectively. An element-wise nonlinear activation function (e.g. a hyperbolic tangent function) is then applied and the resulting output is used as the input to the subsequent layer and the general series of operations is repeated in further layers. Each output element is a learned *feature*, excepting the final layer, whose output is expected to be target result of the training.

Traditional ANNs are also known as fully connected networks (FCONs), and are typically used with unstructured vector input (Figure 2.6). For inputs with regular structure (e.g. a 2D image), convolutional neural networks (CNNs) are a more suitable variant. Here, W and b are applied repeatedly in a sliding window fashion analogous to the standard convolution operation in signal processing (Figure 2.7).

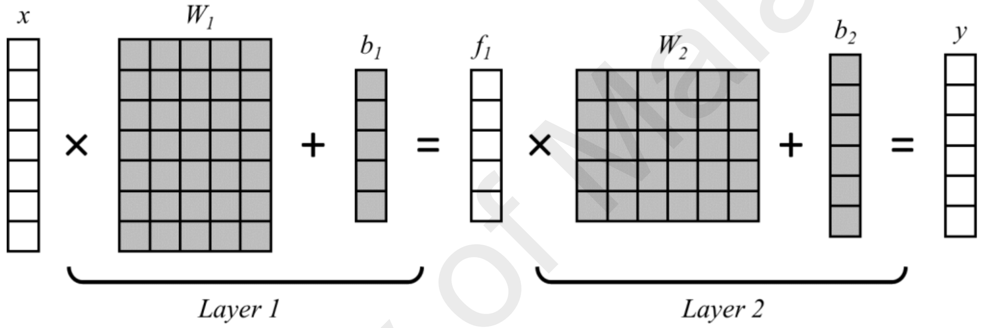


Figure 2.6: Simplified diagram depicting a two-layer fully connected neural network. The input, x, is a seven-element unstructured vector. It is multiplied and summed by the weight matrix W_1 and bias vector b_1, respectively, resulting in a five-element feature vector f_1. f_1 is used as the input to the second layer, resulting in the final six-element output vector, y. The sizes of the weight matrixes and biases vectors are design decisions. The non-linear activation functions are not shown, but would typically be applied as the final step prior to each layers' output. Shaded boxes indicate the network parameters, i.e. these would be the variables being optimized during network training.

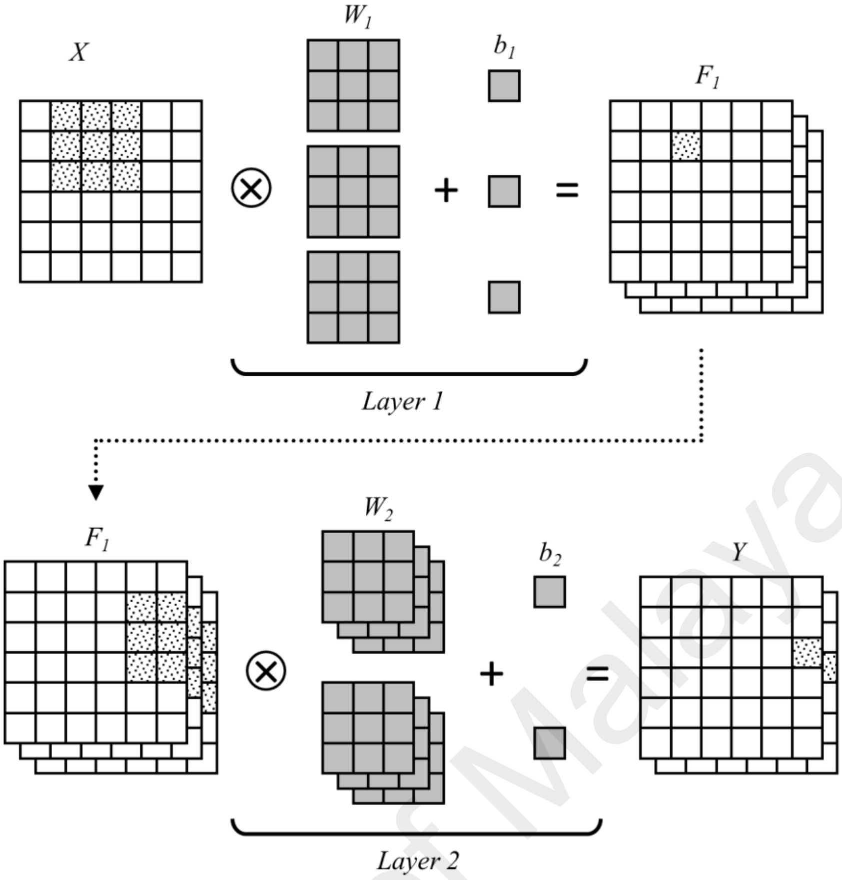


Figure 2.7: Simplified diagram depicting a two-layer convolutional neural network. The input, *X*, is a 6×6×1 gridded matrix. It is multiplied and summed (convolved) by the weight matrix W_1 and bias vector b_1, respectively via sliding window, resulting in a 6×6×3 feature matrix F_1. F_1 is used as the input to the second layer, resulting in the final 6×6×2 output matrix, *Y*. The sizes of the weight matrixes and biases vectors are design decisions. The non-linear activation functions are not shown, but would typically be applied as the final step prior to each layers' output. Shaded boxes indicate the network parameters, i.e. these would be the variables being optimized during network training. Dotted boxes demonstrate how an individual output element is mapped from the corresponding input convolution window.

The *W* and *b* values of all layers are collectively referred to as the network parameters. Starting from a random initialization, the parameters are iteratively updated by feeding random batches of training data through the network, calculating a loss function against the desired output (e.g. mean squared error), then back-propagating the result via an optimization function such as gradient descent. This is repeated until convergence. In recent years various performance-enhancing tweaks have been introduced to the general ANN architecture. Representative examples include rectified linear units (ReLUs) as activation functions (Glorot, Bordes, & Bengio, 2011), max-pooling feature matrixes for local translational invariance and reduced computational load (Krizhevsky, Sutskever, & Hinton, 2012), and random parameter dropout to improve network generalization (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). In general there are few restrictions on allowed operations in the network architecture, save that the final computational graph be differentiable so that the optimization function may be applied.

As previously mentioned, ANNs have a history reaching back to the 1960s. During the resurgence of ANNs in the past few years, a new term known as *Deep Learning* was introduced to differentiate current ANN architectures from previous generations (Bengio, 2009; LeCun et al., 2015). Largely a rebranding exercise, there is no specific definition that describes a particular deep learning architecture. In general, the term implies ANN architectures which are tens to hundreds of layers deep, as opposed to the *shallow* architectures of previous generations; the implication being that these deeper levels of abstraction would allow for fundamentally higher levels of automatic learned data representation.

2.5.2 Neural networks for image segmentation

Most observers date the modern resurgence of ANNs to the year 2012, where a CNN-based approach convincingly won the public ImageNet Large-Scale Visual Recognition Challenge (Krizhevsky et al., 2012; LeCun et al., 2015). ImageNet is an open image processing competition that is run annually since 2010. The primary challenge involves whole image classification of over one million images into 1000 possible classes (e.g. tiger, hamster, restaurant, water bottle, etc.) (Russakovsky et al., 2014). Most CNN approaches to whole image classification involve designing a network architecture that accepts a consistently sized image as input, then through a series of pooling operations and FCON layers, would condense the output to a 1000 length vector of probabilities representing the 1000 unique classes.

Given this existing design of neural network classification systems, the natural adaption to perform image segmentation is to switch from whole image classification to per-pixel classification. I.e., the output would be a 3D matrix where the first two dimensions would match the dimensions of the input image, and the third dimension would be the vector of class probabilities (Figure 2.8). This is also known as *semantic* image segmentation. A common design implementing this is the so-called *encoder-decoder* network or *U-net* architecture, where intermediate layers alternately reduce, then expand the feature matrix (Long, Shelhamer, & Darrell, 2015; Ronneberger, Fischer, & Brox, 2015), e.g., given a 128×128 input image, the feature matrixes in intermediate layers might halve in steps till 32×32 (encoding stage), then double in steps till 128×128 at the final output layer (decoding stage).

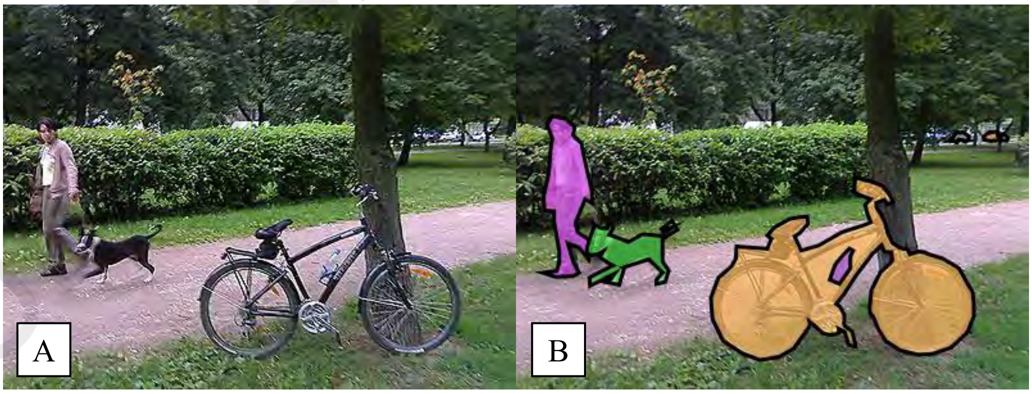


Figure 2.8: Sample images demonstrating per-pixel image segmentation. A: **source input image.** B: **coloured overlays indicate classified objects of interest. At least three different classes are shown, including person (magenta), dog (green), and bicycle (beige). Data sourced from T.-Y. Lin et al. (2014).**

The per-pixel classification method is an extremely flexible design for image segmentation, as there are no inherent restrictions to placement, shape, or connectivity, save for any constraints learned automatically by the network. Indeed, even individual pixel class exclusivity is not a requirement, allowing for overlapping segmentation boundaries. This inherent flexibility is a boon for general image segmentation, but is excessive for segmentation tasks in specialized domains such as medical organ segmentation, where the input images have significantly less variance. An alternate image segmentation method utilizing CNN *regression*, rather than per-pixel classification, is therefore explored and introduced in Chapter 4.

2.5.3 Neural networks for LV image segmentation

Given the extended period of low interest in ANNs as mentioned in Section 2.5.1, the majority of published LV segmentation papers utilizing ANNs are found beginning of 2016 or so. This brief literature review focuses on approaches targeting MR cardiac imaging.

Stalidis, Maglaveras, Efstratiadis, Dimitriadis, & Pappas (2002) is a rare early approach utilizing a three-layer non-convolutional (i.e. fully-connected) neural network variant for per-pixel LV segmentation from SA scans. Their approach was semi-automated, requiring initialization via manual localization of the LV centroid, as well as apical and basal planes. The network input consisted of only three values: the individual pixel intensity, angular position around the LV centrepoint, and SA slice position along the long axis. Unfortunately, the system appeared to have been evaluated against only three subject data, with no indication whether a training-testing data partitioning scheme was used.

Right at the cusp of the ANN resurgence, Ngo & Carneiro (2013) combined level set methods with a two-layer fully-connected ANN to perform LV endocardial wall segmentation. Their approach required manual initialization of the LV bounding box and centroid. The ANN was used as a first pass to produce a 20×20 segmentation matrix; this was then used to compute a distance function for the subsequent level set algorithm to produce the final segmentation image. The system was tested against 45 public datasets from the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2009 challenge, also known as the Sunnybrook database (Radau et al., 2009). Interestingly, the authors labelled their ANN design as a *deep belief* and deep learning architecture despite having only two layers, demonstrating the fuzziness of the term at that time.

More recently, Avendi, Kheradvar, & Jafarkhani (2016) combined CNNs and a deformable model to perform LV endocardial wall segmentation, though only in the ED and ES phases. Three separate networks were trained; consisting of a two-layer CNN for the initial localization of heart region-of-interest, and two separate three-layer fully-connected networks for LV segmentation at basal and mid-ventricular slices, and for apical slices, respectively. The system was tested against the 45-subject Sunnybrook public database.

To better model inter-phase spatial dependencies, Poudel, Lamata, & Montana (2016) utilized a 14-layer CNN encoder-decoder architecture and inserted a recurrent layer at the middle of the encoder-decoder graph, i.e. where the feature matrix was most dense. A recurrent layer is an ANN variant that allows for processing loops, where related data (in this case pixels from neighbouring cardiac phases) are fed into subsequent cycles of the processing loop, allowing data from previous phases to influence evaluation of the current phase. The system only performed LV endocardial wall segmentation, and was tested against the 45-subject Sunnybrook public database as well as a private 234-subject database.

A 15-layer CNN encoder-decoder architecture was used by Tran (2016) to perform complete LV and RV segmentation, including endocardial and epicardial walls. The system was tested against multiple public databases, including the 45-subject Sunnybrook database, as well as the Statistical Atlases and Computational Modelling of the Heart (STACOM) 2011 Left Ventricle Segmentation challenge (LVSC), a public database consisting of 200 subjects.

All the ANN techniques documented above utilize fully supervised training. i.e., all images used for training were paired with gold standard target label maps for calculation of the loss function. In contrast, Bai et al. (2017) tested a two-step semi-supervised method of training: after an initial period of fully supervised training, new images with no gold standard label maps were introduced into the mix. The partially trained network was used to generate estimated label maps, which were then refined via a conditional random field. The refined label maps were used to train the network, and this was repeated until training plateaued. The authors showed a small improvement in the performance of semi-supervised training compared to the standard fully supervised training. The system was tested against a private 340 subject database (100 fully labelled, 240 unlabelled).

Most recently, Baumgartner et al. (2018) investigated four different CNN architectures for complete LV and RV segmentation, including endocardial and epicardial walls. The four tested designs were variants on the basic CNN encoder-decoder architecture, with a notable exception being a variant that used multi-slice (i.e. 3D) images as input. Interestingly, the authors showed the conventional 2D networks having better performance overall, possibly due to limitations of clinical LV cine imaging such as low inter-slice spatial resolution and the presence of inter-slice shift. The system was tested against the Automated Cardiac Diagnosis Challenge 2017, a public database consisting of 150 subjects for training.

All the ANN papers documented in this review perform LV segmentation through per-pixel classification. An alternate approach utilizing CNN *regression* will be discussed in Chapter 4.

Table 2.2: Neural network and deep learning approaches for LV segmentation

Publication	Description	Manual input	No. of subjects
Stalidis et al. (2002)	ANN + image features	LV centroid + select apex and base slices	3
Ngo et al. (2013)	ANN + level set	LV centroid and ROI	45
Avendi et al. (2016)	Dual CNN + deformable model	Select apex slices	45
Poudel et al. (2016)	U-net variant with recurrent design, 2D+time input	None	279
Tran (2016)	U-net variant, 2D LV and RV	None	245
Bai et al. (2017)	Semi-supervised training via conditional random field	None	340
Baumgartner et al. (2018)	U-net variant, both 2D and 3D input	None	150
Zheng et al. (2018)	U-net variant, 3D input (contour propagation)	None	3980

2.6 Chapter Summary

2.6.1 Review summary

Clinically, the LV is the most important chamber of the heart, due to its role in supplying oxygenated blood to the majority of the body. The two primary orientations of the heart are the short- and long-axis, defined by the septum wall. In addition, the LV is typically localized through its apex and base landmarks, being the tip of the heart and the mid of the mitral valve, respectively. Multi-slice cine MRI is used to image the LV; standard acquisition protocols include multiple views or acquisition planes, namely the short-axis, 2-chamber, 4-chamber, and LV outflow tract views. Clinical practice requires the delineation of LV endo- and epicardial contours for quantification of clinically useful diagnostic parameters such as blood volumes and ejection fraction. Automated localization and segmentation techniques are desirable to allow fast and accurate diagnosis of heart diseases.

In this chapter, two general approaches for LV segmentation from multi-slice cine MRI were reviewed: expert knowledge-based techniques and data-driven techniques. The defining differences between them are strong human-introduced assumptions in the former, and weak to no human-introduced assumptions in the latter. For expert knowledge-based techniques, existing approaches were further categorized into three sub-groups: (1) purely image- or pixel-driven methods; (2) methods incorporating statistical or geometric shape methods; and (3) anatomical atlas-based registration. For data-driven techniques, the specific techniques of ANNs and CNNs were focused on.

2.6.2 Research gaps

Due to specific characteristics of MR imaging (namely phase wrap-around), cardiac MRI is usually acquired at a field-of-view that significantly exceeds the size of the heart. It is not uncommon for the heart to only occupy 10% of the acquired image matrix size. This large field-of-view is problematic for many algorithms, as it includes non-cardiac objects that might have a detrimental effect on performance. At a minimum, it increases computational time, especially for algorithms which utilize aspects of global exhaustive search such as registration based methods (Rikxoort et al., 2010). As such, many algorithms require human-assisted LV localization as a first step, e.g. placing a seed point in the centre of the LV blood pool, or defining a crop window. Despite this need, there are few published algorithms for automated LV localization; this is likely due to the research community deeming the problem not important enough to merit its own stand-alone solution. In this light, a fully automatic expert-knowledge based approach for LV blood pool localization was introduced and evaluated in Chapter 3. This approach tackles LV blood pool localization as a sub-problem for enabling subsequent automatic LV segmentation.

In typical short-axis cine MR images, there is usually strong contrast between the blood pool within the LV and the endocardial wall. However, differentiating the epicardial wall from surrounding structures is more difficult, particularly against low-signal lung tissue (Figure 2.3). Delineation is most straightforward at the mid-LV level, where the LV is relatively large and strongly circular. Significant partial volume effects arise in apical slices where the LV tapers off, whereas in basal slices the LV branches off to the ascending aorta and left atrium, resulting in ambiguous boundaries. Delineation is also easiest for the ED and ES phases, when the heart is relatively static and the myocardial wall is well defined. Other phases during contraction exhibit significant motion blurring. Because of these difficulties, most published algorithms only solve for a subset of LV segmentation: endocardial wall only, mid-LV planes only, or ED and ES phases only. In addition, all current published LV segmentation algorithms utilizing ANNs and CNNs carry out the segmentation via per-pixel classification, a highly flexible technique suited to segmentation of generic imagery. It is hypothesized that the specialized domain of LV segmentation may benefit from a more restricted model. An alternate method utilizing CNN regression is therefore introduced and developed for full segmentation of LV myocardium from base-to-apex over all cardiac phases. The details of this method including validation against public challenge data sets are described in more detail in Chapter 4 and Chapter 5.

CHAPTER 3: AUTOMATIC LOCALIZATION OF THE LEFT VENTRICULAR BLOOD POOL CENTROID IN SHORT AXIS CARDIAC CINE MR IMAGES

3.1 Abstract

This chapter presents and validates an expert knowledge-based algorithm for fully automatic localization of the left ventricular (LV) blood pool in short axis cardiac cine magnetic resonance (MR) images; this a sub-problem for enabling subsequent automatic LV segmentation. The algorithm was fully human designed in functionality and parameters; no direct machine learning techniques were used.

The algorithm comprises four steps: (i) quantify motion to determine an initial region of interest surrounding the heart, (ii) identify potential 2D objects of interest using an intensity-based segmentation, (iii) assess contraction / expansion, circularity, and proximity to lung tissue to score all objects of interest in terms of their likelihood of constituting part of the LV, and (iv) aggregate the objects into connected groups and construct the final LV blood pool volume and centroid. This algorithm was tested against 1140 datasets from the Kaggle Second Annual Data Science Bowl, as well as 45 datasets from the STACOM 2009 Cardiac MR Left Ventricle Segmentation Challenge.

Correct LV localization was confirmed in 97.3% of the datasets. The mean absolute error between the gold standard and localization centroids was 2.8 to 4.7 mm, or 12% to 22% of the average endocardial radius. The algorithm has been released as open source and is available online (https://github.com/tanlikuo/localizeLV).

This chapter has been published as (Tan, Liew, et al., 2018); the published text is largely reproduced here excepting minor amendments for consistency and flow.

3.2 Introduction

Cardiovascular diseases (CVDs) are the most common cause of death globally. It is estimated that 31% of all global deaths in 2012 were due to CVDs (Mendis, 2014). Cardiac MRI is considered the gold standard for assessment of various aspects of CVDs, particularly congenital heart diseases (Walsh et al., 2012b). Cardiovascular expert groups now recommend quantification as a standard assessment procedure (Schulz-Menger et al., 2013). These quantifications typically include the full delineation of myocardial borders, which represent a significant time burden when performed manually.

A significant amount of research has focused on automated delineation of the myocardial border, particularly for LV short axis cine images (Petitjean & Dacher, 2011). Such automated quantification usually involves multiple steps; a common initial step is localization of the heart within the large scan field of view (Kadir, Gao, Payne, Soraghan, & Berry, 2012; Petitjean & Dacher, 2011). Localization algorithms vary in their output result, ranging from a single 2D point representing the LV centroid (Pednekar, Muthupillai, Lenge, Kakadiaris, & Flamm, 2006), to approximate delineations of the blood pool or endocardial boundary (Jolly, 2008). For example, LV segmentation based on active contour techniques (Constantinides et al., 2009) or image grey level analysis (Jolly, Xue, Grady, & Guehring, 2009) require a gross initialization in the proximity of the LV endocardium or blood pool, typically via manual placement of a seed point in the LV cavity. Certain segmentation techniques do not directly require LV localization, for example atlas or geometric model registration-based techniques (Tavakoli & Amini, 2013). However, these applications could still benefit from localization as it may reduce computation load (e.g. by cropping the initial volume) or by providing reasonable initial parameter estimates for the registration optimization. As the localization output is designed to be subsequently refined by further segmentation algorithms, robustness in finding the general LV location is more important than precision of any approximate delineations.

Pednekar et al. (2006) described two approaches to LV localization: a dual-contrast method utilizing a black-blood sequence to contrast the bright-blood images of the standard cine scan, and a geometric approach utilizing the intersection line of the vertical long axis and four chamber scout views. Both approaches output an LV centroid, but rely on secondary acquisitions, which may be clinically infeasible or not available (e.g. where limited series are exported for off-site analysis). Lin, Cowan, & Young (2006) utilized the first harmonic of a Fourier transform to mask out non-heart regions, followed by a 2D line profile based segmentation of the endocardium. However, such motion masking can fail in cases with strong non-cardiac motion (e.g. abdominal). Jolly (2008) expanded on Lin et al.'s approach, adding a shape-based scoring scheme that quantifies contraction, roundness, connectedness, and concavity. This performed well on a large sample of 253 datasets, only failing in the presence of a strong aorta and motion artifacts. However, the technique uses direct Otsu thresholding of the blood pool intensity, which the author found to be unreliable in the presence of intensity artifacts (e.g. regions of high signal intensity caused by close proximity to the RF receiver coil). More recently, Zhong, Zhang, Zhao, Tan, & Wan (2014) utilized first and fifth harmonic images to produce edge maps, followed by an anisotropic weighted circular Hough transform to calculate LV centroids. However, this approach was only tested on 10 healthy subjects. While this chapter was in press, Albà et al. (2018) published an approach using random forests regression validated against over 1200 datasets. Notably, none of the aforementioned approaches were benchmarked against publicly available datasets.

Despite these published approaches, many current quantification tools (e.g. Medviso Segment v1.9 R3895, http://segment.heiberg.se) and papers on automatic LV quantification (Hu et al., 2013; Nambakhsh et al., 2013; Yin Wu et al., 2015) still rely on manual specification of the LV centroid, suggesting that automated localization remains an open issue.

In this chapter, an automatic multi-slice LV short axis localisation algorithm utilizing temporal and shape information is proposed. The algorithm provides a robust determination of the blood pool intensity threshold and subsequent binary object segmentation, and making use of the proximity of the LV to low intensity tissue, such as lung, for novel LV localization. The algorithm was tested against 1185 externally published datasets. The code is written in the MATLAB programming language and has been released under a GPL open source license (https://github.com/tanlikuo/localizeLV).

3.3 Materials and Methods

3.3.1 Datasets and Protocol

Following institutional ethics committee approval (989.75), 161 anonymized clinical cine datasets were obtained retrospectively from University Malaya Medical Centre, a tertiary referral hospital. These were routine clinical studies on patients with various cardiac and vascular diseases, acquired over the 12 month period of 2014. No exclusion criteria was used. The algorithm was exclusively designed against these 161 in-house datasets. No machine learning techniques were used.

All in-house clinical studies were performed on a 1.5T MRI scanner (Signa HDxt, GE Healthcare, WI), using an 8-channel cardiac coil. Imaging parameters were: x/y resolution = 1.37 mm/pixel, slice thickness = 8 to 10 mm, cardiac frames/phases = 20.

For primary testing, 1140 datasets from the Kaggle Data Science Bowl Cardiac Challenge (Kaggle & Booz Allen Hamilton, 2015) were utilized; this is a public machine learning challenge in 2015 for determining LV volumes from cardiac cine MRI data. In addition, the algorithm was also tested against 45 published datasets from the STACOM 2009 Cardiac MR Left Ventricle Segmentation Challenge (Radau et al., 2009).

The Kaggle datasets consist of a heterogeneous mix of scanners and protocols, with imaging parameters: x/y resolution = 0.65 to 1.75 mm/pixel, slice thickness = 6 to 10 mm, cardiac frames/phases = 25 to 30. The STACOM datasets were acquired using a 1.5T MRI scanner (Signa, GE Healthcare, WI), with imaging parameters: x/y resolution = 1.37 mm/pixel, slice thickness = 10 mm, cardiac frames/phases = 20 (Radau et al., 2009).

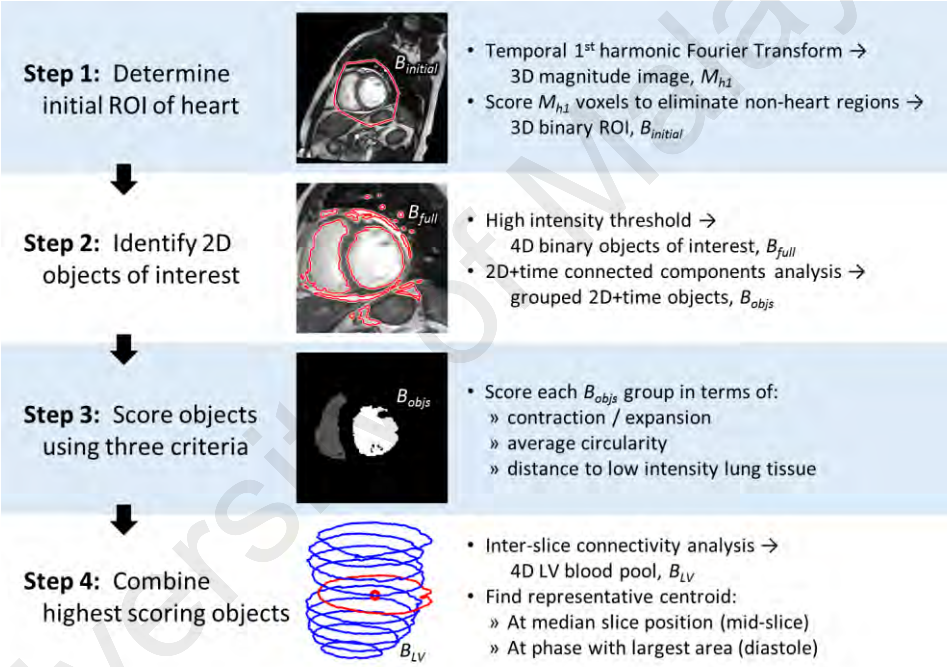


Figure 3.1: Flowchart of the proposed algorithm. Steps 1 through 4 are described in their corresponding chapter sections. Middle column illustrates key results from each respective step. Right column indicates key variables and functions in the source code, corresponding to each respective step.

3.3.2 Automated Localization

The proposed algorithm, as illustrated in Figure 3.1, comprises four main steps: (i) Determining an initial region of interest (ROI) enclosing the heart, while removing background air and non-cardiac tissue; (ii) Identifying high intensity 2D objects of interest within the initial ROI, focusing on the cardiac blood pool; (iii) Use three criteria to score all 2D objects of interest in terms of their likelihood of constituting part of the LV, and; (iv) Combining the 2D objects of interest into connected inter-slice groups, constructing the final volume and identifying the centroid of the LV blood pool.

Due to the complexity of this multi-step algorithm, the reader is advised to refer to the published source code (https://github.com/tanlikuo/localizeLV) as a complement to the textual description.

3.3.2.1 Step 1: Determine Initial Region of Interest

Cardiac MRI scans are gated to capture one complete cycle of the heart. In a typical cardiac scan, it is reasonable to assume that the heart is the primary moving object. A 1D first harmonic Fourier Transform over time (X. Lin et al., 2006) was first calculated, resulting in a magnitude image, M_{h1} (Figure 3.2B), which highlights regions of significant motion, primarily the heart, of each short-axis slice location from base to apex. 2D intensity weighted centroids of M_{h1} were then calculated for each slice, and a 3D line was fitted against all centroids (X. Lin et al., 2006). All voxels were then assigned a score consisting of the product of the inverted M_{h1} intensity and the 2D Euclidean distance of the intersection of the SA plane with the fitted 3D line. The resulting distribution of the scores was fitted to a log-logistic distribution, and voxels with probabilities exceeding 0.9 of the cumulative distribution function were removed (Figure 3.2C). This served to eliminate voxels of low motion located away from the heart. The 3D intensity weighted centroid was then calculated, and all previous steps were repeated iteratively until the Euclidean distance of the 3D centroid between the current and previous iteration fell below one pixel, for a maximum of five iterations.

For each slice in M_{h1}, the remaining (not removed) voxels were consolidated as a single 2D binary object, and the corresponding 2D convex hull was calculated. Finally, all the 2D convex hulls were combined to produce the initial 3D binary ROI, $B_{initial}$ (Figure 3.2B).

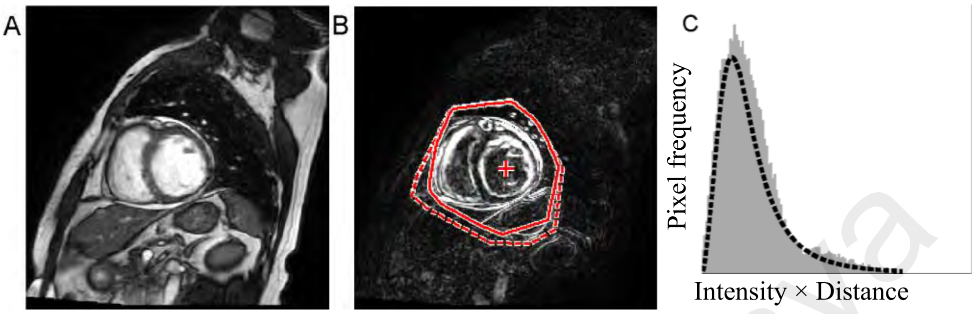


Figure 3.2: Determining an initial ROI targeting the heart, then identifying 2D objects of interest. A: **Original cine image.** B: **Magnitude image of the first harmonic Fourier Transform, M_{h1}, highlighting the convex hull of regions with significant motion. Dotted and solid lines depict the first and final iterations of the binary ROI $B_{initial}$, respectively. C: Gray histogram shows the distribution of the inverted M_{h1} intensity and Euclidean distance product score. Dotted line shows the fitted log-logistic distribution function.**

3.3.2.2 Step 2: Identify 2D Objects of Interest

The algorithm next identifies potential 2D objects of interest within the cine SA slices, masked by the $B_{initial}$ ROI. The target is the LV blood pool, the primary feature being its higher intensity compared to surrounding myocardial tissue. To identify these objects, an appropriate intensity threshold was first determined and applied, then a custom binary connected components analysis was performed to identify and group the individual objects.

Although an intensity threshold could be calculated directly from the cine volume (Jolly, 2008) using Otsu's method, the author found this to be unreliable in the presence of high intensity artifacts which would skew the threshold, i.e. the LV blood pool may be completely omitted due to overestimation of the threshold value. To avoid these difficulties, the first harmonic magnitude image, M_{h1}, from step 1 and the maximum intensity projection of the cine image across time, V_{max}, were combined to determine a more reliable intensity threshold, $t_{intensity}$. Specifically, M_{h1} was first thresholded using Otsu's method to produce a reference binary image, $B_{magnitude}$. Successive intensity thresholds were then applied to V_{max} until the thresholded 3D volume achieved 95% binary voxel intersection with $B_{magnitude}$. This threshold was recorded as $t_{intensity}$. This combination was used for two reasons: (i) $B_{magnitude}$ highlights voxels which fluctuate in intensity over time. The LV blood pool primarily exhibits contraction / expansion motion, resulting in intensity fluctuations at the object edges. (ii) V_{max} maximizes the extent of the LV blood pool, enabling significant overlap with the corresponding LV blood pool edges in M_{h1}, as shown in Figure 3.3. This overlap restricts the selection of $t_{intensity}$ to a range useful for blood pool object identification.

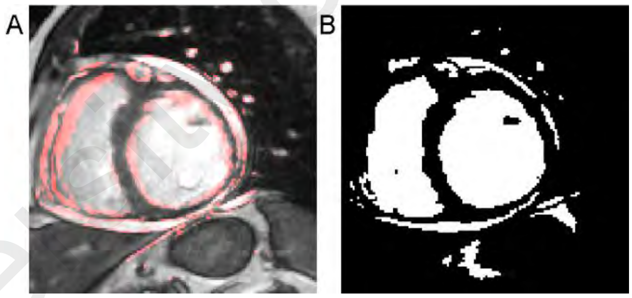


Figure 3.3: Calculating the intensity threshold to identify 2D objects of interest. A: **Thresholded first harmonic magnitude binary image, $B_{magnitude}$ (red tint), overlaid over maximum intensity projection image across time, V_{max}, showing correlated overlap with edges of moving objects.** B: **Result of calculating and applying threshold $t_{intensity}$ using $B_{magnitude}$ and V_{max}.**

To differentiate individual 2D objects of interest, the original cine stack is thresholded using $t_{intensity}$, producing a 4D binary volume B_{full} (Figure 3.3B and Figure 3.4A). The goal is to identify and group individual 2D objects of interest across time, i.e., the LV blood pool at a particular slice should be isolated and grouped as a single 2D+time object. Standard 2D+time (3D) binary connected components analysis is not suitable here due to heart motion causing connectivity “leaks” between cardiac phases, e.g., during systole the right ventricular (RV) blood pool in a particular phase will frequently “overlap” the LV blood pool of a neighbouring phase, resulting in the LV and RV blood pools being erroneously identified as a single 2D+time object.

To tackle this problem, the original cine stack was used to calculate the minimum intensity projection across time, V_{min}, then V_{min} was thresholded using $t_{intensity}$, resulting in the 3D binary volume, B_{core}. This serves to maximize separation between regions of high intensity such as the LV and RV blood pools, and may be thought of as the unchanging “core” of objects that otherwise change shape over time.

2D+time connectivity analysis is carried out between B_{core} and B_{full} with the following criteria: (i) isolated 2D objects in B_{full} that connect across time and overlap a single common B_{core} object, are labelled as a single unit (Figure 3.4A). (ii) If objects in B_{full} overlap with more than one object in B_{core}, then B_{core} objects with area < 10% of the largest object (an empirically selected threshold) were separated out, and the smaller objects were grouped to their larger counterparts based on Euclidean distance. This gracefully handles scenarios where a homogenous region is inadvertently separated into a primary object and surrounding chunks due to imperfect thresholding (Figure 3.4B). (iii) If the previous scenario is not fully resolved, e.g. multiple large B_{core} objects remain connected, then affected objects in B_{full} are split up per-voxel and matched to the closest B_{core} object by Euclidean distance (Figure 3.4C and D). This handles scenarios of intermittent connectivity between objects, such as branching of the RV to the pulmonary artery. The result of the connectivity analysis is a list of 2D+time binary object units, B_{objs}.

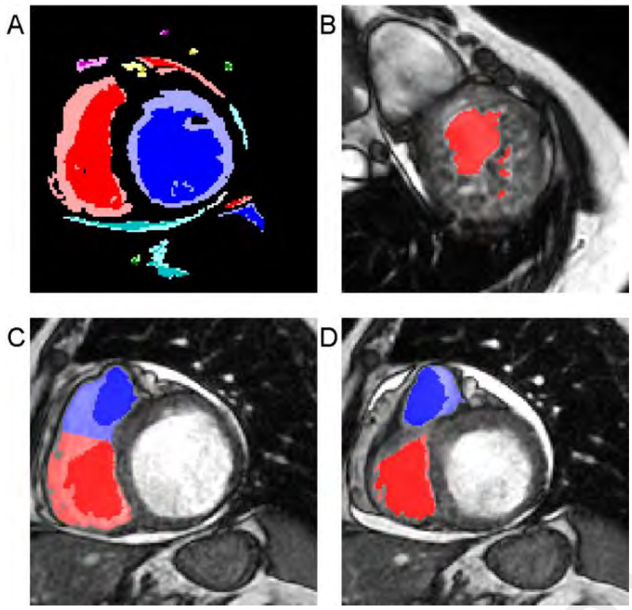


Figure 3.4: Separation into 2D+time connected object units. A: B_{core} **(dark tint) overlaid on** B_{full} **(light tint) at end diastole, demonstrating separation between objects.** B: **multiple** B_{core} **objects (red) shared a common** B_{full} **object in one or more phases, and were grouped together due to their relative sizes.** C & D: **at two particular phases, the RV blood pool demonstrates intermittent connectivity due to cardiac motion along the long axis near the pulmonary artery branch point. The shared** B_{full} **object is split up to enforce separation (**B_{core} **dark tint,** B_{full} **light tint).**

3.3.2.3 Step 3: Scoring the Objects of Interest

Each 2D+time B_{objs} unit was next scored in terms of its likelihood of being a segment of the LV blood pool. The following criteria are used: (i) normalized range of object contraction and expansion over time, (ii) average circularity, and (iii) minimum distance to low intensity weighted centroid.

The first score, s_1, reflects the assumption that the LV volume should be changing over time due to contraction and expansion of the heart. For each connected unit in B_{objs}, the range of the area was calculated, normalized by the average area as shown in Eq. 3.1.

$$s_1 = \left(\max(A) - \min(A)\right) \Big/ \sum_{t=0}^{t=T} \left(a_t \big/ T \right) \tag{3.1}$$

where t = cardiac phase, T = total number of cardiac phases, a_t = 2D area of binary object calculated as sum of pixels, A = set of all 2D areas for all cardiac phases

The second score, s_2, reflects the assumption that the cross sectional LV is generally circular in shape. For each connected unit in B_{objs}, the mean ratio of the major and minor axis lengths of the binary objects was calculated, such that a circle will score 1, and a line 0. This is basically an inversion of the common eccentricity shape measure, where the major and minor axis lengths are derived from the object eigenvalues (Haralick & Shapiro, 1992) as shown in Eq. 3.2.

$$s_2 = 1 - \tfrac{1}{T}\sum_{t=0}^{t=T}\left(2 \times \sqrt{\left(l_{major}/2\right)^2 - \left(l_{minor}/2\right)^2} \Big/ l_{major} \right) \tag{3.2}$$

where t = cardiac phase, T = total number of cardiac phases, l_{major} = major axis length, l_{minor} = minor axis length

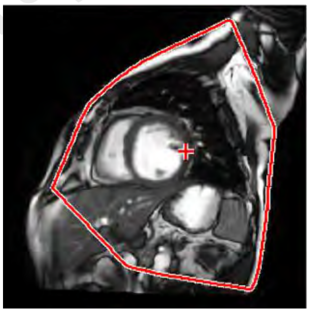


Figure 3.5: The inverted intensity weighted 2D centroid (crosshair), omitting background air and peripheral regions. The centroid was reliably located near the low intensity lung tissue.

The third score, s_3, reflects the assumption that the LV commonly neighbours regions of low intensity, particularly the lung, and therefore objects closer to lung tissue are assigned a higher score. This is done by calculating an inverted intensity weighted centroid on the main body, assuming that the centroid will be biased by the low intensity lung tissue: For each slice of the original cine image, the average intensity image across time is first calculated, thresholded with the same threshold used in earlier steps, $t_{intensity}$, and the area of the convex hull of the resulting binary image is recorded. 2D connected component analysis is applied, and the binary objects are removed one-by-one in inverse order to their Euclidean distance to the 2D centroid, until the resulting area of the combined convex hull drops below 75% of the original starting value. This was found to effectively minimize the effect of irrelevant low intensity regions including background air and semi-connected extremities such as limbs. Within this restricted convex hull, the inverted intensity weighted 2D centroid, and its minimum distance score to each connected unit in B_{objs} was calculated. The score was then inverted so that distinct objects in contact with the centroid will have high values (Figure 3.5).

The final aggregated score is calculated by multiplying the three separate scores. To minimize false matches, the final score for a particular B_{objs} unit is set to zero if the average area is < 400 mm^2, reflecting a minimum threshold for the size of LV blood pool. This cut-off value was obtained empirically from the in-house datasets.

3.3.2.4 Step 4: Combine Objects of Interest into Connected Groups

In the final step, B_{objs} units that are connected across slices were identified (along the LV long-axis) and combined into groups. The group with the highest total score is labelled as the LV blood pool.

To simplify the inter-slice connectivity analysis, each 2D+time B_{objs} unit was first reduced into a single 2D binary object by calculating the binary mode across time. Each object is then analysed in descending order of their score: for each object, neighbouring slices were examined sequentially, searching for objects that (i) have ≤ 15 mm Euclidean distance between their respective 2D centroids, and (ii) have ≥ 0.6 total fractional pixel overlap (normalized by the area of the smaller object). As before, both cut-off values were obtained empirically from the in-house datasets. The search is repeated until all B_{objs} units have been analysed; the group with the highest total score is labelled as the LV blood pool, and the selected B_{objs} units are merged to form a 4D binary object representing the final LV blood pool, B_{LV}. From B_{LV} 2D binary images and centroids may be obtained for each individual slice position and phase. However, in practice it is expected that a single representative 2D centroid will be used, obtained from the median slice position (targeting the LV mid-slice plane) and the phase with the largest blood pool area (targeting diastole).

3.3.3 Validation and Computational Environment

The STACOM datasets include gold standard endocardium contours for both end-diastole (ED) and end-systole (ES) phases. The median slice position was calculated from B_{LV}, and the evaluated 2D binary object and centroid was assessed in both the ED and ES phases against the corresponding gold standard contours. The Kaggle datasets do not include gold standard contours, only clinical measurements of ED and ES blood volume. Therefore, a manual gold standard was constructed by first obtaining the median slice position and the phase with the largest blood pool area from B_{LV}, then assessing the evaluated 2D binary object and centroid against a manual delineation of the LV endocardium wall.

For validation, the absolute error was calculated as the Euclidean distance between the gold standard 2D centroid and the evaluated centroid result. For context, this was converted to a fractional result normalized by the average radius of the gold standard 2D binary object, i.e. this would be the fractional error as normalized by the average endocardium radius. Success was defined where the fractional error is ≤ 0.5, i.e. the evaluated centroid is within 50% of the endocardium radius, and a 1-tailed *t*-test of the mean fractional error being below 0.5 was performed. Finally, the evaluated and gold standard 2D binary blood pool images was compared using the Dice index. This is defined as $2|X \cap Y|/(|X|+|Y|)$, where $|X|$ and $|Y|$ are the number of elements in each respective set (gold standard and evaluated result) of blood pool pixels.

The algorithm was implemented in MATLAB R2014a (MathWorks, Natick, MA), on a workstation with a 4-core 3.4 GHz CPU and 16 GB RAM.

3.4 Results

A total of 1140 datasets (11868 slices) and 45 datasets (572 slices) were analysed from the Kaggle Data Science Bowl Cardiac Challenge and STACOM 2009 challenge, respectively. The average processing time for a single study was approximately 3.0 s. Table 3.1 depicts the validation results achieved with the proposed method. In both the Kaggle and STACOM 2009 challenge sets, the mean fractional error was lower than 0.5 with a statistical significance of $p < 0.001$.

Table 3.1: Validation results of the proposed method on the blinded test datasets from public challenges (n – sample size, ED – end diastole, ES – end systole)

Test Set	Success rate, %	Absolute error, mm (mean ± std. dev)	Fractional error (mean ± std. dev)	Dice index (mean ± std. dev)
Kaggle (n : 1140)	97.5	2.8 ± 3.7	0.12 ± 0.22	0.88 ± 0.10
STACOM (n : 45)	97.8 (ED) 91.1 (ES)	4.1 ± 2.2 (ED) 4.7 ± 2.6 (ES)	0.14 ± 0.07 (ED) 0.22 ± 0.13 (ES)	0.82 ± 0.10 (ED) 0.67 ± 0.19 (ES)

Correct LV localization was confirmed in 1111 (97.5%) datasets in the Kaggle challenge set, 44 (97.8%) ED and 41 (91.1%) ES datasets in the STACOM 2009 challenge set, resulting in a combined success rate of 97.3% (Figure 3.6). The mean absolute error between the gold standard and localization centroids was 2.8 to 4.7 mm, or 12% to 22% of the average endocardial radius. The error was primarily due to the localization algorithm not accounting for papillary muscles, which is also reflected in the mean Dice index of 0.67 to 0.88.

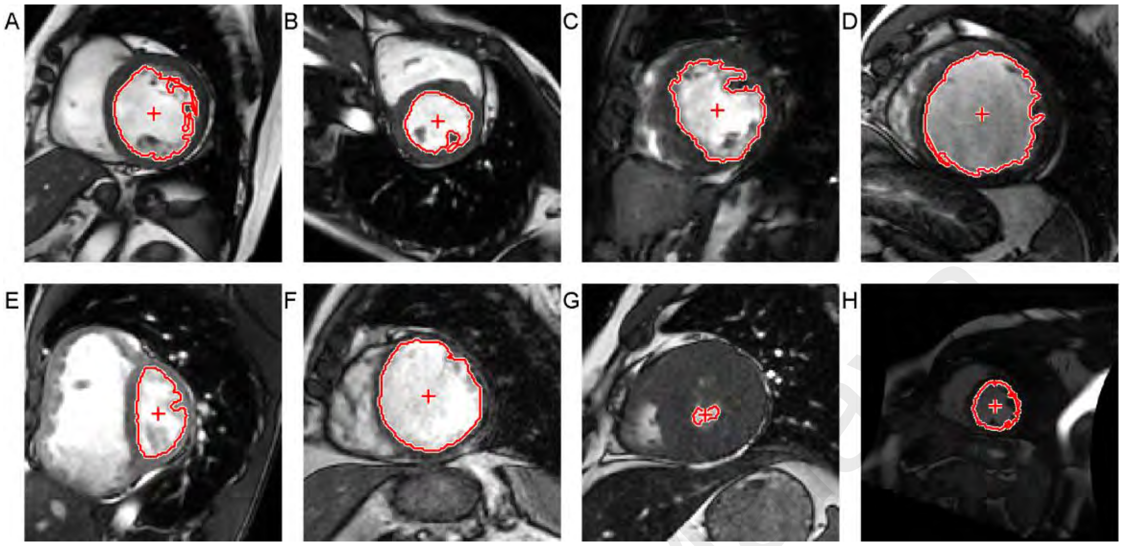


Figure 3.6: A to D: **Sample images from four different datasets demonstrating typical results.** E to H: **Sample images from four different datasets demonstrating robustness of algorithm against suboptimal conditions, namely,** E: **non-circular LV & dilated RV,** F: **thin septum,** G: **thick myocardium / small blood pool,** H: **phase wraparound & high signal intensity caused by close proximity to the RF receiver coil**

The algorithm returned good results even in suboptimal conditions, including cases of non-circular LV, dilated RV, thin septum, thick LV myocardium, strong non-cardiac motion, and various scan artifacts (Figure 3.6). Analysing the failure cases, the most common factors were four cases of the RV being mistakenly localized due to various LV pathologies (Figure 3.7A), and four cases of inadvertent merging of the LV and RV due to indistinct septum boundaries (Figure 3.7B).

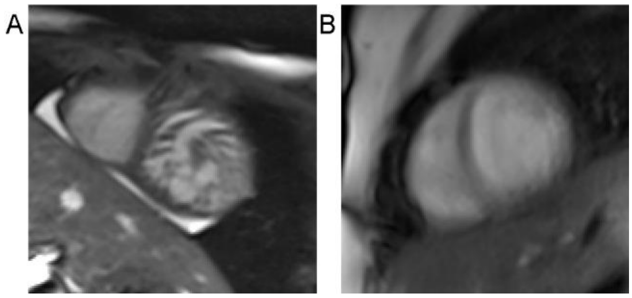


Figure 3.7: Failure cases caused by, A: **LV hypertrabeculation leading to RV being mistakenly localized, and** B: **the LV and RV inadvertently being merged due to severe motion blurring**

3.5 Discussion

In this chapter, an open source, fully automatic algorithm for localizing the LV blood pool centroid in short axis cardiac cine images has been described. The algorithm demonstrated robust performance across multiple publicly-available datasets, and proved to be generalizable across different clinical protocols.

In step 1 of the algorithm, Lin et al.'s technique of obtaining the initial ROI (X. Lin et al., 2006) was utilized, which the author found to be generally robust, but sensitive to motion artifacts and other sources of non-cardiac motion (e.g. intestinal). Lin et al.'s original paper assumes the resulting ROI to only consist of the heart; their subsequent steps relies on the ROI centroid being close to the interventricular septum. However, the author's experience was that this ROI often extended beyond the heart and was best used only as an initial conservative filter for non-cardiac tissue. In step 3 of the algorithm – scoring objects of interest – a novel score reflecting proximity to lung tissue was introduced, on top of contraction and roundness criteria. Removing the lung score would result in a small but significant drop of the combined test success rate from 97.3% to 96.8%. In step 2 of the algorithm – determining an intensity threshold to segment the LV blood pool – the author found that direct application of Otsu's method on the source cine image, as in Jolly (2008), would fail in the presence of high intensity scanning artifacts (e.g. regions of high signal intensity caused by close proximity to the RF receiver coil). Applying Otsu's method directly would result in a significant drop of the combined test success rate from 97.3% to 93.4%. Instead, motion and intensity information were combined to obtain a more reliable threshold.

The proposed algorithm has been shown to be applicable for a wide range of cine MRI images complicated with suboptimal acquisition and clinical conditions, including non-circular LV and dilated RV, thin septum, thick myocardium/small blood pool, phase wraparound and high signal intensity caused by close proximity to the RF receiver coil. The proposed LV localization method is designed to be a precursor algorithm to a further, more complete segmentation system. Although the algorithm attempts to delineate the blood pool as part of the methodology, its direct use for calculating clinical parameters such as ejection fraction (EF) is not recommended, as the emphasis is on reliability of estimation of the LV centroid rather than accurate delineation. A possible future improvement would be the introduction of a goodness metric to flag ambiguous cases and avoid false positives. Several limitations have been noted. In particular, the use of the minimum intensity projection to identify objects of interest makes it possible that the algorithm will miss LV slices at the apex tip, where there is often no consistent presence of the blood pool. The same limitation arises in situations of hyper-contraction, where the derived "core" binary object is very small and risks being broken up or mistaken for noise. The algorithm also has no mechanism for recognizing the LV base, and thus will extend beyond the mitral valve under certain conditions. In addition, accuracy of the algorithm will suffer when individual slices are misaligned due to breathing motion or other motion artefacts, as no inter-slice registration is performed. However, these limitations do not significantly affect the primary utility of the algorithm, i.e. to provide a single robust estimate of the LV blood pool centroid position for segmentation initialization (Kadir et al., 2012).

In conclusion, an open source, fully automatic algorithm for localizing the LV blood pool centroid in short axis cardiac cine images has been described. The algorithm demonstrated excellent performance across 1185 externally published datasets, and is capable of eliminating manual LV centre point selection, thereby enabling the use of follow-on segmentation algorithms.

CHAPTER 4: CONVOLUTIONAL NEURAL NETWORK REGRESSION FOR SHORT-AXIS LEFT VENTRICLE SEGMENTATION IN CARDIAC CINE MR SEQUENCES

4.1 Abstract

This chapter presents and validates a convolutional neural network (CNN)-based method for the semi-automatic segmentation of the left ventricle (LV) in short axis cardiac cine magnetic resonance (MR) images. Manual input is only required for identification of the LV apical and basal slice locations.

In this approach, the complete LV segmentation task (for all short axis slices and phases) is parameterized in terms of the radial distances between the LV centrepoint and the endo- and epicardial contours in polar space. Convolutional neural network *regression* is then utilized to infer these parameters. Utilizing parameter regression, as opposed to conventional per-pixel classification, allows the network to inherently reflect domain-specific physical constraints.

This approach was benchmarked primarily against the publicly-available left ventricle segmentation challenge (LVSC) dataset, which consists of 100 training and 100 validation cardiac MRI cases representing a heterogeneous mix of cardiac pathologies and imaging parameters across multiple centres. This approach attained a 0.77 Jaccard index, which is the highest published overall result in comparison to other automated algorithms. To test general applicability, the algorithm was also evaluated against the Kaggle Second Annual Data Science Bowl, where the evaluation metric was the indirect clinical measures of LV volume rather than direct myocardial contours. This approach attained a Continuous Ranked Probability Score (CRPS) of 0.0124, which would have ranked tenth in the original challenge. With this the effectiveness of convolutional neural

network regression paired with domain-specific features in clinical segmentation has been demonstrated.

This chapter has been published as (Tan et al., 2017) ; the published text is largely reproduced here excepting minor amendments for consistency and flow.

4.2 Introduction

Cardiovascular diseases (CVDs) are the most common cause of death globally; it is estimated that 31% of all global deaths in 2012 were due to CVDs (Low et al., 2014; Mendis, 2014). Cardiac MRI is currently considered the gold standard for the assessment of various aspects of CVDs, particularly congenital heart diseases (Abdul Aziz et al., 2013). Quantification of key cardiac MR images is now recommended as a standard diagnostic procedure by cardiovascular expert groups (Schulz-Menger et al., 2013), including the calculation of LV end-diastolic (ED) and end-systolic (ES) volumes. These quantifications typically require the delineation of LV myocardial borders. Clinically, most physicians restrict delineation to only the ED and ES phases, which requires approximately 20 mins to complete manually (Petitjean & Dacher, 2011). Complete delineation across the entire cardiac cycle would be desirable, but modern 20+ cardiac framerates make this far too tedious and time consuming to be performed manually.

There has been significant research focused on automated segmentation of LV myocardium, particularly for short-axis MR cine images. Techniques to date can generally be categorized as: (1) purely image- or pixel-driven methods such as intensity thresholding or distribution modelling utilizing blood pool to myocardium contrast; (2) methods incorporating statistical or geometric shape methods to model the LV such as truncated ellipsoids; and (3) anatomical atlas-based registration; and combinations of these various approaches (Petitjean & Dacher, 2011; Suinesiaputra et al., 2014; Tavakoli & Amini, 2013).

In typical short-axis steady-state free precession (SSFP) cine MR images, there is strong contrast between the blood pool within the LV and the endocardial wall. However, contrasting the epicardial wall to surrounding structures is more difficult, particularly against low-signal lung tissue. Delineation is most straightforward at the mid-LV level, where the LV is relatively large and strongly circular. Significant partial volume effects arise in apical slices where the LV tapers off, whereas in basal slices the LV branches off to the ascending aorta and left atrium, resulting in ambiguous boundaries. Delineation is also easiest for the ED and ES phases, when the heart is relatively static and the myocardial wall is well defined. Other phases during contraction exhibit significant motion blurring. Because of these difficulties, many published algorithms only solve for a subset of LV segmentation: endocardial wall only, mid-LV planes only, or restricted to the ED or ES phases. Notable automated complete LV segmentation algorithms with apex to base coverage include Lorenzo-Valdés, Sanchez-Ortiz, Elkington, Mohiaddin, & Rueckert (2004) via a 4D probabilistic atlas, Lynch, Ghita, & Whelan (2008) via 3D+time level sets, Cousty et al. (2010) via 4D spatio-temporal watershed graph cuts, Jolly, Guetter, Lu, Xue, & Guehring (2012) via 2D+time intensity distribution modelling and temporal registration, and Queirós et al. (2014) via 3D B-spline active surfaces and temporal optical flow tracking.

In recent years, multi-layer neural networks have been shown to be effective automatic feature extractors for high-dimensional datasets, especially when large quantities of labelled data are available for training (LeCun et al., 2015). Convolutional neural networks (CNN) are a neural network variant modelled to take advantage of data with regular structure, such as the spatial grid matrix of images. CNN approaches have dominated various recent computer vision challenges (Krizhevsky et al., 2012; Szegedy et al., 2014). Initial applications were for whole-image classification, though CNNs are now increasingly being applied to object segmentation. In general, CNN object segmentation is implemented via classification at the pixel level. For example, a CNN used to segment a 64×64 image into five possible classes would have an output matrix size of [64×64]×5. The output values would represent the *per-pixel* probability of belonging to one of the five classes, and individual pixels are assigned to the class with the highest probability value.

Stalidis et al. (2002) is a rare early approach utilizing a three-layer non-convolutional (fully-connected) neural network for per-pixel LV segmentation. The network input consisted of only three values: the individual pixel intensity, angular position around the LV centrepoint, and slice location. More recently, Avendi et al. (2016) combined CNNs and a deformable model to perform LV segmentation, though only for the endocardial wall and only in the ED and ES phases. Three separate networks were trained; consisting of a two-layer CNN for the initial localization of heart region-of-interest, and two separate three-layer fully-connected networks for LV segmentation at basal and mid-ventricular slices, and for apical slices, respectively. Most recently, Tran (2016) used a 15-layer CNN to perform complete left and right ventricular (RV) segmentation. All three approaches referenced here utilized per-pixel classification.

Per-pixel classification enables a high degree of flexibility in object segmentation. There are no restrictions in placement, shape, or quantity of applied labels, save for learned hidden constraints within the trained network. This flexibility is useful for object segmentation in general images, where the problem domain is largely unbounded. For example, categories in the ImageNet general image recognition challenge include animals, plants, geological formations, and man-made objects (Russakovsky et al., 2014). In contrast, medical images have a far more specific problem domain; the complete flexibility of per-pixel classification precludes the ability to enforce physiological constraints. For example, Figure 4 in Tran (2016) depicts LV segmentation failure cases where the CNN placed endo- and epicardial contours wholly separated from each other, a physical impossibility. Other physiological improbabilities include results with more than one myocardial object cluster, or fragmented myocardial volumes.

In this chapter, it is hypothesized that a restricted model is a useful approach to the task of LV segmentation. It is proposed that the task be parameterized by defining the segmentation in terms of radial distances of the LV walls from the centre of the blood pool rather than per-pixel binary image masks, and that the radial distances be computed via neural network *regression* as opposed to per-pixel classification. Specifically, two networks are designed. The first network estimates the 2D LV centrepoint, which is then used to remap the image into polar space. The resulting polar image is input to the second network, which computes the radial distances of the endo- and epicardial wall. Regression based segmentation approaches have been previously proposed, though to date not yet paired with CNNs. Further, most of these algorithms apply regression to the entire shape, typically through an intermediate dimension-reduction procedure such as principal component analysis on the shape parameters (Lay, Birkbeck, Zhang, & Zhou, 2013; Shao, Gao, Wang, Yang, & Shen, 2015; Zhou, 2010). In contrast, the proposed approach parameterizes and regresses LV radial distances point-by-point on a polar coordinate system; this retains the flexibility of the learned network while only incorporating two fundamental constraints: the isocentric relationship of the endo- and epicardium walls, as well as the LV being generally convex in shape.

Finally, it is proposed that there is value in the inclusion of hand-crafted domain-specific features as opposed to the network using hidden learned features from the raw data exclusively. Specifically, the inclusion of a 1st-harmonic Fourier transform (FT-H1) magnitude image is shown to be an efficient way to incorporate information from the entire cardiac cycle and thus improve performance.

4.3 Materials and methods

4.3.1 Dataset

The primary dataset utilized is the left ventricle segmentation challenge (LVSC) dataset (Suinesiaputra et al., 2014), which was made publicly available in 2011 in conjunction with the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2011 conference. The LVSC dataset consists of 200 sets of cardiac MR images of individual patients with coronary artery disease and prior myocardial infarction, sourced from multiple institutions. The primary sequences were short-axis SSFP cine images. Corresponding long-axis SSFP cine images were also available, although only for a subset of the cases. Scanner types and imaging parameters varied between cases, giving a heterogeneous mix of spatial resolutions (0.7 to 2.1 mm/pixel), matrix sizes (156×192 to 512×512), and cardiac phase resolutions (18 to 35 phases).

The LVSC datasets are separated into two groups: 100 fully annotated cases for training and testing, and 100 cases without annotation for validation. The gold standard annotations consist of expert-delineated binary masks representing the LV myocardium, from basal through apical slices for all cardiac phases. The gold standard annotations for the validation set are only available to challenge organizers, and challengers are required to submit their results to the organizers for independent assessment. The long term goal of the LVSC is to establish a consensus-based ground truth for the development, validation and benchmarking of LV segmentation algorithms (Suinesiaputra et al., 2014); the consensus ground truth being continually estimated from the results of participating algorithms, using the simultaneous truth and performance level estimate (STAPLE) method (Warfield, Zou, & Wells, 2004).

The 100 training cases were split into individual 2D+time datasets. The datasets were manually filtered, removing duplicates and omitting data that would mislead the training algorithm. For example, the gold standard was based on an expert-driven 3D surface finite element cardiac model, and this resulted in certain basal slices containing only partial myocardium contours due to the model having a slanted surface. Such slices were omitted from training. The final training set consisted of 866 2D+time datasets, or 20030 individual images. These were randomly divided into two groups: 85% for direct training of the network and 15% for cross-validation and hyper-parameter tuning. An 80:20 training to cross-validation ratio would be more conventional, but in this case a larger training fraction was used given the relatively small size of the database.

As a secondary test, the trained network was evaluated against the Kaggle Data Science Bowl Cardiac Challenge, a public machine learning challenge in 2015 for determining LV volumes from cardiac cine MRI data (Kaggle & Booz Allen Hamilton, 2015). The Kaggle data consists of 700 datasets for training and cross-validation, as well as 440 test datasets for final scoring. Unlike the LVSC dataset, the Kaggle dataset has no gold standard LV contours available; the objective and evaluation metric is solely based on the predicted LV volume at end-diastole (ED) and end-systole (ES).

4.3.2 Neural networks

Neural networks for supervised machine learning are generally implemented as a cascading series of matrix operations. The values in the matrixes are the parameters of the network, often called weights, w and biases, b. These values are initially randomized, and neural network training is carried out repeatedly to adjust these values such that a particular loss function, L, is minimized. Pairs of w and b matrices are commonly grouped as a *layer*, and the result of a particular layer is forwarded to the next layer to be used as input. For example, a basic fully-connected layer assigns data in a 1D vector, and might be implemented as $x_1 = x_0 \cdot w_1 + b_1$, where x_0 is the result from the previous layer. Convolutional layers operate on 2D data with regular structure, and also contain w

and b matrix pairs. The w and b matrices are typically smaller than the input data, and the matrix multiplications and additions are carried out via a sliding window in a manner analogous to the convolution operation in signal and image processing. Further implementation details including activations, pooling, regularization, gradient optimization, and hyperparameters may be found in Goodfellow et al. (2016).

4.3.3 Overview

Applications of neural networks in image segmentation commonly utilize a per-pixel classification model, where the trained network classifies each pixel into two or more classes, e.g. myocardium and background. In this chapter a *linear regression* based model is utilized, where the LV myocardium is parameterized in terms of radial distance from the LV centrepoint. The motivations are that this implicitly enforces useful physiological constraints in the model: that there be only a single connected object; that the blood pool or endocardial wall be a subset of the epicardial wall; that the LV is generally convex in shape; and that the endocardial and epicardial contours share a common centrepoint. This requires two separate neural networks: the first for LV centrepoint localization (CPL) (referred to as the CPL network) (Figure 4.1A-C), and the second for the delineation of the LV myocardial borders (MB) (referred to as the MB network) (Figure 4.1D-F).

In brief, the algorithm was initiated with a gross, automated estimate of the LV centrepoint. The source image was then cropped and fed into the CPL network (Figure 4.1A-B), which generated a more accurate location of the LV centrepoint (Figure 4.1C). The source image was re-cropped based on this centrepoint and mapped into polar space, before being fed into the MB network (Figure 4.1D-E). The final outputs of the MB network were the endo- and epicardial wall radii (Figure 4.1F). Implementation is illustrated in Figure 4.1 and detailed in the following sections.

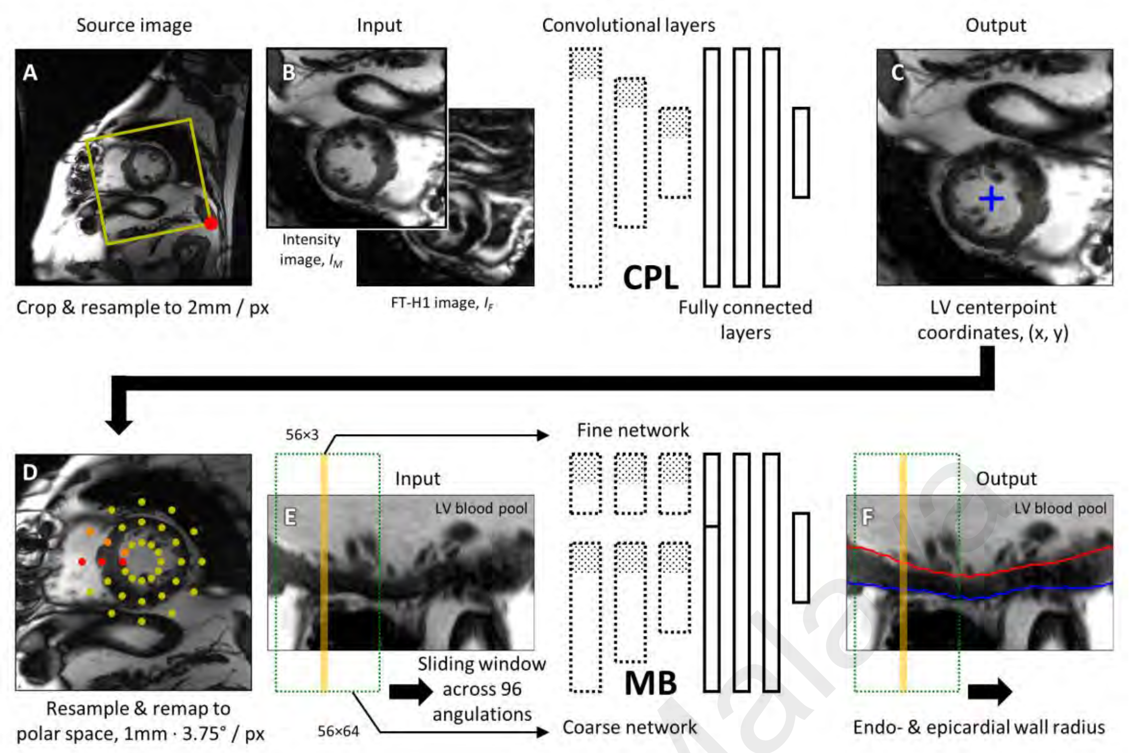


Figure 4.1: Block diagram of dual neural network architecture used for complete LV segmentation. (Upper row) **CPL network for centrepoint localization.** A: **2D source image is resampled and cropped. Yellow outline indicates crop bounding box, and red dot indicates top-left corner of cropped image.** B: **Preprocessed intensity I_M and 1st harmonic Fourier transformation (FT-H1) I_F images fed into CPL network.** C: **CPL network regression outputs two values representing LV centrepoint coordinates, (x, y) at blue crosshair.** (Lower row) **MB network for LV segmentation.** D: **2D source image is resampled, cropped, and remapped to polar space using the LV centrepoint result from the CPL network as origin. Yellow dots indicate polar space sampling coordinates, red and orange dots indicate polar space at theta = 0° and theta = 30°, respectively.** E: **Preprocessed magnitude and FT-H1 (not shown) images. Separate crops are fed into the fine and coarse MB sub-networks via sliding window, respectively.** F: **MB network regression outputs two ρ values (per angulation θ) representing LV endo- (red) and epicardial (blue) wall radius.**

4.3.4 Data preparation and augmentation

Both the CPL and MB networks operate on 2D images. Specifically, the input to both networks consisted of one cine intensity image, I_M from a unique slice and cardiac phase, paired with the corresponding 1st harmonic image, I_F as indicated in Figure 4.1 (B). I_F

was derived from the discrete 1D Fourier Transform of the image, applied on each (x, y) pixel location across the full cardiac cycle using Eq. 4.1.

$$h_{x,y} = \left|\sum_{t=0}^{T-1} v_{x,y,t} \cdot \left(\cos\left(-2\pi \frac{t}{T}\right) + i \sin\left(-2\pi \frac{t}{T}\right)\right)\right| \qquad (4.1)$$

where t = cardiac phase, T = total number of cardiac phases, $v_{x,y,t}$ = pixel value at a particular (x, y) location and phase, $h_{x,y}$ = resulting 1st harmonic magnitude at a particular (x, y) location. Eq. 1 is applied across all (x, y) locations to form I_F. The 1st harmonic was chosen as this correlates most closely to the standard gated single cardiac cycle acquisition. Figure 4.2 indicates the significant improvement in network performance with the inclusion of I_F apart from the standard cine intensity image.

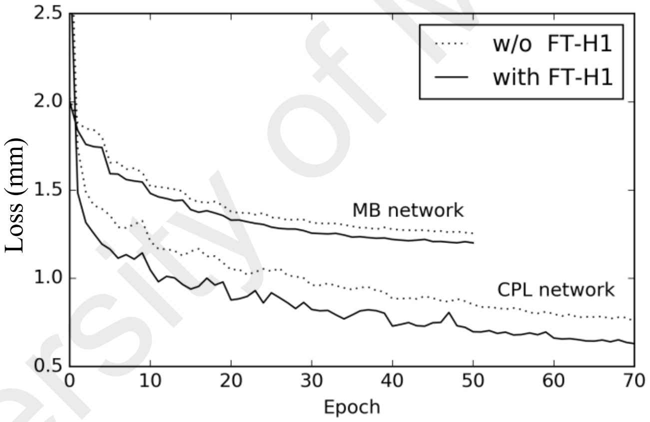


Figure 4.2: The inclusion of the 1st harmonic image (FT-H1) brings a small but significant improvement in performance for both networks. Losses shown are the mean over each epoch.

For the localization of the LV centrepoint, an initial estimate, C_0, was first determined by calculating the unique intersection point between each short axis (SA) image and two long axis (LA) images, either the vertical long axis (two-chamber), four-chamber, or left ventricular outflow tract views. If insufficient LA images were available, C_0 would simply be initialized to the center of the SA image.

For the CPL network, I_M and I_F were resampled to 2 mm/pixel using linear interpolation, and cropped to 84×84 pixels with C_0 in the center, resulting in I_{M_CPL} and I_{F_CPL}, respectively (Figure 4.1A-B). Both I_{M_CPL} and I_{F_CPL} are individually normalized to a mean and standard deviation (SD) of 0 and 1, respectively to remove inter-image contrast variation, before being input to the network. To improve model generalization, during training random augmentation was performed (Goodfellow et al., 2016), which includes displacing the centrepoint up to ±35 mm, scaling the image up to ±15%, performing rotation up to ±180°, distorting the mean and SD of image intensity up to ±0.15, and adding Gaussian noise with 0.15 SD (Figure 4.1A-B). The output of the CPL network is a new and more accurate LV centrepoint, C_1 (Figure 4.1C).

For the MB network, I_M and I_F were resampled to 1 mm/pixel using linear interpolation, then cropped and remapped both images into polar space using C_1 as the origin. These resulting polar images I_{M_MB} and I_{F_MB} have a radius of 56 mm and 96 angular sectors, resulting in a size of 56×96 pixels. Both images were then individually normalized to a mean and SD of 0 and 1, respectively. Unlike the CPL network, I_{M_MB} and I_{F_MB} were not utilized directly. Instead, separate 56×64 and 56×3 windows were taken and fed into the network, representing one particular angular section in polar space (Figure 4.1D-E) using a sliding window over the angle. During training random augmentation were performed, by displacing the centrepoint up to ±50% of the radial distance to the gold standard endocardial wall, scaling up to ±15%, distorting the mean and SD up to ±0.15, and adding Gaussian noise with 0.15 SD (Figure 4.1D-E). The outputs of the MB network are two ρ values representing the endo- and epicardial radius for a particular angular location. The MB network is repeated 96 times for a single image, where the inputs are taken from I_{M_MB} and I_{F_MB} via sliding window across the wrapped and buffered polar images (Figure 4.1F). The resulting radial points (96 endo- and 96 epicardial ρ values) are remapped back to Cartesian space to form the final myocardium contour.

4.3.5 Network architecture and parameters

The convolutional neural network architecture for the CPL network is summarized in Table 4.1, and for the MB network in Table 4.2. All input matrices were 3D because the magnitude (I_{M_CPL} and I_{M_MB}) and 1st harmonic (I_{F_CPL} and I_{F_MB}) images were stacked in the 3rd dimension and treated as a single input matrix. The MB network utilized a dual-resolution design, where two parallel sub-networks operated on a “coarse” and “fine” view of the input images, respectively (Figure 4.1E-F). The fine sub-network has a coverage of 56×3 (radial × angular), located in the centre of the input window. It was expected to provide specific, detailed information for segmentation, and pooling was avoided so as to retain maximum spatial detail. The coarse sub-network has a coverage of 56×64, though a stride of 2 was immediately executed on the angular dimension prior to the 1st layer input (thus reducing it to 56×32). This sub-network was expected to provide overall, contextual information about the surrounding sectors, and is heavily pooled to reduce operational load. The choice of input matrix size (56×64) and (56×3) was a compromise between desired field-of-view and operational efficiency.

Table 4.1: CPL network architecture for LV centrepoint localization. The outputs are the (x, y) coordinates of the LV centrepoint. CN[*]: convolutional layer, FC[*]: fully connected layer

Layer	In	Weights	Pool	Out*
CN1†	84×84×2	5×5×72	2×2	40×40×24
CN2†	40×40×24	5×5×72	2×2	18×18×24
CN3†	18×18×24	5×5×72	2×2	7×7×24
FC1	7×7×24	7×7×768	–	1×1×256
FC2	256	256×768	–	256
FC3	256	256×768	–	256
Out	256	256×2	–	2

† Valid padding was used in all convolutional layers, i.e., the output matrix size is reduced by 4 in both image dimensions after convolution

* Maxout activations (3 maxout units) are used in all layers except for the final linear output. This reduces the output weight / convolutional channels dimension to ⅓ of the original size.

Table 4.2: MB network architecture for LV segmentation. The output is the endocardial and epicardial radiuses or distance ρ from LV centrepoint. There are two parallel sub-networks (coarse and fine) whose outputs are concatenated at the input of layer FC2. CN[*]: convolutional layer, FC[*]: fully connected layer

	Coarse sub-network				**Fine sub-network**			
Layer	**In**	**Weights**	**Pool**	**Out***	**In**	**Weights**	**Pool**	**Out***
CN1	56×32×2	3×3×18	2×2	28×16×6	56×3×2	3×3×54	–	56×3×18
CN2	28×16×6	3×3×18	2×2	14×8×6	56×3×18	3×3×54	–	56×3×18
CN3	14×8×6	3×3×18	2×2	7×4×6	56×3×18	3×3×54	–	56×3×18
FC1	7×4×6	7×4×192	–	1×1×64	56×3×18	56×3×576	–	1×1×192
FC2	64+192	256×768	–	256	← merge			
FC3	256	256×768	–	256				
Out	256	256×2	–	2				

* Maxout activations (3 maxout units) are used in all layers except for the final linear output. This reduces the output weight / convolutional channels dimension to ⅓ of the original size.

Maxout activations (3 maxout units) (Goodfellow, Warde-farley, Mirza, Courville, & Bengio, 2013) were used in all cases due to their higher efficiency in parameter utilization. Dropout (Srivastava et al., 2014) was used for regularization (0.33 dropout probability), but only in the fully connected layers. All pooling operations were max-pool. All hyper-parameters including the overall network designs were determined empirically. The CPL and MB networks consist of around 1.4 million and 2.2 million parameters to be estimated, respectively.

The Adam stochastic optimization algorithm (Kingma & Ba, 2014) was used to minimize a standard mean-squared error loss function calculated from the predicted and gold standard radius values. Stochastic training was performed with a mini-batch size of 24 and a base learning rate of 0.001, which was annealed by half each five epochs. Training was stopped when the cross-validation loss was manually observed to have ceased improving for at least 10 consecutive epochs.

The neural network architecture was designed using the Python v2.7 scientific development environment (Python Software Foundation, Delaware, U.S.) and the TensorFlow r0.8 machine learning framework (Google Inc., California, U.S.). The network was trained and executed on an Intel Core i7-4770 CPU based workstation with NVIDIA GeForce GTX 980 4 GB GPU.

4.3.6 Post-processing

The following operations were carried out on the MB network output: all radii were truncated at a minimum of 0 (i.e., no negative values), and the endocardial radius was truncated at a maximum equal to the corresponding epicardial radius (i.e., endocardial radius ≤ epicardial radius).

In addition, for each contour, the distance between each point and their immediate neighbors was calculated. Points whose neighboring distance exceeded 2 mm were assumed to be errors and removed; this threshold was obtained from analysis of the gold standard training data. The contour was then smoothed and interpolated using a cubic smoothing spline (Barry, 1973).

When applied to the LVSC validation set, this neighboring distance criteria affected 1.5% and 2.6% of the endo- and epicardium points, respectively, and was heavily weighted to the LV apex and base. Approximately 45% and 20% of the affected points were located in the apex and base end slices, respectively, defined here as the lower and upper 10% of the total slice coverage. The overall effect of the post-processing was a shift of 1.5 ± 1.3 mm and 2.0 ± 1.7 mm (mean ± standard deviation) of the endo- and epicardium points, respectively.

4.3.7 Supplementary Training

The Kaggle dataset is an order of magnitude larger than the LVSC dataset, and as such is likely to contain images that the LVSC trained network does not generalize well against. The CPL network that was trained against the LVSC dataset was thus run against the Kaggle training dataset. From this 13 subjects were manually identified, where the LV centrepoint prediction was deemed unacceptable. The LV centrepoints for these 13 datasets were then manually annotated and added to the LVSC dataset pool for re-training. The resulting CPL* network with this supplementary training data was only used exclusively for evaluation against the Kaggle dataset. It was deemed impractical to perform similar supplementary training for the MB network due to the difficulty to produce full sets of 4D endo- and epicardial gold standard delineations.

4.3.8 Validation

The start and end slices reflecting the apex and base locations, respectively were manually identified; other image slices not within the coverage range were omitted from validation. Binary image masks were generated from the myocardial contours and validated against the corresponding gold standards using the Jaccard similarity index, as well as sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV). These evaluations were carried out by an independent party – the LVSC challenge organizers. Of particular note, the specificity and NPV values were calculated based on a region of interest (ROI), which was obtained by dilating the gold standard binary mask by 1.5×. This is because the full image is dominated by non-myocardium tissue, which would heavily bias the specificity and NPV values (Suinesiaputra et al., 2014).

Three gold standard clinical measurements in the LVSC dataset were also available for separate assessment: end-diastolic volume (EDV), end-systolic volume (ESV), and LV mass. The volumes were calculated using the trapezoidal rule, i.e. by summing the relevant binary areas of each slice, and multiplying the result with the mean slice thickness.

The Kaggle challenge utilizes a continuous ranked probability score (CRPS) for evaluation (Kaggle & Booz Allen Hamilton, 2015); this basically involves determining a cumulative distribution function (CDF) for the LV volume as opposed to a single value prediction. Thus, in addition to calculation of the LV volumes as described previously, the Kaggle training dataset was also used to fit a linear regression model against the initial predicted LV volume (ED and ES), the scan length, and the subject age and gender. The CDF for each subject was determined by building a Gaussian distribution with mean and standard deviation directly obtained from the regression model prediction and single-observation confidence bounds.

4.4 Results

In Table 4.3, the results between the proposed convolutional network regression model (CNR) were compared against other published results using the same LVSC CS* consensus validation dataset (Suinesiaputra et al., 2014; Tran, 2016). For most of the evaluation measures including Jaccard index, the CNR model outperforms all other published methods except for AU. It is important to note that the AU method uses manual guide-point modeling to assist the fitting of a finite element cardiac model to the image data, and requires human expert approval of all slices and frames (B. Li, Liu, Occleshaw, Cowan, & Young, 2010).The AU method was used by the challenge organizer to estimate the gold standard global priors in the original 2011 LVSC study.

Table 4.3: Comparison of segmentation performance between the proposed convolutional network regression model (CNR, marked by arrow) and other techniques tested against the LVSC validation set using the CS* consensus. AU (B. Li et al., 2010), AO (Fahmy, Al-Agamy, & Khalifa, 2012), SCR (Jolly et al., 2012), DS, and INR (Margeta, Geremia, Criminisi, & Ayache, 2012) values are taken from Table 2 of (Suinesiaputra et al., 2014). FCN values are taken from Table 3 of (Tran, 2016). Values are provided as mean (standard deviation), and in descending order by Jaccard index.

Method	Manual input	Jaccard Index	Sensitivity	Specificity	PPV	NPV
AU	Interactive 4D guide point placement	.84 (.17)	.89 (.13)	.96 (.06)	.91 (.13)	.95 (.06)
→ CNR	Identify basal and apical slice	.77 (.11)	.88 (.09)	.95 (.04)	.86 (.11)	.96 (.02)
FCN	None	.74 (.13)	.83 (.12)	.96 (.03)	.86 (.10)	.95 (.03)
AO	Delineate first frame	.74 (.16)	.88 (.15)	.91 (.06)	.82 (.12)	.94 (.06)
SCR	None	.69 (.23)	.74 (.23)	.96 (.05)	.87 (.16)	.89 (.09)
DS	Delineate first frame	.64 (.18)	.80 (.17)	.86 (.08)	.74 (.15)	.90 (.08)
INR	None	.43 (.10)	.89 (.17)	.56 (.15)	.50 (.10)	.93 (.09)

The LVSC validation set categorizes individual images into apex, mid, and base levels, in a rough 2:5:3 ratio. The mean ± SD Jaccard index was 0.71 ± 0.13, 0.79 ± 0.09, and 0.77 ± 0.12 for the apex, mid, and base levels, respectively. Regarding image parameters, there was a weak positive correlation between x/y resolution and Jaccard Index (Spearman's rank correlation, $p = .021$, $r = .238$), i.e. higher x/y resolution was correlated with better Jaccard Index. No similar significant correlations were found for slice thickness and number of cardiac phases.

Figure 4.3, Figure 4.4, and Figure 4.5 illustrate good results from both typical and difficult datasets. These include cases with severe image quality issues, as well as cases where the LV shapes were non-circular. Figure 4.6 illustrates a known problem with neural networks, where visually indistinguishable images can return significantly different results (Szegedy et al., 2013). The interconnected complexity of the networks –

particularly at the fully connected layers – makes such issues difficult to troubleshoot. Figure 4.7 depicts erroneous LV contours below the apex, illustrating an issue arising from the 2D nature of the model; the model is unable to utilize information in neighbouring slices to apply 3D physiological constraints. Figure 4.7 also illustrates an issue of the strong dependency of the MB network to the CPL network; in slices close to the apex where the blood pool is intermittent and/or small, poor estimation of the centrepoint directly leads to poor endo- and epicardial contours.

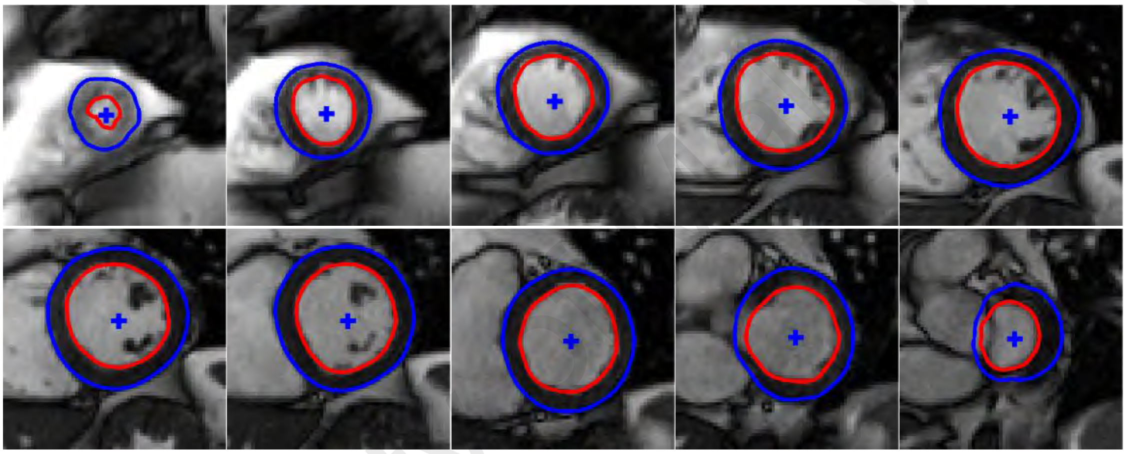


Figure 4.3: (Left to right, top to bottom) **Sample case from the validation set, demonstrating good delineation from apex to base. Blue crosshair represents LV centrepoint from CPL network, red and blue contours represent LV endo- and epicardial contours, respectively, from MB network.** (Bottom row, 2nd and 3rd images from left) **The 2D nature of the model allows the networks to be insensitive to inter-slice misalignment.**

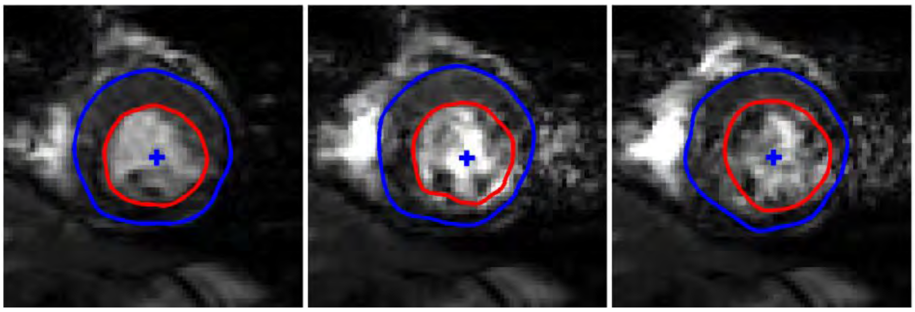


Figure 4.4: (Left to right) **Consecutive phases of a sample case from the validation set, demonstrating reasonable delineation results despite severe image quality issues.**

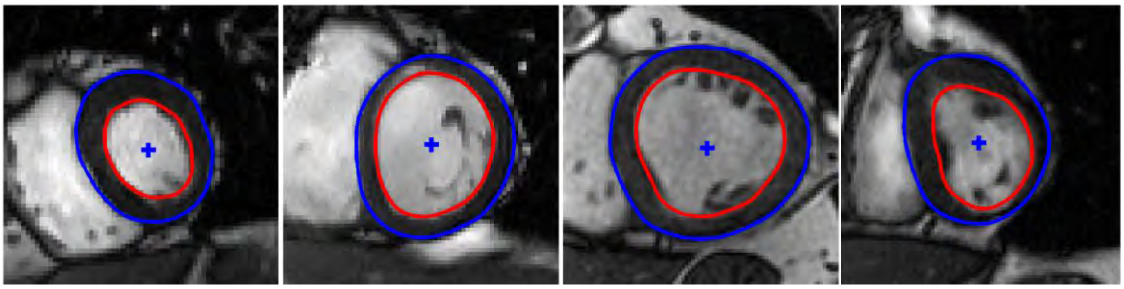


Figure 4.5: Sample cases from the validation set, demonstrating good delineation on LV shapes deviating from absolute circularity.

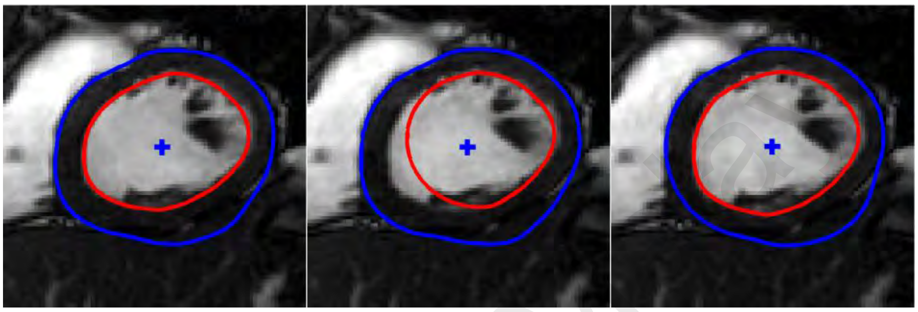


Figure 4.6: (Left to right) **Consecutive phases of a sample case from the validation set. Despite the three images appearing largely similar visually, the MB network nevertheless returned dramatically different endocardial contour results for the middle image. Due to the black box-like nature of neural networks, issues such as this can be hard to troubleshoot.**

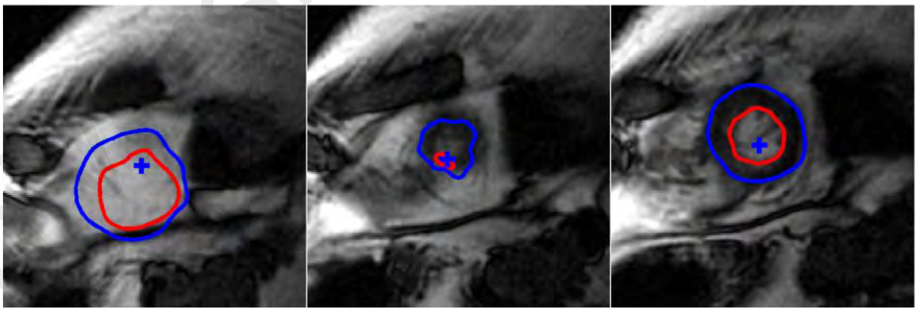


Figure 4.7: (Left to right) **A sample case from the validation set showing neighbouring slices from below the apex towards the mid-cavity. It is physiologically improbable for the LV area to increase dramatically towards the apex, but the 2D nature of the model precludes encoding for such physiological constraints.** (Middle) **The CTR network made a poor estimation of the LV centrepoint, directly affecting the quality of the subsequent endo- and epicardial contour.**

Comparing clinical parameters to the gold standard, mean ± SD of absolute difference and signed difference of ejection fraction (EF) was 4.8% ± 4.5% and +0.3% ± 6.6%, respectively, demonstrating no significant bias towards over- or under-estimation of EF. Large overestimations for EDV, ESV and LV mass were observed, at +46.5 ± 38.0 ml, +29.9 ± 29.7 ml and +49.1 ± 31.0 g, respectively. It is hypothesized that this is primarily due to the elementary method of calculating volumes used, as well as differences in determining LV coverage. The basal and apical slices were manually identified from short axis images exclusively, and included 2.1 additional slices on average compared to the gold standard, which represents around 31% extra LV coverage not accounted for by the gold standard. More accurate volume measurements might be attained via 3D modelling and the use of long axis images (Liew et al., 2015), which is not covered in the current chapter.

For the Kaggle challenge, a mean CRPS of 0.0124 was obtained, which would have placed tenth out of 192 in the original challenge rankings; a relatively strong result considering the MB network was not trained against the Kaggle data. The lack of gold standard contours makes it hard to identify issues in detail; it was observed that some of the weakest scores were obtained from contours which appear reasonable visually. As per comparison of the LVSC clinical parameters, it is hypothesized that the error is primarily due to differences in determining LV coverage. The mean ± SD of absolute and signed difference of EF was 4.4% ± 4.1% and −0.6% ± 6.0%, respectively. The absolute difference for EDV and ESV were 11.8 ± 9.8 ml and 8.7 ± 7.6 ml, respectively.

CPL and MB network training required approximately 4 and 5 hours, respectively. Mean ± SD of the time taken to process an entire 4D volume is 1.9 ± 1.0 s for the CPL network (around 10 ms per image), and 6.2 ± 2.9 s for the MB network (around 30 ms per image). The impact of omitting use of the GPU and only executing the networks on the CPU was also assessed; the resulting execution time taken was approximately 3× longer, or approximately 30 s per 4D volume.

4.5 Discussion

Neural networks in supervised deep learning represents a new generation of data-driven techniques, where the underlying training data provides the bulk of the technique performance rather than hand-crafted features or algorithms. The roots of deep learning lie in early neural network research developed in the 1950s-60s, but practical utilization remained constrained until the development of GPU implementations, which allow for multiple order-of-magnitude increases in training speed (Schmidhuber, 2015). Due to the relatively small seven-layer, three million parameters network used, the author hesitated to label this approach as deep learning, especially when compared with the 20+ layers (Szegedy et al., 2014), 50+ million parameters (Krizhevsky et al., 2012) convolutional behemoths introduced in recent years.

Comparison is restricted primarily to other published approaches that were also validated against the LVSC dataset (Table 4.3). The proposed approach calls for additional manual or automated (X. Lu & Jolly, 2013) identification of the basal and apical slices; however, identification can be carried out post-calculation as the network does not use slice position information, i.e., basal and apical slice identification is not expected to be a bottleneck in automated batch operations.

Of all automated and semi-automated algorithms, the semi-automated AU method (B. Li et al., 2010) has demonstrated the best performance. However, as noted previously, the AU method uses manual guide-point modelling to assist the fitting of a finite element cardiac model to the image data. I.e., its operation requires significant manual input. The AO (Fahmy et al., 2012) and FCN (Tran, 2016) methods present results that are comparable, but inferior, to those presented here. The AO method uses optical flow tracking to propagate a contour across all cardiac phases. However, it requires manual delineation of the first frame for all slices. Special comparison is made to the FCN method, as, like the method presented in this chapter, it is based on CNNs.

The FCN method utilizes a 15 layer fully convolutional neural network comprising 11 million parameters for per-pixel classification. The proposed approach utilizes two seven-layer mixed convolutional and fully connected neural networks comprising three million total parameters for LV parameter regression. Despite being approximately one third of the size, it is notable that the smaller architecture proposed here performed better overall when tested against the LVSC CS* validation set. It is difficult to pinpoint the primary factor in this, as there are multiple differences in both approaches: per-pixel classification vs. myocardial radius regression, single vs. dual (CPL and MB) networks, Cartesian vs. polar space, the inclusion of 1^{st} harmonic Fourier transform (I_F) images, data preparation and augmentation techniques, and rectified linear units (ReLU) vs. maxout activations amongst others. More experimentation will be needed to identify which factors provide the most significant impact, although care should be taken to avoid overfitting to the validation set. Smaller networks provide an additional advantage: the ability to utilize less powerful GPUs, facilitating adoption particularly in lower income settings. The results demonstrate that larger and deeper networks do not necessarily equate to better performance, and that carefully introduced domain-specific models and features can still be valuable as opposed to fully automated hidden feature learning.

For the Kaggle challenge, most of the strongest-scoring entrants used CNN-based per-pixel segmentation ensemble approaches trained with hand-labelled data, though only for endocardium delineation. The supplementary results against the Kaggle challenge demonstrate that the proposed approach is generally applicable against foreign datasets and secondary clinical measures including LV volume. The discrepancy seen in certain cases with visually reasonable contours but poor scores suggests that the primary error lies in the determination of LV coverage. Inclusion of LA volumes would probably improve estimation significantly. Indeed, the winner of the Kaggle challenge utilized CNN per-pixel segmentation on both SA and LA volumes (T. Lee & Liu, 2016).

It has been shown that the 1st harmonic magnitude image provides an efficient way of incorporating information from the entire temporal cycle; similar approaches may be useful in other applications with inherently cyclical data, e.g. respiration. The utilization of polar space allows the MB network to be inherently rotationally invariant. Similar polar space remapping combined with radial parameter regression models, may be suitable for related segmentation problems where the object of interest is circular or convex in shape, e.g. large vessel lumen segmentation.

Some limitations and potential improvements remain. Rudimentary 2D post-processing was utilized; additional filtering across slice depth and cardiac phase could be applied. For example, the erroneous result in Figure 4.7 might have been avoided by building a probabilistic regression function to estimate fractional slice area from apex-base slice position. The regression function could be tuned to reject results with slice areas deviating too far from the norm, although the required inclusion of an additional threshold may impact the generality of this implementation. An extension of this work would be to apply the model to RV segmentation. However, it is suspected that a similar polar space model would be challenging because of the interventricular septum defining the RV being somewhat concave in shape.

In conclusion, in this chapter an automated method for the complete short axis + time segmentation of LV MR cine images was presented. The proposed method is based on convolutional neural network *regression*, and utilizes a two-step approach of LV centrepoint localization followed by determination of the LV endocardial and epicardial radiuses in polar space. The approach was benchmarked against the LVSC dataset, and exhibits the best overall results to date in comparison to other automated algorithms. With this it has been demonstrate that convolutional neural network regression paired with carefully introduced domain-specific models and features is effective for the specific clinical task.

CHAPTER 5: FULLY AUTOMATED SEGMENTATION OF THE LEFT VENTRICLE IN CINE CARDIAC MRI USING NEURAL NETWORK REGRESSION

5.1 Abstract

This chapter presents and validates a convolutional neural network (CNN) regression-based method for the fully automatic segmentation of the left ventricle (LV), utilizing both short axis (SA) and long axis (LA) cardiac cine magnetic resonance (MR) images, with full coverage from apex to base across all cardiac phases.

In this chapter, the use of neural network regression as introduced in Chapter 4 is further extended. Three independent networks were designed and trained for the inference of LV landmarks in LA volumes, as well as LV centrepoints and myocardial contours in SA volumes, respectively. The algorithm was validated against 100 datasets from the publicly-available Left Ventricle Segmentation Challenge (LVSC) database, 440 datasets from the publicly-available Kaggle Second Annual Data Science Bowl database, and 10 subjects from in-house data, representing a heterogeneous mix of scanner types, imaging protocols and parameters.

Tested against the LVSC database, a small but statistically significant improvement in both the Jaccard index (0.77 ± 0.11) and modified Hausdorff distance metrics (1.33 ± 0.71 mm) was obtained, when compared to the previous work in Chapter 4. This represents the best published LVSC performance to date for a fully automated algorithm. Tested against the Kaggle database, the signed difference in evaluated blood volume is +7.2 ± 13.0 mL and −19.8 ± 18.8 mL for the end-systolic (ES) and end-diastolic (ED) phases, respectively. This performance is comparable to published inter-reader variability values from multiple independent expert readers. The in-house data was used to measure scan-rescan reproducibility; there was no significant difference between the algorithm

and a manual quantification for both ES and ED phases ($p > 0.05$). Execution time per case was approximately 12s, including SA and LA volumes. The strong performance overall is suggestive of practical clinical utility.

This chapter has been published as (Tan, McLaughlin, et al., 2018); the published text is largely reproduced here excepting minor amendments for consistency and flow.

5.2 Introduction

Cardiovascular diseases (CVD) are the primary cause of death globally, accounting for approximately 30% of all deaths in 2013 (Roth et al., 2015). Ultrasound is the primary imaging modality for CVD diagnosis due to its portability and low cost. However, cardiac MR is recognized as the reference standard for the assessment of cardiac volumes and regional functions due to its greater accuracy and reproducibility (Abdul Aziz et al., 2013; Gardner, Bingham, Allen, Blatter, & Anderson, 2009). Most standard cardiac MR protocols begin with assessing the LV structure and functions via steady state free precession (SSFP) gated cine imaging due to its high signal-to-noise ratio and excellent contrast between the myocardium and blood pool (Kramer et al., 2013). Standard acquisitions include LA images captured along the plane passing through the LV apex and mitral valve, and a stack of SA images orthogonal to the LA plane, captured between the LV apex and mitral valve.

In standard clinical practice, quantification of LV function is performed via manual delineation of the LV myocardium (endo- and epicardium) within the SA images, for the end-diastole (ED) and end-systole (ES) cardiac phases. This allows the evaluation of standard clinical measurements such as LV ED and ES blood volumes, ejection fraction, and LV mass (Schulz-Menger et al., 2013). Despite delineating only two cardiac phases, such manual tracing can take up to 20 minutes by a radiologist. Full delineation across all cardiac phases would enable useful quantification of motion parameters to identify regional LV dysfunction. However, the excessive effort required for manual full delineation makes it impractical for clinical adoption.

Many LV segmentation algorithms have been published to date, with an extensive review provided by Petitjean & Dacher (2011). Techniques used include general image-processing based methods such as intensity distribution modelling of the LV tissue and blood pool (Jolly et al., 2012, p. 201); deformable models such as active contours targeting the myocardium boundaries (Berbari et al., 2007); statistical shape and appearance models (Mitchell et al., 2001); and anatomical atlas-based registration (X. Zhuang et al., 2010). The attributes of published algorithms vary, ranging from semi-automated (Bricq et al., 2016) to fully automated (Margeta et al., 2012); endocardium only (Avendi et al., 2016) to complete myocardium (B. Li et al., 2010); mid-slices only (Mitchell et al., 2001) to full coverage from apex to base (mitral valve) (Jolly et al., 2012); and dual ED/ES phases only (Avendi et al., 2016) to all cardiac phases (Tran, 2016). Fully automated algorithms are inherently superior in terms of convenience, as well as their elimination of subjective inter-observer variability. However, it is still difficult for these algorithms to provide comparable performance to semi-automated algorithms which utilize human expert input. In addition, most published segmentation algorithms – both fully and semi-automated – are only validated against non-public, single institution datasets (Petitjean & Dacher, 2011). This makes comparison between such published results infeasible due to the differing test conditions: differing quantity of test images, the pathological status (or none) of the patients, as well as the imaging parameters and hardware.

In Chapter 4, CNN regression was introduced for the segmentation of LV myocardium, including full coverage from LV apex to base, across all cardiac phases (Tan et al., 2017). However, the approach was only semi-automatic (i.e. required user input) and was solely trained on SA images; manual intervention was still required to identify the basal and apical SA slices to constrain the volumetric quantification within the LV. In addition, the algorithm did not combine segmentation results across neighbouring phases, i.e. each 2D slice was processed in isolation, leading to inconsistent results in the contour from one phase to the next. In this chapter, the aim is to develop and validate a fully automated algorithm for segmentation of the LV in cardiac MRI. The work in this chapter improves over Chapter 4 through: (i) redesigning the network architecture to incorporate LA images for localizing apex and base landmarks, which enables full automation of the SA segmentation task, (ii) utilizing LA landmarks to stabilize SA segmentation in challenging apical and basal slices by restricting the input SA field-of-view (FOV), and (iii) implementing 2D+time post processing to improve segmentation consistency.

5.3 Material and Methods

5.3.1 Data

Data was utilized from three sources – two sources for the training and primary assessment of the algorithm, and the third source solely for the assessment of scan-rescan reproducibility. The first was data from the 2011 LVSC – a segmentation competition initiated during the 2011 Statistical Atlases and Computational Modelling of the Heart (STACOM) workshop (Suinesiaputra et al., 2014). The LVSC database consists of 200 publicly accessible cardiac MR cine cases; 100 with ground truth myocardial contours in all slices and phases for training and cross-validation, and 100 unlabelled cases for test validation. The database includes SA and LA volumes. The LA volumes are typically a mix of LA two-chamber (LA2C), LA four-chamber (LA4C), and LV outflow tract (LVOT) single slice 2D+time views. The subjects were comprised of patients with coronary artery diseases and regional wall motion abnormalities due to prior myocardial infarction. Patient characteristics were 76.8%/23.2% male/female, with mean age

62.7 years (range 34–84 years). By convention, the ground truth contours included trabeculae and papillary muscles in the blood pool, excluding them from the LV mass.

The second data source was the Kaggle Second Annual Data Science Bowl – a competition held in 2016 to evaluate ES and ED blood volumes (Kaggle & Booz Allen Hamilton, 2015). The Kaggle database consists of 1140 publicly accessible cardiac MR cine cases; 700 with ground truth ES and ED blood volume measurements for training and cross-validation, and 440 unlabelled cases for test validation. The database includes SA and LA volumes for most subjects, the LA volumes having a mix of two-chamber and four-chamber single slice 2D+time views (<1% of subjects also included LVOT views, but these were omitted for ease of processing). The subjects were comprised of a mix of patients with both normal and abnormal cardiac functions. Patient characteristics were 58.8%/41.2% male/female, with mean age 42.1 years (range 2 weeks – 88 years). The Kaggle ground truth only contains clinical blood volume measurements, i.e., no myocardial contours, centrepoints, or LA landmarks were available. To address this limitation, the training dataset was manually assessed and 104 LA views were manually labelled for additional LA landmark localization training, and 15 cases manually labelled for additional SA centrepoint localization training.

The combined LVSC training data and the manually labelled Kaggle data was initially split 85:15 by subject for training and cross validation, respectively, during hyperparameter optimization. The entire training and cross-validation dataset was then combined for the final training run after all hyperparameters had been finalized. There are a total of 26,069 and 9,860 individual images for SA and LA training, respectively (Table 5.1).

Table 5.1: Datasets used for training, cross-validation, and test validation. Descriptions of the three networks (LM, CTR, MB) are in the "Segmentation System Overview" section

		Training & cross-validation		Test validation	
Networks		**LVSC**	**Kaggle**	**LVSC**	**Kaggle**
LM network *	− LA2C	98 (2,275)	52 (1,560)	98 (2,310)	434 (13,020)
	− LA4C	99 (2,295)	52 (1,560)	98 (2,315)	429 (12,870)
	− LVOT	93 (2,170)	None	93 (2,210)	None
CTR network †		100 (22,259)	15 (3,810)	100 (28,115)	440 (136,620)
MB network †		100 (22,259)	None	100 (28,115)	440 (136,620)

* Values shown are: number of 2D+time LA views (number of images);
LA2C – two chamber, LA4C – four chamber, LVOT – outflow tract
† Values shown are: number of 3D+time SA volumes (number of images)

The third source of data consists of in-house scan-rescan data from a previous study on LV motion correction (Liew et al., 2015). Ten healthy subjects were recruited and scanned three times during the same session, and their respective ES and ED blood volumes quantified via a manually delineated, motion-corrected 3D surface model. All subjects had SA and LA scans, the LA volumes being two-chamber, four-chamber, and LVOT single slice 2D+time views. Subject characteristics were 30%/70% male/female, with mean age 48.4 years (range 39−61 years)

Patient information from the public LVSC and Kaggle databases were anonymized by their respective providers; data usage agreements were obtained for use in this study. This study also received institutional review board approval (ref: 989.75) for the reproducibility analysis.

5.3.2 MR protocol

All MR images from the three databases were acquired using a gated SSFP pulse sequence. The LVSC and Kaggle databases were sourced from a variety of imaging centres and scanner types, leading to a heterogeneous mix of imaging protocols and parameters. For the LVSC database, the range (mode) of imaging parameters were: echo time [TE] 0.96 – 2.98 (1.13) ms, repetition time [TR] 2.50 – 79.1 (59.2) ms, flip angle 25°

– 90° (45°), in-plane (x/y) resolution 0.68 – 2.14 (1.56) mm, slice thickness 6 – 10 (8) mm, number of slices 5 – 17 (11), number of cardiac phases 18 – 35 (25). For the Kaggle database, the range (mode) of imaging parameters were: echo time [TE] 1.04 – 1.54 (1.19) ms, repetition time [TR] 14 – 54.7 (34.2) ms, flip angle 35° – 79° (50°), in-plane (x/y) resolution 0.59 – 1.95 (1.41) mm, slice thickness 5 – 11 (8) mm, number of slices 2 – 21 (10), number of cardiac phases 25 – 30 (30). The in-house scan-rescan data was acquired on a single 1.5T MRI system (Signa HDxt 1.5T, GE Healthcare, WI). The imaging parameters were: echo time [TE] 1.6 ms, repetition time [TR] 3.7 ms, flip angle 55°, in-plane (x/y) resolution 1.37 mm, slice thickness 8 mm, number of slices 10 – 15, number of cardiac phases 20.

5.3.3 Automated segmentation

5.3.3.1 Neural networks

Artificial neural networks are a family of mathematical functions with numerous recent successes in tackling artificial intelligence problems, including image processing and recognition (LeCun et al., 2015). Though originated in the 1960s, neural networks have seen a strong resurgence in recent years, thanks largely to improvements in computing hardware and the availability of large quantities of training data (Goodfellow et al., 2016).

Neural networks can be understood as a chain of linear operations interspersed with various nonlinear *activation* functions. Each group in the chain is more commonly known as a *layer*, which consists of a matrix of weights, W, and a vector of biases, b. For each individual layer the input vector is multiplied and summed against W and b, respectively. An element-wise nonlinear *activation* function (e.g. a hyperbolic tangent function) is then applied and the resulting output is used as the input to the subsequent layer and the general series of operation is repeated in further layers.

Traditional neural networks are also known as fully connected networks (FCON), and are typically used with unstructured vector input. For inputs with regular structure (e.g. a 2D image), CNNs are a more suitable variant. Here, W and b are applied repeatedly in a sliding window fashion analogous to the standard convolution operation in signal processing.

The W and b values of all layers are referred to as the network parameters. Starting from a random initialization, the parameters are iteratively updated by calculating a loss function (e.g. mean squared error) and back-propagating the result via an optimization function such as gradient descent, until convergence. Further information may be found in Goodfellow et al. (2016).

5.3.3.2 Segmentation System Overview

The proposed system primarily operates on 2D intensity images, as well as 2D first harmonic magnitude images, $H1_{mag}$, obtained by applying a 1D Fourier transform across the temporal dimension of a 2D+time slice. In Chapter 4, $H1_{mag}$ was found to efficiently incorporate temporal information for SA volumes (Tan et al., 2017).

Three separate neural networks were trained (Figure 5.1): (i) LV Landmarks (LM) network, where the LV base plane (mid of mitral valve) and the LV apex tip were localized in LA images. (ii) Centrepoint (CTR) network, where the LV centrepoint was localized in SA images, (iii) Myocardial Boundaries (MB) network, where the myocardial boundaries were delineated in SA images. In all three networks, the inferred result is obtained through neural network regression. This is particularly notable for the MB network, where the myocardial boundaries are delineated as individual radial points inferred from a polar transform of the input image centered on the LV centroid, as opposed to the more common technique of myocardial segmentation by per-pixel classification. In Chapter 4, this regression technique was found to be superior to other state-of-the-art per-pixel classification networks; it implicitly enforces useful physiological constraints in the model, such as there being only a single connected object, and that the endo- and epicardium contours share a common centrepoint (Tan et al., 2017).

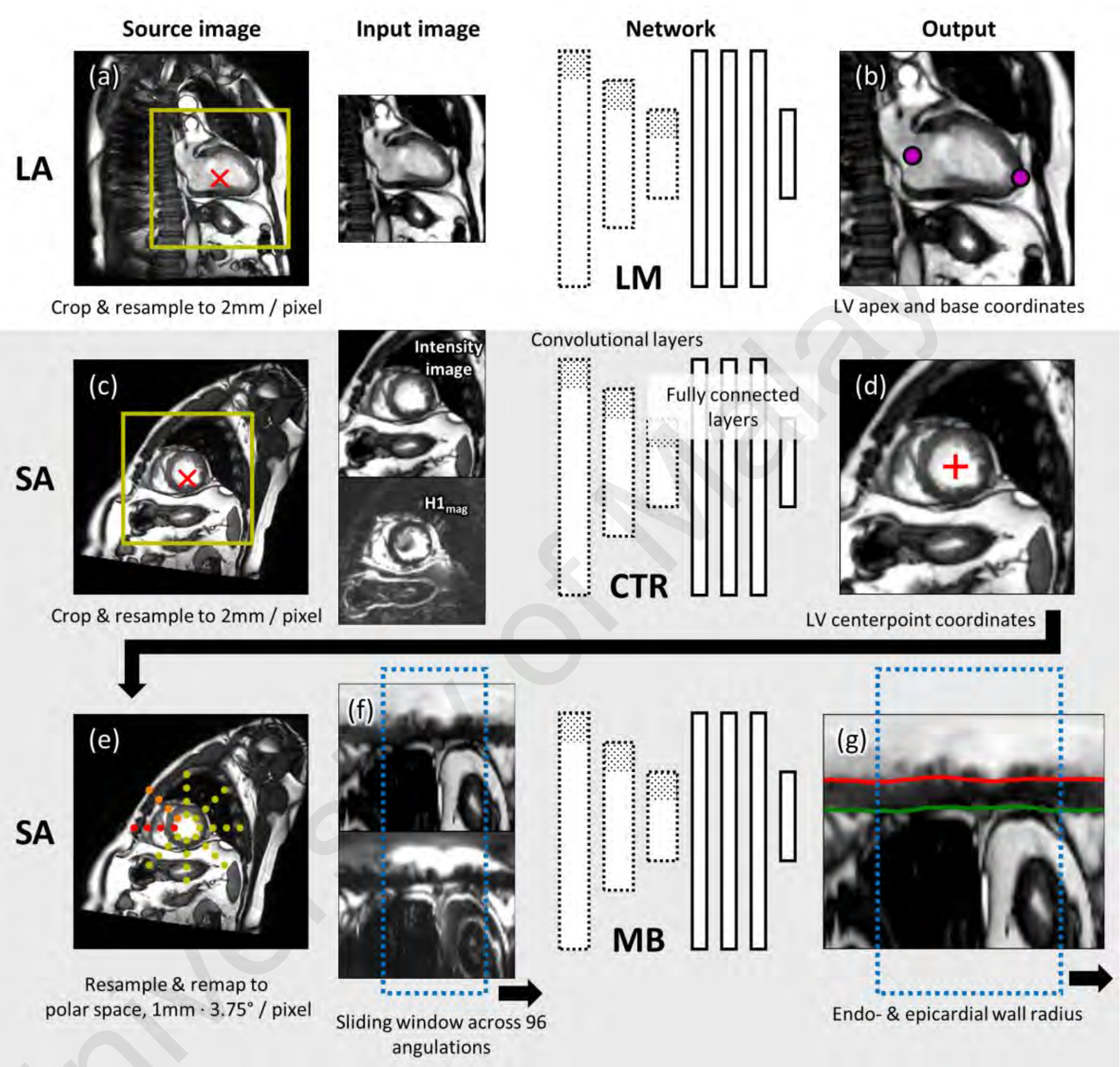


Figure 5.1: Overview of the segmentation system. Each row represents one of the three independent networks (LM, CTR, and MB). Columns illustrate the left-to-right sequential flow of the system: initial source image → pre-processed input images → neural network inference → evaluated output. The network imagery in the third column is representative for the purpose of illustration, refer to Table 5.2 for the detailed network architecture.

For both SA and LA images, an initial estimate of the location of the LV centrepoint, C_0, was first determined by calculating the intersection point between the SA and LA images.

If insufficient LA images were available, C_0 was initialized at the centre of each image instead.

In the LM network, individual LA intensity images were first resampled to a standard resolution of 2mm/pixel, with intensity values normalized to zero mean and a standard deviation (SD) of 1, and a 96×96 pixel crop was performed centred on C_0 (Figure 5.1a – red x-mark) to include the entire LV. The cropped images were input to the LM network, which outputs four values via regression: the evaluated (x, y) coordinates of the LV base and LV apex tip (Figure 5.1b – magenta dots). These coordinates were used to determine LV longitudinal coverage within the stack of corresponding SA images. This is primarily used to identify which SA slices to be included when evaluating clinical measurements such as ES and ED blood volume, as well as to identify basal and apical SA slices for additional processing, as described in the Pre- and Post-processing section.

For the input of the CTR network, the same process of resampling (2mm/pixel), intensity-normalization (zero mean, SD of 1) and cropping (96x96 pixels) was applied to both the SA intensity and $H1_{mag}$ images with centre at C_0 (Figure 5.1c – red x-mark). Given this input, the CTR network outputs two values via regression: the evaluated (x, y) coordinates of the LV centrepoint, C_1 (Figure 5.1d – red crosshair). These coordinates were used for the polar remapping of the SA images for input into the MB network.

For the input of the MB network, individual SA intensity and $H1_{mag}$ images were first scaled to 1mm/pixel, normalized to zero mean and SD of 1, and the images were remapped to polar coordinate space centred on C_1, with radius 80 pixels and 96 angular sections (Figure 5.1e – circular dot pattern represents the polar coordinate space and bounds, red dots indicate zero θ position and orange dots indicate positive angular direction). The finer resampling was chosen to improve accuracy of delineation of the myocardial boundary. The resulting remapped images were 80×96 in size with cyclical buffering on both ends of the angular dimension. From this, individual 80×64 crops were taken along the angular dimension and input to the MB network, in the manner of a sliding window operation with unit step size (Figure 5.1f & 1g – dotted blue box indicates the size of sliding window). Each individual pass outputs two values via regression: the evaluated radius of the endo- and epicardial wall. Thus, for a single SA slice, the MB network was evaluated 96 times, resulting in 96 endo- and epicardial radius values (Figure 5.1g – red and green lines, respectively). Finally, these values were remapped back to Cartesian coordinate space to form the myocardial boundaries.

5.3.3.3 Real-time random augmentation

Where there are small numbers of training datasets, random augmentation has been shown to improve generalization of the network by artificially increasing the number of training datasets (Goodfellow et al., 2016). Specifically, this involves distorting the original training input data and target output result to create a new training sample, e.g. by displacing the image slightly and calculating the new corresponding target centrepoint. Random augmentation is only performed during network training.

For the LM and CTR networks, input images were augmented by random rotation (by ±180°), flipping, displacement (by 35 mm) and scaling (by ±15%). For the MB network, input images were randomly rotated by ±180°, flipped, centrepoint perturbed up to 75% of the minimum endocardial radius, and scaled by ±15%. In addition, for the CTR and MB networks, 10% of the time a circular crop mask was applied to mask out arbitrary non-LV portions of the image, emulating a post-processing field-of-view (FOV) reduction operation. Finally, for all networks random Gaussian noise (0 mean, 0.15 SD) was added. The aforementioned random FOV reduction was implemented to train the CTR and MB networks to handle the associated FOV reduction task during inference, as described later in the Pre- and Post-processing – CTR and MB network section

(Section 5.3.3.6). Interestingly, during training a small improvement in cross-validation loss with the addition of the random FOV reduction was observed, even though the cross-validation inference did not apply any corresponding FOV reduction (Figure 5.2). It is hypothesized that the random FOV reduction influenced the network to de-emphasize non-LV information.

Real-time random augmentation was performed as opposed to pre-generating a fixed number of augmented samples, i.e., the random distortions were generated and applied continuously during training.

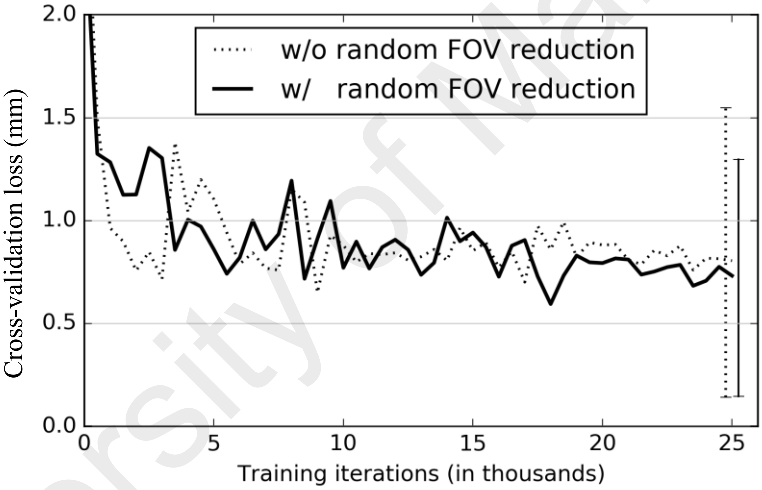


Figure 5.2: A small improvement in cross-validation loss is seen with the addition of random FOV reduction when training the CTR network. Dotted and solid lines are the absolute error averaged over 500 iterations. Each error bar indicates the 10th and 90th percentiles.

5.3.3.4 Network architecture

In Chapter 4, two networks (CPL/CTR and MB) with differing architecture were utilized (Tan et al., 2017). In particular, the earlier MB network utilized a specific "coarse and fine" dual sub-network design with varying input windows; the total number of parameters for both networks was around three million. In this chapter, the network architectures were simplified significantly and made consistent across all three networks, while maintaining the total number of parameters at three million despite the inclusion of an additional new network (LM) for processing the long axis views.

All three networks (LM, CTR, MB) now use a single architecture: 8 CNN layers + 4 FCON layers, including the final output layer. Each network consists of approximately 1 million parameters. All intermediate layers had parameter quantities of comparable orders of magnitude (right column of Table 5.2).

Table 5.2: Basic architecture of all three networks (LM, CTR, MB)

Layer	Size	Parameters (000's)
1. CNN *	5×5×64	3
2. CNN	5×5×64	102
3. CNN *	5×5×64	102
4. CNN	5×5×64	102
5. CNN *	3×3×96	55
6. CNN	3×3×96	83
7. CNN *	3×3×96	83
8. CNN	3×3×96	83
9. CNN → FCON †	6×6×960	95
10. FCON	320	103
11. FCON	320	103
12. FCON → OUT	4 (LM) or 2 (CTR, MB)	0.6

* x = 2, y = 2 stride applied on input data

† separable (separate depthwise and pointwise convolution) CNN. Maxout activation with 3 units used.

Exponential linear units (ELU) (Clevert, Unterthiner, & Hochreiter, 2016) were used as activation functions for all layers, except the final output layer (no activation function used) and layer nine (maxout activation (Goodfellow et al., 2013) with three units used). Pooling was not used; striding of size two produced similar accuracy at lower operational cost. Standard CNNs were used in all convolutional layers except layer nine, where a separable convolution was used instead; this is an operation where the spatial convolution (depth-wise) is performed independently from the channel convolution (pointwise). The use of maxout activation and separable convolution in layer nine was primarily motivated by the need to control the number of parameters in this layer to be of comparable magnitude to the other layers.

Although the quantity of training data here is relatively small, the author opted not to use any regularization techniques such as dropout, as there was little benefit in the observed cross-validation loss. The real-time augmentation used during training appears to provide sufficient regularization for the task.

The Adam stochastic optimizer (Kingma & Ba, 2014) was used to minimize a standard mean squared error loss function, using a mini-batch size of 64. The initial learning rate was set to 0.001, and annealed by half every ten thousand training runs. The network was designed using the TensorFlow r1.0 machine learning framework (Google Inc., California, U.S.), and executed on a 3.4GHz Intel processor based workstation with a single NVIDIA GTX980 graphics processing unit (GPU). Each network took approximately two hours to complete training. During inference, complete execution of all three networks took approximately 12s per study, including SA and LA volumes.

5.3.3.5 Adjustment for paediatric cases

There were at least 99 subjects in the Kaggle database with age below 12 years, including 17 subjects below one year old. The median spatial resolution for subjects ≤12 years and >12 years of age were 0.7 and 1.4 mm, respectively. The standard 2 mm (LM & CTR) and 1 mm (MB) rescaling used during data preparation would result in a loss of spatial resolution for images captured at finer resolutions. Despite being acceptable for adult-sized hearts, such loss is detrimental for paediatric subjects, particularly infants.

Acquisition FOV was utilized as a surrogate measurement for heart size, calculated as $\sqrt{m_x \times m_y} \times \sqrt{s_x \times s_y}$, where $m_{x,y}$ is the matrix size, and $s_{x,y}$ is the pixel spacing. From analysis of the Kaggle database, a median reference FOV value of 310mm, and a "small heart” threshold of 250mm was determined empirically. For datasets with FOV metric below the threshold, the image is scaled up via bilinear interpolation to match the reference metric. E.g., a dataset with a 200mm FOV metric would be scaled up by 1.55× during data preparation.

5.3.3.6 Pre- and Post-processing

(a) LM network

If a single subject had multiple LA volumes, the multiple evaluated landmark points were consolidated via the arithmetic mean. If an individual set of points were separated from their counterparts by >15 mm Euclidean distance, they were assumed to be errors and discarded. If only two LA volumes were available but their landmark points disagreed by >15 mm, preference was applied in this order: LA2C → LVOT → LA4C. This order was chosen as the LA2C acquisition is the most straightforward longitudinal view for LV landmark localization, whereas the LVOT and LA4C acquisitions are more complicated to process as they may have the aorta and the right ventricle, respectively, in view.

A single set of (x, y) coordinates was obtained for each LA landmark by calculating the median across the cardiac phase. This provides a representative position of the LV apex tip and base (mid of mitral valve). These coordinates were then mapped to their corresponding positions in each SA slice to estimate z-dimension proximity to the LV apex and base position, the z-dimension here being the perpendicular dimension with respect to the SA plane. At least 20% LV coverage was expected (i.e., at least 20% of SA slices classified as between apex and base positions). If not, the points were assumed invalid and discarded. In all cases, the threshold constants were determined from analysis of the LVSC training data.

(b) CTR and MB network

Using results from the LM network, each SA slice was categorized as apex, mid-level, or basal, using the criteria <0.2, 0.2 – 0.9, and >0.9 fractional z-position, respectively. The mid-level slices were assumed to produce more reliable results and were processed through the CTR and MB networks as-is. From these results, the various LV mid-level centrepoints were obtained, and used to calculate the global 95th percentile of the epicardial radius, r_{95}, to be subsequently used as contextual information for processing the apical and basal slices.

In the CTR network: for apical and basal slices, each initial centrepoint estimate C_0 (the intersection between SA and LA images) is replaced by its evaluated C_1 counterpart (the output of the CTR network) from the neighbouring medial slice, with the assumption that the neighbouring medial C_1 is a better starting estimate of the current centrepoint under evaluation, particularly for apical slices. A FOV reduction (i.e., a circular crop mask) was also applied using r_{95} (the mid-level 95th percentile epicardial radius) as the base value, with 1.2× for basal slices, and 0.5× – 1× for apex-tip to apex-mid slices. The LV is especially small relative to the full image when close to the apex, potentially causing the network to be confused by other high intensity objects. This FOV reduction can be thought of as a conservative crop to exclude non-LV objects, which forces the network to only consider image data within the reduced FOV.

For the MB network, directly applying a similar reduced FOV tended to reduce the segmentation accuracy of good quality images due to the elimination of surrounding contextual data. Instead, a two-pass run is performed: in the first pass the uncropped image is processed. A second-pass with reduced FOV is applied for images with >15% outlier points. Outlier points were determined by the filtering and smoothing process described below.

For post-processing, CTR and MB results were filtered and smoothed using a periodic cubic spline filter. MB results were smoothed across spatial and temporal dimensions (i.e. 2D+time contour smoothing), while CTR results were smoothed temporally only. For the CTR network: outliers were identified by analysing point-to-point Euclidean distances between neighbouring phases; points >5.9 mm distance (which correspond to 99.99 percentile of LVSC training data) were filtered out. For the MB network: outliers for each slice and time frame were identified by analysing the point-to-point Euclidean distance between neighbouring radial points; the threshold was determined using a standard sigmoid function (Eq. 5.1), with the median radial distance as input, x. Points exceeding the threshold were filtered out. In all cases, the threshold, categorization, and multiplier constants were empirically determined from analysis of the LVSC training data.

$$d/\Big(1+\exp\Big(-\big(b\times(x+a)\big)\Big)\Big)+c \tag{5.1}$$

5.3.4 Validation and Testing

For the LVSC database, the validation ground truth was based on a merged, consensus dataset (identified as CS*) built from multiple automatic and semi-automatic raters (Suinesiaputra et al., 2014). The results were benchmarked to the LVSC ground truth in two ways: on an averaged point-by-point basis via the modified Hausdorff distance (MHD) (Dubuisson & Jain, 1994); and in the form of binary myocardium images via the Jaccard index and Dice index. In addition, the images were subdivided to apical, mid-level, and basal slice locations, and their MHD metrics analysed separately (the Dice and Jaccard index can be unreliable metrics for apical slices due to the small size of the LV binary image).

For the Kaggle database, only the ground truth ED and ES blood volumes are provided. Blood volume was calculated via trapezoidal rule integration across the identified LV slices, adjusting the result to compensate for LV slice coverage, i.e. in cases where the

SA slices did not cover the full extent of the LV, the integrated volume result was adjusted to compensate via a truncated ellipsoid function simulating the generic shape of the endocardium. The Kaggle challenge utilizes a continuous ranked probability score (CRPS) for evaluation, which necessitates building a cumulative distribution function (CDF) for the LV volume as opposed to a single value prediction (Kaggle & Booz Allen Hamilton, 2015). A linear regression model was fit against the Kaggle training set, with the predicted LV volumes and subject age and gender as regressors. The CDF was built as a Gaussian distribution with mean and standard deviation obtained from the regression model. For the LVSC and Kaggle evaluations, a paired *t*-test was performed comparing Chapter 4 (Tan et al., 2017) to the current results.

For the scan-rescan reproducibility experiment using in-house datasets from 10 healthy subjects, the ED and ES blood volumes were calculated in a similar manner to the Kaggle evaluation, then the standard deviation (SD) of the volumes across the three scans for each subject were calculated, and the overall mean SD across subjects obtained. A paired *t*-test was utilized to compare the automated and manual results.

By design, the CTR and MB networks are tightly coupled. The results from the CTR network directly feed into the MB network; whereby the CTR network infers the LV centrepoints, and the MB network infers the myocardium as radial distances from the inferred centrepoint to the myocardium boundary. To test the independent errors of both networks, gold standard LV centrepoints were generated from the LVSC CS* reference binary images. For the CTR network, the independent error was calculated as the Euclidean distance between the evaluated C_1 result and the gold standard centrepoints. For context, this was also converted to a fractional result normalized by the average radius of the endocardium (in images with no blood pool, the epicardium radius was used instead). For the MB network, the LVSC validation test and metrics were repeated, but centred on the gold standard centrepoints (as opposed to the C_1 centrepoints).

Unfortunately, the LVSC CS* dataset does not include reference delineations for the LA volumes. Thus a similar independent error analysis for the LM network was unable to be carried out.

Finally, the added effect of the pre- and post-processing was tested. The LVSC validation test and metrics were repeated, but with all pre- and post-processing disabled for both the Chapter 4 (Tan et al., 2017) and current architectures. Specifically, the special processing for apical and basal slices (FOV reduction) was disabled, as well as all filtering and smoothing functions as described in the Pre- and Post-processing section. The results reflect the raw, unfiltered output from the three networks.

5.4 Results

Evaluated against the LVSC database, there is a small but statistically significant improvement in all metrics when compared to the earlier, semi-automated architecture in Chapter 4 (Tan et al., 2017) (Table 5.3). Notably, there is an approximately 2% improvement in the MHD despite the change to a fully-automated algorithm. To place this metric in perspective, over 95% of slices have MHD values ≤2× the in-plane resolution (i.e. ≤2 pixels). To the best of the author's knowledge, this is the highest overall performance to date for a fully-automated algorithm tested against the LVSC database (Table 5.4). Figure 5.3 illustrates results of a representative case from apex to base, diastole to systole.

Table 5.3: Comparison of results between the previous semi-automated algorithm of (Tan et al., 2017), and the fully-automated algorithm presented here. ES = end-systole, ED = end-diastole.

	Semi-automated (Chapter 4)	**Fully-automated (this work)**	**Difference ‡**
LVSC database			
Jaccard index (JI) *	0.765 ± 0.111	0.769 ± 0.109	0.003 ± 0.053
Dice index (DI) *	0.862 ± 0.083	0.864 ± 0.080	0.003 ± 0.041
Modified Hausdorff distance (MHD) (mm) *	1.355 ± 0.718	1.329 ± 0.710	0.026 ± 0.495
Apical	1.720 ± 1.002	1.769 ± 1.126	-0.049 ± 1.103
Mid-level *	1.250 ± 0.557	1.212 ± 0.493	0.037 ± 0.287
Basal *	1.963 ± 1.192	1.913 ± 1.168	0.050 ± 0.352
Std. deviation of JI between phases †	0.051	0.049	0.002
Std. deviation of DI between phases *	0.035	0.034	0.002
Std. deviation of MHD between phases (mm) †	0.352	0.329	0.023
Kaggle database			
ES blood volume (absolute diff.) (mL) *	9.9 ± 9.1	11.4 ± 9.5	−1.5 ± 9.1
ES blood volume (signed diff.) (mL) *	+1.9 ± 13.4	+7.2 ± 13.0	−5.3 ± 9.8
ED blood volume (absolute diff.) (mL) *	26.8 ± 16.3	21.7 ± 16.5	5.1 ± 13.2
ED blood volume (signed diff.) (mL) *	−25.0 ± 19.0	−19.8 ± 18.8	5.2 ± 15.1
Continuous ranked probability score	0.0124	0.0122	0.0002

* Statistically significant difference at $p<0.001$ using paired *t*-test

† Statistically significant difference at $p<0.01$ using paired *t*-test

‡ Direction of comparison chosen such that positive values indicate improvement

Table 5.4: Comparison of results between the proposed algorithm and other published techniques. AU (B. Li et al., 2010), AO (Fahmy et al., 2012), SCR (Jolly et al., 2012), DS, and INR (Margeta et al., 2012) values are taken from Table 2 of (Suinesiaputra et al., 2014). FCN values are taken from Table 3 of (Tran, 2016). Values are of mean (standard deviation).

Method	**Manual input**	**Jaccard Index**
AU	Interactive 4D guide point placement	.84 (.17)
CNR (this work)	**None**	**.77 (.11)**
FCN	None	.74 (.13)
AO	Delineate first frame	.74 (.16)
SCR	None	.69 (.23)
DS	Delineate first frame	.64 (.18)
INR	None	.43 (.10)

Comparing the Chapter 4 (Tan et al., 2017) semi-automated results and current automated results visually, the perceived improvement appears to come from the adoption of 2D+time contour smoothing, which results in higher phase-to-phase consistency. This is measured by calculating the standard deviation of the Jaccard index, Dice index, and MHD metrics across all cardiac phases for each individual slice, demonstrating approximately 4% to 7% average reduced variation in performance on a phase-to-phase basis (Table 5.3).

Another notable change is the additional pre- and post-processing applied to slices at the base and apex. In particular, apical slices are very challenging due to the relative small size of the blood pool (including disappearance during systolic phases). No statistically significant difference was found between the Chapter 4 (Tan et al., 2017) and current architecture in the MHD metric for apical slices ($p = 0.027$). However, it is suspected that this is because the LVSC consensus validation dataset (CS*) does not include ground truth data for many apical slices. Consensus images were only generated for slices with valid results from at least three contributing raters; since apical slices are problematic for many algorithms, this likely resulted in invalid results for many raters. Around 47% of slices categorized as apical by the LM network were not included in the LVSC consensus validation dataset (Figure 5.4 – low result density at both ends). This included slices that demonstrated clear visual improvement when compared to the Chapter 4 architecture (Figure 5.5). To explore this further, the MHD metric was calculated directly between the Chapter 4 (Tan et al., 2017) and current architecture for all apical slices (i.e., as opposed to calculating MHD against the CS* reference). It was found that slices missing from CS* had significantly higher mean MHD compared to slices in CS* (2.59 vs. 1.23 mm, $p < 0.001$), i.e., slices missing from CS* showed larger differences in results between the previous and current architectures. This strongly suggests that apical slices missing from

CS* may be more challenging (and are thus affected by the additional pre- and post-processing in the current architecture).

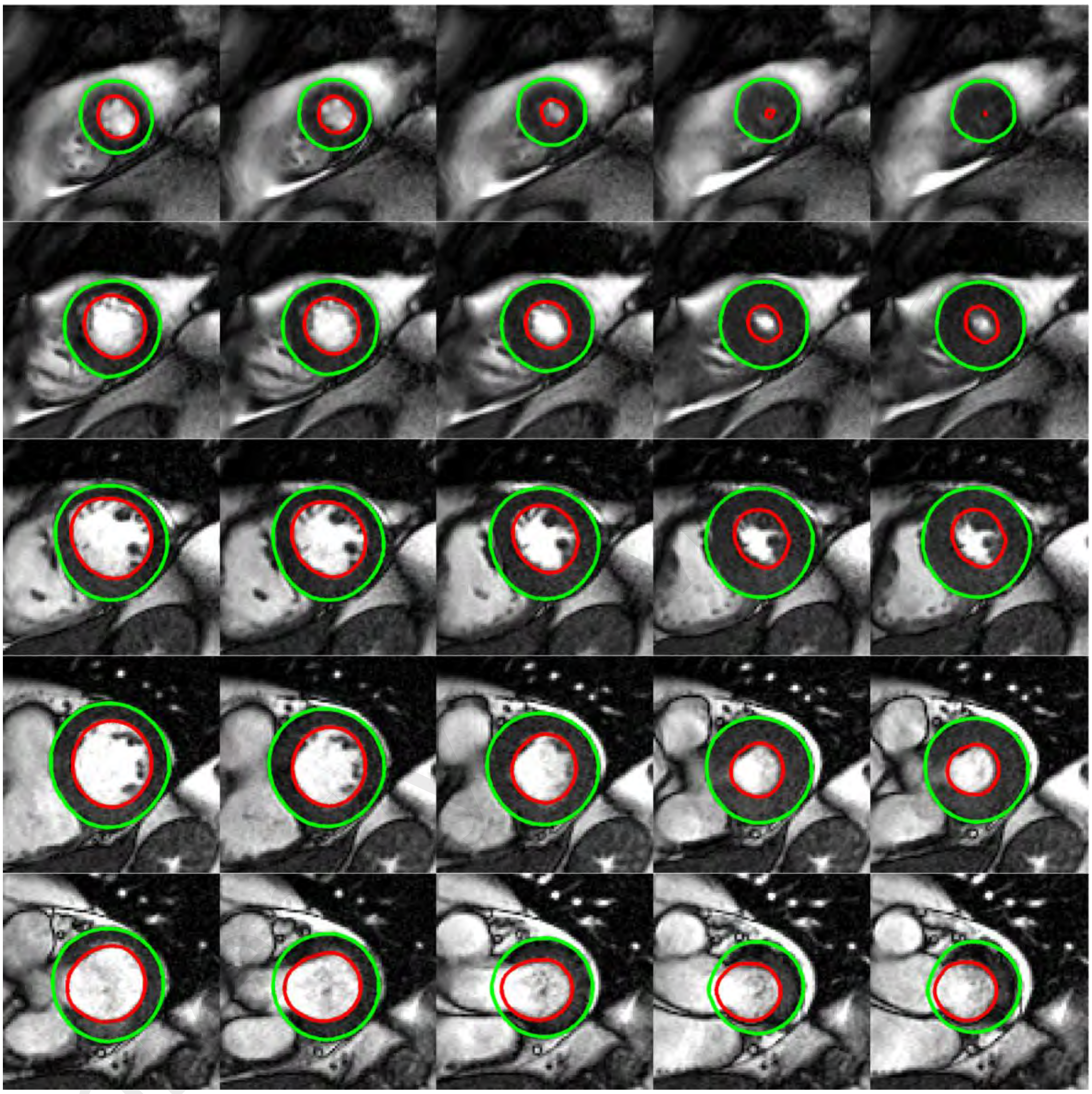


Figure 5.3: Representative segmentation result from the LVSC validation dataset. (Top to bottom) Representative slices from apex to base. (Left to Right) Representative cardiac phases from diastole to systole. Red and green contours are endo- and epicardium results from the MB network, respectively.

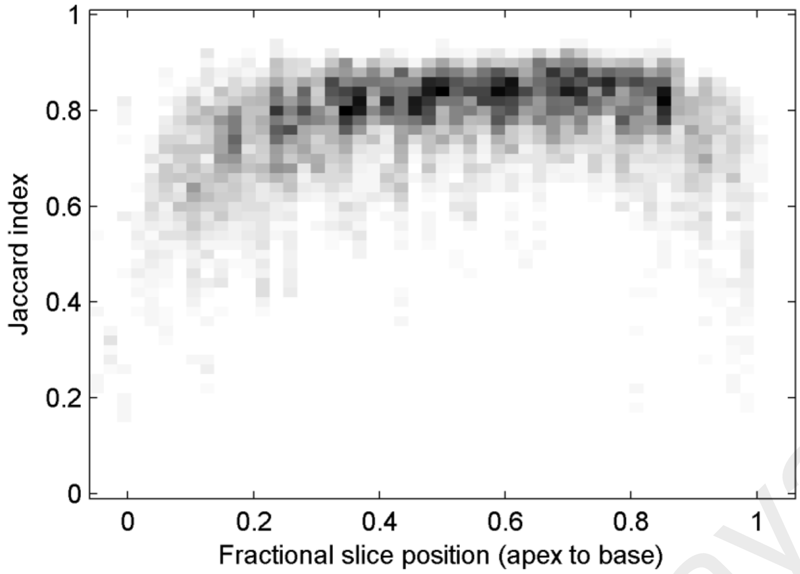


Figure 5.4: Segmentation quality as a function of fractional slice position along LV apex (zero) to base (one). Performance is strongest in the mid-LV, with noticeable drop-offs towards the apex and base ends. In addition, towards the apex and base ends there is a lack of data in the consensus gold standard for evaluation.

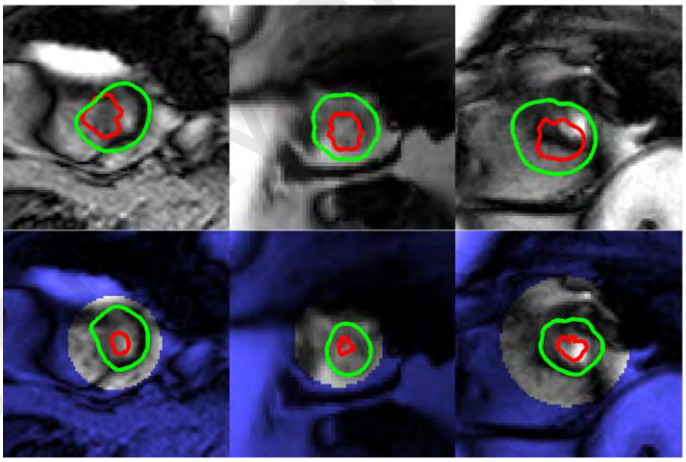


Figure 5.5: Sample images from LVSC validation dataset demonstrating improved stability due to FOV reduction for apical slices. (Top row) Results from Chapter 4 architecture (Tan et al., 2017), (bottom row) results for current architecture. Red and green contours are endo- and epicardial contours resulted from MB network. Bottom row dark blue tint areas indicate the FOV reduction crop masks.

Evaluated against the Kaggle database, there is a similar statistically significant improvement in blood volume estimation for the ED phase, though the ES results are more mixed (Table 5.3). Notably, these results are comparable to reported inter-reader variability values for multiple independent expert readers (Suinesiaputra et al., 2015): bias (mean signed difference) up to ±13 / ±19 mL for ES/ED, and precision (standard deviation of signed difference) up to 13 / 13 mL for ES/ED. In comparison, the algorithm bias is +7.2 / −19.8 mL for ES/ED, and the algorithm precision is 13.0 / 18.8 mL for ES/ED. Despite this, no statistically significant difference was found for the CRPS ($p = 0.67$). This is likely because the CRPS score is strongly affected by the separate linear regression used for its calculation; the regression technique used was unchanged between the Chapter 4 (Tan et al., 2017) and current architectures.

Evaluating the scan-rescan reproducibility, no statistically significant difference was found between the automated and manual methods. The mean variability (standard deviation across three scans, averaged across all subjects) for the ES phase was 2.43 ± 1.10 mL and 3.35 ± 3.63 mL, $p = 0.41$ for the automated and manual methods, respectively. The mean variability for the ED phase was 3.20 ± 2.26 mL and 4.40 ± 3.17 mL, $p = 0.32$ for the automated and manual methods, respectively.

For the independent error analysis, the CTR network was found to perform well, with a mean error of around 1.8 mm or 8% of the endocardial radius (Table 5.5). As expected, apical slices performed the worst and mid-level slices the best. The performance of basal slices was almost similar to apical slices in terms of absolute error, though it is suspected this is partially due to limitations of the gold standard centrepoints used; some of the LVSC CS* reference basal slices included only partial binary coverage of the myocardium, affecting calculation of the blood pool centrepoint. Given the low error of the CTR network, the independent error of the MB network was found to be largely similar to its end-to-end performance shown in Table 5.3. The independent error is around 1% better for all metrics, the biggest increase being the MHD for apical slices (around 8% improvement).

Table 5.5: Independent error analysis of the CTR and MB networks. For the CTR network, the fractional error was determined by normalizing the absolute error against the average endocardium radius. For the MB network, the rightmost column is the end-to-end network error analysis reproduced from Table 5.3 for convenience of comparison.

CTR network	**Independent Error**	**End-to-End**
Absolute error (mm) (fractional error)	1.782 ± 1.271 (0.083)	
Apical	2.108 ± 1.987 (0.150)	
Mid-level	1.704 ± 1.053 (0.071)	
Basal	2.086 ± 1.622 (0.092)	
MB network		
Jaccard index (JI)	0.771 ± 0.108	0.769 ± 0.109
Dice index (DI)	0.866 ± 0.079	0.864 ± 0.080
Modified Hausdorff distance (MHD) (mm)	1.310 ± 0.621	1.329 ± 0.710
Apical	1.692 ± 0.710	1.769 ± 1.126
Mid-level	1.201 ± 0.489	1.212 ± 0.493
Basal	1.930 ± 1.142	1.913 ± 1.168

Finally, the added effect of pre- and post-processing (PPP) was tested, and its overall effect was found to be modest. For example, the mean Jaccard, Dice, and MHD metrics for the current architecture worsened from 0.769 / 0.864 / 1.329 mm to 0.767 / 0.863 / 1.338 mm with PPP disabled. There was still a statistically significant difference ($p < 0.001$) for all three metrics comparing the Chapter 4 (Tan et al., 2017) to current architecture, demonstrating improvements in network training and architecture independent from the additional PPP. In contrast, with PPP disabled, there were no statistically significant differences for all three metrics when testing phase-to-phase consistency (i.e. standard deviation between phases) between the Chapter 4 (Tan et al., 2017) to current architecture, further demonstrating the real effect of PPP 2D+time contour smoothing.

5.5 Discussion

In this chapter a fully automated algorithm has been presented, utilizing both SA and LA information concomitantly for the segmentation of LV myocardium in SA cardiac MR images, with full coverage from apex to base, for all cardiac phases. Despite being a fully-automated operation, a small but statistically significant improvement in segmentation performance as compared to the previous semi-automated approach (Chapter 4) (Tan et al., 2017) has been shown, while significantly simplifying and making consistent the network architecture.

To the best of the author's knowledge, the mean 0.77 Jaccard index represents the best performance to date for a fully automated algorithm as evaluated against the public LVSC database. The only approach exceeding this performance is the semi-automated AU rater (Table 5.4), which requires significant manual input through the interactive placement of guide points in 4D (B. Li et al., 2010). Additionally, the performance scores for AU are slightly advantaged due to its results being used to build the consensus validation dataset; i.e., the results for AU are not fully independent of the consensus ground truth.

In the evaluation against the Kaggle Second Annual Data Science Bowl challenge, a small but significant improvement in ED blood volume estimation compared to the previous, semi-automated approach (Chapter 4) (Tan et al., 2017) was shown. Notably, the performance of the current algorithm is comparable to reported variability values for human raters (Suinesiaputra et al., 2015), despite being tested against an order of magnitude more studies. The Kaggle CRPS of 0.0122 would have placed tenth position out of the 192 original challengers, a relatively strong result considering the MB network used for segmentation was not trained against any of the Kaggle data. Notably, the top three competitors in the original challenge all utilized CNNs in some way or form: the champion (T. Lee & Liu, 2016) and second runner up (Wit, 2016) utilized per-pixel CNN segmentation of the blood pool (endocardium only), whereas the first runner up (Korshunova, 2016) utilized direct CNN regression of blood volume (no delineations).

LV delineation is inferred through the use of neural network regression. The proposed design necessitates the use of the polar transform so that the myocardium contour can be parameterized as radial distances from the LV centroid. The polar transform may introduce errors where the blood pool is small; the endocardium contour approaches the LV centroid and may lead to significant interpolation artefacts. This is most apparent in apical slices, and is likely a significant reason for the reduced performance there. Nevertheless, the proposed approach has been shown to be effective overall.

The proposed approach is dependent on three notable assumptions. First, to handle pediatric cases, it is assumed that small FOVs are a reliable proxy for small hearts, where an extra zoom factor can be triggered to compensate. This assumption can fail in cases where an inappropriate large FOV was used during acquisition (i.e., excessive inclusion of empty space), in patients with a small heart but a large body (would not trigger the small FOV threshold), or in images that have been cropped beforehand (smaller FOV than expected, leading to inadvertent trigger of the zoom factor). Nevertheless, none of these situations were seen to occur in the extensive collection of test data.

The second assumption is a reliance on consistent patient positioning metadata in the DICOM tags; these are used to correlate the SA and LA scans together, enabling the initial centrepoint estimate, C_0, as well as for mapping the localized LA landmarks to the SA images, defining the LV apex-to-base extent and thus the proper calculation of the clinical measurements, and enabling the FOV reduction in apical and basal slices. Unlike the first assumption of FOV size, a small number of cases in the Kaggle database were identified where the patient positioning metadata between the SA and LA volumes were inconsistent (e.g. where SA and LA volumes in the same study appeared to have different frames of reference). However, these situations were always easily detectable via out-of-bounds C_0 or landmark coordinates, allowing for straightforward flagging for manual intervention.

Finally, the majority of the proposed approach is purely based on 2D or 2D+time. This allows the algorithm as a whole to be insensitive to inter-slice shifts due to patient movement between slice acquisitions. The FOV reduction for apical and basal slices are exceptions to this, where they depend on consistent inter-slice positioning. The FOV reduction is based on a relatively conservative value – the 95th percentile of the mid-level epicardium radiuses – nevertheless it may fail in situations of extreme inter-slice shift, though no evidence of that was seen in the independent test data.

There were some limitations in this study design. The study design was only retrospective in nature; no new datasets were collected. In addition, reference data is lacking for some aspects of the evaluation: the Kaggle dataset contains reference clinical volume measurements but no myocardium delineations, while the LVSC dataset is lacking gold standard delineations in a significant fraction of apical slices.

In conclusion, a fully automated algorithm utilizing SA and LA information for the segmentation of LV myocardium in SA cardiac MR images has been presented, with full coverage from apex to base, for all cardiac phases. This is the best performing fully automated algorithm to date as evaluated by the public LVSC challenge, while demonstrating performance comparable to human readers in both absolute variability of clinical parameters, as well as in scan-rescan reproducibility. This overall performance is a strong indicator of practical clinical utility.

CHAPTER 6: LINEAR-REGRESSION CONVOLUTIONAL NEURAL NETWORK FOR FULLY AUTOMATED CORONARY LUMEN SEGMENTATION IN INTRAVASCULAR OPTICAL COHERENCE TOMOGRAPHY

6.1 Abstract

This chapter presents the convolutional neural network (CNN) regression technique introduced in Chapter 4 and Chapter 5, as applied to a new, unrelated task – automatic segmentation of vessel lumen wall in intravascular optical coherence tomography (OCT) – thus demonstrating the generalizability of the technique that was presented in Chapter 5.

Intravascular OCT is an optical imaging modality commonly used in the assessment of coronary artery diseases during percutaneous coronary intervention. Manual segmentation to assess luminal stenosis from OCT pullback scans is challenging and time consuming. A linear-regression CNN is developed to automatically perform vessel lumen segmentation, parameterized in terms of radial distances from the catheter centroid in polar space. Benchmarked against gold standard manual segmentation, the proposed algorithm achieves average locational accuracy of the vessel wall of 22 microns, and 0.985 and 0.970 in Dice coefficient and Jaccard similarity index, respectively. The average absolute error of luminal area estimation is 1.38%. The processing rate is 40.6 ms per image, suggesting the potential to be incorporated into a clinical workflow and to provide quantitative assessment of vessel lumen in an intra-operative timeframe.

This chapter has been published as (Yong et al., 2017); the published text is largely reproduced here excepting minor amendments for consistency and flow. For the work described in this chapter, this authors' role was on the conception and implementation of the neural network architecture. Data collection and analysis of the intravascular OCT scans were led by the first author, Yan Ling, Yong.

6.2 Introduction

Cardiovascular disease is the leading cause of death globally (Mendis, 2014). Atherosclerosis of the coronary artery disease results in remodelling and narrowing of the arteries that supply oxygenated blood to the heart, and thus may lead to myocardial infarction. Common interventional approaches include percutaneous coronary intervention and coronary artery bypass graft surgery (American Heart Association, 2017). The choice of treatment will vary depending on a range of clinical factors, including morphology of the vessel wall, and degree of stenosis as quantified by cross-sectional luminal area.

Imaging of the vasculature, specifically coronary arteries, plays a critical role in assessment of these treatment options. X-ray computed coronary angiography and cardiac magnetic resonance imaging (MRI) allow non-invasive imaging, but are very limited in their ability to assess the structure of the artery walls (Nikolaou, Alkadhi, Bamberg, Leschka, & Wintersperger, 2010). Invasive techniques, such as intravascular ultrasound (IVUS) (De Franco & Nissen, 2001), provide cross-sectional imaging of the artery walls, but with limited spatial resolution (Garcìa-Garcìa, Gogas, Serruys, & Bruining, 2011). Intravascular optical coherence tomography (IVOCT) lacks the image penetration depth of IVUS, but provides far higher resolution imaging, allowing visualization and quantification of critical structures such as the fibrous cap of atherosclerotic plaques and delineation of the arterial wall layers (Bezerra, Costa, Guagliumi, Rollins, & Simon, 2009; Jang et al., 2002; Stamper, Weissman, & Brezinski, 2006). In addition, IVOCT is finding application in imaging coronary stents to assess vascular healing and potential restenosis (Jaguszewski & Landmesser, 2012; Karanasos et al., 2014).

Delineation of the vessel lumen in IVOCT images enables quantification of the luminal cross-sectional area. Such delineation has also been used as the first step towards plaque segmentation (Celi & Berti, 2014; Wang et al., 2010) and the assessment of stent apposition (Adriaenssens et al., 2014). However, manual delineation is impractical due to the high number of cross-sectional scans acquired in a single IVOCT pullback scan, typically >100 images. Automatic delineation of the lumen wall is challenging due to various reasons. Non-homogenous intensity, blood residue, the presence and absence of different types of stents, irregular lumen shapes, image artifacts, and bifurcations are some of these challenges (Tearney et al., 2012).

Previous delineation approaches have employed edge detection filters (Sihan et al., 2008) and spline-fitting to segment the lumen boundary and stent struts (Gurmeric, Isguder, Carlier, & Unal, 2009). Other approaches have included the use of wavelet transforms and mathematical morphology (Moraes, Cardenas, & Furuie, 2013), Otsu's automatic thresholding and intersection of radial lines with lumen boundaries (Celi & Berti, 2014; Wang et al., 2010), Markov-Random fields models (Tsantis et al., 2012) and light back-scattering methods (Guha Roy et al., 2016).

Deep learning is a type of machine learning algorithm utilizing artificial neural networks (ANNs), which in recent years has been found useful for medical image processing. Input features are processed through a multi-layered network, defined by a network of weights and biases, to produce a non-linear output. During training, these weight and bias values are optimized by minimizing a loss function, mapping training input to known target output values. CNNs are a particular subset of ANN that operate on input with regular structure: they apply convolutional filters to the input of each layer, and have proven to be highly effective in image classification tasks (Havaei et al., 2017; Krizhevsky et al., 2012; C. S. Lee et al., 2017).

Most neural network applications in image processing are image-based classification models where the network is trained to classify each pixel in the input image into one of several classes. The use of this technique has been extended into a variety of medical image segmentation applications. For example, CNNs has been used to classify lung image patches in interstitial lung disease (Q. Li et al., 2014) as well as head and neck cancer in hyperspectral imaging (Halicek et al., 2017). CNNs have also been applied in retinal layer and microvasculature segmentation of retinal OCT images (Fang et al., 2017; Prentašic et al., 2016), and arterial layers segmentation in patients with Kawasaki disease (Abdolmanafi, Duong, Dahdah, & Cheriet, 2017). These CNN methods employ the commonly used feature classification approach.

An alternative approach is to train the network to perform linear-regression, in contrast to feature classification. In Chapter 4 and Chapter 5, a linear-regression CNN model has been demonstrated to outperform conventional CNN in cardiac left ventricle segmentation (Tan et al., 2017; Tan, McLaughlin, et al., 2018). CNN regression was used to infer the radial distances between the left ventricle centrepoint and the endo- and epicardial contours in polar space. This indicates the possibility of an alternative application of CNNs for image segmentation in comparable medical applications.

In this chapter, a novel method of coronary lumen segmentation for clinical assessment and treatment planning of coronary artery stenosis using a linear-regression CNN is proposed. The algorithm is tested on in vivo clinical images and assessed against gold-standard manual segmentations. This is the first use of a linear-regression CNN approach to the automated delineation of the vessel lumen in IVOCT images. This chapter is structured as follows: Section 6.3 provides experimental details and an explanation of the CNN architecture and implementation; Section 6.4 provides accuracy results benchmarked against inter-observer variability of manual segmentation, and an assessment of the impact of varying the amount of training data; and Section 6.5 concludes with a discussion of the potential clinical impact and limitations of such an approach.

6.3 Materials and method

6.3.1 IVOCT data acquisition and preparation for training and testing

The data used for this study comprises IVOCT-acquired images of patients diagnosed with coronary artery disease. The IVOCT images were acquired from the University of Malaya Medical Centre catheterization laboratory using two standard clinical systems: Illumien and Illumien Optis IVOCT Systems (St. Jude Medical, USA). Both systems have an axial resolution of 15 µm and a scan diameter of 10 mm. The Ilumien system and the Illumien Optis system have maximum frame rates of 100 fps and 180 fps, respectively. The study was approved by the University of Malaya Medical Ethics Committee (Ref: 20158-1554), and all patient data were anonymized.

In total 64 pullbacks were acquired from 28 patients (25%/75% male/female, with mean age 59.71 (±9.61) years) using Dragonfly^{TM} Duo Imaging Catheter with 2.7 F crossing profile when the artery was under contrast flushing (Iopamiro® 370). The internal rotating fibre optic imaging core performed rotational motorized pullback scans for a length of 54 mm or 75 mm in under 5 sec. These scans include multiple pre- and post-stented images of the coronary artery at different locations. These pullbacks were randomly assigned to one of two groups with a ratio of 7:3, i.e. 45 pullbacks were randomly designated as training sets and the remaining 19 as test sets. Excluding images depicting only the guide catheter, each pullback contains between 155 to 375 polar images. These images contain a heterogeneous mix of images with the absence or presence of stent struts (metal stents or bioresorbable stents or both), fibrous plaques, calcified plaques, lipid-rich plaques, ruptured plaques, thrombus, dissections, motion artifacts, bifurcations and blood artifacts. The original size of each pullback frame was 984 × 496 pixels (axial × angular dimension), and was subsampled in both dimensions to 488 × 248 pixels to reduce training and processing time. For each image, raw intensity values were converted from linear scale to logarithmic scale before normalizing by mean and standard deviation.

Gold-standard segmentations were generated on both training and test sets by manual frame-by-frame delineation using ImageJ (Schindelin, Rueden, Hiner, & Eliceiri, 2015) in Cartesian coordinates, according to the document of consensus (Tearney et al., 2012), whereby a contour was drawn between the lumen and the leading edge of the intima. The contour was also manually drawn across the guidewire shadow and bifurcation at locations that best represent the underlying border of the main lumen, gauged by the adjacent slices. The manual contour of the lumen border for each image was subsequently converted to polar coordinates, smoothed and interpolated to 100 points using cubic *B*-spline interpolation method for CNN training and testing.

6.3.2 CNN regression architecture & implementation details

Using our linear-regression CNN model, in each polar image the radius parameter of the vessel wall is inferred at 100 equidistant radial locations, rather than the more conventional approach of classifying each pixel within the image. This has the advantage of avoiding the physiologically unrealistic results that may arise from segmentation of individual pixels. The lumen segmentation was parameterized in terms of radial distances from the centre of the catheter in polar space.

The general flow of the proposed CNN model is illustrated in Figure 6.1. Our network consists of a simple structure with 4 convolutional layers and 3 fully-connected layers, including the final output layer. All polar images were padded circularly left and right before being windowed for input. The window dimension was 488 ×128 pixels centered on each individual radial point, therefore yielding 100 inputs and 100 evaluated radial distances per image.

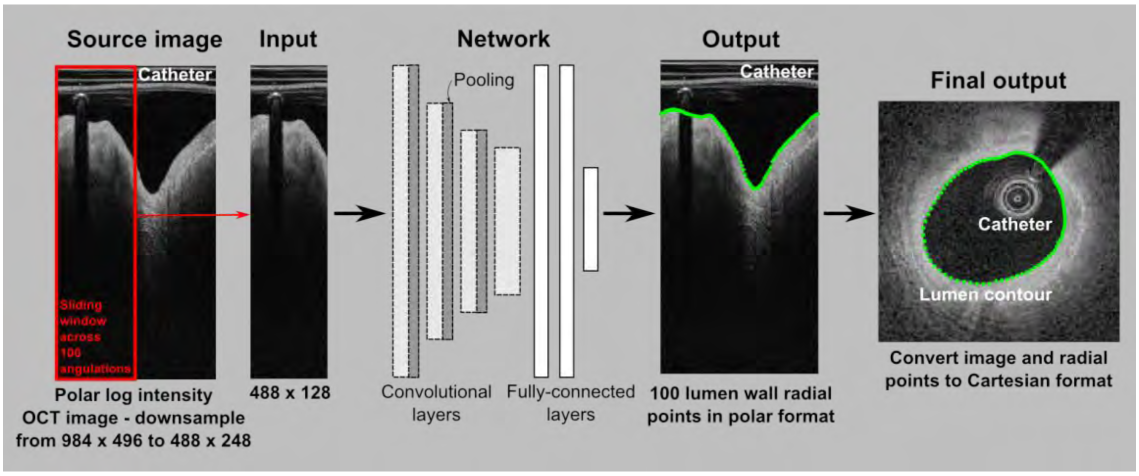


Figure 6.1: Overview of the linear-regression CNN segmentation system (refer to text for details).

The details of the network architecture are described in Table 6.1. In the network architecture, a filter kernel of size 5 × 5 × 24 with boundary zero-padding was applied for all convolutional layers, yielding 24 feature maps at each layer. In the first layer, a stride of 2 was also applied along the angular dimension to reduce computational load. The first three layers were also max-pooled by size 2 × 2. Each fully-connected layer contains 512 nodes. Exponential linear units (ELU) (Clevert et al., 2016) were used as the activation functions for all layers, including both convolutional and fully-connected layers, except the final layer. Dropout with keep probability of 0.75 was applied to the fully-connected layers FC1 and FC2, to improve the robustness of the network (Rokach & Maimon, 2005). The final layer outputs a single value representative of the radial distance between the lumen border and the centre of the catheter for the radial position being evaluated.

The objective function used for the network training is the standard mean-squared error. Starting from a random initialization, the weight and bias parameters are iteratively minimized by calculating the mean squared error between the gold standard radial distance and the output of the CNN training. The Adam stochastic gradient algorithm was used to perform the optimization, i.e. minimization, of the objective function (Kingma & Ba, 2014). The network was trained stochastically with a mini-batch size of 100 at a base-learning rate of 0.005. The learning rate was halved every 50,000 runs. The training was stopped at 400,000 runs where convergence was observed (i.e. when the observed losses had ceased to improve for at least 100,000 runs). The trained weights and biases of the network, amounting to approximately 6.3 million parameters, are subsequently used to predict the lumen contour on the test sets.

The neural network was designed in a Python (Python Software Foundation, Delaware, USA) environment using the TensorFlow v1.0.1 machine learning framework (Google Inc., California, USA). The execution of the network was performed on a Linux-based Intel i5-6500 CPU workstation with NVIDIA GeForce GTX1080 8GB GPU. The training time for 45 train sets was 13.8 hours and the complete inference time for each test image was 40.6 ms.

Table 6.1: Linear-regression CNN architecture for lumen segmentation at each windowed image. The output is the radial distance at the lumen border from the center of the catheter. CN: convolutional layer, FC: fully-connected layer.

Layer	In	Weights	Pooling	Out
CN1*	488×128×1	1×5×5×24	2×2	244×32×24
CN2	244×32×24	24×5×5×24	2×2	122×16×24
CN3	122×16×24	24×5×5×24	2×2	61×8×24
CN4	61×8×24	24×5×5×24	-	61×8×24
FC1	11712	11712×512	-	512
FC2	512	512×512	-	512
Out	512	512×1		1

*A stride of size 2 was applied on the angular dimension to reduce computational load

6.3.3 Validation

The accuracy of our proposed linear-regression CNN lumen segmentation was validated against the gold standard segmentation of the test data pullback acquisitions, which were the aforementioned 19 manually delineated pullbacks. These pullbacks contain in total 5685 images. The accuracy was assessed in three ways: (1) on a point-by-point basis via distance error measure; (2) in the form of binary image overlaps and (3) based on luminal area.

The first assessment involves point-by-point analysis on the 100 equidistant radial contour points from all images, whereby the mean absolute Euclidean distance error between the gold standard and predicted contours was computed for each image.

The second assessment was performed to evaluate the regions delineated as lumen. The amount of overlap between the binary masks as generated from the predicted contours and the corresponding gold standards were computed using the Dice coefficient and Jaccard similarity index.

The third assessment targeted at the luminal area, which is one of the clinical indices to locate and grade the extent of coronary stenosis for treatment planning. Luminal area was computed from the binary mask produced from the predicted contours and compared against the corresponding gold standard. A 1-tailed Wilcoxon signed ranks test is also performed on the errors of the estimated luminal areas at significance level of 0.001. Three-dimensional surface models of the lumen wall were also generated for all pullbacks to facilitate visual comparison of the segmentation by manual contouring and by automated contouring using the proposed CNN regression model.

6.3.4 Dependency of network performance on training data quantity

To understand the dependency of the network performance to the amount of training data required, the variation in accuracy of the 19 test pullbacks is assessed against different numbers of training data sets. Tests were performed with 10, 15, 20, 25, 30, 35, 40 and 45 pullbacks. The training pullbacks for each group were selected randomly. The number of training runs with different training sets was kept constant at 400,000 runs, with a similar base learning rate and learning rate decay protocol.

6.3.5 Inter-observer variability against CNN accuracy

To quantify the allowable variation in segmentation, an experiment is performed to assess variation in the manual gold standard that would be generated by three independent observers.

One hundred images were selected randomly from five pullbacks of the test sets and the lumen manually delineated by three independent observers. The interobserver variability was assessed through Bland-Altman analyses, consistent with Celi and Berti in their study on the segmentation of coronary lesions (Celi & Berti, 2014). Specifically, the signed differences between all possible corresponding pairs of luminal areas from all three observers were plotted against their mean area differences. Bland-Altman analyses were also performed on luminal areas evaluated by the CNN against the corresponding evaluation by all observers. These analyses provide an understanding of the total bias and limits of agreement (i.e. 95% confidence interval or 1.96×standard deviation of the signed differences from the mean) among all observers themselves as well as between the CNN and the observers.

6.4 Results

6.4.1 Dependency of network performance on training data quantity

The results assessing the impact of training data quantity on CNN accuracy are shown in Figure 6.2. The value reported here is the mean positional accuracy of each point along the vessel wall. There was notable improvement in CNN accuracy with increases in the training data quantity up until 25 training data sets. Beyond that, the mean absolute error per image varied little with increased data. However, the optimal CNN segmentation was obtained from training with the highest sample size, i.e. 45 pullbacks consisting of 13,342 training images, as summarized in Table 6.2. At 45 training pullbacks, the median of the mean absolute error per image as quantified using point-by-point analysis was 21.87 microns, whereas Dice coefficient and Jaccard similarity index were calculated as 0.985 and 0.970, respectively.

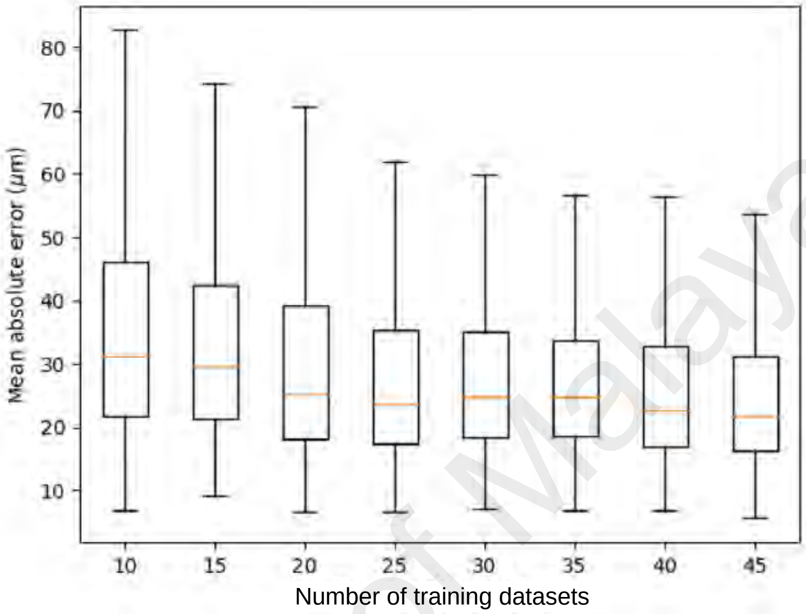


Figure 6.2: Mean absolute error against different numbers of training datasets.

Table 6.2: Accuracy of CNN segmentation with 45 training pullbacks (n = 13,342). The values are obtained based on the segmentation on 19 test pullbacks (n = 5,685).

Measure	**Median (interquartile range)**
Mean absolute error per image (point-by-point analysis), μm	21.87 (16.28, 31.29)
Dice coefficient	0.985 (0.979, 0.988)
Jaccard similarity index	0.970 (0.958, 0.977)

Representative segmentation results are shown in Figure 6.3. Apart from performing well on images with clear lumen border contrast Figure 6.3(a), linear-regression CNN segmentation has shown robustness in segmenting images with inhomogeneous lumen intensity (b), severe stenosis (c), blood residue due to suboptimal flushing (d)-(f), multiple reflections (g), embedded stent struts (h)-(i), malapposed metallic stent struts (j), malapposed bioresorbable stent struts (k), and minor side branches ((c), (i) & (l)). Acceptable lumen segmentation was found at the shadow behind the guide wire and metallic stent struts across all images. Errors were observed to occur most frequently at major bifurcations (angle spanning > approximately 90˚), where the appropriate boundary for segmenting the main vessel was ambiguous (Figure 6.4(c)-(d)). 72% of the 100 worst performing segmentation were found to contain major bifurcations and, at these locations, overestimation of the area of the main vessels was noted.

Based on the results obtained with the optimal training quantity (45 pullback data sets), luminal area estimates in all 19 test pullbacks were calculated, as tabulated in Table 6.3. CNN segmentation yields median (interquartile range) luminal area of 5.28 (3.88, 7.45) mm^2 matching well with the results of manual segmentation of 5.26 (3.93, 7.45) mm^2 (i.e. gold standard). The median (interquartile range) absolute error of luminal area was 1.38%, which is statistically significantly below 2% ($p<0.001$) as tested by the 1-tailed Wilcoxon signed rank test. Figure 6.5 shows two representative examples of the 3D reconstructed vessel wall from two different pullbacks for visual comparison of CNN regression (middle column) against gold standard manual (left column) segmentation. The vessel wall was color-coded with the cross-sectional luminal area. Difference in luminal area between CNN regression and gold standard segmentation are color-coded on the vessel wall on the right column.

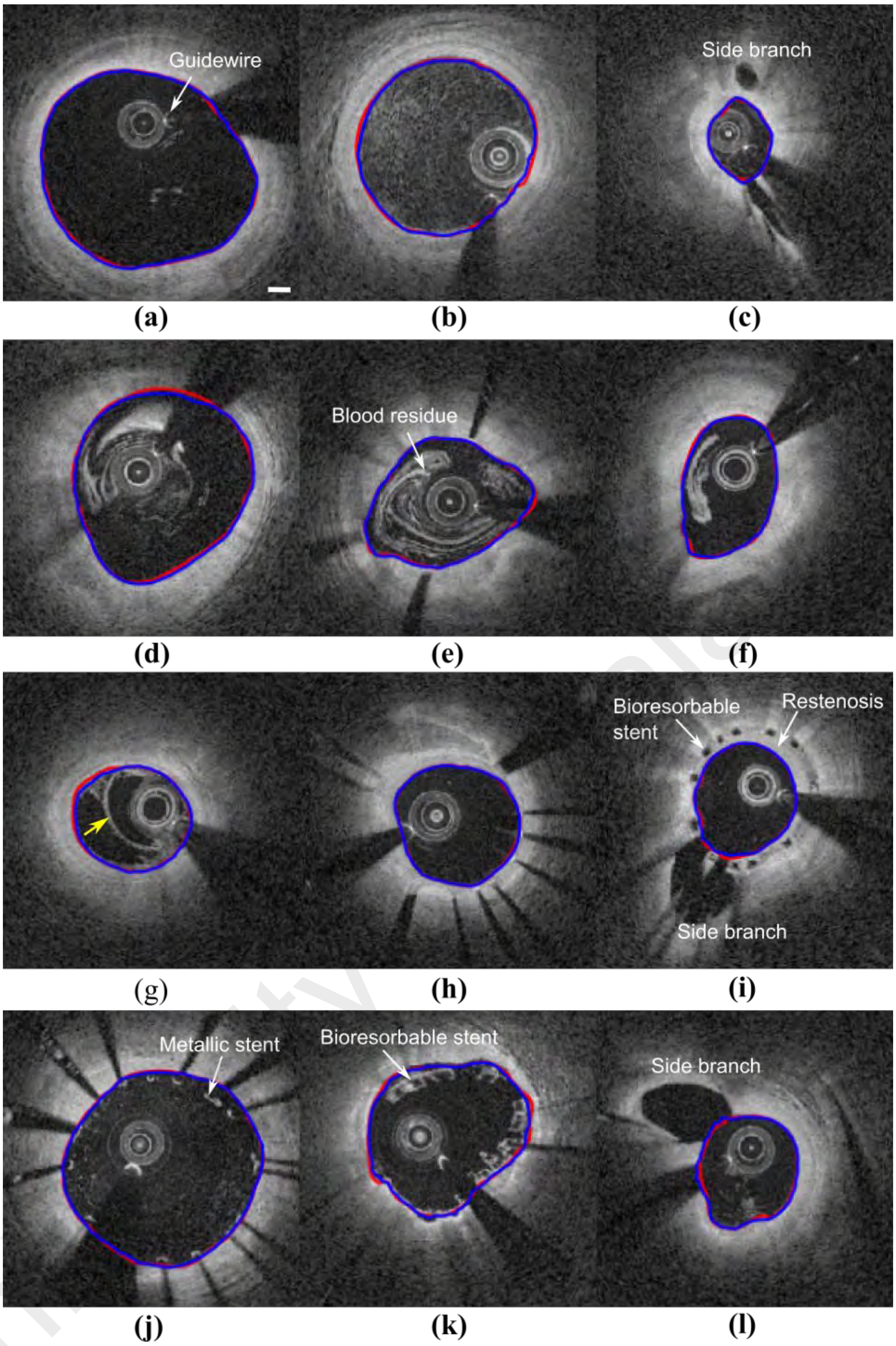


Figure 6.3: Representative results from the test sets, showing good segmentation from linear-regression CNN on images with good lumen border contrast (a), inhomogenous lumen intensity (b), severe stenosis (c), blood swirl due to inadequate flushing (d)-(f), multiple reflections (indicated by yellow arrow) (g), embedded metallic and bioresorbable stent struts due to restenosis in (h) & (i), respectively, malapposed metallic stent struts (j), malapposed bioresorbable stent strut (k), and minor side branch (l). Blue and red contours represent CNN segmentation and gold standard, respectively. Scale bar (a) represents 500 microns.

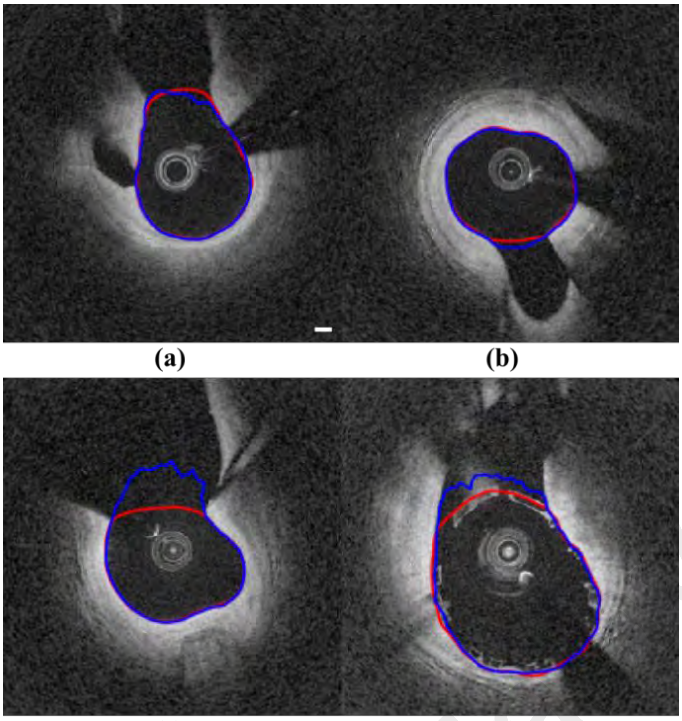


Figure 6.4: Representative cases from the test sets, showing reasonable lumen segmentation from linear-regression CNN on images with medium-sized bifurcations (a-b). Poorer results were seen at major bifurcations (c-d), where the appropriate boundary for segmenting the main vessel was ambiguous. Blue and red contours represent CNN segmentation and gold standard, respectively. Scale bar (a) represents 500 microns.

Table 6.3: Luminal area in 19 test pullbacks with optimal training.

Method	**Median (Interquartile range)**
Luminal Area (mm^2)	
Manual segmentation area	5.28 (3.88, 7.45)
CNN segmentation area	5.26 (3.93, 7.45)
Percentage Error[a] (%)	
Signed percentage error	0.06 (-1.24, 1.53)
Absolute percentage error[b]	1.38 (0.63, 2.62)

[a] Normalized by manual segmentation area
[b] Significantly below 2%, p<0.001

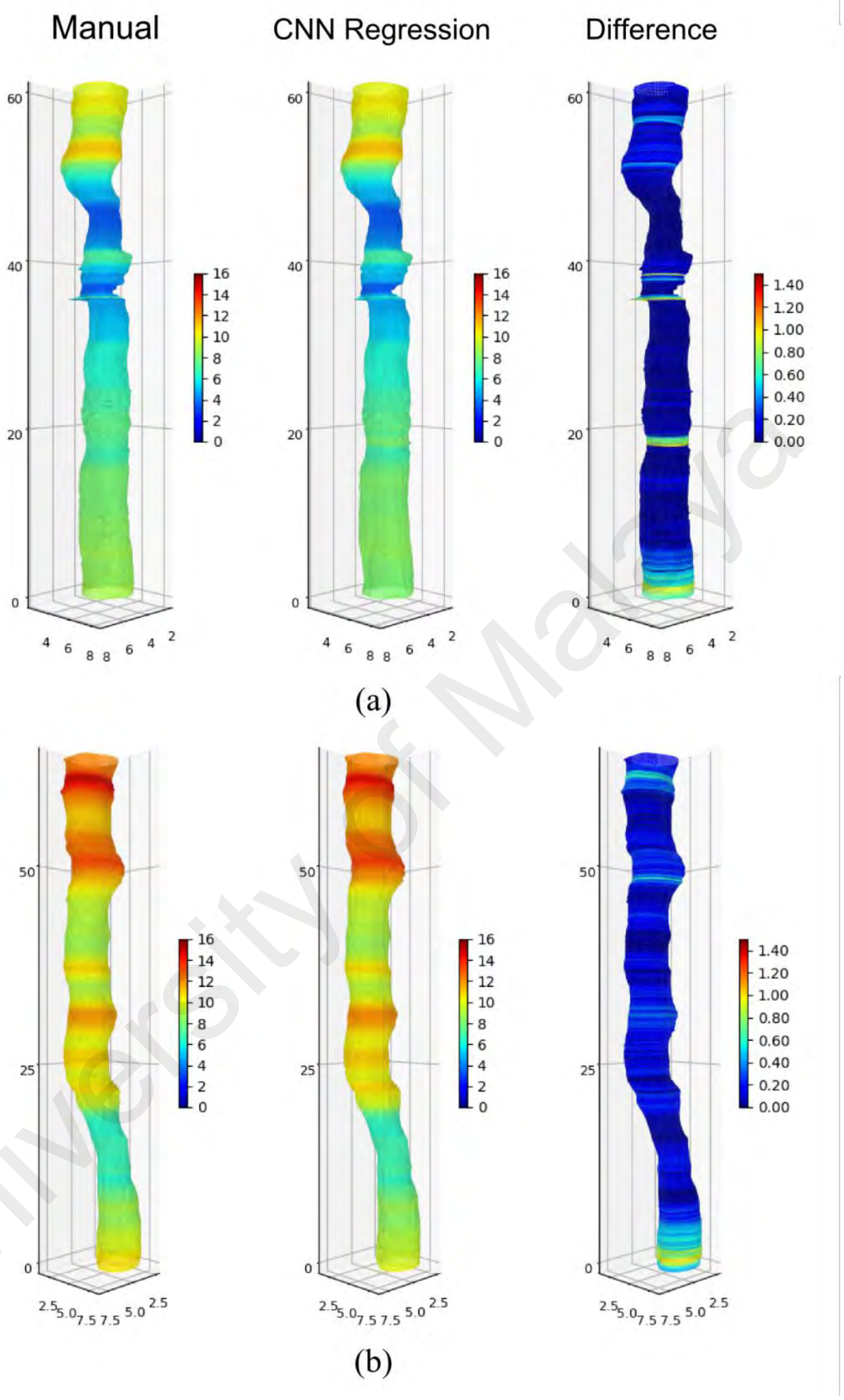
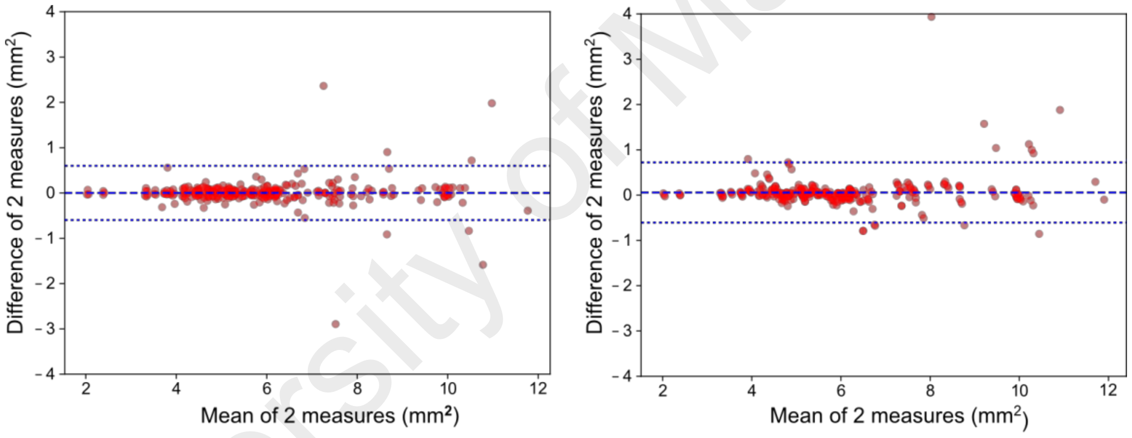


Figure 6.5: Reconstruction of vessel wall from two different pullbacks for visual comparison of CNN regression segmentation against the gold standard manual segmentation. Vessel walls (left and middle columns) are color-coded with cross-sectional luminal area. Difference in luminal area is displayed on the right. The axis is in mm and colourbar indicates luminal area in mm^2.

6.4.2 Inter-observer variability against CNN accuracy

The Bland-Altman analysis between all three observers showed a bias (mean signed difference) of 0.0 mm^2 and limits of agreement of ±0.599 mm^2 in terms of luminal area estimation (Figure 6.6(a)). Comparing the CNN to all observers, the bias was 0.057 mm^2 and the variability in terms of limits of agreement was comparable at ±0.665 mm^2 (Figure 6.6(b)). These results suggest that automated segmentation had sub-100 micron bias to over-estimate luminal area, and that the variation between automated and manual estimates of luminal area was only slightly greater than the inter-observer variability between human observers.



(a) **(b)**

Figure 6.6: Bland-Altman plot analysis of luminal area for all possible pair-comparisons between different observers (a) and between CNN and observers (b) for the 100 randomly selected images from the test set.

6.5 Discussion

Lumen dimension is an important factor in the optimization of percutaneous coronary intervention. This measure allows the clinician to localize and measure the length of lesions along the vessel wall before making an optimum selection of stent for deployment. It also allows one to indirectly assess the quality of stenting (i.e. based on total expansion of the narrowed artery) and is the first step towards quantifying the amount of stent malapposition. Misinterpretation of lesion location and length results in both clinical and financial consequences as additional stents are required for redeployment, and overlapping of multiple stents are often associated with increase incidences of restenosis, thrombosis and adverse clinical outcomes (Suzuki, 2014).

Manually quantifying coronary lumen dimension from IVOCT images over the entire extent of the imaged segment is currently not clinically feasible in view of the number of sample images available per pullback (i.e. >100 images). Automatic segmentation is desirable but challenging due to the significant variety of image features and artifacts obtained in routine scanning, restricting the operation of most image processing algorithms to a specific subset of good quality images. Deep learning techniques have been shown to be more robust in a pool of heterogeneous input images, and this has also been demonstrated in our results (Tan et al., 2017). Our study represents the first study employing such a technique, combined with a linear regression approach, to the automatic segmentation of lumen from IVOCT images.

Our results showed a notable increase in CNN accuracy up to 25 training pullbacks, and incremental improvements thereafter. The median accuracy in luminal radius at each radial location, against a manual gold standard, was 21.87μm at optimal training with 45 training pullbacks, which is comparable to the OCT system's axial resolution (15μm). The median luminal area was marginally greater by manual segmentation in comparison to CNN segmentation (i.e. 5.28 mm^2 vs 5.26 mm^2), yielding a median error of 1.38% (i.e. significantly <2% at $p = 0.001$). The CNN also has good limits of agreement against all observers (±0.665 mm^2), which is comparable with the limit of agreement among all observers (±0.599 mm^2).

Published algorithms have required the prior removal of guide-wires or blood artifacts in the images as well as interpolation of output contours across guidewire shadow and bifurcation (Abdolmanafi et al., 2017; Celi & Berti, 2014; Giovanni J Ughi, Adriaenssens, Desmet, & D'hooge, 2012; Wang et al., 2010) in order to complete an accurate segmentation. Our linear-regression CNN algorithm did not require additional pre- and post-processing of the data, with the behaviour across these features arising implicitly from the training data. In addition, the proposed method works on a wide spectrum of IVOCT images whether in presence or absence of stent struts. This approach was found to be of utility in assessing patient both pre- and post-stenting. Furthermore, the CNN segmentation was able to segment images regardless of stent types and no prior information on implanted type is needed, as can be required by some other segmentation techniques (G. J. Ughi et al., 2012), making it applicable in a wider range of clinical settings.

While training time was significant (13.8 hours for 45 training pullbacks), this is all pre-computed prior to clinical usage. The subsequent time to process a test image was extremely small (40.6 ms). Thus, the use of linear-regression CNNs offers the potential of intra-operative assessment of the vessel lumen during an intervention.

Limitations of the algorithm occur at areas with highly irregular lumen shapes, and at major bifurcations, where vessel lumen of the main branch is ambiguous even for manual segmentation. This implementation of the algorithm has adopted a 2D processing approach where each image is processed independently. Extending this to a volumetric approach, where adjacent slices influence the segmentation of each image, may result in more stable results in these situations. Alternatively, some form of energy minimization approach may be incorporated into the CNN cost function to enforce additional regularization of the lumen shape.

In conclusion, this chapter has demonstrated a novel linear-regression CNN for the segmentation of vessel lumen in IVOCT images. The algorithm was tested on clinical data and compared against a manual gold-standard. Results suggested that the CNN provided accurate estimates of the lumen boundary, with errors only slightly greater than the inter-observer variability between multiple human observers. In addition, the algorithm was fast, processing test images at a rate of 40.6 ms per image. Our results suggest that linear-regression CNN-based approach has the potential to be incorporated into a clinical workflow and provide quantitative assessment of vessel lumen in an intra-operative timeframe.

CHAPTER 7: CONCLUSION

In this thesis, the development and implementation of two fully automatic algorithms has been presented: one for the localization of left ventricle (LV) blood pool in cardiac cine magnetic resonance imaging (MRI) (Chapter 3), and another for the segmentation of LV myocardium from basal to apical slice locations across all cardiac phases (Chapter 4 and Chapter 5). The LV localization algorithm was developed using expert knowledge-based techniques, while the LV segmentation algorithm was developed using a data-driven technique, specifically convolutional neural network (CNN) regression. In addition, the underlying CNN regression technique has been demonstrated to generalize to another, unrelated application: fully automatic segmentation of vessel lumen wall in intravascular optical coherence tomography (OCT) (Chapter 6).

7.1 Research contributions and significance

For the LV localization task, the proposed algorithm showed significant improvement to existing solutions (Jolly, 2008; X. Lin et al., 2006), particularly in the presence of non-cardiac motion and high intensity scanning artifacts. The test against 1185 externally published datasets also represents the largest published validation to date for automatic LV localization. Correct LV localization was confirmed in 97.3% of the datasets, with an average processing time of 3.0 s per case. At time of writing, there are no known freely available LV localization implementations published online, whether open source or otherwise. The open source release of the proposed LV localization code will benefit the research community, as an easily accessible benchmark for future algorithmic improvements in automatic LV localization.

For the LV segmentation task, the novel use of neural network regression for image segmentation was introduced, by parameterizing the task in the form of radial distances from the LV centrepoint. When compared to the standard practice of image segmentation by per-pixel classification, it was hypothesized that this alternate approach would improve performance by encoding basic physiological constraints such as the LV endo- and epicardium contour having a common point of origin. The results demonstrated that this approach achieved superior performance when compared against a more conventional architecture with 3× larger parameter count (Tran, 2016). In addition, restricted field-of-view (FOV) processing was introduced for challenging slice locations near the LV apex, and demonstrated significant improvement in performance thereof. Benchmarked against the public Left Ventricle Segmentation challenge (LVSC), this approach achieved the best performance published to date (0.77 ± 0.11 Jaccard index) for a fully automatic algorithm. Execution time was approximately 12 s per case.

7.2 Study limitations and future work

For the LV localization task, a complete direct comparison against other published LV localization algorithms was not performed. This is because at the time of development: (1) there was no freely available implementation of said algorithms; (2) none of the published algorithms were tested against publicly available or standard databases of images; and (3) the primary published algorithms of interest (Jolly, 2008; X. Lin et al., 2006) were missing details hindering full reimplementation, e.g. threshold values. As such a comprehensive comparison between competing approaches is still lacking, though the open source release of the proposed LV localization code should improve this situation for future researchers.

For the LV segmentation task, the primary benchmark was the public LVSC database; at time of writing it remains the largest public database for full LV segmentation including both endo- and epicardial contours for all cardiac phases. This allowed for direct comparison to other published algorithms targeting the same database. In addition, the evaluation is semi-blinded as the results are calculated by the independent third-party

LVSC organizers. Nevertheless, there are quirks in the way the final LVSC performance is calculated: importantly, participating groups are allowed to test against a subset of the LVSC test set, and their final, averaged performance is only reflective of this subset rather than of the whole. This can include subsets by subject, slice position, phase, and even individual images. Thus, it is unlikely that the final result tables reflect an identical playing field (i.e., competitors can opt to skip images they deem problematic). This author considered implementing competing CNN per-pixel classification architectures for direct comparison, but at time of development did not have access to hardware that could support the size of state-of-the-art CNN architectures. Indeed, the significantly lower parameter count of the CNN regression architecture was partially an aspect of this hardware limitation.

At time of writing, newer public LV segmentation databases have been released (e.g. the 2017 Automated Cardiac Diagnosis Challenge). Due to the data-driven nature of CNNs, additional training data would likely improve performance further; the new training data should be integrated into the training pool for future study.

Assessed visually, the current performance of the LV segmentation algorithm at mid-level slice locations approaches human parity, especially when normal human inter-rater variability is taken into account. However, the LV contour radius CNN regression architecture necessitates an initial polar transformation of the image centred on the LV blood pool, and this produces strong image artifacts close to the polar centroid. Though targeted improvements such as restricted FOV processing were introduced (Chapter 5), the final segmentation performance remains noticeably weaker at slice locations near the LV apex where image artifacts are strongest. It is suspected that hybrid architectures may be useful here, where a CNN per-pixel classification network is used for slice locations near the apex, and the CNN regression network used for all other slice locations. This is because a CNN per-pixel classification architecture would not require the polar transformation and its subsequent image distortion.

In summary, the following research directions might be promising for future work:

(1) Integrate data from newly released public LV segmentation databases to improve the network performance.

(2) Develop a hybrid architecture for LV segmentation: regression networks for base-to-mid level positions, and per-pixel networks for apical positions.

7.3 Final remarks

Both project objectives have been met successfully. For the first objective, an open source algorithm for automatic LV blood pool localization has been introduced and published. For the second objective, a novel application of neural network regression was introduced and utilized to develop a fully automated LV segmentation algorithm for cardiac cine MRI. This was validated against a diverse set of publicly available and in-house cardiac cine MRI data. The strong performance overall is suggestive of practical clinical utility.

REFERENCES

Abdolmanafi, A., Duong, L., Dahdah, N., & Cheriet, F. (2017). Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. *Biomedical Optics Express*, *8*(2), 1203–1220. https://doi.org/10.1364/BOE.8.001203

Abdul Aziz, Y. F., Fadzli, F., Rizal Azman, R., Mohamed Sani, F., Vijayananthan, A., & Nazri, M. (2013). State of the Heart: CMR in Coronary Artery Disease. *Current Medical Imaging Reviews*, *9*(3), 201–213.

Adriaenssens, T., Ughi, G. J., Dubois, C., Onsea, K., De Cock, D., Bennett, J., … Desmet, W. (2014). Automated detection and quantification of clusters of malapposed and uncovered intracoronary stent struts assessed with optical coherence tomography. *The International Journal of Cardiovascular Imaging*, *30*(5), 839–848. https://doi.org/10.1007/s10554-014-0406-z

Albà, X., Lekadir, K., Pereañez, M., Medrano-Gracia, P., Young, A. A., & Frangi, A. F. (2018). Automatic initialization and quality control of large-scale cardiac MRI segmentations. *Medical Image Analysis*, *43*, 129–141. https://doi.org/10.1016/j.media.2017.10.001

Albà, X., Ventura, R. M. F. i, Lekadir, K., Tobon-Gomez, C., Hoogendoorn, C., & Frangi, A. F. (2014). Automatic cardiac LV segmentation in MRI using modified graph cuts with smoothness and interslice constraints. *Magnetic Resonance in Medicine*, *72*(6), 1775–1784. https://doi.org/10.1002/mrm.25079

American Heart Association. (2017, March 27). Cardiac Procedures and Surgeries. Retrieved January 16, 2018, from http://www.heart.org/HEARTORG/Conditions/HeartAttack/TreatmentofaHeartAttack/Cardiac-Procedures-and-Surgeries_UCM_303939_Article.jsp#.Wl1wTq6Wbcs

Assen, H. C. van, Danilouchkine, M. G., Frangi, A. F., Ordás, S., Westenberg, J. J. M., Reiber, J. H. C., & Lelieveldt, B. P. F. (2006). SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis*, *10*(2), 286–303. https://doi.org/10.1016/j.media.2005.12.001

Avendi, M. R., Kheradvar, A., & Jafarkhani, H. (2016). A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical Image Analysis*, *30*, 108–119. https://doi.org/10.1016/j.media.2016.01.005

Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., … Rueckert, D. (2017). Semi-supervised Learning for Network-Based Cardiac MR Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017* (pp. 253–260). Springer, Cham. https://doi.org/10.1007/978-3-319-66185-8_29

Bai, W., Shi, W., Ledig, C., & Rueckert, D. (2015). Multi-atlas segmentation with augmented features for cardiac MR images. *Medical Image Analysis*, *19*(1), 98–109. https://doi.org/10.1016/j.media.2014.09.005

Barry, J. M. (1973). *SPLINS and SMOOTH: two FORTRAN smoothing routines*. Australian Nuclear Science and Technology Organisation.

Baumgartner, C. F., Koch, L. M., Pollefeys, M., & Konukoglu, E. (2018). An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation. In M. Pop, M. Sermesant, P.-M. Jodoin, A. Lalande, X. Zhuang, G. Yang, … O. Bernard (Eds.), *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges* (pp. 111–119). Springer International Publishing.

Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning*, *2*(1), 1–127. https://doi.org/10.1561/2200000006

Berbari, R. E., Bloch, I., Redheuil, A., Angelini, E., Mousseaux, E., Frouin, F., & Herment, A. (2007). 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). https://doi.org/10.1109/IEMBS.2007.4353341

Bezerra, H. G., Costa, M. A., Guagliumi, G., Rollins, A. M., & Simon, D. I. (2009). Intracoronary Optical Coherence Tomography: A Comprehensive Review. *JACC. Cardiovascular Interventions*, *2*(11), 1035–1046. https://doi.org/10.1016/j.jcin.2009.06.019

Bitar, R., Leung, G., Perng, R., Tadros, S., Moody, A. R., Sarrazin, J., … Roberts, T. P. (2006). MR Pulse Sequences: What Every Radiologist Wants to Know but Is Afraid to Ask. *RadioGraphics*, *26*(2), 513–537. https://doi.org/10.1148/rg.262055063

Bricq, S., Frandon, J., Bernard, M., Guye, M., Finas, M., Marcadet, L., … Lalande, A. (2016). Semiautomatic detection of myocardial contours in order to investigate normal values of the left ventricular trabeculated mass using MRI. *Journal of Magnetic Resonance Imaging*, *43*(6), 1398–1406. https://doi.org/10.1002/jmri.25113

Celi, S., & Berti, S. (2014). In-vivo segmentation and quantification of coronary lesions by optical coherence tomography images for a lesion type definition and stenosis grading. *Medical Image Analysis*, *18*(7), 1157–1168. https://doi.org/10.1016/j.media.2014.06.011

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2016). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). In *International Conference on Learning Representations (ICLR) 2016*. Retrieved from http://arxiv.org/abs/1511.07289

Constantinides, C., Chenoune, Y., Kachenoura, N., Roullot, E., Mousseaux, E., Herment, A., & Frouin, F. (2009). Semi-automated cardiac segmentation on cine magnetic resonance images using GVF-Snake deformable models. *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*.

Cousty, J., Najman, L., Couprie, M., Clément-Guinaudeau, S., Goissen, T., & Garot, J. (2010). Segmentation of 4D cardiac MRI: Automated method based on spatio-temporal watershed cuts. *Image and Vision Computing*, *28*(8), 1229–1243. https://doi.org/10.1016/j.imavis.2010.01.001

De Franco, A. C., & Nissen, S. E. (2001). Coronary intravascular ultrasound: implications for understanding the development and potential regression of atherosclerosis. *The American Journal of Cardiology*, *88*(10A), 7M-20M.

Dubuisson, M. P., & Jain, A. K. (1994). A modified Hausdorff distance for object matching. In *Proceedings of 12th International Conference on Pattern Recognition* (Vol. 1, pp. 566–568). https://doi.org/10.1109/ICPR.1994.576361

Eslami, A., Karamalis, A., Katouzian, A., & Navab, N. (2013). Segmentation by retrieval with guided random walks: Application to left ventricle segmentation in MRI. *Medical Image Analysis*, *17*(2), 236–253. https://doi.org/10.1016/j.media.2012.10.005

Fahmy, A. S., Al-Agamy, A. O., & Khalifa, A. (2012). Myocardial Segmentation Using Contour-Constrained Optical Flow Tracking. In O. Camara, E. Konukoglu, M. Pop, K. Rhode, M. Sermesant, & A. Young (Eds.), *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges* (pp. 120–128). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-28326-0_12

Fang, L., Cunefare, D., Wang, C., Guymer, R. H., Li, S., & Farsiu, S. (2017). Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomedical Optics Express*, *8*(5), 2732–2744. https://doi.org/10.1364/BOE.8.002732

Frangi, A. F., Niessen, W. J., & Viergever, M. A. (2001). Three-dimensional modeling for functional analysis of cardiac images, a review. *IEEE Transactions on Medical Imaging*, *20*(1), 2–5. https://doi.org/10.1109/42.906421

Garcìa-Garcìa, H. M., Gogas, B. D., Serruys, P. W., & Bruining, N. (2011). IVUS-based imaging modalities for tissue characterization: similarities and differences. *The International Journal of Cardiovascular Imaging*, *27*(2), 215–224. https://doi.org/10.1007/s10554-010-9789-7

Gardner, B. I., Bingham, S. E., Allen, M. R., Blatter, D. D., & Anderson, J. L. (2009). Cardiac magnetic resonance versus transthoracic echocardiography for the assessment of cardiac volumes and regional function after myocardial infarction: an intrasubject comparison using simultaneous intrasubject recordings. *Cardiovascular Ultrasound*, *7*, 38. https://doi.org/10.1186/1476-7120-7-38

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. In *PMLR* (pp. 315–323). Retrieved from http://proceedings.mlr.press/v15/glorot11a.html

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Goodfellow, I., Warde-farley, D., Mirza, M., Courville, A., & Bengio, Y. (2013). Maxout Networks. In S. Dasgupta & D. Mcallester (Eds.), *Proceedings of the 30th International Conference on Machine Learning (ICML-13)* (Vol. 28, pp. 1319–1327). JMLR Workshop and Conference Proceedings. Retrieved from http://jmlr.org/proceedings/papers/v28/goodfellow13.pdf

Guha Roy, A., Conjeti, S., Carlier, S. G., Dutta, P. K., Kastrati, A., Laine, A. F., … Sheet, D. (2016). Lumen Segmentation in Intravascular Optical Coherence Tomography Using Backscattering Tracked and Initialized Random Walks. *IEEE Journal of Biomedical and Health Informatics*, *20*(2), 606–614. https://doi.org/10.1109/JBHI.2015.2403713

Gurmeric, S., Isguder, G. G., Carlier, S., & Unal, G. (2009). A new 3-D automated computational method to evaluate in-stent neointimal hyperplasia in in-vivo intravascular optical coherence tomography pullbacks. *Medical Image Computing and Computer-Assisted Intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, *12*(Pt 2), 776–785.

Halicek, M., Lu, G., Little, J. V., Wang, X., Patel, M., Griffith, C. C., … Fei, B. (2017). Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. *Journal of Biomedical Optics*, *22*(6), 60503. https://doi.org/10.1117/1.JBO.22.6.060503

Hall, J. E. (2015). The Heart. In *Guyton and Hall Textbook of Medical Physiology, 13e* (13 edition). Philadelphia, PA: Saunders.

Haralick, R. M., & Shapiro, L. G. (1992). Appendix A. In *Computer and Robot Vision* (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., … Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, *35*, 18–31. https://doi.org/10.1016/j.media.2016.05.004

Heart. (2017, October 4). Retrieved from https://en.wikipedia.org/w/index.php?title=Heart&oldid=803758538

Hu, H., Liu, H., Gao, Z., & Huang, L. (2013). Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming. *Magnetic Resonance Imaging*, *31*(4), 575–584. https://doi.org/10.1016/j.mri.2012.10.004

Jaguszewski, M., & Landmesser, U. (2012). Optical Coherence Tomography Imaging: Novel Insights into the Vascular Response After Coronary Stent Implantation. *Current Cardiovascular Imaging Reports*, *5*(4), 231–238. https://doi.org/10.1007/s12410-012-9138-4

Jang, I.-K., Bouma, B. E., Kang, D.-H., Park, S.-J., Park, S.-W., Seung, K.-B., … Tearney, G. J. (2002). Visualization of coronary atherosclerotic plaques in patients using optical coherence tomography: comparison with intravascular ultrasound. *Journal of the American College of Cardiology*, *39*(4), 604–609.

Jolly, M.-P. (2006). Automatic Segmentation of the Left Ventricle in Cardiac MR and CT Images. *International Journal of Computer Vision*, *70*(2), 151–163. https://doi.org/10.1007/s11263-006-7936-3

Jolly, M.-P. (2008). Automatic Recovery of the Left Ventricular Blood Pool in Cardiac Cine MR Images. In D. Metaxas, L. Axel, G. Fichtinger, & G. Székely (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008* (pp. 110–118). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-85988-8_14

Jolly, M.-P., Guetter, C., Lu, X., Xue, H., & Guehring, J. (2012). Automatic Segmentation of the Myocardium in Cine MR Images Using Deformable Registration. In O. Camara, E. Konukoglu, M. Pop, K. Rhode, M. Sermesant, & A. Young (Eds.), *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges* (pp. 98–108). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-28326-0_10

Jolly, M.-P., Xue, H., Grady, L., & Guehring, J. (2009). Combining Registration and Minimum Surfaces for the Segmentation of the Left Ventricle in Cardiac Cine MR Images. In G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, & C. Taylor (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009* (pp. 910–918). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-04271-3_110

Kadir, K., Gao, H., Payne, A., Soraghan, J., & Berry, C. (2012). LV wall segmentation using the variational level set method (LSM) with additional shape constraint for oedema quantification. *Physics in Medicine and Biology*, *57*(19), 6007–6023. https://doi.org/10.1088/0031-9155/57/19/6007

Kaggle, & Booz Allen Hamilton. (2015, December 14). Second Annual Data Science Bowl | Kaggle. Retrieved December 7, 2016, from https://www.kaggle.com/c/second-annual-data-science-bowl

Karanasos, A., Simsek, C., Gnanadesigan, M., van Ditzhuijzen, N. S., Freire, R., Dijkstra, J., … Regar, E. (2014). OCT assessment of the long-term vascular healing response 5 years after everolimus-eluting bioresorbable vascular scaffold. *Journal of the American College of Cardiology*, *64*(22), 2343–2356. https://doi.org/10.1016/j.jacc.2014.09.029

Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. Retrieved from http://arxiv.org/abs/1412.6980

Korshunova, I. (2016, March 15). Diagnosing Heart Diseases with Deep Neural Networks. Retrieved November 14, 2017, from http://irakorshunova.github.io/2016/03/15/heart.html

Kramer, C. M., Barkhausen, J., Flamm, S. D., Kim, R. J., & Nagel, E. (2013). Standardized cardiovascular magnetic resonance (CMR) protocols 2013 update. *Journal of Cardiovascular Magnetic Resonance*, *15*(1), 1–10. https://doi.org/10.1186/1532-429X-15-91

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Lay, N., Birkbeck, N., Zhang, J., & Zhou, S. K. (2013). Rapid Multi-organ Segmentation Using Context Integration and Discriminative Models. In J. C. Gee, S. Joshi, K. M. Pohl, W. M. Wells, & L. Zöllei (Eds.), *Information Processing in Medical Imaging* (pp. 450–462). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-38868-2_38

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, C. S., Tyring, A. J., Deruyter, N. P., Wu, Y., Rokem, A., & Lee, A. Y. (2017). Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomedical Optics Express*, *8*(7), 3440–3448. https://doi.org/10.1364/BOE.8.003440

Lee, T., & Liu, Q. (2016, March 16). Automatic Left ventricle volume calculation in cardiac MRI using Convolutional Neural Network. Retrieved December 7, 2016, from https://github.com/woshialex/diagnose-heart

Li, B., Liu, Y., Occleshaw, C. J., Cowan, B. R., & Young, A. A. (2010). In-line Automated Tracking for Ventricular Function With Magnetic Resonance Imaging. *JACC: Cardiovascular Imaging*, *3*(8), 860–866. https://doi.org/10.1016/j.jcmg.2010.04.013

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014). Medical image classification with convolutional neural network. In *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)* (pp. 844–848). https://doi.org/10.1109/ICARCV.2014.7064414

Liew, Y. M., McLaughlin, R. A., Chan, B. T., Aziz, Y. F. A., Chee, K. H., Ung, N. M., … Lim, E. (2015). Motion corrected LV quantification based on 3D modelling for improved functional assessment in cardiac MRI. *Physics in Medicine & Biology*, *60*(7), 2715. https://doi.org/10.1088/0031-9155/60/7/2715

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., … Dollár, P. (2014). Microsoft COCO: Common Objects in Context. *ArXiv:1405.0312 [Cs]*. Retrieved from http://arxiv.org/abs/1405.0312

Lin, X., Cowan, B. R., & Young, A. A. (2006). Automated Detection of Left Ventricle in 4D MR Images: Experience from a Large Study. In R. Larsen, M. Nielsen, & J. Sporring (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* (pp. 728–735). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/11866565_89

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR) 2015* (pp. 3431–3440). IEEE.

Lorenzo-Valdés, M., Sanchez-Ortiz, G. I., Elkington, A. G., Mohiaddin, R. H., & Rueckert, D. (2004). Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis*, *8*(3), 255–265. https://doi.org/10.1016/j.media.2004.06.005

Low, W.-Y., Lee, Y.-K., & Samy, A. L. (2014). Non-communicable diseases in the Asia-Pacific region: Prevalence, risk factors and community-based prevention. *International Journal of Occupational Medicine and Environmental Health*, *28*(1), 20–26. https://doi.org/10.2478/s13382-014-0326-0

Lu, X., & Jolly, M.-P. (2013). Discriminative Context Modeling Using Auxiliary Markers for LV Landmark Detection from a Single MR Image. In O. Camara, T. Mansi, M. Pop, K. Rhode, M. Sermesant, & A. Young (Eds.), *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges* (pp. 105–114). Springer Berlin Heidelberg.

Lu, Y.-L., Connelly, K. A., Dick, A. J., Wright, G. A., & Radau, P. E. (2013). Automatic functional analysis of left ventricle in cardiac cine MRI. *Quantitative Imaging in Medicine and Surgery*, *3*(4), 200–209. https://doi.org/10.3978/j.issn.2223-4292.2013.08.02

Lynch, M., Ghita, O., & Whelan, P. F. (2006). Automatic segmentation of the left ventricle cavity and myocardium in MRI data. *Computers in Biology and Medicine*, *36*(4), 389–407. https://doi.org/10.1016/j.compbiomed.2005.01.005

Lynch, M., Ghita, O., & 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), 195–203. https://doi.org/10.1109/TMI.2007.904681

Margeta, J., Geremia, E., Criminisi, A., & Ayache, N. (2012). Layered Spatio-temporal Forests for Left Ventricle Segmentation from 4D Cardiac MRI Data. In O. Camara, E. Konukoglu, M. Pop, K. Rhode, M. Sermesant, & A. Young (Eds.), *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges* (pp. 109–119). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-28326-0_11

Mendis, S. (2014). *Global status report on noncommunicable diseases 2014*. World Health Organization.

Mitchell, S. C., Lelieveldt, B. P. F., Geest, R. J. van der, Bosch, H. G., Reiver, J. H. C., & Sonka, M. (2001). Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. *IEEE Transactions on Medical Imaging*, *20*(5), 415–423. https://doi.org/10.1109/42.925294

Moraes, M. C., Cardenas, D. A. C., & Furuie, S. S. (2013). Automatic lumen segmentation in IVOCT images using binary morphological reconstruction. *BioMedical Engineering OnLine*, *12*, 78. https://doi.org/10.1186/1475-925X-12-78

Nachtomy, E., Cooperstein, R., Vaturi, M., Bosak, E., Vered, Z., & Akselrod, S. (1998). Automatic assessment of cardiac function from short-axis MRI: procedure and clinical evaluation. *Magnetic Resonance Imaging*, *16*(4), 365–376. https://doi.org/10.1016/S0730-725X(98)80019-2

Nambakhsh, C. M. S., Yuan, J., Punithakumar, K., Goela, A., Rajchl, M., Peters, T. M., & Ayed, I. B. (2013). Left ventricle segmentation in MRI via convex relaxed distribution matching. *Medical Image Analysis*, *17*(8), 1010–1024. https://doi.org/10.1016/j.media.2013.05.002

Ngo, T., & Carneiro, G. (2013). Left Ventricle Segmentation from Cardiac MRI Combining Level Set Methods with Deep Belief Networks. In *ICIP 2013 proceedings* (pp. 695–699). Melbourne, Vic. Australia: IEEE. https://doi.org/10.1109/ICIP.2013.6738143

Nikolaou, K., Alkadhi, H., Bamberg, F., Leschka, S., & Wintersperger, B. J. (2010). MRI and CT in the diagnosis of coronary artery disease: indications and applications. *Insights into Imaging*, *2*(1), 9–24. https://doi.org/10.1007/s13244-010-0049-0

O'Brien, S. P., Ghita, O., & Whelan, P. F. (2011). A Novel Model-Based 3D+Time Left Ventricular Segmentation Technique. *IEEE Transactions on Medical Imaging*, *30*(2), 461–474. https://doi.org/10.1109/TMI.2010.2086465

Pednekar, A. S., Muthupillai, R., Lenge, V. V., Kakadiaris, I. A., & Flamm, S. D. (2006). Automatic identification of the left ventricle in cardiac cine-MR images: Dual-contrast cluster analysis and scout-geometry approaches. *Journal of Magnetic Resonance Imaging*, *23*(5), 641–651. https://doi.org/10.1002/jmri.20552

Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S. E., & Frangi, A. F. (2016). A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *29*(2), 155–195. https://doi.org/10.1007/s10334-015-0521-4

Petitjean, C., & Dacher, J.-N. (2011). A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis*, *15*(2), 169–184. https://doi.org/10.1016/j.media.2010.12.004

Pluempitiwiriyawej, C., Moura, J. M. F., Wu, Y.-J. L., & Ho, C. (2005). STACS: new active contour scheme for cardiac MR image segmentation. *IEEE Transactions on Medical Imaging*, *24*(5), 593–603. https://doi.org/10.1109/TMI.2005.843740

Poudel, R. P. K., Lamata, P., & 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. https://doi.org/10.1007/978-3-319-52280-7_8

Prentašic, P., Heisler, M., Mammo, Z., Lee, S., Merkur, A., Navajas, E., … Loncaric, S. (2016). Segmentation of the foveal microvasculature using deep learning networks. *Journal of Biomedical Optics*, *21*(7), 75008. https://doi.org/10.1117/1.JBO.21.7.075008

Queirós, S., Barbosa, D., Heyde, B., Morais, P., Vilaça, J. L., Friboulet, D., … D'hooge, J. (2014). Fast automatic myocardial segmentation in 4D cine CMR datasets. *Medical Image Analysis*, *18*(7), 1115–1131. https://doi.org/10.1016/j.media.2014.06.001

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A. J., & Wright, G. A. (2009). Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI. *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*. Retrieved from http://hdl.handle.net/10380/3070

Rikxoort, E. M. van, Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M. A., … Ginneken, B. van. (2010). Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. *Medical Image Analysis*, *14*(1), 39–49. https://doi.org/10.1016/j.media.2009.10.001

Rokach, L., & Maimon, O. (2005). Clustering Methods. In *Data Mining and Knowledge Discovery Handbook* (pp. 321–352). Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_15

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Roth, G. A., Huffman, M. D., Moran, A. E., Feigin, V., Mensah, G. A., Naghavi, M., & Murray, C. J. L. (2015). Global and Regional Patterns in Cardiovascular Mortality

From 1990 to 2013. *Circulation*, *132*(17), 1667–1678. https://doi.org/10.1161/CIRCULATIONAHA.114.008720

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2014). ImageNet Large Scale Visual Recognition Challenge. *ArXiv:1409.0575 [Cs]*. Retrieved from http://arxiv.org/abs/1409.0575

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: an open platform for biomedical image analysis. *Molecular Reproduction and Development*, *82*(7–8), 518–529. https://doi.org/10.1002/mrd.22489

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schulz-Menger, J., Bluemke, D. A., Bremerich, J., Flamm, S. D., Fogel, M. A., Friedrich, M. G., … Nagel, E. (2013). Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) Board of Trustees Task Force on Standardized Post Processing. *Journal of Cardiovascular Magnetic Resonance*, *15*(1), 35. https://doi.org/10.1186/1532-429X-15-35

Shao, Y., Gao, Y., Wang, Q., Yang, X., & Shen, D. (2015). Locally-constrained boundary regression for segmentation of prostate and rectum in the planning CT images. *Medical Image Analysis*, *26*(1), 345–356. https://doi.org/10.1016/j.media.2015.06.007

Sihan, K., Botha, C., Post, F., Winter, S. de, Regar, E., Hamers, R., & Bruining, N. (2008). A novel approach to quantitative analysis of intravascular optical coherence tomography imaging. In *2008 Computers in Cardiology* (pp. 1089–1092). https://doi.org/10.1109/CIC.2008.4749235

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.*, *15*(1), 1929–1958.

Stalidis, G., Maglaveras, N., Efstratiadis, S. N., Dimitriadis, A. S., & Pappas, C. (2002). Model-based processing scheme for quantitative 4-D cardiac MRI analysis. *IEEE Transactions on Information Technology in Biomedicine*, *6*(1), 59–72. https://doi.org/10.1109/4233.992164

Stamper, D., Weissman, N. J., & Brezinski, M. (2006). Plaque characterization with optical coherence tomography. *Journal of the American College of Cardiology*, *47*(8 Suppl), C69-79. https://doi.org/10.1016/j.jacc.2005.10.067

Suinesiaputra, A., Bluemke, D. A., Cowan, B. R., Friedrich, M. G., Kramer, C. M., Kwong, R., … Nagel, E. (2015). Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours. *Journal of Cardiovascular Magnetic Resonance*, *17*, 63. https://doi.org/10.1186/s12968-015-0170-9

Suinesiaputra, A., Cowan, B. R., Al-Agamy, A. O., Elattar, M. A., Ayache, N., Fahmy, A. S., … Young, A. A. (2014). A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Medical Image Analysis*, *18*(1), 50–62. https://doi.org/10.1016/j.media.2013.09.001

Suzuki, K. (2014). Mining of Training Samples for Multiple Learning Machines in Computer-Aided Detection of Lesions in CT Images. In *2014 IEEE International Conference on Data Mining Workshop* (pp. 982–989). https://doi.org/10.1109/ICDMW.2014.111

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2014). Going Deeper with Convolutions. *ArXiv:1409.4842 [Cs]*. Retrieved from http://arxiv.org/abs/1409.4842

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *ArXiv:1312.6199 [Cs]*. Retrieved from http://arxiv.org/abs/1312.6199

Tan, L. K., Liew, Y. M., Lim, E., Abdul Aziz, Y. F., Chee, K. H., & McLaughlin, R. A. (2018). Automatic localization of the left ventricular blood pool centroid in short axis cardiac cine MR images. *Medical & Biological Engineering & Computing*, *56*(6), 1053–1062. https://doi.org/10.1007/s11517-017-1750-7

Tan, L. K., Liew, Y. M., Lim, E., & McLaughlin, R. A. (2017). Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Medical Image Analysis*, *39*, 78–86. https://doi.org/10.1016/j.media.2017.04.002

Tan, L. K., McLaughlin, R. A., Lim, E., Abdul Aziz, Y. F., & Liew, Y. M. (2018). Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression. *Journal of Magnetic Resonance Imaging*, *48*, 140–152. https://doi.org/10.1002/jmri.25932

Tavakoli, V., & Amini, A. A. (2013). A survey of shaped-based registration and segmentation techniques for cardiac images. *Computer Vision and Image Understanding*, *117*(9), 966–989. https://doi.org/10.1016/j.cviu.2012.11.017

Tearney, G. J., Regar, E., Akasaka, T., Adriaenssens, T., Barlis, P., Bezerra, H. G., … International Working Group for Intravascular Optical Coherence Tomography (IWG-IVOCT). (2012). Consensus standards for acquisition, measurement, and reporting of intravascular optical coherence tomography studies: a report from the International Working Group for Intravascular Optical Coherence Tomography Standardization and Validation. *Journal of the American College of Cardiology*, *59*(12), 1058–1072. https://doi.org/10.1016/j.jacc.2011.09.079

Tran, P. V. (2016). A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI. *ArXiv:1604.00494 [Cs]*. Retrieved from http://arxiv.org/abs/1604.00494

Tsantis, S., Kagadis, G. C., Katsanos, K., Karnabatidis, D., Bourantas, G., & Nikiforidis, G. C. (2012). Automatic vessel lumen segmentation and stent strut detection in intravascular optical coherence tomography. *Medical Physics*, *39*(1), 503–513. https://doi.org/10.1118/1.3673067

Ughi, G. J., Adriaenssens, T., Onsea, K., Kayaert, P., Dubois, C., Sinnaeve, P., … D'hooge, J. (2012). Automatic segmentation of in-vivo intra-coronary optical coherence tomography images to assess stent strut apposition and coverage. *The International Journal of Cardiovascular Imaging*, *28*(2), 229–241. https://doi.org/10.1007/s10554-011-9824-3

Ughi, Giovanni J, Adriaenssens, T., Desmet, W., & D'hooge, J. (2012). Fully automatic three-dimensional visualization of intravascular optical coherence tomography images: methods and feasibility in vivo. *Biomedical Optics Express*, *3*(12), 3291–3303. https://doi.org/10.1364/BOE.3.003291

Üzümcü, M., van der Geest, R. J., Swingen, C., Reiber, J. H. C., & Lelieveldt, B. P. F. (2006). Time Continuous Tracking and Segmentation of Cardiovascular Magnetic Resonance Images Using Multidimensional Dynamic Programming. *Investigative Radiology*, *41*(1), 52. https://doi.org/10.1097/01.rli.0000194070.88432.24

Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, *45*(1, Supplement 1), S61–S72. https://doi.org/10.1016/j.neuroimage.2008.10.040

Walsh, R., Fang, J., Fuster, V., & O'Rourke, R. (2012a). Functional Anatomy of the Heart. In *Hurst's the Heart - Manual of Cardiology* (13th ed.). New York: McGraw-Hill Education / Medical.

Walsh, R., Fang, J., Fuster, V., & O'Rourke, R. (2012b). Magnetic resonance imaging of the heart: introduction. In *Hurst's the Heart - Manual of Cardiology* (13th ed.). New York: McGraw-Hill Education / Medical.

Wang, Z., Kyono, H., Bezerra, H. G., Wang, H., Gargesha, M., Alraies, C., … Rollins, A. M. (2010). Semiautomatic segmentation and quantification of calcified plaques in intracoronary optical coherence tomography images. *Journal of Biomedical Optics*, *15*(6), 061711. https://doi.org/10.1117/1.3506212

Warfield, S. K., Zou, K. H., & Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, *23*(7), 903–921. https://doi.org/10.1109/TMI.2004.828354

Wit, J. de. (2016, March 15). 3rd place solution for the second national datascience bowl. Retrieved November 14, 2017, from http://juliandewit.github.io/kaggle-ndsb/

Woo, J., Slomka, P. J., Kuo, C.-C. J., & Hong, B.-W. (2013). Multiphase segmentation using an implicit dual shape prior: Application to detection of left ventricle in cardiac MRI. *Computer Vision and Image Understanding*, *117*(9), 1084–1094. https://doi.org/10.1016/j.cviu.2012.11.012

Wu, Yin, Jiang, K., Zhang, N., Gao, Y., Chen, Y., Zheng, H., … Chung, Y.-C. (2015). Efficient method for analyzing MR real-time cines: Toward accurate quantification of left ventricular function. *Journal of Magnetic Resonance Imaging*, *42*(4), 972–980. https://doi.org/10.1002/jmri.24869

Wu, Yuwei, Wang, Y., & Jia, Y. (2013). Segmentation of the left ventricle in cardiac cine MRI using a shape-constrained snake model. *Computer Vision and Image Understanding*, *117*(9), 990–1003. https://doi.org/10.1016/j.cviu.2012.12.008

Yong, Y. L., Tan, L. K., McLaughlin, R. A., Chee, K. H., & Liew, Y. M. (2017). Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography. *Journal of Biomedical Optics*, *22*(12), 126005. https://doi.org/10.1117/1.JBO.22.12.126005

Zhang, H., Wahle, A., Johnson, R. K., Scholz, T. D., & Sonka, M. (2010). 4-D Cardiac MR Image Analysis: Left and Right Ventricular Morphology and Function. *IEEE Transactions on Medical Imaging*, *29*(2), 350–364. https://doi.org/10.1109/TMI.2009.2030799

Zheng, Q., Delingette, H., Duchateau, N., & Ayache, N. (2018). 3-D Consistent and Robust Segmentation of Cardiac Images by Deep Learning With Spatial Propagation. *IEEE Transactions on Medical Imaging*, *37*(9), 2137–2148. https://doi.org/10.1109/TMI.2018.2820742

Zhong, L., Zhang, J.-M., Zhao, X., Tan, R. S., & Wan, M. (2014). Automatic Localization of the Left Ventricle from Cardiac Cine Magnetic Resonance Imaging: A New Spectrum-Based Computer-Aided Tool. *PLoS ONE*, *9*(4), e92382. https://doi.org/10.1371/journal.pone.0092382

Zhou, S. K. (2010). Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram. *Medical Image Analysis*, *14*(4), 563–581. https://doi.org/10.1016/j.media.2010.04.002

Zhuang, X., Hawkes, D. J., Crum, W. R., Boubertakh, R., Uribe, S., Atkinson, D., … Hill, D. L. G. (2008). Robust registration between cardiac MRI images and atlas for segmentation propagation. In *Medical Imaging 2008: Image Processing* (Vol. 6914, p. 691408). International Society for Optics and Photonics. https://doi.org/10.1117/12.769445

Zhuang, X., Rhode, K. S., Razavi, R. S., Hawkes, D. J., & Ourselin, S. (2010). A Registration-Based Propagation Framework for Automatic Whole Heart Segmentation of Cardiac MRI. *IEEE Transactions on Medical Imaging*, *29*(9), 1612–1625. https://doi.org/10.1109/TMI.2010.2047112

Zhuang, Xiahai. (2013). Challenges and Methodologies of Fully Automatic Whole Heart Segmentation: A Review [Research article]. https://doi.org/10.1260/2040-2295.4.3.371