11-2019

Computational Methods for Segmentation of Multi-Modal Multi-Dimensional Cardiac Images

Shusil Dangi
sxd7257@rit.edu

Computational Methods for Segmentation of Multi-Modal Multi-Dimensional Cardiac Images

by

Shusil Dangi

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology

November 2019

Computational Methods for Segmentation of Multi-Modal Multi-Dimensional Cardiac Images

by

Shusil Dangi

B.E. Pulchowk Campus, Institute of Engineering, Tribhuvan University, 2011

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology

2019

Signature of the Author _____

Accepted by _____
Coordinator, Ph.D. Degree Program Date

Rochester Institute of Technology
Imaging Science

CERTIFICATE OF EXAMINATION

Advisor

Dr. Cristian A. Linte

Committee

Dr. Nathan D. Cahill

Dr. Christopher Kanan

Dr. Ziv Yaniv

The dissertation by

Shusil Dangi

entitled

Computational Methods for Segmentation of Multi-Modal Multi-Dimensional Cardiac Images

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Imaging Science

Date

Dr. Daniel Phillips, Outside Chair

Dedicated to my family, friends, and mentors who have supported me along this long journey and helped me find the guidance, patience and strength to carry on.

Abstract

Segmentation of the heart structures helps compute the cardiac contractile function quantified via the systolic and diastolic volumes, ejection fraction, and myocardial mass, representing a reliable diagnostic value. Similarly, quantification of the myocardial mechanics throughout the cardiac cycle, analysis of the activation patterns in the heart via electrocardiography (ECG) signals, serve as good cardiac diagnosis indicators. Furthermore, high quality anatomical models of the heart can be used in planning and guidance of minimally invasive interventions under the assistance of image guidance.

The most crucial step for the above mentioned applications is to segment the ventricles and myocardium from the acquired cardiac image data. Although the manual delineation of the heart structures is deemed as the gold-standard approach, it requires significant time and effort, and is highly susceptible to inter- and intra-observer variability. These limitations suggest a need for fast, robust, and accurate semi- or fully-automatic segmentation algorithms. However, the complex motion and anatomy of the heart, indistinct borders due to blood flow, the presence of trabeculations, intensity inhomogeneity, and various other imaging artifacts, makes the segmentation task challenging.

In this work, we present and evaluate segmentation algorithms for multi-modal, multi-dimensional cardiac image datasets. Firstly, we segment the left ventricle (LV) blood-pool from a tri-plane 2D+time trans-esophageal (TEE) ultrasound acquisition using local phase based filtering and graph-cut technique, propagate the segmentation throughout the cardiac cycle using non-rigid registration-based motion extraction, and reconstruct the 3D LV geometry. Secondly, we segment the LV blood-pool and myocardium from an open-source 4D cardiac cine Magnetic Resonance Imaging (MRI) dataset by incorporating average atlas based shape constraint into the graph-cut framework and iterative segmentation refinement. The developed fast and robust framework is further extended to perform right ventricle (RV) blood-pool segmentation from a different open-source 4D cardiac cine MRI dataset. Next, we employ convolutional neural network based multi-task learning framework to segment the myocardium and regress its area, simultaneously, and show that segmentation based computation of the myocardial area is significantly better than that regressed directly from the network, while also being more interpretable. Finally, we impose a weak shape constraint via multi-task learning framework in a fully convolutional network and show improved segmentation performance for LV, RV and myocardium across healthy and pathological cases, as well as, in the challenging apical and basal slices in two open-source 4D cardiac cine MRI datasets.

We demonstrate the accuracy and robustness of the proposed segmentation methods by comparing the obtained results against the provided gold-standard manual segmentations, as well as with other competing segmentation methods.

Keywords: image segmentation; image registration; cardiac ultrasound; cine MRI; cardiac function; atlas segmentation; graph-cut; convolutional neural network; fully convolutional network; multi-task learning

Acknowledgments

Firstly, I am indebted to my advisor Dr. Cristian A. Linte for his support and guidance throughout this long journey. As one of his first PhD students, his enthusiasm, dedication, and perseverance, during his early academic career has been a constant source of inspiration for me. I appreciate the long nights he committed during paper submission deadlines, countless moments of discussion on exciting new ideas, hours he spent on editing the papers and helping perfect my presentation, and the funding contributions he has made to make my Ph.D. experience productive and stimulating. I am grateful for his support during my internship opportunities to help me better prepare for my career. I am truly honored to have Cristian as my advisor and wish to have him as a mentor for the rest of my career.

Besides my advisor, I would like to thank the rest of my dissertation committee members: Dr. Ziv Yaniv, Dr. Nathan D. Cahill, Dr. Christopher Kanan, and Dr. Daniel Philips, for their constructive criticisms and invaluable advise. The committees feedback during the thesis proposal was highly valuable in improving this dissertation. Furthermore, the faith this committee has put on my potential has helped me excel further and push myself towards excellence.

I would like to give a special thanks to Dr. Ziv Yaniv for providing me an opportunity of a summer internship at National Library of Medicine (NLM). I am also very grateful to have received his weekly technical guidance and encouragement following the internship. Ziv has been an exceptional mentor in my PhD journey, and I wish to seek his guidance moving forward in my career.

I am grateful to Dr. Andinet Enquobahrie for providing me an internship opportunity at Kitware. It was a great learning experience about open-source softwares, corporate culture, and team work. Six months I spent during the internship helped be grow both personally and professionally.

I would also like to thank Dr. Nathan D. Cahill for his technical guidance during the early days of my Ph.D. I am also very thankful to Dr. Karl Q. Schwarz for providing his expertise in cardiac diagnosis and treatment, and giving me an opportunity to visit his hospital to get a first-hand experience of several medical imaging

and diagnosis techniques in clinical practice.

I am very thankful to my lab mates for their continued support. Specially, Dr. Yehuda Kfir Ben-Zikri provided me with help and support in and out of the lab during my first few years of Ph.D., before his graduation. The discussions and collaboration with Aditya Daryanani, Dawei Liu, and Roshan Reddy Upendra have been enjoyable and fruitful. I would also like to thank other lab mates Golnaz Jalalahmadi and S. M. Kamrul Hasan for making the graduate school exciting and fun.

I would also like to thank the members of our reading group: Sandesh Ghimire, Sushant Kafle, Manoj Acharya, Prasanna K. Gyawali, Kishan KC, and Aayush Chaudhary. The discussions in the group were helpful to brainstorm new project ideas with a mix of intellectual fun. I am also thankful to my roommate Utsav Gewali for his help and support throughout the Ph.D. program.

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R35GM128877, the Office of Advanced Cyber infrastructure of the National Science Foundation under Award No. 1808530, and the Intramural Research Program of the U.S. National Institutes of Health, National Library of Medicine. Hence, I would like to express my sincere gratitude to the funding agencies that made this work possible.

Last but not the least, I would like to express my deepest gratitude to my father Mani Raj Dangi and mother Saraswati Dangi for always believing in me. I would also like to thank my younger sister Sunita Dangi and younger brother Sujan Dangi for their constant love and support. I am also very grateful for the encouragement and assistance from my cousins Kalpana Thapa and Sanchita Thapa during this journey. Finally, I would like to thank all my family and friends without whose unconditional love and continuous support this dissertation would not have been possible.

Shusil Dangi
Rochester Institute of Technology
November 2019

Contents

List of Tables

List of Figures

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
4D	Four-Dimensional
AAM	Active Appearance Model
ACDC	Automated Cardiac Diagnosis Challenge
API	Application Programming Interface
ASM	Active Shape Model
BOLD	Blood Oxygen Level Dependent
BP	Blood Pool
BG	Background
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CRF	Conditional Random Field
CT	Computed Tomography
CTA	Computed Tomography Angiography
DCM	Dilated Cardiomyopathy
DFCN	Densely Connected Fully Convolutional Network
DICOM	Digital Imaging and Communications in Medicine
DM	Dice Metric
DM	Distance Map
DMR	Distance Map Regularized
DMTRL	Deep Multitask Relationship Learning Network
ECG	Electrocardiogram

ED	End-diastole
EDV	End-diastolic Volume
EF	Ejection Fraction
EM	Expectation-Maximization
ES	End-systole
ESV	End-systolic Volume
FA	Fully-Automatic
FCN	Fully Convolutional Network
FIMH	Function Imaging and Modeling of the Heart
FOV	Field of View
GMM	Gaussian Mixture Model
GPU	Graphics Processing Unit
GVF	Gradient Vector Flow
HCM	Hypertrophic Cardiomyopathy
HD	Hausdorff Distance
HU	Hounsfield Unit
ICE	Intracardiac Echocardiography
ICM	Iterated Conditional Mode
IoU	Intersection-over-Union
IPP	Image Orientation Patient
ITK	Insight Segmentation and Registration Toolkit
JM	Jaccard Metric
KHz	Kilohertz
LA	Long-Axis
LoA	Limits of Agreement
LoG	Laplacian of Gaussian
LRW	Lazy Random Walk
LV	Left Ventricle
LVSC	Left Ventricle Segmentation Challenge
MAD	Mean Absolute Distance

MAD	Mean Absolute Difference
MAP	Maximum A Posteriori
MHz	Megahertz
MICCAI	Medical Image Computing and Computer Assisted Intervention
MINF	Myocardial Infarction
MR	Magnetic Resonance
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
MSD	Mean Squared Difference
MSD	Mean Surface Distance
MST	Minimal Spanning Tree
MTL	Multi-Task Learning
MTN	Multi-Task Network
NCC	Normalized Cross-Correlation
NMR	Nuclear Magnetic Resonance
NPV	Negative Predictive Value
PC	Personal Computer
PET	Positron Emission Tomography
PPV	Positive Predictive Value
RAM	Random Access Memory
ReLU	Rectified Linear Unit
RF	Radio Frequency
ROI	Region of Interest
RV	Right Ventricle
RVSC	Right Ventricle Segmentation Challenge
SA	Simulated Annealing
SA	Semi-Automatic
SA	Short-Axis
SD	Standard Deviation
SOS	Stochastic Outlier Selection

SPECT	Single Photon Emission Computed Tomography
SPIE	The International Society for Optics and Photonics
SSFP	Steady State Free Precision
STACOM	Statistical Atlases and Computational Models of the Heart
STAPLE	Simultaneous Truth and Performance Level Estimation
TE	Echo Time
TEE	Transesophageal Echocardiogram
TR	Repetation Time
TRE	Target Registration Error
TTE	Transthoracic Echocardiogram
US	Ultrasound
WHO	World Health Organization

Chapter 1

Introduction, Background, and Thesis Overview

The goal of this chapter is to provide the reader with an overview of the clinical background for the proposed work, along with a brief and timely review of the literature on medical image segmentation, while identifying current clinical challenges.

1.1 Medical Imaging

Medical imaging is a noninvasive technique of creating the visual representation of the structure and function of interior organs of a body. When a body is exposed to some form of energy that can penetrate through and interact with the tissues, the detected signal containing information about the anatomical interaction can be used to construct an image [1]. Hence, medical imaging can be interpreted as a solution to a mathematical inverse problem, where the properties of tissue (cause) is inferred from the observed energy signal (effect).

Visible light energy is used mostly outside the radiology department in light microscopy [2], endoscopy [3], and optical coherence tomography [4], due to its limited ability to penetrate tissues at depth; whereas the electromagnetic spectrum outside of the visible regime is typically used in diagnostic radiology. Depending on the type of energy and the acquisition technology used, different modalities of medical

images can be acquired. Some common imaging modalities routinely used in clinical radiology are x-rays, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), and nuclear medicine — single photon emission computed tomography (SPECT), positron emission tomography (PET) [5–8]. There is no single best imaging modality; rather, different imaging modalities are suitable for different applications and provide complementary information about the patient.

1.1.1 X-Ray Imaging

X-ray is the first imaging modality discovered in 1895 by German physicist Conrad Roentgen, who received the Nobel Prize in 1901 as he demonstrated its diagnostic abilities for imaging human body. Diagnostic X-rays have a wavelength between 0.01nm and 0.1nm (corresponding to 123 keV to 12.3 keV energy range) with reasonable attenuation to discriminate bone, soft tissue, and air. Since X-ray photons carry enough energy to ionize atoms and disrupt molecular bonds, it is an ionizing modality, and is harmful to living tissue with increased risk of radiation-induced cancer [9].

During imaging, the X-rays are transmitted through the body and collected on a film or an array of detectors. Since the X-rays are attenuated more by bones than soft tissues or air, the collected two-dimensional (2D) attenuation map serves as an image with excellent spatial resolution. Despite its harmful effects, X-ray is extensively used in the diagnosis of broken bones, lung cancer, breast cancer, etc. when the risk is greatly out-weighted by the benefits of the examination.

1.1.2 Computed Tomography Imaging

Two-dimensional images produced by X-ray are not adequate for many diagnostic applications requiring three-dimensional (3D) quantitative and qualitative information about the anatomical structures. This led to the development of Computed Tomography (CT); tomography referring to a picture (graph) of a slice (tomo). The X-ray source and detector are mounted on a gantry that rotates around the patient capturing 2D projections from multiple angular view-points, which are then reconstructed into a 3D axial slice through the patient via back-projection algorithm

[7, 10, 11]. The patient lying on a movable bed is moved axially to acquire multiple axial slices that are stacked together to produce a 3D CT image.

The reconstructed image represents the linear attenuation coefficient map of the scanned object, which is converted to the standard Hounsfield unit (HU) corresponding to the actual intensity values of the CT image.

$$CTnumber = \frac{\mu - \mu_{water}}{\mu_{water}} \times 1000 \tag{1.1}$$

In a CT image, the intensity value for water is 0 HU, air is -1000 HU, and soft tissue, bones have values from several hundred to several thousand HU, respectively.

The capability of modern helical and multislice CT scanners to produce very high quality (less than 0.5mm isotropic resolution) full body scans in less than 1 minute acquisition time has established it as the most widespread diagnostic imaging modality. CT scans are routinely used in clinic for head/full-body scan for diagnosis of accident injuries, dental planning, detection of lung nodules, diagnosis of lung emphysema, etc. Similarly, contrast enhanced CT is used for perfusion analysis of brain, liver, and tumors, as well as analysis of vessels for stenoses and aneurysms. Furthermore, the fast acquisition time of CT is suitable for cardiac imaging, hence, is used for quantification of coronary artery calcification and examining the dynamic motion of the heart muscles to detect abnormalities.

As CT scan involves multiple x-ray acquisitions, the patient is exposed to high ionizing radiation, increasing the risk of cancer [12]. Hence, since the development of the first clinical CT scanner in early 1970s, the field is mostly driven by the motivation of reducing the acquisition time and lowering the radiation dose, while maintaining the quality of images.

1.1.3 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a tomographic imaging technique that produces 3D images of the human body based on the Nuclear Magnetic Resonance (NMR) phenomenon. As the human body is mostly composed of fat and water, consisting of hydrogen atoms, the NMR signals from the nucleus of these hydrogen atoms

can be recorded by applying an external magnetic field gradient and appropriate radio frequency (RF) pulse sequence to record k-space data, which is inverse fourier transformed to construct an MRI image [13, 14].

The density of hydrogen atoms for fat, muscle, and other tissues are different, hence rendering MRI as a powerful soft-tissue contrast imaging method. It is a versatile modality, as different tissues can be highlighted in the images by changing the acqusition parameters such as RF pulse sequence, repetation time (TR), and echo time (TE). Similarly, the image slices can be acquired in any direction by changing the external magnetic field gradients. Furthermore, with the faster acquisition of images by exploiting the mathematical properties of the k-space, parallel acquisition techniques, and the advent of new RF pulse sequences, it is possible to perform MRI angiography, diffusion imaging, as well as functional imaging based on the blood oxygen level dependent (BOLD) response.

MRI has been a popular modality for medical diagnosis due to the lack of ionizing radiation and high contrast sensitivity to soft tissues. However, it is limited by slower acquisition speed, high equipment and siting cost, operational complexity, significant imaging artifacts, and MR safety concerns.

MRI is the preferred cardiac imaging modality due to its capability of tissue characterization. Specifically, the steady state free precision (SSFP) pulse sequence [15], which reduces the acquisition time while maintaining a good signal-to-noise ratio as well as good blood-myocardium contrast, is coupled with electrocardiogram (ECG) gating to produce movie of a heart slice throughout the cardiac cycle, known as cine MRI acquisition. The short-axis cine MR slices covering the whole heart are stacked together to generate a pseudo four-dimensional (4D) volume, which can be used to perform quantitative analysis of cardiac indices [16]. However, since very few slices are acquired due to higher acquisition time, and the slices might be misaligned due to breathing/patient-motion, 3D analysis of cine MR images is challenging. In this thesis, we use open-source cine MRI datasets for cardiac image segmentation.

1.1.4 Ultrasound Imaging

Sound waves with frequencies above 20 kilohertz (KHz), the upper audible limit of human hearing, are termed as ultrasound (US). US imaging devices constitute of piezoelectric crystal-based transducer that produces the ultrasound beam (typically 1-15 megahertz (MHz)) as well as receive the returned echo from the tissue [17].
A-mode imaging is the simplest form of US imaging where the transducer transmits US pulses through the patient body such that the time and amplitude of the received echos provide the location information and the tissue structures and interfaces along the path, respectively. Hence, a vector image with spatial dimension representing the location and the intensity representing the amplitude of received echo is generated.
M-mode images are formed by arranging A-mode vector images, along a fixed US beam at different time instances, in columns of a 2D matrix. These images are useful to analyze moving objects inside the body.
B-Mode images are constructed by pivoting the transducer at a point about an axis acquiring several A-mode vector images, along a V-shaped imaging region, and combining them into a 2D matrix. This is the most commonly used imaging mode.

Real-time 3D US imaging is now possible using 2D array of transducer with beam-forming electronic in the transducer handle. Similarly, doppler US can be used to measure the flow velocities based on the shift of frequency in an ultrasound wave due to blood/liquid flow. Furthermore, gas-filled encapsulated microbubbles can be used as contrast agents to enhance the echo amplitudes and hence the image contrast.

Due to low-cost, portability, real-time, and ionizing-radiation-free acquisition, US is used in various clinical applications such as breast, cardiac, gynecologic, obstetrics, pediatrics, and vascular imaging. However, the images suffer from various artifacts such as *speckle noise* due to the random alignment of sound waves reflected on microscopic tissue inhomogeneties, *shadows* casted behind a strong reflecting object, *multiple reflections* between two strong reflectors displayed as multiple echos, and *mirroring* when an object is placed between the transducer and a strongly reflecting layer [6].

Cardiac US imaging is routinely used in clinics to evaluate the cardiac contractile

functions, blood-flow measurement using doppler imaging, and myocardial deformation evaluation using strain imaging [18]. Transthoracic echocardiogram (TTE) is the most common non-invasive assessment, where the heart is imaged by placing the US probe on the chest or abdomen of the patient. However, the quality of acquired image is low due to the signal attenuation by layer of fat and muscle on the US path. Transesophageal echocardiography (TEE) produces better images by passing a specialized probe containing an ultrasound transducer at its tip into the patient's esophagus and imaging the heart from close. Similarly, Intracardiac echocardiography (ICE) performed through a venous or arterial sheath is able to produce high resolution images from within the heart without the need for general anesthesia. However, it has limited field-of-view and high associated cost because of single-use catheters. Hence, TEE is generally the preferred US protocol in the clinic. In this thesis, we use the multi-plane TEE image sequences for LV segmentation and 3D reconstruction.

1.1.5 Nuclear Medicine Imaging

Nuclear medicine [19] refers to a branch of radiology where a patient is injected with substances containing radioactive isotope such that the x- and/or gamma rays emitted during radioactive decay detected by a radiation detector is used to make projection images. The acquired multiple projection images are back projected to reconstruct a tomographic slice, similar to CT imaging. It is a functional imaging modality since it provides physiological information of imaged organs, as the emissivity of a healthy tissue is different from a diseased one. Hence, the nuclear images are usually coupled with the structural CT images to analyze both the structure and function of organs.

Most popular nuclear medicine imaging techniques are SPECT and PET. In SPECT, the tomographic images are reconstructed from the x- or gamma-ray emissions from the patient detected by a nuclear camera at multiple angles around the patient. In PET, the positron (e^+) emitted during the decay of an isotope combines with an electron (e^-) to produce an annihilation radiation emitting two photons in opposite directions; the detected photon pairs give information about a straight line

along which the annihilation event took place and can be used to compute the 3D distribution of the PET agent and hence generate a tomographic emission image. PET is more sensitive to small physiological changes in tissues than SPECT at an expense of higher imaging cost.

Typically the functional images acquired from SPECT and PET are combined with structural CT images; SPECT-CT and PET-CT imaging are often used in cardiology, oncology, neurology, and imaging of infection and inflammation.

1.2 Image Segmentation

Image segmentation refers to the grouping of pixels/voxels in an image based on common properties such as intensity, color, texture, and location [20, 21]. The segmentation task assigns a label to each pixel/voxel in an image based on its features. Clinicians often require to delineate an organ, tumor, vessels, etc from the acquired X-ray, MRI, CT, or US images for diagnostics, planning and guidance. Although the manual delineation is referred to as the gold-standard, it is time intensive (specially for 3D/4D images), and is prone to intra- and inter-observer variability. Hence, it is often desirable to perform semi-/fully-automatic segmentation that can be manually adjusted if desired.

Based on the amount of prior knowledge used, image segmentation algorithms can be broadly classified into (i) No prior, (ii) Weak prior, (iii) Strong prior, and (iv) Machine learning based methods.

1.2.1 No Prior Based Algorithms

These are the simplest kind of segmentation algorithms that do not use any prior information about the shape/geometry of the object and rely only on the pixel/voxel intensity information.

1.2.1.1 Thresholding

Thresholding algorithms determine the intensity thresholds, T_1 and T_2, based on the intensities of the input image, $I(\boldsymbol{x})$, such that the foreground/background separated binary image, $S(\boldsymbol{x})$, can be obtained as:

$$S(\boldsymbol{x}) = \begin{cases} 1 & \text{if } T_1 \leq I(\boldsymbol{x}) \leq T_2 \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

Otsu's method [22] provides an optimal global threshold that minimizes the intra-class intensity variance while maximizing the inter-class intensity variance for an image with bi-modal histogram. Multi-level thresholding can be used to delineate multiple foreground objects from the background. Similarly, adaptive thresholding based on the local image statistics is preferred when the image statistics vary significantly in different image regions, possibly due to imaging artifacts.

1.2.1.2 Edge Detection and Linking

Edges can be detected from an image using a discrete approximation of the derivative operator along each spatial dimension, based on the intensity differences between the neighboring pixels. The first order derivative operator is approximated by several discrete filters such as Roberts, Prewitt, and Sobel. Similarly, the second order derivative operator is approximated by a discrete Laplacian filter.

Due to the high sensitivity of derivative operator to noise, usually the image is smoothed by a Gaussian filter before edge detection. Laplacian of Gaussian (L0G) filtering can produce fine edges based on zero-crossings. The obtained discontinuous edges can be linked based on the proximity, edge strength, and edge direction, to generate a continuous boundary. Canny edge detection [23] algorithm uses a multi-stage pipeline to produce fine boundaries of the objects in an image. Similarly, Hough transform [24] can be used to detect the geometric objects such as line, circle, and ellipse from the edges.

1.2.1.3 Region Growing

Region growing algorithm groups pixels or subregions into a larger region based on a predefined criteria. The algorithm requires manual/automatic "seed" points, such that the neighboring pixels with features (intensity, texture) similar to the seed are appended to grow the region. Region growing stops when no more neighboring pixels satisfy the inclusion criteria, resulting in segmented image with multiple foreground regions.

1.2.1.4 Morphological Watershed

In the morphological watershed algorithm [25], the images are interpreted as a topographical map with image intensity representing an additional dimension. For example, a 2D image can be viewed as a 3D topographical map with the third dimension representing the pixel intensity. The pixels with minimum intensity within a topographical region are termed as regional minimum, which can be obtained via image smoothing followed by local/global thresholding.

The basic principle of the watershed algorithm is to punch a hole in each regional minimum and flood the entire topography by letting water rise through the holes at a uniform rate. When the rising water in distinct catchment basins is about to merge, a dam is built to prevent the merging. At the end of flooding, when only the top of the dams are visible above the water line, the dam boundaries dividing the watersheds corresponds to the continuous boundary between the segmented regions.

1.2.2 Weak Prior Based Algorithms

These algorithms pose segmentation as a global energy minimization problem in an image, while imposing weak constraints such as piecewise continuity of segmented regions, and smooth curvature of the segmented boundary. The use of global image context and weak prior knowledge generates superior segmentation results compared to the no prior based algorithms, at an expense of higher computation cost.

1.2.2.1 Deformable Models

These models pose the segmentation problem as an energy minimization problem in variational framework.

Active Contours or Snakes

Active contours are the parameterized spline curves that are evolved based on the internal spline energy and the external image forces, to lock onto nearby edges of the objects to be segmented in an image [26]. The snake is represented parametrically by $\boldsymbol{v}(s) = (x(s), y(s))$, where $0 \leq s \leq 1$, with the total energy functional represented as the sum of internal and external energy functionals:

$$E^*_{snake} = \int_0^1 E_{int}(\boldsymbol{v}(s)) + E_{ext}(\boldsymbol{v}(s)) \tag{1.3}$$

Where the internal spline energy can be written as:

$$E_{int}(\boldsymbol{v}(s)) = \frac{1}{2}\left(\alpha(s)|\boldsymbol{v}_s(s)|^2 + \beta(s)|\boldsymbol{v}_{ss}(s)|^2\right) \tag{1.4}$$

The first- and second-order terms control the amount of stretch and the amount of curvature in the snake and are weighted by $\alpha(s)$ and $\beta(s)$, respectively. The external energy term, $E_{ext}(\boldsymbol{v}(s))$, is formulated as the weighted combination of the image intensity, negative image gradient magnitude, and the curvature of level contours in the original paper [26]. Starting with an initial curve selected manually/automatically, the Euler equations corresponding to the energy functional of equation **1.3** are solved iteratively until convergence to obtain the stationary point, yielding the final segmentation result.

Due to the internal spline energy forces, the snake would collapse into a point or a line (depending on if the curve is closed or open, respectively) if placed in a region with uniform intensity. Cohen [27] proposed a modification to the external energy to include a "balloon" (pressure) force acting outward (or inward) in the normal direction of the curve, allowing the curve to inflate (or deflate) and hence avoid the collapse. The direction of the pressure force could be changed depending on the requirement of expanding/shrinking the snake.

The snake formulation still suffers from two major limitations: the curve has to be initialized fairly close to the final solution due to the limited capture range of the image gradient force, and the convergence to boundary concavities is poor. These limitations were solved by the introduction of gradient vector flow (GVF) [28] as the external energy term in the snake formulation. The GVF, computed as a diffusion of the gradient vectors of a gray-level or binary edge map derived form the image, increases the capture range of snakes as well as help them move into the boundary concavities. Several modifications for the external energy term have been proposed in the literature to further improve the performance of the snakes.

Level Sets

The level set framework [29] can handle topology changes during contour evolution, process multiple contours simultaneously, and allow cusps and corners, which is not possible with the parametric representation of curve in active contour model. The interface, $d\Omega$, (a curve in 2D or a surface in 3D) is represented implicitly by zero-contour of a higher dimension Lipschitz continuous function, ϕ, as $d\Omega(t) = \{\boldsymbol{x}|\phi(t,\boldsymbol{x}) = 0\}$. In practice the signed distance function is chosen as the level-set function:

$$\phi(\boldsymbol{x}) = \begin{cases} -d \text{ for } \boldsymbol{x} \in \Omega^- \\ +d \text{ for } \boldsymbol{x} \in \Omega^+ \\ 0 \text{ for } \boldsymbol{x} \in d\Omega \end{cases} \tag{1.5}$$

where d is the Euclidian distance to $d\Omega$. The evolution equation of the level set function can be written as:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0, \quad \phi(0,\boldsymbol{x}) = \phi_0(\boldsymbol{x}) \tag{1.6}$$

where, the function F is called the speed function and the set $\{\boldsymbol{x}|\phi_0(\boldsymbol{x} = 0)\}$ defines the initial interface. The level set function can be used to compute the normal to the interface $\boldsymbol{n}$ and the interface's mean curvature κ as:

$$\boldsymbol{n} = \nabla\phi/|\nabla\phi| \quad \text{and} \quad \kappa = \nabla \cdot \boldsymbol{n}$$

For the motion by mean curvature, the speed function can be selected as $F = \kappa$. Similarly, the speed function can be modified by a monotonously decreasing function of the image gradient magnitude, $g(|\nabla I_0|)$, to stop the level set evolution on the desired boundaries [30].

Since the level set function does not retain its signed distance function properties as it evolves in time through equation 1.6, it needs to be reinitialized periodically after every few iterations [31]. The reinitialization might move the level set incorrectly, while increasing the computation cost. Hence, Li et al.[32] proposed a variational framework that penalized the movement of level set away form the signed distance function and did not require reinitialization during the evolution.

Caselles et al.[33] formulated the classical energy-based active contour model (1.2.2.1) as a problem of finding a geodesic curve in a Riemannian space derived from the image content; equivalent to finding a curve of minimal weighted length in certain framework. This level set based geodesic active contour framework exploited the connection between the classical energy-based model and the intrinsic level set model. The geodesic formulation introduced a new term that further attracted the deforming curve to the boundary and hence improved the convergence, while also reducing the number of hyper-parameters.

The early active contour and level set formulations used image gradients as the external force to determine the object boundaries, hence, they struggled with noisy images and failed to segment objects with weak/smooth boundaries. Chan and Vese [34] proposed a region-based segmentation model based on the minimization of Mumford-Shah functional [35] for segmentation. The corresponding energy minimization problem is equivalent to the minimal partition problem, and is solved via level set evolution using finite difference approximation. Since the method does not depend on image gradients, the energy landscape is smoother, resulting in better convergence, even with poor initializations. Most recent level-set segmentation methods employ both gradient- and region-based external energy terms to obtain more robust segmentation results.

1.2.2.2 Graph Theoretical Models

Graph theoretical models represent an image as a discrete graph and pose the segmentation as a combinatorial graph partitioning problem. For a graph $G = (V, E)$, $V = \{v_1, ..., v_n\}$ is a set of vertices representing the image elements (pixels/voxels/super-pixels), and E is a set of edges connecting pairs of neighboring vertices. Each edge $(v_i, v_j) \in E$ is weighted by a weight $w(v_i, v_j)$ based on the properties of the two vertices connected by the edge, hence introducing weak constraint into the framework. The image segmentation problem is formulated as a partitioning of the graph G into mutually exclusive connected sub-graphs $G_s = (V_s, E_s)$, where $V_s \subseteq V$, $E_s \subseteq E$, and $s = \{1, 2, ..., k\}$, such that $\{G_i \cap G_j = \phi : i, j \in \{1, 2, ...k\}, i \neq j\}$ and $G = \bigcup_{s=1}^{k} G_s$. Where the properties of vertices (intensity, texture, color, etc) within a sub-graph should be similar, while that between different sub-graphs should be dissimilar.

The degree of dissimilarity between the subgraphs can be computed as a graph cut. A cut partitions the graph into disjoint connected sub-graphs G_a and G_b and its value is defined as:

$$cut(G_a, G_b) = \sum_{v_a \in G_a, v_b \in G_b} w(v_a, v_b) \tag{1.7}$$

which is the sum of weights of all the edges connecting G_a and G_b. Hence, image segmentation is equivalent to finding the optimal cut in the graph. A comprehensive survey of graph theoritical models for image segmentation can be found in [36]

Minimal spanning tree based methods

A spanning tree is an acyclic sub-graph (tree) including all vertices of the connected graph, G, but with a single path between any two vertices. Out of multiple spanning trees in a graph, the minimal spanning tree (MST) is the one with smallest edge weights; it can be computed using several algorithms [37–39]. For example, Kruskal's algorithm [37] list all of the edges in ascending order and adds edge connected to the MST with smallest weight, without generating any cycles. Next, if we consider a binary image segmentation problem, the task is to find and remove a single edge dividing the MST into two connected sub-graphs of descent sizes, such that the

intra-class variance is minimized while the inter-class variance is maximized. In case of multi-class labeling, multiple edges can be removed to obtain multiple connected sub-graphs representing segmented regions. MST based image segmentation can be found in several papers [40–42]

Shortest path based methods

Shortest path based methods obtain the segmentation boundary by finding the shortest path between vertices in a weighted graph. Most of these methods rely on manual input points to guide the segmentation boundary to the desired location. *Intelligent Scissors* [43] framework allows objects in a digital image to be extracted quickly and accurately using mouse gestures. An user starts tracing from a point and as s/he move the cursor closer to the boundary of the object, the live-wire boundary snaps to the object boundary. The live-wire boundary position is the shortest path from the previous boundary point to the current cursor location computed using optimal graph search (Dijkstra [44]) and dynamic programming in a 2D weighted graph, with edge weights defined as the weighted combination of Laplacian zero-crossing, gradient magnitude, and gradient direction. Similar algorithm based on the user interaction using mouse cursor has been proposed in [45]. Furthermore, 3D extensions of the shortest path based segmentation methods can be found in [46, 47].

Geodesic shortest path formulation computes the geodesic distance of each pixel to the labeled forground or background pixels and assigns a label to the pixel depending on the shortest geodesic path [48]. The geodesic paths are weighted according to the image contents.

Random walk based methods

Random walk based methods represent an image as a graph with edge weights proportional to the image gradients. An interactive multi-class segmentation of the image can be obtained based on the seed points provided by an user, such that the probability of a particular pixel assigned to class k is determined by the probability of a random walker starting at that pixel first reaching the k^{th} seed point. Hence, for each

pixel a tuple of k probabilities are computed corresponding to k-classes, and the final segmentation is obtained by assigning a pixel to the class with highest probability. Grady [49] presented an algorithm to compute random walk probabilities by solving a set of sparse linear systems, obtaining good segmentation results on synthetic and real images.

Shen et al.[50] proposed a lazy random walk (LRW) algorithm [51] for superpixel segmentation. Adding a self-loop over the graph vertex makes the random walk process lazy and help make full use of the global relationship between the pixel and all the seeds. A vertex with heavy self-loop is more likely to absorb its neighboring pixels than the one with light self-loop, which enables the vertex to absorb and capture both the weak boundary and texture information. Furthermore, since the LRW algorithm computes the commute time from seed-point to other pixels, as opposed to starting from the pixels to the seed-point in original random walk algorithm, it produces better probability maps, hence generating excellent superpixel results.

Several modifications of random walk algorithm (sometimes in conjunction with other methods) have been proposed for segmentation of various structures from medical images [52–54].

Graph cut based on Spectral Clustering

Spectral clustering is a dimensionality reduction technique which uses specturm (eigenvectors) of the graph similarity matrix for data clustering. It has an equvalent interpretation as a graph cut and can be applied for image segmentation.

The weighted *adjacency matrix* of the undirected graph, $G = (V, E)$, is the symmetric $n \times n$ matrix $W = w(v_i, v_j)$. The degree of a vertex $v_i \in V$ is $d_i = \sum_{j=1}^{n} w(v_i, v_j)$. The *degree matrix* D is defined as the diagonal matrix with the degrees $d_1, ..., d_n$ on the diagonal. The unnormalized graph *Laplacian matrix* is defined as $L = D - W$. If the partition of the graph is represented by $\boldsymbol{x} = [x_1, ..., x_n]^T$, such that $x_i = 1$ or -1 if the vertex v_i belongs to partition G_a or G_b, respectively, then

the value of cut is given by:

$$\begin{aligned} cut(G_a, G_b) &= \sum_{v_a \in G_a, v_b \in G_b} w(v_a, v_b) \\ &= \frac{1}{4} \sum_{i,j=1}^{n} w(v_i, v_j)(x_i - x_j)^2 \\ &= \frac{1}{2} \boldsymbol{x}^T L \boldsymbol{x} \end{aligned} \tag{1.8}$$

To avoid the trivial solution $x_i = 0$, for all i, the following quadratic constraint is imposed: $\boldsymbol{x}^T \boldsymbol{x} = n$. If we relax the requirement on $\boldsymbol{x}$ such that it can be real valued, the approximate solution to this constrained optimization problem can be obtained from the real eigenvector of the *Laplacian matrix* L with smallest eigenvalue.

$$(D - W)\boldsymbol{x} = \lambda \boldsymbol{x} \tag{1.9}$$

However, the smallest eigenvalue (λ_1) is 0, yielding the noninteresting solution $\boldsymbol{x} = [1, ..., 1]$. Therefore the second smallest eigenvalue (λ_2) and the sign of the associated eigenvector yields the optimal solution [55].

Since the minimum cut favors cutting small sets of isolated nodes in the graph, several modifications have been proposed to maintain resonably large clusters. Ratio cut [56] tries to optimize the following cost function:

$$RatioCut(G_a, G_b) = \frac{cut(G_a, G_b)}{|G_a|} + \frac{cut(G_a, G_b)}{|G_b|} \tag{1.10}$$

where, $|.|$ represents the number of vertices in the graph. To obtain the ratio cut, the elements of eigenvector corresponding to λ_2 are assigned to two clusters based on the threshold that minimizes equation 1.10. Heuristically, the elements of the eigenvector, $x_i \in \mathbb{R}$, can be clustered into two groups using *k-means clustering* algorithm to determine the cluster of each vertex [57] .

Shi and Malik [58] introduced the normalized cut cost function to improve the balance between the cluster sizes after the graph cut:

$$Ncut(G_a, G_b) = \frac{cut(G_a, G_b)}{vol(G_a)} + \frac{cut(G_a, G_b)}{vol(G_b)} \tag{1.11}$$

where, $vol(G_a) = \sum_{i \in G_a, j \in G} w(v_i, v_j)$ is the total connection from vertices in G_a to all vertices in the graph, $vol(G_b)$ is defined similarly. The graph partition minimizing equation 1.11 can be obtained by solving the eigenvalue problem

$$(D - W)\boldsymbol{x} = \lambda D\boldsymbol{x} \tag{1.12}$$

The eigenvector corresponding to λ_2 can be used to bipartition the graph. Each sub-graph can be further partitioned recursively if necessary.

Several objective functions for graph partitioning have been proposed to maximize the inter-cluster similarity while minimizing the intra-cluster similarity [59, 60] to improve the clustering performance.

Graph cut on Markov random field models

Undirected graphical model with markov properties, also known as Markov random field (MRF), can encode the local spatial interaction between image pixels. MRF allows probabilistic interpretation of image segmentation, where an image, $\mathbf{x} = \{\mathbf{x_1}, ..., \mathbf{x_n}\}$, has a random variable, $\mathbf{x_i}$, associated with each pixel. Each $\mathbf{x_i}$ needs to be assigned a label, $\mathcal{L}$ (foreground/background), during image segmentation.

The joint distribution of image label, $\mathbf{x}$, can be written as a product of distributions of *maximal cliques*. Where a *clique* refers to a subset of nodes in a graph with an edge beween each node; *maximal clique* is a clique with maximum number of nodes. Let C be a clique and $\mathbf{x}_C$ be the set of variables in that clique, then the joint distribution can be written as:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C) \tag{1.13}$$

where $\psi_C(\mathbf{x}_C)$ is potential function of maximal clique C. Z is a normalizing constant called the *partition function* and is given by:

$$Z = \sum_{\mathbf{x}} \prod_C \psi_C(\mathbf{x}_C) \tag{1.14}$$

which ensures $p(\mathbf{x})$ is a probability distribution. The potential functions are restricted to be non negative, $\psi_C(\mathbf{x}_C) \geq 0$, ensuring $p(\mathbf{x}) \geq 0$ (Hammersley and Clifford [61]).

The potential functions can be expressed as *Gibbs Distribution* [62, 63]:

$$\psi_C(\mathbf{x}_C) = \exp\{-E(\mathbf{x}_C)\} \tag{1.15}$$

where $E(\mathbf{x}_C)$ is the *energy function* of maximal clique $\mathbf{x}_C$. Hence, the joint distribution defined as the product of clique potentials is equivalent to the Gibbs distribution of the total energy, which is the sum of energies of all maximal cliques. The maximum *a posteriori* (MAP) estimate of the joint distribution yields optimum labeling of the image, which is equivalent to the minimum energy configuration (equation 1.15).

For a first order MRF, the total energy can be represented as:

$$E(f) \;=\; \sum_{\{p,q\}\in\mathcal{N}} V_{p,q}(f_p, f_q) \;+\; \sum_{p\in\mathcal{P}} D_p(f_p), \tag{1.16}$$

where the first term represents smoothness energy which enforces spatial smoothness between pixels in a set of interacting pairs, $\mathcal{N}$, and the second term represents the data energy defined for each pixel based on its likelihood of being labeled f_p.

Gemen and Geman [62, 63] first introduced the Bayesian image restoration using MRF image model. They used the Gaussian distribution likelihood and ising model prior to obtain the posterior distribution in the Bayesian framework. The MAP inference of the posterior distribution was performed by serial Gibbs sampling with simulated annealing (SA) [64] schedule. Besag [65] later proposed MAP estimation using *Iterated Conditional Mode* (ICM). ICM finds the mode for each node conditioned on all the neighbors based on the current estimate of variables, and synchronously updated the whole image repetadely until convergence. ICM can obtain a local maxima significantly faster with no guarantee of global convergence, while SA has a guaranteed global convergence provided a slow cooling schedule.

Greig et al.[66] discovered the minimum energy configuration of the MRF (correspondingly the MAP) is equivalent to the minimum cut in the graph. Hence, they computed the exact MAP estimate for a binary image restoration problem using the Ford-Fulkerson algorithm [67], which states that the minimum cut in a graph can be obtained from the maximum flow. Similarly, Wu and Leahy [68] performed data clustering using the maxflow-mincut approach on a specially constructed graph, and applied it on image segmentation.

The traditional SA and ICM optimization algorithms are slow because they allow small moves where only one pixel changes its label at a time. Boykov et al.[69] proposed two algorithms based on graph cuts that efficiently find a local minimum with respect to two types of large moves (large set of pixels changing labels) called expansion and swap moves. These algorithms have strong convergence guarantee within a known factor of the global minimum for two general classes of interaction penalty V: metric and semimetric. V is called a *metric* on the space of labels $\mathcal{L}$ if it satisfies:

$$V(\alpha, \beta) = 0 \leftrightarrow \alpha = \beta, \tag{1.17a}$$

$$V(\alpha, \beta) = V(\beta, \alpha) \geq 0, \tag{1.17b}$$

$$V(\alpha, \beta) \leq V(\alpha, \gamma) + V(\gamma, \beta) \tag{1.17c}$$

for any labels $\alpha, \beta, \gamma \in \mathcal{L}$. If V satisfies only (1.17a) and (1.17b), it is called a *semimetric*[1]. Later, Kolmogorov and Zabih [70] characterized the energy functions that can be minimized by graph cuts and also provided respective graph construction techniques. Recently, Krähenbühl and Koltun [71] proposed an efficient approximate inference algorithm for fully connected conditional random field (CRF) model with pairwise edge potentials defined by a linear combination of Gaussian kernels.

1.2.3 Strong Prior Based Algorithms

These algorithms constraint the solution space based on the training examples and hence yield good segmentation results even with ill-defined or missing object boundaries. However, the imposed constraints are very strict, causing the algorithms to fail if the training set is not representative of the population.

1.2.3.1 Active shape and appearance models

The active shape model (ASM) learns the pattern of shape variability from the training set of correctly annotated images, such that, the result of a segmentation

[1]For directed graphs, the convergence guarantee hold for $V(\alpha, \beta) \neq V(\beta, \alpha)$

algorithm can be constrained to plausible solutions [72]. ASM can be generated as follows:

- User marks multiple (n) corresponding landmarks, $\boldsymbol{x} = (x_1, y_1, ..., x_n, y_n)^T \in \mathbb{R}^{2n}$, from N training examples.
- Align the training shapes by scaling, rotating, and translating, using Procrustes method [73].
- Compute the mean shape, $\overline{\boldsymbol{x}}$, using

$$\overline{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x_i} \tag{1.18}$$

- Compute the $2n \times 2n$ covariance matrix $\boldsymbol{S}$, using

$$\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^T (\boldsymbol{x_i} - \overline{\boldsymbol{x}}) \tag{1.19}$$

- Perform eigen-decomposition of the covariance matrix, $\boldsymbol{S}$, such that the columns of matrix $\boldsymbol{P}$ represent the eigenvectors, and the diagonal elements of $\boldsymbol{\Lambda}$ represent the eigenvalues (λ_i)

$$\boldsymbol{S} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T \tag{1.20}$$

- A shape in training set can now be approximated using the mean shape and the weighted sum of k eigenvectors with largest eigenvalues, represented by a $2n \times k$ matrix $\boldsymbol{P}_k$ as:

$$\boldsymbol{x} = \overline{\boldsymbol{x}} + \boldsymbol{P}_k \boldsymbol{b} \tag{1.21}$$

  with the corresponding weight vector $\boldsymbol{b} = (b_1, ..., b_k)^T$.

- New example shapes similar to those in training sets can now be generated by varying the parameters, b_i's, within suitable limits. The parameters are linearly independent, though there may be nonlinear dependencies still present. The variance of b_k over training set is λ_k (k^{th} eigenvalue), hence the suitable limits for plausible shapes lies within three standard deviations of the mean:

$$-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k} \tag{1.22}$$

The built ASM model can be used to constraint the evolution of the active contours (**1.2.2.1)** after each contour update. The residual deformation, $d\boldsymbol{x}$, obtained after compensating for the translation, scale, and rotation of the model yields the parameter update, $d\boldsymbol{b}$, as:

$$d\boldsymbol{b} = \boldsymbol{P}_k^T d\boldsymbol{x} \tag{1.23}$$

such that the model parameters can be updated as:

$$\boldsymbol{b}_{t+1} \leftarrow \boldsymbol{b}_t + \boldsymbol{W}_b d\boldsymbol{b} \tag{1.24}$$

where, $\boldsymbol{W}_b$ is a diagonal matrix of weights, which can be identity, or each weight can be proportional to the variance of the corresponding shape parameter over the training set, to allow rapid movement of parameters with higher variance. Further, the shape parameters, b_i's, can be restricted withing three standard deviations as in **(1.22)**, to allow the evolved shape to be within a plausible range. Hence, the final segmentation, consistent with the training shapes, can be obtained via iterative updates upon convergence. The imposed shape constraint combined with the image information overcomes the segmentation challenges with missing object boundaries.

An Active Appearance Model (AAM) extends the ASM to include gray-level appearance of the object of interest. The built statistical model of the shape and appearance is more robust, and can fit a test image even from poor starting estimates. Model fitting refers to finding the model parameters to minimize the sum of squared difference between the test image and the synthesized model. Since the gradient descent optimization for model fitting requires expensive gradient computation in each iteration, the original paper by Cootes et al.[74] learns the linear relationship between model parameter displacements and the residual errors (between a training image and a synthesized model example), such that, during model fitting, the current residual is used to predict the model parameter displacements leading to a better fit. To speed up the model fitting, several optimization algorithms have been proposed [75], including warping the model to the test image via piecewise affine transformation [76] using the inverse compositional image alignment [77] algorithm. A comprehensive review of statistical shape models for 3D medical image segmentation can be found in [78].

1.2.3.2 Atlas-based models

Atlas-based models solve the image segmentation problem via image registration. In the context of cardiac image segmentation, an atlas is a cardiac US/MR image coupled with the corresponding manual segmentation of the heart chambers. The optimum transform, registering the atlas image to a test image, can be applied to the atlas segmentation, to obtain the segmentation of the test image. Based on the number of atlases used, atlas-based models can be divided into — single- , probabilistic-(average-), or multi-atlas based approach [79]. Atlas-based models have been extensively used for medical image segmentation [80, 81].

Single-atlas model

In this model, a single segmented image with good resolution and contrast is selected as an atlas. The atlas image is first registered to a test image via global similarity/affine transform and the alignment is further refined using a deformable transform. The obtained optimum transform (global+deformable) applied to the atlas label yields the test image segmentation.

Due to the large variability in test images, usually a single atlas in not sufficient to produce good results. Hence, multiple atlases can be registered to the test patient, such that, the segmentation obtained from the best atlas is selected. The best atlas can be chosen based on one of the two criteria — the atlas producing best similarity metric after global (and optionally deformable) registration, or the atlas requiring least deformation to register to the test patient.

Probabilistic atlas model

Multiple atlases can be registered to the same reference coordinate system, such that, the pixel-wise average intensity (after normalization) and labels provide the average appearance of the anatomy and probabilistic-atlas, respectively. The probabilistic-atlas represents the probability of each pixel belonging to a specific label in the reference coordinate system. Hence, the probabilistic-atlas for a specific group: gender, age, ethnicity etc. can be generated to better represent the variability in that group.

To segment a test image, the optimum transformation from the average appearance image (in the reference coordinate system) to the intensity normalized test image is obtained via registration, and hence the probabilistic-label is transferred. Furthermore, the transferred label can be used as a prior probability and combined with a likelihood term in a Bayesian framework to obtain the maximum a posteriori (MAP) probability of each pixel belonging to a specific label. This framework is attractive as it is fast, requiring a single image registration step, while also incorporating the variability in the multiple training atlases.

Multi-atlas model

The multi-atlas model registers multiple atlases to the test image, hence, transferring multiple labels. The label fusion step combines multiple labels to obtain the test image segmentation. Various label fusion strategies have been proposed in the literature [81, 82]. The simplest strategy is *majority voting*, where the label for each pixel is assigned according to the most frequent label obtained from multiple atlases. This strategy can be refined by assigning higher weights to the labels obtained form atlases with higher similarity to the test image (after registration). Furthermore, better segmentation performance can be obtained by weighting the atlases based on local image similarity. As the image registration is prone to errors, multi-atlas model is more robust compared to the single-/probabilistic-atlas models, at an expense of increased computational cost.

1.2.4 Machine Learning Based Algorithms

Machine learning algorithms recognize a pattern in the data using statistical techniques without being explicitly programmed. Supervised methods learn the mapping from the image intensity to the labels using the training image and corresponding segmentation, whereas the unsupervised methods try to find patterns in the image intensities to group similar pixels in the same class.

1.2.4.1 Unsupervised Methods

Unsupervised methods cluster the image pixels based on image features such as: intensity, color, and texture. Provided the number of clusters k, the k-means clustering algorithm [83] assigns each pixel to a cluster by reducing the within-class variance. Although attractive due to its simplicity, the k-means clustering algorithm assumes similar sized spherical clusters, and performs hard clustering (each data is assigned to a single class). The fuzzy c-means algorithm [84] extends k-means to produce soft clustering by providing the membership of each data point to multiple clusters. Similarly, the Gaussian mixture model (GMM) is a generalization of k-means algorithm, which can model non-spherical clusters, with soft assignment of data to multiple clusters. The Gaussian mixture model is estimated using the iterative Expectation-Maximization algorithm [85].

In contrast to the above mentioned clustering algorithms, the mean shift algorithm [86] does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. The mean shift algorithm creates a window around each data point, finds the weighted mean of data within each window, shifts the window to the mean, and repeat until convergence. The windows that end up near the same peak/mode of the data density are merged and assigned to the same cluster. Hence, a robust cluster of data is generated at an expense of higher computational cost.

When the pixel location is used alongside the image features for unsupervised clustering, perceptually similar pixels can be grouped together to create *super-pixels* [87]. The obtained over-segmentation can be used as a preprocessing step for the subsequent image segmentation task.

1.2.4.2 Supervised Methods

Pixel-wise Classification

For pixel-wise classification, a set of features is extracted from each pixel, such as: image intensities in the fixed neighborhood patch, color information, Gabor filter output, image gradients, pixel location etc., and a discriminative model (Support Vector Machine [88], Random Forest [89] or Neural Network classifier [90]) is trained

on these pixel features to predict the corresponding labels [91–94]. These methods do not encode the global context, as only the local image information is used for label prediction, hence, producing noisy results. The obtained probabilistic result can either be thresholded to yield the segmentation or can be incorporated as a prior probability in a Bayesian framework to combine with a likelihood to produce a MAP estimate.

Convolutional Neural Networks

Neural networks are machine learning models that loosely mimic the neural system in the animal brain. The model is composed of hierarchical layers. Each layer i consists of a weight matrix W_i applied to the input vector x_i along with a bias vector b_i, typically passed through a non-linear activation function $\sigma(\cdot)$ to generate an output vector $x_{i+1} = \sigma(W_i \cdot x_i + b_i)$. Hence, each layer learns the data representation at a higher, slightly more abstract level. Composing a large number of such transformations enables the network to learn a very complex function

$$x_n = \mathcal{F}(x_0) = \sigma(\sigma(W_{n-1} \cdot \sigma(W_{n-2}...\sigma(W_1 \cdot \sigma(W_0 \cdot x_0 + b_0) + b_1)... + b_{n-2}) + b_{n-1})$$

transforming the input into a data representation suitable for a particular application in hand. Specifically, Hornik [95] showed that standard multilayer feedforward networks with as few as a single hidden layer and arbitrary bounded and nonconstant activation function are universal approximators, provided that sufficiently many hidden units are available. However, in practice, deep networks with multiple hidden layers are preferred, as they are able to approximate complex functions with manageable number of hidden units, showing excellent performance in various applications [96].

Although the first neural network was first proposed in 50's and its first training algorithm was proposed in 70's, recent advances in computing power, availability of large datasets, and discovery of new non-linear functions (Rectified Linear Unit (ReLU)) overcoming the vanishing gradient problem, have led to resurgence of deeper, more powerful neural networks in 2010's, reintroduced as Deep Learning [96].

Given a training input/output pair (I, O), the input I is passed through the network $\mathcal{F}(\cdot)$ to generate an output $\mathcal{F}(I)$. The network output is compared against the reference output O to compute a loss $\mathcal{L}\{\mathcal{F}(I), O\}$, where the loss $\mathcal{L}\{\cdot, \cdot\}$ computes the dissimilarity between the two inputs. The goal of network training is to adjust the network parameters W_i's and b_i's, such that the loss $\mathcal{L}\{\mathcal{F}(I), O\}$ is minimized. The gradient of the loss with respect to the network output is computed and back-propagated [97] (via chain-rule) to obtain the gradients with respect to each network weight. Typically, the network weights are updated based on the gradients computed for a random batch of training examples, performing a stochastic gradient descent optimization [98]. The network is trained sufficiently long until the training loss stabilizes. A separate validation dataset is appropriated to evaluate the generalization performance of the network. The trained network yielding the lowest validation error is used as the final model to test it on held-out test datasets.

Convolutional Neural Network (CNN) [99] is a special type of neural network designed to discover spatially invariant patterns in the data. The weight matrix W is smaller than the input data matrix, and is learned across the input data via a sliding window operation called convolution. The parameter sharing across the spatial locations help reduce the number of network parameters, leverage the spatial relationship in the input data, and introduce spatial invariance [100]. Typically, a pooling layer (e.g. max/average pooling) is used following several convolutions and non-linear activations (e.g. ReLU), to introduce invariance to small translations, perform multi-resolution analysis, as well as reduce the size of intermediate features, ultimately reducing the computational complexity and memory footprint of the model.

The unique capability of CNNs to learn problem specific hierarchical features in an end-to-end manner have established them as a powerful general purpose supervised machine learning tool that can be deployed for various computer vision problems yielding state-of-the-art performance. CNNs [101] significantly improved image classification performance on a large scale visual recognition challenge (ImageNet challenge [102]) with hundreds of object categories and millions of images. This success led to its wide adoption in various other applications.

Although CNNs have shown incredible performance, these networks have been

found to be vulnerable to small perturbations of the input, imperceptible to the humans, termed as adversarial inputs [103]. Similarly, in [104], CNN based classification is reported to be sensitive to small image transformations (e.g. one pixel translation), due to the violation of spatial-invariance properties of the convolution, as a result of the sub-sampling (pooling) operations performed in the network. Furthermore, in [105], for the CNN based object detectors, it has been shown that slight changes of one object's position in the image can impact the detection of all the other objects present in the image. Nonetheless, CNN based methods produce excellent results for various computer vision and natural language processing applications, rendering them as the first choice methods for these applications.

In the context of image segmentation, Long et al.[106] proposed the first fully convolutional network for semantic image segmentation, adapting the contemporary classification networks fine-tuned for the segmentation task obtaining state-of-the-art performance. This performance motivated the use of CNNs in medical image analysis. However, their initial adoption in the medical domain was challenging, due to the limited availability of medical imaging data and associated costly manual annotation. These challenges were later circumvented by patch-based training, data augmentation, and transfer learning techniques [107, 108]. Notably, the U-Net architecture [109] with data augmentation has been very successful in medical image segmentation.

1.3 Cardiac Image Segmentation

World Health Organization (WHO) [2] facts estimated 17.9 million deaths from cardiovascular diseases in 2016, representing 31% of all mortalities, rendering cardiovascular conditions the main cause of death globally. Hence, the timely diagnosis and treatment follow-up of these pathologies is crucial. Cardiac diagnosis indices such as: systolic and diastolic volume, ejection fraction, myocardial mass and thickness can be computed by segmenting the heart chambers from the cardiac images. Furthermore, high quality anatomical models can be generated and used for treatment planning and

[2] https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

image-guided interventions. Ultrasound and MRI are the two most common cardiac imaging modalities.

1.3.1 Cardiac Ultrasound Segmentation

Over the past three decades, ultrasound (US) imaging has evolved as the preferred, standard-of-care imaging modality for the diagnosis, screening, and monitoring of several conditions. Specifically, thanks to its real-time capabilities, relatively inexpensive cost (compared to other modalities), and lack of exposure to ionizing radiation, US imaging has become the "first-line" modality for patient screening, diagnosis, and cardiac function assessment.

While 2D US has been the clinical standard, developments in 3D image acquisition and transducer design and technology have revolutionized echocardiography imaging, enabling both real-time 3D trans-esophageal and intra-cardiac image acquisition. However, in most cases the clinicians do not access the entire 3D image volume when analyzing the data, but rather focus on several key views that render the cardiac anatomy of interest during the US imaging exam, enabling image acquisition at a much higher spatial and temporal resolution. Two such common approaches are the bi-plane and tri-plane data acquisition protocols; as their name states, the former comprises two orthogonal image views, while the latter depicts the cardiac anatomy based on three co-axially intersecting views spaced at 60° to one another.

Trans-esophageal echocardiography (TEE) enables heart imaging while minimizing signal attenuation and optimizing field-of-view. As such, TEE is not only used for screening and diagnosis, but also for intra-operative therapy monitoring and/or image-guided cardiac interventions. Since the mid-2000s, TEE technology has accommodated 3D image acquisition and visualization of the cardiac anatomy in lieu of simple 2D renderings. However, despite the added bonus of 3D and 4D (3D + time) displays, the inherent trade-off between frame rate, and extent of anatomy covered, has determined clinicians to resort to the acquisition and visualization of multi-planar (orthogonal bi-plane or tri-plane) images to estimate the required parameters to assess cardiac function (i.e., ejection fraction) or identify critical features for image-guided

therapy.

Since cardiac anatomy is continuously changing, the intra-operative anatomy depicted using real-time US imaging also needs to be updated by tracking the key features of interest and endocardial LV boundaries. Therefore, near real-time feature tracking in US images is critical for two reasons: 1) to identify location of surgical targets for accurate tool to target navigation and on-target instrument positioning; and 2) to enable pre- to intra-op image registration as a means to fuse pre-op CT or MR images used during planning with intra-op images for enhanced guidance.

Several approaches for LV segmentation in echocardiography [110] have been popularly formulated as a contour finding problem, with the active contour method [111–114] being extensively used. Given its edge-based energy approach, the active contour method often produces many local minima and is sensitive to the initialization. Inspired by the active contours, the level set method [115–118] uses both edge- and region-based energy, making it more robust and less sensitive to initialization. Active shape and active appearance models [78, 119] incorporate knowledge of the LV shape and appearance from manually segmented training sets (assuming a Gaussian distribution of shape and appearance) into the snakes/levelset framework, hence, suffer from the initialization issues and might be too restrictive in some cases (e.g. diseased outlier cases). On the other hand, database-guided segmentation [120] overcome the initialization problem by implicitly encoding prior knowledge from the expert-annotated databases, yet at the expense of a highly complex search process. Other supervised learning techniques, such as artificial neural networks , have been used to detect endocardial border pixels using expert annotated training sets, but require large training sets and are unable to handle cases well outside of the training set.

1.3.2 Cardiac MRI Segmentation

High image quality, good tissue contrast, and no ionizing radiation has established MRI as a standard clinical modality for non-invasive assessment of cardiac performance. Cardiac contractile function quantified via the systolic and diastolic volumes, ejection fraction, and myocardial mass represents a reliable diagnostic value

and can be computed by segmenting the left (LV) and right (RV) ventricles from cardiac cine-MRI. Moreover, since the MR image acquisition technology has evolved to the extent that it enables the acquisition of peri-operative (just in time) images of the patient within the interventional suite moments before the procedure, reconstructed high quality LV models depicting full cardiac morphology can be used to precisely localize pathologies during an image guided intervention. Although manual delineation of the ventricle is deemed as the gold-standard approach, it requires significant time and effort and is highly susceptible to inter- and intra-observer variability. These limitations suggest a need for fast, robust, and accurate semi- or fully-automatic segmentation algorithms.

Various segmentation techniques for cardiac MR images have been proposed in the literature [121–123]. The image-based approaches with weak or no prior information [124, 125], such as thresholding, edge-based and region-based approaches, or pixel-based classifications methods, require user interaction for proper segmentation of the ill-defined regions. On the other hand, shape prior deformation models [126], active shape and appearance models [127–129], and atlas-based approaches [130–132] are more likely to overcome this problem at the expense of manually building a training set. While many methods focus on segmentation of a single ventricle, some methods perform joint segmentation of both ventricles [126, 133–135], exploiting the geometry information, relative position of ventricles, and similar intensity characteristic of blood pool cavities.

Multi-atlas based approaches have shown promising results in biomedical image segmentation [81]. However, they rely on a number of computationally demanding and time limiting nonrigid image registration steps followed by label fusion. Hence despite its accuracy, it has experienced minimal to no adoption in actual clinical applications primarily due to its complexity, high dependence on parameters variability, and computational demands. On the other hand, combinatorial optimization based graph-cut techniques [124] were also shown to serve as powerful tools for image segmentation; these techniques are fast and guarantee convergence within a known factor of the global minimum (in a binary labeling problem) for special class of functions (i.e., regular functions) [70]. In an attempt to further refine these techniques, adding

shape constraints to the graph cut framework has been shown to significantly improve cardiac image segmentation results [127, 136, 137].

Convolutional neural network based methods have been the most popular methods for image segmentation in recent years [106]. Convolution filters learned by the fully convolutional network (FCN) [138] naturally encode the spatial relationship between the pixels, as well as the relationship between various classes, producing piece-wise smooth segmentation results. Training the network with a variety of healthy and pathological MR images and corresponding manual segmentations, supplemented by artificial data-augmentation (e.g. rotation, translation, non-rigid deformation), produces excellent segmentation results for a new patient within a fraction of seconds on a Graphics Processing Unit (GPU). There have been attempts to leverage inter-slice spatial dependencies between the 2D cine MR slices [139] and impose 3D anatomical constraints [140] into the FCN framework. Furthermore, the output of the FCN is used to initialize a deformable model in [141] to improve LV segmentation accuracy. Several modifications to the FCN architecture and various post-processing schemes have been proposed to improve the semantic segmentation results as summarized in [142]. Nevertheless, the focus has primarily been on the LV segmentation and very few works have been disseminated on the RV segmentation.

1.3.3 Clinical Indices Estimation

Clinical indices such as LV/RV blood-pool volume, LV myocardium area and thickness, can either be computed from the corresponding segmentation or estimated directly from the features extracted from a given US/MR image. In [143], the LV volume for a new test patient is computed as the weighted combination of the LV volumes of training patients, such that the weights are computed according to their closeness to each of the LV in the training set, assessed using a specially designed likelihood function. Further, they exploited the linear relationship between the LV and RV volumes to obtain a better estimate of the RV volume.

Similarly, there have been attempts to train a machine learning model to predict the LV and RV volumes based on the learned unsupervised representation of a test

image [144]. Furthermore, other works [145, 146] regress the clinical indices directly from a given image using convolutional neural networks. Nevertheless, the direct estimation based methods are less interpretable, hence more challenging to troubleshoot in case of erroneous results, than their segmentation based counterparts.

1.3.4 Segmentation and Clinical Index Evaluation

Overlap and surface distance measures provide synergistic information regarding the accuracy of the obtained segmentation with respect to the reference segmentation. Furthermore, the clinical indices associated with the obtained segmentation is also evaluated against that computed from the reference.

1.3.4.1 Dice and Jaccard Coefficients

Given two binary segmentation masks, A and B, the Dice and Jaccard coefficient are defined as:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}, \quad \text{Jaccard} = \frac{|A \cap B|}{|A \cup B|} \tag{1.25}$$

where, $|\cdot|$ gives the cardinality (i.e. the number of non-zero elements) of each set. Maximum and minimum values (1.0 and 0.0, repectively) for Dice and Jaccard coefficient occur when there is 100% and 0% overlap between the two binary segmentation masks, respectively.

However, these overlap measures are less sensitive when evaluated on large solid objects (e.g. LV/RV blood-pool in mid-slice) compared to smaller objects (e.g. LV/RV blood-pool in apical slices) and objects with complex shapes (e.g. myocardium with a donut shape). Hence, we also evaluate the surface distances for comprehensive segmentation evaluation.

1.3.4.2 Mean Surface Distance and Hausdorff Distance

Let, S_A and S_B, be surfaces (with N_A and N_B points, respectively) corresponding to two binary segmentation masks, A and B, respectively. The mean surface distance

(MSD) is defined as:

$$\text{MSD} = \frac{1}{2}\left(\frac{1}{N_A}\sum_{p\in S_A} d(p, S_B) + \frac{1}{N_B}\sum_{q\in S_B} d(q, S_A)\right) \tag{1.26}$$

Similarly, Hausdorff Distance (HD) is defined as:

$$\text{HD} = \max\left(\max_{p\in S_A} d(p, S_B), \max_{q\in S_B} d(q, S_A)\right) \tag{1.27}$$

where,

$$d(p, S) = \min_{q\in S} d(p, q)$$

is the minimum Euclidean distance of point p from the points $q \in S$. Hence, MSD computes the mean distance between the two surfaces, whereas, HD computes the largest distance between the two surfaces, and is sensitive to outliers.

1.3.4.3 Ejection Fraction and Myocardial Mass

Ejection Fraction (EF) is an important cardiac parameter quantifying the cardiac output. EF is defined as:

$$\text{EF} = \frac{\text{EDV} - \text{ESV}}{\text{EDV}} \times 100\% \tag{1.28}$$

where, EDV is the end-diastolic volume, and ESV is the end-systolic volume. Similarly, the myocardial mass can be computed from the myocardial volume as:

$$\text{Myo-Mass} = \text{Myo-Volume } (cm^3) \times 1.06\ (gram/cm^3) \tag{1.29}$$

The correlation coefficients for the EF and myocardial mass computed from the ground-truth versus those computed from the automatic segmentation is reported. Correlation coefficient of $+1$ (-1) represents perfect positive (negative) linear relationship, whereas that of 0 represents no linear relationship between two variables.

1.3.4.4 Limits of Agreement

To compare the clinical indices computed from the ground-truth versus those obtained from the automatic segmentation, we take the difference between each pair of

the two observations. The mean of these differences is termed as *bias*, and the 95% confidence interval, mean $\pm 1.96 \times$standard deviation (assuming a Gaussian distribution), is termed as *limits of agreement* (LoA).

1.4 Challenges in Cardiac Image Segmentation

1.4.1 Ultrasound Images

US images suffer from various artifacts due to the echo based image acquisition procedure. Images contain characteristic granular texture due to the constructive and destructive interference of sound waves scattered by microscopic tissue inhomogeneities, also termed as speckle noise. The image contrast is low due to similar strength of echo received from different tissues. In addition, if regions with different attenuation properties appear at the same depth, the image intensity within the same tissue type appear inhomogeneous. These artifacts make the LV boarder indistinct and sometimes misleading.

Signal dropout can occur due to backscatter (reflection of signals back to the direction from which they came) if the ultrasound beam is parallel to the tissue boundary. Similarly, the regions behind strongly reflecting objects do not produce any echo due to the shadow artifact. This can lead to missing anatomical boundaries.

The cardiac motion and the deformation of LV boundaries, specially, fast motion during the systolic phase poses a big challenge for segmentation. Furthermore, large variations in appearance, configuration, and shape of the LV within the cardiac phase and across patients creates additional challenges.

Despite the growing availability of 3D US imaging techniques, the wider field-of-view comes at the expense of temporal resolution. Hence, clinicians prefer multi-plane 2D images of the heart, as their acquisition does not compromise the frame rate, which is the case for 3D and 4D imaging. These multi-plane 2D images, typically consisting of the 4-, 3- and 2-chamber views, are used to compute the LV-RV blood pool volumes and ejection fraction, although these parameters are deemed not as reliable as those computed from 3D imaging.

1.4.2 MRI Images

The short-axis cine MR slices can be acquired throughout the cardiac cycle using the SSFP sequence coupled with the echocardiogram (ECG) gating. Multiple short-axis cine MR slices covering the whole heart are stacked together to generate a pseudo-4D volume, thus, allowing analysis of cardiac functions throughout the cardiac cycle.

Although the SSFP pulse sequence makes blood-pool appear bright whereas the myocardium darker, the intensity values don't have a standardized physical meaning as in X-Ray or CT (Hounsfield Unit) and can vary between different acquisitions. Similarly, since the recorded slice signal is obtained from a volume with some slice-thickness, a voxel can be a mixture of signals from several tissue types. This phenomenon termed as the *partial volume effect* causes fuzziness in intensity, which is further exacerbated due to the blood flow. Furthermore, the papillary muscles and trabeculations inside the heart chambers have same intensity profile as the myocardium and hence are difficult to distinguish. All of the mentioned artifacts causes indistinct borders making the segmentation of heart chambers challenging.

Although LV remains close to the ring shape, the RV has a crescent shape, which varies a lot across the apex-base axis. Hence, the segmentation of RV is more challenging than the LV. In addition, the thickness of the RV myocardium is small and close to the MR image resolution, making the segmentation of the epicardial wall very difficult.

In short-axis cine MR acquisition, typically 8-10 slices are acquired along the apex-base axis across the heart. Hence the resolution in long-axis is very poor (around 10mm). Further, the respiration and patient motion during the acquisition results in misaligned slices. This makes 3D processing of the acquired volumes difficult. Thus, most of the existing algorithms treat the short-axis slices independently, resulting in poor segmentation results in the basal and apical regions.

1.5 Contributions of this Dissertation

1.5.1 Ultrasound Image Segmentation

Due to the inherent noise in ultrasound images, intensity based methods usually perform sub-optimally. The local phase-based filter computes the image response to even and odd quadrature filters at multiple scales, which are considered as the real- and imaginary- components, respectively, therefore allowing the computation of the magnitude and phase of local signals. The difference in response of the odd filter compared to the even filter across different scales, normalized by the total magnitude, provides a contrast invariant measure of local-phase asymmetry. Here we use the local-phase asymmetry measure to detect the endocardium border, hence overcoming the challenges of noisy ultrasound images.

We use the graph-cut framework to minimize the region-based energy with a smoothness constraint. The Gaussian intensity-likelihood term for the foreground and background regions overcome the issue of intensity inhomogeneity. Similarly, the smoothness constraint forced by the graph neighborhood structure provide consistent segmentation in the regions with missing anatomical boundaries.

Since the LV features and boundaries are more prominent in the end-diastole images, we first segment the LV from the end-diastole frame and subsequently propagate the segmentation using non-rigid registration algorithm to overcome the difficulties due to large deformation of LV boundaries. The LV anatomy is modeled as a two compartment model consisting of muscle — linear elastic, isotropic, and incompressible, and blood-pool , with prescribed smoothness constraints to allow rapid motion of the endocardial contour.

Clinicians compute the cardiac contractile function for each view separately, assuming isotropy in 3D, such that, the final metric is the average of the indices from tri-plane views. Hence, the computed indices do not represent the actual 3D geometry of the LV. Here, we rearrange the segmentation from tri-plane views in their original 3D orientation (60° apart) and reconstruct the 3D LV geometry, therefore achieving a more accurate and precise estimate of the cardiac indices.

1.5.2 MRI Image Segmentation

To overcome the challenges with the non-standardized intensity values in the MR images, as well as the fuzziness due to the partial volume effect and blood flow, we use the Gaussian mixture model intensity likelihood for the foreground and background regions in the Graph-cut framework. This regional energy term coupled with the graph neighborhood smoothness constraints provide robustness against the intensity fuzziness in MR images.

To overcome the segmentation difficulty due to the distinct shapes of LV and RV and their variability across patients as well as along the apex-base axis, we incorporate the average atlas based shape prior into the graph-cut framework. Furthermore, the proposed iterative refinement procedure helps overcome the initial average atlas to test patient registration error. This also allows the segmentation to be guided by prominent image features, relaxing the atlas constraints, in case of diseased/outlier test patients.

Due to the large slice-spacing and motion artifacts caused by breathing and patient movements, the short-axis cine MRI slices are misaligned. Hence, most of the cine MRI segmentation algorithms treat each slice separately to perform a 2D slice-wise segmentation. Although 2D algorithms obtain good segmentation results in mid-ventricular slices, there is limited information in the apical and basal regions, hence, the segmentation results are poor in these regions. We realign the misaligned slices based on the LV center information to generate a coherent 3D volume, and hence utilize the 3D context during segmentation. This improves the segmentation results in the problematic apical and basal slices.

Finally, we leverage the success of fully convolutional neural networks in the semantic image segmentation task to perform end-to-end segmentation of the heart chambers. We employ a multi-task learning network to segment the myocardium as well as regress its area. Furthermore, we employ the multi-task learning framework to impose a shape-constraint into a fully convolutional network, hence improving the segmentation performance and the computed clinical indices, as well as, the generalization performance of the network when trained on a dataset acquired at one

institution and tested on a dataset acquired at a different institution. We show that our proposed network is able to produce excellent segmentation results (for LV blood-pool, RV blood-pool, and LV myocardium) from the MR images of both normal and pathological cases, while also improving the segmentation performance in the challenging apical and basal slices.

1.5.3 Clinical Indices Estimation

To evaluate the accuracy and precision of the clinical indices obtained from segmentation-based and direct estimation-based methods, we trained a fully convolutional network to segment the myocardium and regress its area, simultaneously. We demonstrate that segmentation-based area estimate is significantly better than that obtained from direct estimation. Furthermore, as the quality of the obtained segmentation can be visually verified by the clinician, segmentation-based clinical index estimation is interpretable and more reliable, as opposed to that obtained from direct estimation-based methods, hence showing better promise for clinical adoption. However, it should be noted that it is easier to obtain the reference clinical index for direct-estimation, compared to the reference per-pixel segmentation required for segmentation-based method.

1.6 Dissertation Overview

The thesis chapters provide a detailed description of the proposed methods to overcome the challenges with the segmentation of 2D, 3D, and 4D cardiac images from Ultrasound and Cine MRI modalities. Chapters 2-4 employ traditional Atlas and Graph-cut based segmentation techniques. In Chapter 5, we make a transition towards CNN based approach by training it for slice mis-alignment correction and extending previously developed method for full 3D segmentation. Finally, to leverage the recent success of CNN based methods, we train a fully convolutional network on large dataset in Chapters 6-7, obtaining highly accurate segmentation of the heart chambers compared to the traditional segmentation approaches.

Chapter 2

This chapter provides a detailed description of the left Ventricle segmentation from tri-plane TEE image sequences. The proposed methodology encompasses three steps: 1) endocardial left ventricle (LV) feature extraction and blood-pool segmentation from the raw 2D multi-plane image sequences, 2) frame-to-frame feature tracking and propagation through the cardiac cycle using non-rigid image registration, and 3) 3D reconstruction of the LV blood pool geometry at the desired cardiac phases using spline-based interpolation and convex hull fitting. The materials presented in this chapter are adapted from the manuscripts published in 2015 in SPIE Medical Imaging and Function Imaging and Modeling of the Heart (FIMH) in Springer's Lecture Notes in Computer Science series.

Chapter 3

This chapter provides a detailed description of the developed probabilistic atlas prior based graph cut segmentation method. The proposed 2D segmentation methods are applied for the segmentation of the left Ventricle from an open-source cardiac cine MRI dataset with provided manual segmentation. The materials presented in this chapter are adapted from a manuscript published in Springer's Lecture Notes in Computer Science as part of the 2016 MICCAI Workshop on Statistical Atlases and Computational Models of the Heart (STACOM).

Chapter 4

The developed probabilistic atlas prior based graph cut segmentation method is applied for the 2D segmentation of the right ventricle from an open-source cardiac cine MRI dataset with provided manual segmentation. The materials presented in this chapter are adapted from the work published in SPIE Medical Imaging 2017 and Springer's lecture Notes in Computer Science as part of the proceedings of the 2017 Functional Imaging and Modeling of the Heart conference.

Chapter 5

This chapter provides a detailed description of the proposed convolutional neural network (CNN) based method to predict the LV centers from 2D cine MRI images and correct the slice-misalignment to create a coherent 3D volume. Finally, full 3D graph-cut segmentation with probabilistic atlas shape prior yields the segmentation. The materials presented in this chapter are adapted from a manuscript published in the 2018 SPIE medical Imaging conference.

Chapter 6

This chapter provides a detailed description of a proposed multi-task learning method used to segment the myocardium and regress its area, simultaneously. The uncertainty of the two tasks are predicted by the network and used to weigh the losses. Using this architecture, we not only show improvement in segmentation performance, but also conclude that segmentation-based myocardium area estimate is better than the area directly regressed by the network. The materials presented in this chapter have been adapted from a manuscript on the same topic published in Springer's Lecture Notes in Computer Science as part of the 2018 MICCAI Workshop on Statistical Atlases and Computational Models of the Heart.

Chapter 7

This chapter provides a detailed description of a proposed distance map regularized fully convolutional network, featuring multi-task learning. The proposed method is extensively evaluated on two open-source cardiac cine MRI datasets, demonstrating significant improvement in segmentation performance. The materials presented in this chapter have been adapted from a manuscript published in Medical Physics.

Chapter 8

This chapter summarizes the contributions of this thesis in the field of cardiac image segmentation from both clinical and algorithmic perspectives. Further, the

proposed future work following this dissertation is provided.

Bibliography

[1] J.T. Bushberg, J.A. Seibert, and E.M. Leidholdt. *The Essential Physics of Medical Imaging*. Wolters Kluwer Health, 2011.

[2] D.B. Murphy and M.W. Davidson. *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley, 2012.

[3] Vani JA Konda and Irving Waxman. *Endoscopic Imaging Techniques and Tools*. Springer, Cham, 2016.

[4] W. Drexler and J.G. Fujimoto. *Optical Coherence Tomography: Technology and Applications*. Optical Coherence Tomography. Springer International Publishing, 2015.

[5] T. Peters and K. Cleary. *Image-Guided Interventions: Technology and Applications*. Springer US, 2008.

[6] W. Birkfellner. *Applied Medical Image Processing, Second Edition: A Basic Course*. CRC Press, 2016.

[7] J.L. Prince and J.M. Links. *Medical Imaging Signals and Systems*. Pearson, 2014.

[8] A.P. Dhawan. *Medical Image Analysis*. IEEE Press Series on Biomedical Engineering. Wiley, 2011.

[9] Amy Berrington de Gonzlez and Sarah Darby. Risk of cancer from diagnostic x-rays: estimates for the uk and 14 other countries. *The Lancet*, 363(9406):345 – 351, 2004.

[10] J. Hsieh. *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. Press Monograph Series. SPIE, 2015.

[11] T.M. Buzug. *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer Berlin Heidelberg, 2010.

[12] David J. Brenner and Eric J. Hall. Computed tomography an increasing source of radiation exposure. *New England Journal of Medicine*, 357(22):2277–2284, 2007. PMID: 18046031.

[13] J.P. Hornak. *The Basics of MRI*. Rochester Institute of Technology, 1996.

[14] R.W. Brown, E.M. Haacke, Y.C.N. Cheng, M.R. Thompson, and R. Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley, 2014.

[15] HY Carr. Steady-state free precession in nuclear magnetic resonance. *Physical Review*, 112(5):1693, 1958.

[16] P. Libby and R.Y. Kwong. *Cardiovascular Magnetic Resonance Imaging*. Contemporary Cardiology. Humana Press, 2008.

[17] T.L. Szabo. *Diagnostic Ultrasound Imaging: Inside Out*. Academic Press Series in Biomedical Engineering. Academic Press, 2013.

[18] Jaydev K. Dave, Maureen E. Mc Donald, Praveen Mehrotra, Andrew R. Kohut, John R. Eisenbrey, and Flemming Forsberg. Recent technological advancements in cardiac ultrasound imaging. *Ultrasonics*, 84:329 – 340, 2018.

[19] A. Giussani and C. Hoeschen. *Imaging in Nuclear Medicine*. Springer Berlin Heidelberg, 2013.

[20] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Pearson/Prentice Hall, 2008.

[21] R. Szeliski. *Computer Vision: Algorithms and Applications*. Texts in Computer Science. Springer London, 2010.

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

[23] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986.

[24] P.V.C. Hough. Method and means for recognizing complex patterns, Dec 1962.

[25] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, Jun 1991.

[26] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[27] Laurent D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53(2):211 – 218, 1991.

[28] Chenyang Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, Mar 1998.

[29] Stanley Osher and James A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, November 1988.

[30] Vicent Caselles, Francine Catté, Tomeu Coll, and Françoise Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31, Dec 1993.

[31] Mark Sussman, Peter Smereka, and Stanley Osher. A level set approach for computing solutions to incompressible two-phase flow. *Journal of Computational Physics*, 114(1):146 – 159, 1994.

[32] Chunming Li, Chenyang Xu, Changfeng Gui, and M. D. Fox. Level set evolution without re-initialization: a new variational formulation. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 430–436 vol. 1, June 2005.

[33] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, Feb 1997.

[34] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb 2001.

[35] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, July 1989.

[36] Bo Peng, Lei Zhang, and David Zhang. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 46(3):1020 – 1038, 2013.

[37] Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.

[38] R. C. Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, Nov 1957.

[39] E.W. Dijkstra. Some theorems on spanning subtrees of a graph. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen: Series A: Mathematical Sciences*, 63(2):196–199, 1960.

[40] Ying Xu and Edward C. Uberbacher. 2d image segmentation using minimum spanning trees. *Image and Vision Computing*, 15(1):47 – 57, 1997.

[41] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, Sep 2004.

[42] Ali Saglam and Nurdan Akhan Baykan. Sequential image segmentation based on minimum spanning tree representation. *Pattern Recognition Letters*, 87:155 – 162, 2017. Advances in Graph-based Pattern Recognition.

[43] Eric N. Mortensen and William A. Barrett. Intelligent scissors for image composition. In *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '95, pages 191–198, New York, NY, USA, 1995. ACM.

[44] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numer. Math.*, 1(1):269–271, December 1959.

[45] Alexandre X. Falco, Jayaram K. Udupa, Supun Samarasekera, Shoba Sharma, Bruce Elliot Hirsch, and Roberto de A. Lotufo. User-steered image segmentation paradigms: Live wire and live lane. *Graphical Models and Image Processing*, 60(4):233 – 260, 1998.

[46] Ghassan Hamarneh, Johnson Yang, Chris McIntosh, and Morgan Langille. 3d live-wire-based semi-automatic segmentation of medical images. In *Medical Imaging 2005: Image Processing*, volume 5747, pages 1597–1604. International Society for Optics and Photonics, 2005.

[47] L. Grady. Minimal surfaces extend shortest path segmentation methods to 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2):321–334, Feb 2010.

[48] X. Bai and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007.

[49] Leo Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (11):1768–1783, 2006.

[50] J. Shen, Y. Du, W. Wang, and X. Li. Lazy random walks for superpixel segmentation. *IEEE Transactions on Image Processing*, 23(4):1451–1462, April 2014.

[51] David Aldous and James Allen Fill. Reversible markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014, available at `http://www.stat.berkeley.edu/$\sim$aldous/RWG/book.html`.

[52] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen. Random walk and graph cut for co-segmentation of lung tumor on pet-ct images. *IEEE Transactions on Image Processing*, 24(12):5854–5867, Dec 2015.

[53] Abouzar Eslami, Athanasios Karamalis, Amin Katouzian, and Nassir Navab. Segmentation by retrieval with guided random walks: Application to left ventricle segmentation in mri. *Medical Image Analysis*, 17(2):236 – 253, 2013.

[54] Vasileios G. Kanas, Evangelia I. Zacharaki, Christos Davatzikos, Kyriakos N. Sgarbas, and Vasileios Megalooikonomou. A low cost approach for brain tumor segmentation based on intensity modeling and 3d random walker. *Biomedical Signal Processing and Control*, 22:19 – 30, 2015.

[55] Kenneth M. Hall. An r-dimensional quadratic placement algorithm. *Management Science*, 17(3):219–229, 1970.

[56] L. Hagen and A. B. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11(9):1074–1085, Sep 1992.

[57] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, Dec 2007.

[58] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, August 2000.

[59] C. H. Q. Ding, Xiaofeng He, Hongyuan Zha, Ming Gu, and H. D. Simon. A min-max cut algorithm for graph partitioning and data clustering. In *Proceedings 2001 IEEE International Conference on Data Mining*, pages 107–114, 2001.

[60] Mari C.V. Nascimento and Andr C.P.L.F. de Carvalho. Spectral methods for graph clustering a survey. *European Journal of Operational Research*, 211(2):221 – 231, 2011.

[61] J. M. Hammersley and P. E. Clifford. Markov random fields on finite graphs and lattices. Unpublished manuscript, 1971.

[62] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov 1984.

[63] Stuart Geman and Christine Graffigne. Markov random field image models and their applications to computer vision. In *Proceedings of the international congress of mathematicians*, volume 1, page 2, 1986.

[64] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[65] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.

[66] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(2):271–279, 1989.

[67] Lester R Ford and Delbert R Fulkerson. Maximal flow through a network. *Canadian journal of Mathematics*, 8(3):399–404, 1956.

[68] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1101–1113, Nov 1993.

[69] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[70] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, Feb 2004.

[71] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 109–117. Curran Associates, Inc., 2011.

[72] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38 – 59, 1995.

[73] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, Mar 1975.

[74] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, Jun 2001.

[75] X. Gao, Y. Su, X. Li, and D. Tao. A review of active appearance models. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(2):145–158, March 2010.

[76] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, Nov 2004.

[77] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, Feb 2004.

[78] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4):543 – 563, 2009.

[79] Torsten Rohlfing, Robert Brandt, Randolf Menzel, Daniel B Russakoff, and Calvin R Maurer. Quo vadis, atlas-based segmentation? In *Handbook of Biomedical Image Analysis*, pages 435–486. Springer, 2005.

[80] Mariano Cabezas, Arnau Oliver, Xavier Llad, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158 – e177, 2011.

[81] Juan Eugenio Iglesias and Mert R. Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205 – 219, 2015.

[82] P. Aljabar, R.A. Heckemann, A. Hammers, J.V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726 – 738, 2009.

[83] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press.

[84] J. C. Bezdek, R. Ehrlich, and W. Full. FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10:191–203, 1984.

[85] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[86] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.

[87] David Stutz, Alexander Hermans, and Bastian Leibe. Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding*, 166:1 – 27, 2018.

[88] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995.

[89] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[90] Christopher M Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.

[91] Xiang-Yang Wang, Ting Wang, and Juan Bu. Color image segmentation using pixel wise support vector machine classification. *Pattern Recognition*, 44(4):777 – 787, 2011.

[92] G. Stalidis, N. Maglaveras, S. N. Efstratiadis, A. S. Dimitriadis, and C. Pappas. Model-based processing scheme for quantitative 4-d cardiac mri analysis. *IEEE Transactions on Information Technology in Biomedicine*, 6(1):59–72, March 2002.

[93] Victor Lempitsky, Michael Verhoek, J. Alison Noble, and Andrew Blake. Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography. In Nicholas Ayache, Hervé Delingette, and Maxime Sermesant, editors, *Functional Imaging and Modeling of the Heart*, pages 447–456, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[94] K. Z. Mao, Peng Zhao, and Puay-Hoon Tan. Supervised learning-based cell image segmentation for p53 immunohistochemistry. *IEEE Transactions on Biomedical Engineering*, 53(6):1153–1163, June 2006.

[95] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257, 1991.

[96] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 05 2015.

[97] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[98] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[99] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[100] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[101] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[102] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[103] X. Yuan, P. He, Q. Zhu, and X. Li. Adversarial examples: Attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9):2805–2824, Sep. 2019.

[104] Aharon Azulay and Yair Weiss. Why do deep convolutional networks generalize so poorly to small image transformations? *CoRR*, abs/1805.12177, 2018.

[105] Amir Rosenfeld, Richard S. Zemel, and John K. Tsotsos. The elephant in the room. *CoRR*, abs/1808.03305, 2018.

[106] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[107] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 06 2017.

[108] Geert Litjens et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88, 2017.

[109] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[110] J.A. Noble and D. Boukerroui. Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging*, 25(8):987–1010, Aug 2006.

[111] A. Mishra, P.K. Dutta, and M.K. Ghosh. A ga based approach for boundary detection of left ventricle with echocardiographic image sequences. *Image and Vision Computing*, 21(11):967–976, 2003.

[112] M. Mignotte and J. Meunier. A multiscale optimization approach for the dynamic contour-based boundary detection issue. *Comput Med Imaging Graph.*, 25(3):265–275, 2001.

[113] Auzuir Ripardo de Alexandria, Paulo Csar Cortez, Jessyca Almeida Bessa, John Hebert da Silva Flix, Jos Sebastio de Abreu, and Victor Hugo C. de Albuquerque. psnakes: A new radial active contour model and its application in the segmentation of the left ventricle from echocardiographic images. *Computer Methods and Programs in Biomedicine*, 116(3):260 – 273, 2014.

[114] Pedro Pedrosa Rebouas Filho, Paulo Csar Cortez, Antnio Carlos da Silva Barros, and Victor Hugo C. de Albuquerque. Novel adaptive balloon active contour method based on internal force for image segmentation a systematic evaluation on synthetic and real images. *Expert Systems with Applications*, 41(17):7707 – 7721, 2014.

[115] O. Bernard, D. Friboulet, P. Thevenaz, and M. Unser. Variational b-spline level-set: A linear filtering approach for fast deformable model evolution. *IEEE Trans Image Process.*, 18(6):1179–1191, June 2009.

[116] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int J Comput Vision*, 69(3):335–351, September 2006.

[117] A. Belaid, D. Boukerroui, Y. Maingourd, and J. F. Lerallut. Phase-based level set segmentation of ultrasound images. *IEEE Transactions on Information Technology in Biomedicine*, 15(1):138–147, Jan 2011.

[118] Xulei Qin, Zhibin Cong, and Baowei Fei. Automatic segmentation of right ventricular ultrasound images using sparse matrix transform and a level set. *Physics in Medicine & Biology*, 58(21):7609, 2013.

[119] J.G. Bosch, S.C. Mitchell, B.P.F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J.H.C. Reiber. Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Trans Med Imaging*, 21(11):1374–1383, Nov 2002.

[120] B. Georgescu, X.S. Zhou, D. Comaniciu, and A. Gupta. Database-guided segmentation of anatomical structures with complex appearance. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 429–436, June 2005.

[121] Caroline Petitjean and Jean-Nicolas Dacher. A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis*, 15(2):169 – 184, 2011.

[122] Caroline Petitjean, Maria A. Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jrme Caudron, Su Ruan, Ismail Ben Ayed, M. Jorge Cardoso, Hsiang-Chou Chen, Daniel Jimenez-Carretero, Maria J. Ledesma-Carbayo, Christos Davatzikos, Jimit Doshi, Guray Erus, Oskar M.O. Maier,

Cyrus M.S. Nambakhsh, Yangming Ou, Sbastien Ourselin, Chun-Wei Peng, Nicholas S. Peters, Terry M. Peters, Martin Rajchl, Daniel Rueckert, Andres Santos, Wenzhe Shi, Ching-Wei Wang, Haiyan Wang, and Jing Yuan. Right ventricle segmentation from cardiac MRI: A collation study. *Medical Image Analysis*, 19(1):187 – 202, 2015.

[123] Peng Peng, Karim Lekadir, Ali Gooya, Ling Shao, Steffen E. Petersen, and Alejandro F. Frangi. 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, Apr 2016.

[124] O. M. O. Maier, D. Jimnez, A. Santos, and M. J. Ledesma-Carbayo. Segmentation of rv in 4d cardiac mr volumes using region-merging graph cuts. In *2012 Computing in Cardiology*, pages 697–700, Sept 2012.

[125] Ching-Wei Wang, Chun-Wei Peng, and Hsiang-Chou Chen. A simple and fully automatic right ventricle segmentation method for 4-dimensional cardiac mr images. *Workshop in Medical Image Computing and Computer Assisted Intervention*, pages 1–8, 2012.

[126] Damien Grosgeorge, Caroline Petitjean, Jérôme Caudron, Jeannette Fares, and Jean-Nicolas Dacher. Automatic cardiac ventricle segmentation in MR images: a validation study. *International Journal of Computer Assisted Radiology and Surgery*, 6(5):573–581, 2011.

[127] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan. Graph cut segmentation with a statistical shape model in cardiac {MRI}. *Computer Vision and Image Understanding*, 117(9):1027 – 1035, 2013.

[128] Julien Abi-Nahed, Marie-Pierre Jolly, and Guang-Zhong Yang. *Robust Active Shape Models: A Robust, Generic and Simple Automatic Segmentation Tool*, pages 1–8. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[129] Mohammed S. ElBaz and Ahmed S. Fahmy. Active shape model with inter-profile modeling paradigm for cardiac right ventricle segmentation. In *Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I*, MICCAI'12, pages 691–698, Berlin, Heidelberg, 2012. Springer-Verlag.

[130] Wenjia Bai, Wenzhe Shi, Haiyan Wang, Nicholas S Peters, and Daniel Rueckert. Multiatlas based segmentation with local label fusion for right ventricle MR images. *image*, 6:9, 2012.

[131] Yangming Ou, Jimit Doshi, Guray Erus, and Christos Davatzikos. Multi-atlas segmentation of the cardiac mr right ventricle. *Proceedings of 3D Cardiovascular Imaging : A MICCAI Segmentation Challenge*, 2012.

[132] MA Zuluaga, MJ Cardoso, and Sebastien Ourselin. Automatic right ventricle segmentation using multi-label fusion in cardiac MRI. 2012.

[133] Jochen Peters, Olivier Ecabert, Carsten Meyer, Hauke Schramm, Reinhard Kneser, Alexandra Groth, and Jürgen Weese. *Automatic Whole Heart Segmentation in Static Magnetic Resonance Image Volumes*, pages 402–410. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[134] Maria A. Zuluaga, M. Jorge Cardoso, Marc Modat, and Sébastien Ourselin. *Multi-atlas Propagation Whole Heart Segmentation from MRI and CTA Using a Local Normalised Correlation Coefficient Criterion*, pages 174–181. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[135] W. Bai, W. Shi, D. P. O'Regan, T. Tong, H. Wang, S. Jamil-Copley, N. S. Peters, and D. Rueckert. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac mr images. *IEEE Transactions on Medical Imaging*, 32(7):1302–1315, July 2013.

[136] Daniel Freedman and Tao Zhang. Interactive graph cut based segmentation with shape priors. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, I:755–762, 2005.

[137] Dwarikanath Mahapatra. Cardiac Image Segmentation from Cine Cardiac MRI Using Graph Cuts and Shape Priors. *Journal of Digital Imaging*, 26(4):721–730, 2013.

[138] Phi Vu Tran. A fully convolutional neural network for cardiac segmentation in short-axis MRI. *CoRR*, abs/1604.00494, 2016.

[139] Rudra P. K. Poudel, Pablo Lamata, and Giovanni Montana. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 83–94, Cham, 2017. Springer International Publishing.

[140] O. Oktay et al. Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, Feb 2018.

[141] M.R. Avendi, Arash Kheradvar, and Hamid Jafarkhani. 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, 2016.

[142] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and José García Rodríguez. A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857, 2017.

[143] Z. Wang, M. B. Salah, B. Gu, A. Islam, A. Goela, and S. Li. Direct estimation of cardiac biventricular volumes with an adapted bayesian formulation. *IEEE Transactions on Biomedical Engineering*, 61(4):1251–1260, April 2014.

[144] Xiantong Zhen, Zhijie Wang, Ali Islam, Mousumi Bhaduri, Ian Chan, and Shuo Li. Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. *Medical Image Analysis*, 30:120 – 129, 2016.

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

[146] Wufeng Xue, Gary Brahm, Sachin Pandey, Stephanie Leung, and Shuo Li. Full left ventricle quantification via deep multitask relationships learning. *Medical Image Analysis*, 43:54 – 65, 2018.

Chapter 2

Left Ventricle Segmentation, Tracking, and 3D Reconstruction from Multi-plane 2D TEE Image Sequences

Multi-plane 2D TEE [1] *images constitute the clinical standard of care for assessment of left ventricle function, as well as for guiding various minimally invasive procedure that rely on intra-operative imaging for real-time visualization. We propose a framework that enables automatic, rapid and accurate endocardial left ventricle feature identification and blood-pool segmentation using a combination of image filtering, graph cut, non-rigid registration-based motion extraction, and 3D LV geometry reconstruction techniques applied to the TEE image series. We evaluate our*

[1] This chapter is adapted from:
Dangi S. et al., "Endocardial left ventricle feature tracking and reconstruction from tri-plane transesophageal echocardiography data," Proc. SPIE 9415, Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling, 941505 (18 March 2015)
Dangi S. et al., "Automatic LV Feature Detection and Blood-Pool Tracking from Multi-plane TEE Time Series". In: van Assen H., Bovendeerd P., Delhaas T. (eds) Functional Imaging and Modeling of the Heart. FIMH 2015. Lecture Notes in Computer Science, vol 9126. Springer, Cham

proposed framework using several retrospective patient tri-plane TEE image sequences and demonstrate comparable results to those achieved by expert manual segmentation using clinical software.

2.1 Introduction

While 2D US has been the clinical standard for over two decades, developments in 3D image acquisition and transducer design and technology have revolutionized echocardiography imaging, enabling both real-time 3D trans-esophageal and intra-cardiac image acquisition. However, in most cases the clinicians do not access the entire 3D image volume when analyzing the data, but rather focus on several key views that render the cardiac anatomy of interest during the US imaging exam, enabling image acquisition at a much higher spatial and temporal resolution. Two such common approaches are the bi-plane and tri-plane data acquisition protocols; as their name states, the former comprises two orthogonal image views, while the latter depicts the cardiac anatomy based on three co-axially intersecting views spaced at 60° to one another. But since the LV is a 3D structure, there is a critical need for analysis tools that are automated (to remove used bias and variability) to reconstruct the 3D geometry from 2D time sequence images, and characterize its parameters throughout the cardiac cycle. The cardiac parameters such as systolic and diastolic volumes and ejection fraction, constitute critical clinical biomarkers, and are currently traditionally assessed from 2D views following manual annotation.

In this work we propose the implementation and clinical validation of an automatic workflow that encompasses well-evaluated filtering, segmentation, registration, and volume reconstruction techniques as a means to provide a rapid, robust and accurate framework for feature tracking from multi-plane ultrasound image sequences. The proposed computational framework was developed in close collaboration with our echocardiography colleagues, motivated by the need to reduce user-dependent and

user-induced bias and reduce the uncertainty associated with the process of manually identifying features from US image sequences. The impact and contribution of the proposed work is the integration of several image processing techniques (i.e., phase-based filtering, segmentation, registration and volume reconstruction) into a streamlined workflow that utilizes traditional standard of care images and fits seamlessly within the current workflows associated with both cardiac function assessment and intra-operative cardiac intervention guidance and monitoring. The accuracy of the proposed workflow is assessed against gold-standard results from the *GE Echopac PC* clinical software.

2.2 Methodology

Speckle noise and signal dropouts inherent in US images render intensity based approaches unreliable; rather, local-phase based approaches [1], theoretically invariant to the intensity magnitude, have been preferred for detecting endocardium. Here we exploit the robustness of phase-based feature detection and combine it with the power of graph cut-based techniques [2] that use both region and boundary regularization, to obtain a rapid, automatic piecewise smooth segmentation of the LV blood pool and muscle regions. In addition, we conducted a study using retrospective clinical patient data consisting of tri-plane (60° to one another) TEE image sequences through the cardiac cycle to validate the proposed tools and demonstrate their clinical utility and performance against commercial, clinical-grade, clinician-operated software.

The proposed methodology encompasses three steps: 1) endocardial left ventricle (LV) feature extraction and blood-pool segmentation from the raw 2D multi-plane image sequences, 2) frame-to-frame feature tracking and propagation through the cardiac cycle using non-rigid image registration, and 3) 3D reconstruction of the LV blood pool geometry at the desired cardiac phases using spline-based interpolation and convex hull fitting.

2.2.1 LV Feature Extraction and Blood-pool Segmentation

Although expert manual segmentation is deemed as gold-standard, the specular noise specific to all US images gives rise to significant uncertainty when extracting features of interest, highly dependent on the expert's experience and expertise. To improve consistency, we propose the use an automatic filtering and feature extraction technique that identifies the blood-pool and tissue regions from 2D US images using a two-step process.

2.2.1.1 Image Preprocessing via Monogenic Filtering

Intensity-based edge detection algorithms are inefficient in identifying features from US images, whereas the intensity invariant local phase-based techniques have shown promising results [1], where a local phase of $\pm\pi/2$ signifies high symmetry, while a local phase of 0 or π signifies high asymmetry [3]. The local phase computation of a 1D signal uses a complex analytic signal comprised of the original signal as the real part and its corresponding Hilbert transform as the imaginary part. However, since the Hilbert transform is mathematically restricted to 1D with no straightforward extension to 2D and 3D, we used the method described in [4] to extend the concept of the analytic signal to higher dimensions using a monogenic signal. The higher dimension monogenic signal is generated by combining a bandpass Gaussian-derivative filter with a vector-valued odd filter (i.e., a Reisz filter). The low frequency variations in the local phase are extracted using a high spread (σ) Gaussian-derivative filter, while the high frequency components are extracted using a low spread (σ) Gaussian-derivative filter.

The first step in our workflow is to highlight low frequency variation in local phase of the 2D+T image volume, by computing feature asymmetry measure from the monogenic signals generated using bandpass filter with low center frequency. Since the innermost step edges represent blood pool (BP) region with very high confidence,

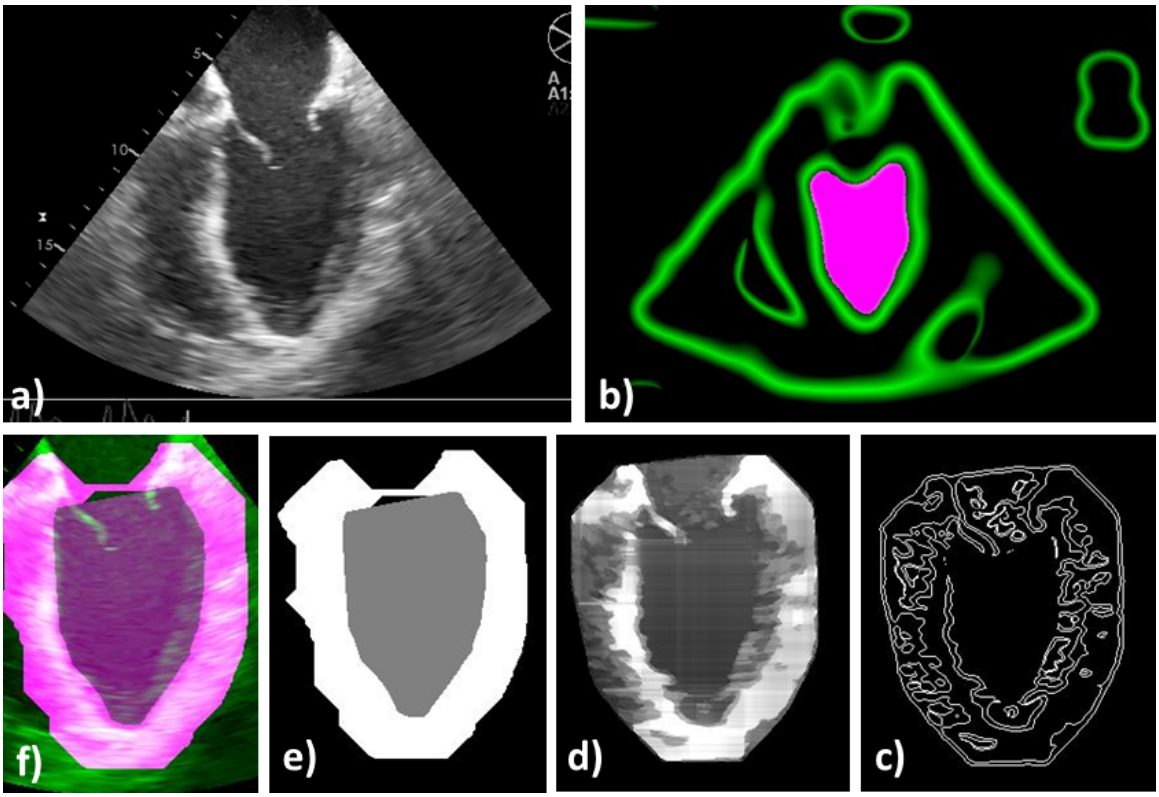


Figure 2.1: Segmentation Workflow: a) original US image, (b) high spread (σ) low frequency monogenic filter applied to the "2D + time" image dataset shown with the high confidence blood pool mask, (c) low spread (σ) high frequency monogenic filter output with blood pool removed, (d) "cartoon" image with enhanced regions, and (e) graph cut segmentation output (f) superimposed onto the original image.

we obtain these innermost points by switching to polar coordinates and hence generate a BP mask (**Fig. 2.1b**) using spline interpolation [5].

2.2.1.2 Image Smoothing

We perform frame averaging to reduce the noise present in the US image sequence. We then obtain the high frequency variations (edges) in the local phase of the frame-averaged 2D image ROI by applying the monogenic filter with high center frequency. The edges obtained within the high confidence BP mask are removed for better smoothing of the BP region (**Fig. 2.1c**). For each row and column, we replace

the intensity of the pixels between two edge pixels by their local mean, and combine them to obtain an edge-preserved smooth image, which we refer to as a "cartoon image" (**Fig. 2.1d**).

2.2.1.3 Graph Cut Segmentation of End-Diastole Image

The resulting "cartoon" image is used to construct a four neighborhood graph structure in which each pixel is connected to its east, west, north and south neighbors. Three special nodes called terminals are added, which represent three classes (labels): background, blood pool, and myocardium. The segmentation can be formulated as an energy minimization problem to find the labeling f, such that it minimizes the energy:

$$E(f) \;=\; \sum_{\{p,q\}\in\mathcal{N}} V_{p,q}(f_p, f_q) \;+\; \sum_{p\in\mathcal{P}} D_p(f_p), \tag{2.1}$$

where the first term represents the smoothness energy which forces pixels p and q, defined by the set of interacting pair of pixels $\mathcal{N}$, towards same label, and the second term represents the data energy that reduces the disagreement between the labeling f and the observed data.

The links between each pixel and the terminals (i.e., t-links) are formulated as the negative logarithm of the normal distribution:

$$D_p(f_p) = -ln\left(\frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(i_p-\mu)^2}{2\sigma^2}\right)\right), \tag{2.2}$$

where i_p is the pixel intensity, and μ and σ are the mean and standard deviation (SD) of pixel intensities for different classes, respectively. Since the background region is mostly dark, the intensity mean and SD for the background region are empirically selected as 0 and 0.01, respectively, whereas that for the BP and muscle regions are computed using the BP mask obtained from the monogenic filter output.

The links between neighboring pixels, called n-links, are weighted according to

their intensity similarity to formulate the smoothness energy:

$$V_{p,q}(f_p, f_q) = \begin{cases} 2K \cdot T(f_p \neq f_q) & \text{if } |I_p - I_q| \leq C \\ K \cdot T(f_p \neq f_q) & \text{if } |I_p - I_q| > C \end{cases} \tag{2.3}$$

where $T(\cdot)$ is 1 if its argument is true, and otherwise 0, K is a constant, and C is a intensity threshold that forces the neighboring pixels within the threshold towards the same label.

A cut in a graph separates the vertices such that they are left connected with a single terminal. Cut cost equals the sum of edge weights through the cut. Minimum cut is defined as the cheapest cut among all cuts, which can be found by computing the maximum flow between the terminals. The corresponding labeling minimizes the overall energy of the graph. We use the expansion algorithm [2] to find the minimum cut yielding the optimal segmentation of background, blood-pool, and myocardium (**Fig. 2.1e**).

2.2.2 Frame-to-frame Feature Tracking and Propagation

2.2.2.1 Image Preprocessing

Once the end-diastole image is segmented using the procedure outlined in **2.2.1**, the extracted features are tracked and propagated throughout the cardiac cycle using non-rigid registration. Prior to registration, a region of interest spanning the entire LV including blood-pool, myocardium, and surrounding region is obtained as the dilated convex hull of the BP mask (**Fig. 2.2b**). Moreover, the mitral valve region of the BP segmentation is "trimmed" using a straight line joining the leaflet hinges (**Fig. 2.1e**).

2.2.2.2 Non-rigid Registration Algorithm

The employed registration algorithm is a modified version of the biomechanics-based algorithm proposed by Lamash et al.[6], adapted to operate on 2D rather than

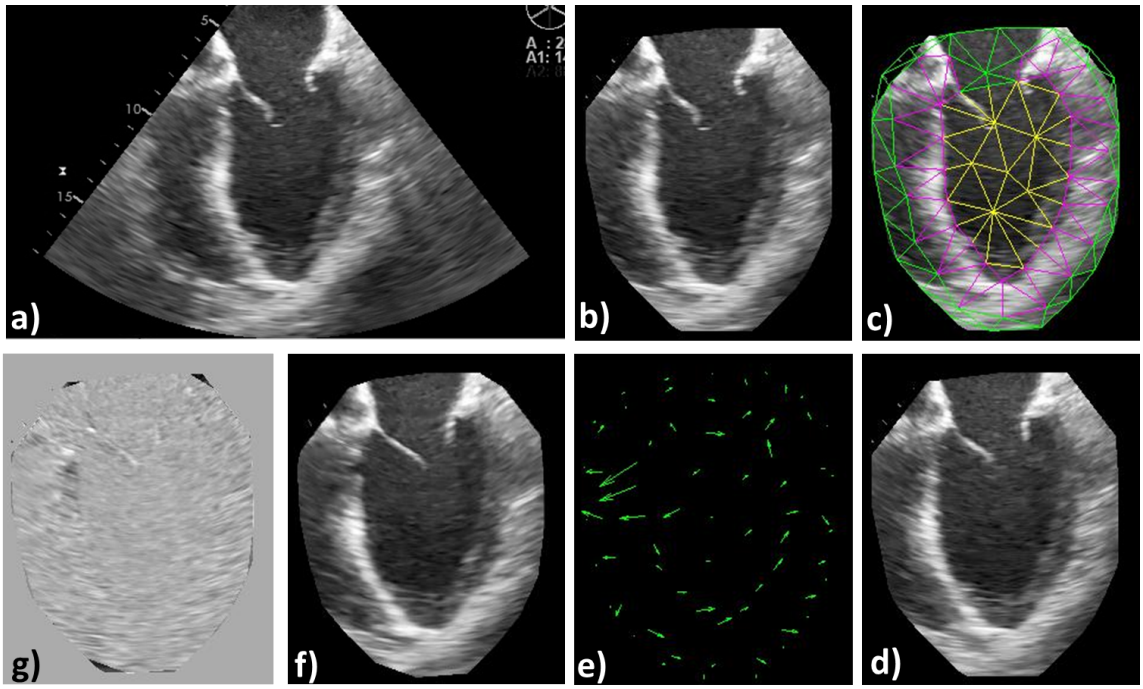


Figure 2.2: Registration workflow: a) the original image is "prepared" by automatically identifying an LV-centered ROI (b) onto which the mesh is applied (c), then registered to the target image (d); the resulting displacement field (e) is applied to the pre-registered image (b) to obtain the registered image (f), which can be compared to the target image (d) by visualizing the digitally subtracted image (g).

3D images. The LV anatomy is modeled as a two compartment model consisting of muscle — linear elastic, isotropic, and incompressible, and blood-pool, with prescribed smoothness constraints to allow rapid motion of the endocardial contour.

We initialize the algorithm by first discretizing the endocardial and epicardial contours, then constructing a mesh of the blood-pool and myocardium (**Fig. 2.2c**). Rather than resorting to a rectangular grid, we account for the local curvature of the endocardial border using a finite-element like mesh defined via linear shape functions. The algorithm deforms the mesh by estimating the required deforming forces that minimize the sum of the squared difference between the initial and target images. To avoid large deformations and ensure a smooth displacement field, a linear elastic regularization approach [7] is utilized.

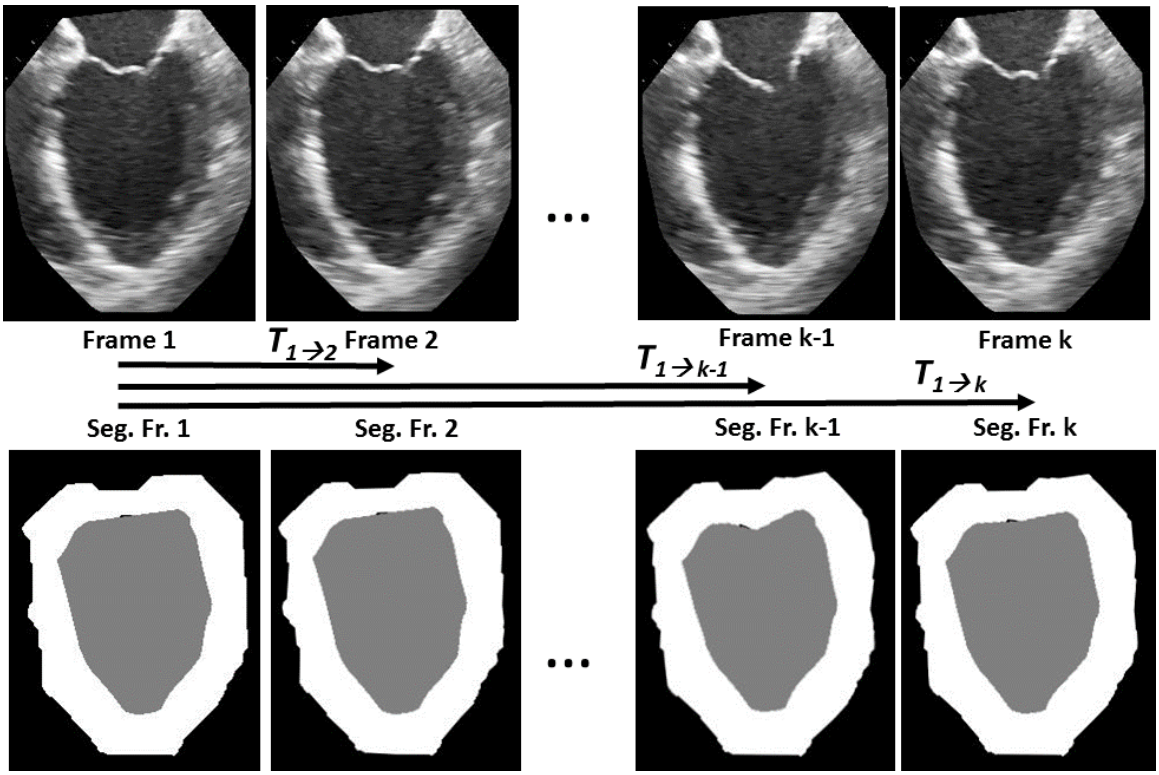


Figure 2.3: The frame-to-frame motion transforms ($T_{(k-1)\to k}$) are estimated by non-rigidly registering adjacent images in the sequence, then concatenated ($T_{1\to k} = T_{1\to 2} \cdot \ldots \cdot T_{(k-1)\to k}$) and applied to the segmented end-diastolic (ED) frame ($F_k = T_{1\to k} \cdot F_1 = T_{(k-1)\to k} \cdot \ldots \cdot T_{1\to 2} \cdot F_1$).

The frame-to-frame displacement field $T_{(k-1)\to k}$ is estimated by non-rigidly registering adjacent images in the cardiac cycle. The computed transforms are then concatenated ($T_{1\to k} = T_{1\to 2} \cdot T_{2\to 3} \cdot \ldots \cdot T_{(k-1)\to k}$) and applied in reverse order to the segmented end-diastole frame ($F_k = T_{1\to k} \cdot F_1 = T_{(k-1)\to k} \cdot \ldots \cdot T_{2\to 3} \cdot T_{1\to 2} \cdot F_1$), enabling feature propagation through the cardiac cycle (**Fig. 2.3**). This frame-to-frame registration procedure allows us to use the most reliable (and easy to accurately segment) end-diastole frame as an anchor and propagate the segmentation via small deformations at a lower computational cost, compared to the group-wise registration of the frames performed using computationally expensive single joint optimization.

2.2.3 3D LV Volume Reconstruction

Following the segmentation of each of the tri-plane views at end-diastole using the technique in **2.2.1** and their propagation throughout the cardiac cycle using the registration technique in **2.2.2**, the resulting images are re-inserted into a pseudo-3D image volume along the same orientation at which they were originally acquired (i.e., 60° apart) corresponding to each cardiac phase. The boundary points of each segmented contour at the same elevation are then fitted using the parametric variational cubic spline technique in [5]. The spline interpolated data is used to generate a convex hull using the algorithm proposed in [8]. The schematic illustration of the 3D LV reconstruction is shown in **Fig. 2.4**.

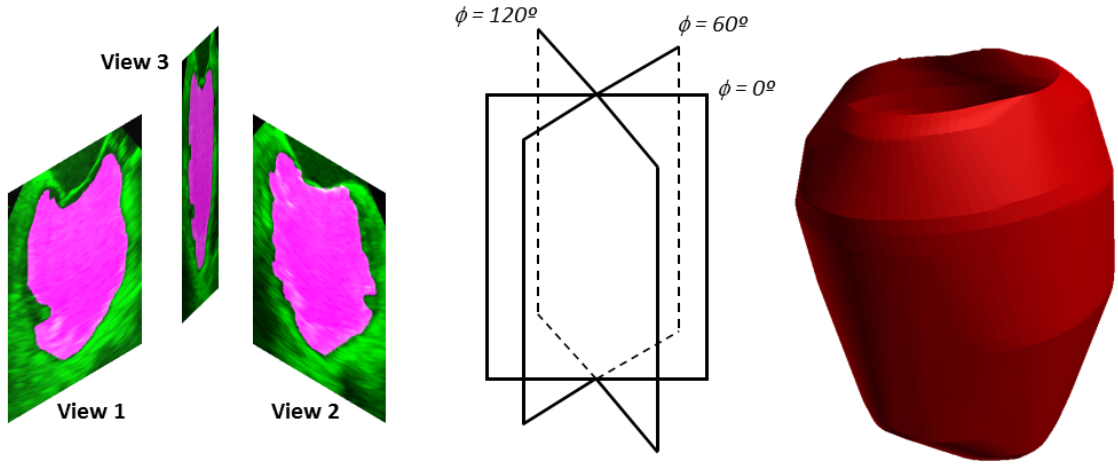


Figure 2.4: Schematic illustration of the 3D LV reconstruction: the tri-plane views at 60° (a) are inserted at their appropriate orientation (b), followed by spline interpolation and convex hull generation (c).

2.3 Evaluation and Results

We conducted a study using retrospective tri-plane time series data spanning multiple cardiac cycles from patients who underwent TEE imaging for cardiac function assessment. Since the proposed framework encompasses three different components

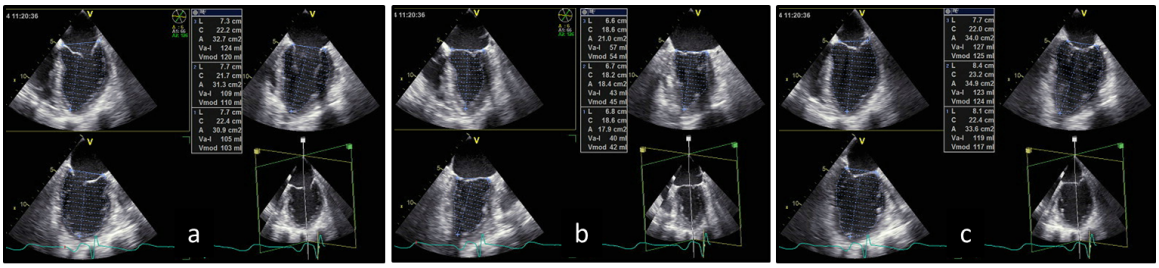


Figure 2.5: Manual segmentation of the LV performed by an expert cardiologist. Each figure consists of three views from the tri-plane TEE acquisition at: (a) mid-diastole, (b) end-systole and (c) end-diastole phases. The quantitative measurements of the three views calculated by the *GE EchoPac PC* commercial software can be seen in the table for each figure, the order being 1,2 and 3 for top-left, top-right and bottom-left panels, respectively. Also shown is the 3D orientation of the three views in the bottom-right of each image.

— automatic extraction of endocardial features, registration-based feature tracking and propagation, and volume reconstruction — we assessed the performance of each component against the ground truth, which consists of the blood-pool representation annotated manually by the expert clinician, using the EchpPac PC clinical software (**Fig. 2.5**). In addition, we also evaluated the performance at each stages of our application running in MATLAB on an Intel® Xenon® 3.60 GHz 32GB RAM PC.

Table 2.1: Comparison between the blood-pool area measurements (Mean ± Std. Dev. [cm^2]) annotated by the expert (Ground Truth) and the area obtained via A — automatic feature detection from individual frames; B — single phase automated feature detection + registration-based propagation; and C — single phase expert manual annotation + registration-based propagation. Measurements are evaluated at two cardiac phases —- end-diastole (ED) and end-systole (ES) — and averaged across all views and cardiac cycles spanned by the acquired data.

Blood-pool Area [cm^2] vs. Method	**ED**	**ES**
Ground Truth: Multi Phase Expert Manual Seg	52.1 ± 3.2	50.4 ± 4.6
Method A: Multi Phase Auto Seg	51.2 ± 3.5	48.9 ± 4.3
Method B: Single Phase Auto Seg + Reg	50.1 ± 4.0	48.3 ± 4.6
Method C: single Phase Manual Expert Seg + Reg	49.8 ± 4.6	48.2 ± 5.1

2.3.1 Automatic Direct Frame LV Segmentation Evaluation

We first evaluated the accuracy of our automatic, direct frame endocardial feature extraction component against expert manual annotation of the same features from the same frames performed by a cardiologist using the GE EchoPac PC clinical software. **Table 2.1** summarizes the blood-pool area measurements annotated by the expert (Ground Truth) and the area obtained via A — automatic feature detection from individual frames; B — single phase automated feature detection + registration-based propagation; and C — single phase expert manual annotation + registration-based propagation. Measurements are evaluated at two cardiac phases — end-diastole (ED) and end-systole (ES) — and averaged across all views and multiple cardiac cycles spanned by the acquired sequences. Our automatic blood-pool extraction technique required 26.5s to segment a "2D + time" 15 frame TEE tri-plane sequence.

Table 2.2: Mean ± Std. Dev. of several metrics — Dice Coefficient [%], Hausdorff Distance [mm], Mean Absolute Distance (MAD) Error [mm], and Endocardial TRE [mm] — used to compare the expert clinicians' blood-pool annotations (Ground Truth) with the blood-pool annotation obtained via A — automatic feature detection from individual frames; B — single phase automated feature detection + registration-based propagation; and C — single phase expert manual annotation + registration-based propagation. Measurements are evaluated at two cardiac phases — end-diastole (ED) and end-systole (ES).

Comparison Metrics	**Expert vs. A**		**Expert vs. B**		**Expert vs. C**	
	ED	**ES**	**ED**	**ES**	**ED**	**ES**
DICE Coeff [%]	94.9±0.7	94.7±1.4	93.8±0.9	94.6±1.0	95.1±1.0	95.2±1.8
Haussdorf Dist [mm]	4.7±0.9	5.2±1.3	7.9±3.5	5.9±1.3	6.4±1.7	5.4±2.1
MAD Error [mm]	1.5±0.3	1.6±0.6	1.9±0.4	1.7±0.5	1.7±0.2	1.8±0.7
Endocardial TRE [mm]	1.9±0.2	2.0±0.7	2.6±0.7	2.1±0.5	2.2±0.2	2.2±0.8

2.3.2 Registration-based Blood-pool Propagation Evaluation

To evaluate the accuracy with which the non-rigid registration algorithm propagates the extracted features throughout the cardiac cycle, we employed several met-

rics, including the Dice correlation, Hausdorff distance, mean absolute distance error and endocardial target registration error (TRE) computed between the ground truth blood-pool manually annotated by the expert and the blood-pool depicted via three other methods under consideration, as shown in **Table 2.2**.

Fig. 2.6 visually compares the ground truth blood-pool annotation performed by the expert clinician to that extracted via direct frame feature identification, as well as registration-based propagation of the single-frame blood-pool annotated either manually by the expert or automatically using the first component of our proposed framework. The segmentation propagation technique required $162s$ to run through a 15 frame tri-plane TEE sequence.

2.3.3 3D Volume Reconstruction Evaluation

Lastly, we assessed the accuracy of the 3D LV reconstruction procedure by comparing the reconstructed LV volume to that estimated by the GE EchoPac PC clinical software following expert manual segmentation. The end-diastolic and systolic volume measurements are summarized in **Table 2.3**, along with the corresponding ejection fraction measurements. Performance-wise, the LV volume reconstruction from a tri-plane sequence requires $11.6s$.

Table 2.3: Comparison between the LV blood-pool volume and Ejection Fraction (EF) between expert manual annotations (Ground Truth) and A — automatic feature detection from individual frames; B — single phase automated feature detection + registration-based propagation; and C — single phase expert manual annotation + registration-based propagation. Measurements were evaluated at two cardiac phases — end-systole (ES) and end-diastole (ED).

LV Assessment Metric	**EchoPac**		**Auto A**		**Manual + Reg B**		**Auto + Reg C**	
	ED	**ES**	**ED**	**ES**	**ED**	**ES**	**ED**	**ES**
Mean Vol [mL]	249.0	223.0	247.6	220.8	242.0	217.7	232.0	209.6
Std Dev Vol [mL]	3.5	10.8	3.5	3.8	2.0	1.5	10.4	9.8
LV EF (%)	10.4 ± 5.6		10.9 ± 2.0		10.0 ± 0.8		9.6 ± 0.4	

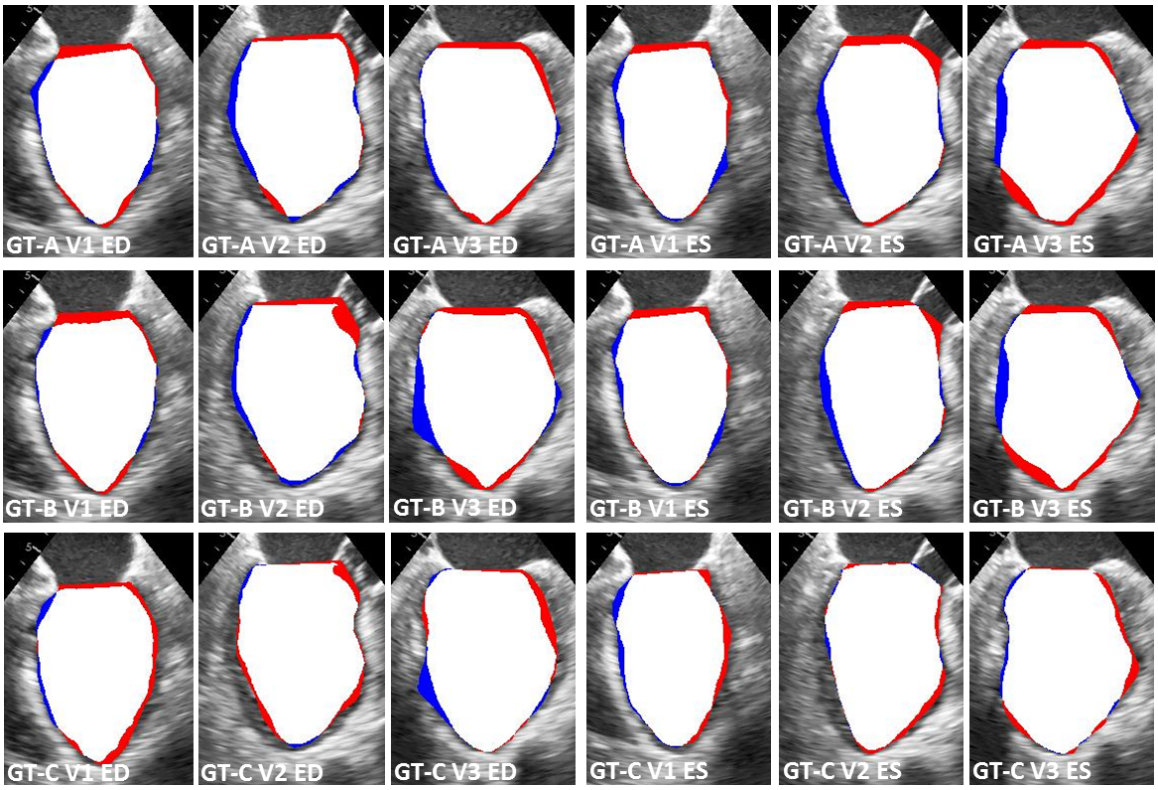


Figure 2.6: Visual comparison of the blood-pool annotations achieved via *A — automatic feature detection from individual frames; B — single phase automated feature detection + registration-based propagation; and C — single phase expert manual annotation + registration-based propagation* vs. the ground truth expert manual blood-pool annotation (GT) quantified at end-diastole (ED) and end-systole (ES) for the three tri-plane views (V1, V2 and V3). White regions are common between the GT and each of the three A, B and C blood-pool estimates, red regions belong to the expert annotated blood-pool (GT), while the blue regions belong to the blood-pool area depicted by each of the three annotation methods A, B or C under comparison. Panels are named according to the same convention — i.e., the panel labeled *GT-B V2 ES compares the ground truth expert-annotated blood-pool (GT) to the blood-pool annotated using Method B displayed in View 2 at end-systole* (Color version available online).

2.4 Discussion

We described the implementation and clinical data evaluation of a rapid, automatic framework that encompasses well-evaluated filtering, segmentation, registration, and volume reconstruction techniques as a means to provide a rapid, robust and

accurate framework for feature tracking from multi-plane ultrasound image sequences. All components of the proposed technique — segmentation, registration-based feature tracking and propagation, and 3D blood-pool volume reconstruction — were assessed against expert manual segmentation at both the systolic and diastolic cardiac phases and demonstrated accurate and consistent performance, while significantly minimizing user-induced variability. Furthermore, unlike other techniques that operate on 3D datasets, this technique enables rapid and consistent analysis of multi-plane, 2D US image sequences — the standard format for acquisition, interpretation, and analysis of cardiac US images.

As the proposed work-flow integrates multiple algorithms, the influence of different parameters in the segmentation result is an important consideration. The frequency specific to the monogenic filter operates over a wide range of values and yields a good quality "cartoon image" for further segmentation. Similarly, for the graph cut algorithm, the mean and standard deviations for the blood pool, muscle and background regions are adaptively extracted from the image content, while the threshold 'C' that constraints the pixels towards same label can span a sufficiently wide range without significantly effecting the segmentation result. Furthermore, Lamash *et al.* [6] have thoroughly studied the effects of various regularization parameters in the biomechanics-based registration; for our purpose we selected the optimal parameters as suggested by the paper [6]. In summary, the proposed work-flow yields a consistent segmentation result over a wide range of parameter values.

Unlike expert manual segmentation that is highly sensitive to intra- and inter-observer variability, the proposed technique provides a consistent result for each dataset, which can be reviewed and improved, if needed, by expert clinicians. The single-phase feature extraction, followed by tracking and propagation via registration further reduces uncertainty, avoiding the need to segment each frame independently by using the *a priori* frame information along with the image sequence to achieve optimal segmentation. Hence, should the expert clinician choose to perform any ad-

justments to the single-phase segmentation, their precise tracking and propagation throughout the cardiac cycle is guaranteed by the registration-based implementation.

2.5 Summary and Future Work

The impact and contribution of the proposed work is the integration of several image processing techniques (i.e., phase-based filtering, segmentation, registration and volume reconstruction) into a streamlined work-flow that utilizes traditional standard-of-care 2D multi-plane TEE image sequences and fits seamlessly within the current work-flows associated with both cardiac function assessment and intra-operative cardiac intervention guidance and monitoring.

Ongoing and future efforts include further evaluation and demonstration of how the proposed technique can cater to dynamically reconstructing 3D endocardial LV representations that facilitate computer-assisted assessment of stroke volume and ejection fraction, as well as employing intra-operative multi-plane 2D TEE data to dynamically update and animate CT and/or MRI anatomy depicted pre-operatively to better represent the intra-operative conditions.

Lastly, although we believe the most meaningful assessment is still against the expert clinicians analysis of the same input data, we acknowledge the importance of assessing the output of our proposed framework against the output of other techniques and extend the analysis to a large dataset of multi-plane image sequences acquired across multiple cardiac cycles.

Besides its direct application to computer-aided cardiac function assessment, the proposed framework is readily adaptable to the guidance and monitoring of image-guided cardiac interventions, most of which involve the use of real-time ultrasound imaging — the clinical standard of care for cardiac procedures.

Bibliography

[1] M. Mulet-Parada and J. A. Noble. 2D + T acoustic boundary detection in echocardiography. *Med Image Anal.*, 4:21–30, 2000.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[3] P. Kovesi. Symmetry and asymmetry from local phase. In *Proc 10th Australian Joint Conf Artificial Intelligence*, pages 2–4, 1997.

[4] K. Rajpoot, V. Grau, and J. A. Noble. Local-phase based 3d boundary detection using monogenic signal and its application to real-time 3-D echocardiography images. In *Proc. IEEE Int Symp Biomed Imaging*, pages 783–6, 2009.

[5] E. T. Y. Lee. Choosing nodes in parametric curve interpolation. *Comput Aided Des.*, 21:363–70, 1989.

[6] Y. Lamash, A. Fischer, S. Carasso, and J. Lessick. Strain analysis from 4D cardiac CT image data. *IEEE Trans Biomed Eng.*, 62:511–21, 2015.

[7] B. Zitova and J. Flusser. Image registration methods: a survey. *Image Vision Comput.*, 21:977–1000, 2003.

[8] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Trans Math Software*, 22:469–83, 1996.

Chapter 3

Probabilistic Atlas Prior based Graph Cut Segmentation: Application and Validation on Left Ventricle Slice-wise Segmentation from Cine Cardiac MRI

We propose a framework for left ventricle (LV) segmentation from cardiac cine MRI [1]*. First, we segment the LV blood pool using iterative graph cuts, and subsequently use this information to segment the myocardium. We formulate the segmentation procedure as an energy minimization problem in a graph subject to the shape prior obtained by label propagation from an average atlas using affine registration. The proposed framework has been validated on 30 patient cardiac cine MRI datasets*

[1]This chapter is adapted from Dangi S. et al., "Integrating Atlas and Graph Cut Methods for Left Ventricle Segmentation from Cardiac Cine MRI." In: Mansi T., McLeod K., Pop M., Rhode K., Sermesant M., Young A. (eds) Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges. STACOM 2016. Lecture Notes in Computer Science, vol 10124. Springer, Cham

available through the STACOM LV segmentation challenge and yielded fast, robust, and accurate segmentation results.

3.1 Introduction

Various segmentation techniques for cardiac MR images have been proposed in the literature [1]. Among them, multi-atlas based approaches have shown promising results in biomedical image segmentation [2]. However, they rely on a number of computationally demanding and time limiting nonrigid image registration steps followed by label fusion. Hence despite their accuracy, most existing segmentation techniques have experienced minimal adoption in the actual clinical applications primarily due to their complexity, high dependence on parameters variability, and computational demands.

On the other hand, combinatorial optimization based graph-cut techniques are fast and guaranteed to produce results within a known factor of the global minimum, for some special classes of functions (termed as regular functions) [3] and have proved to be powerful tools for image segmentation. Moreover, adding a shape constraint into the graph cut framework has been shown to improve the cardiac image segmentation results significantly [4–6]. However, these methods require a manual input to introduce a shape constraint at the right location in the image.

In this work, we leverage the performance of the graph cut framework and augment it by incorporating shape constraints [7, 8] in the form of an average atlas-based segmentation of the anatomy whose label was generated and propagated using a single affine registration. Subsequently, we iteratively refine the segmentation to obtain an accurate and robust segmentation of the myocardium. Hence, we do not require any manual input to introduce shape constraint into the graph-cut framework and simultaneously take advantage of the prior knowledge in the form of atlas-based segmentation requiring affine as opposed to nonrigid registration, which is more computationally

efficient and less sensitive to parameters variability.

3.2 Methodology

Whole heart cine MRI images are generated by stacking 2D+T short-axis slices acquired during a single breath hold. Since this acquisition approach introduces an intensity difference between the slices, as well as slice misalignments, we can follow one of two approaches to segment the data: one approach is to implement a slice motion correction protocol to realign the slices into a coherent 3D volume. The other approach, also implemented here, resorts to slice-wise processing and segmentation instead of a 3D segmentation.

Another challenge is the ill-defined contrast of the LV myocardium in MR images, which makes the image-driven segmentation difficult. As such, to obtain better segmentation of the apical and basal regions, we exploit the prior knowledge in the form of an average atlas. The proposed methodology formulates the segmentation problem in the context of a graph based energy minimization framework. The blood pool is first segmented using an iterative graph cut technique; then, this information is used to segment the myocardium.

3.2.1 Data Preprocessing

This study is conducted on 30 cardiac cine-MR images taken from the DETERMINE [9] cohort available as a part of the STACOM Cardiac Atlas Segmentation Challenge Project database[2]. The semi-automatically segmented images obtained by applying the method described in [10] accompany the dataset and serves as the gold-standard for assessing the proposed segmentation technique.

We select a reference patient volume with good contrast (via visual inspection), average size (104mm apex-base; dataset range: 80mm to 140mm), and horizontal

[2] http://www.cardiacatlas.org

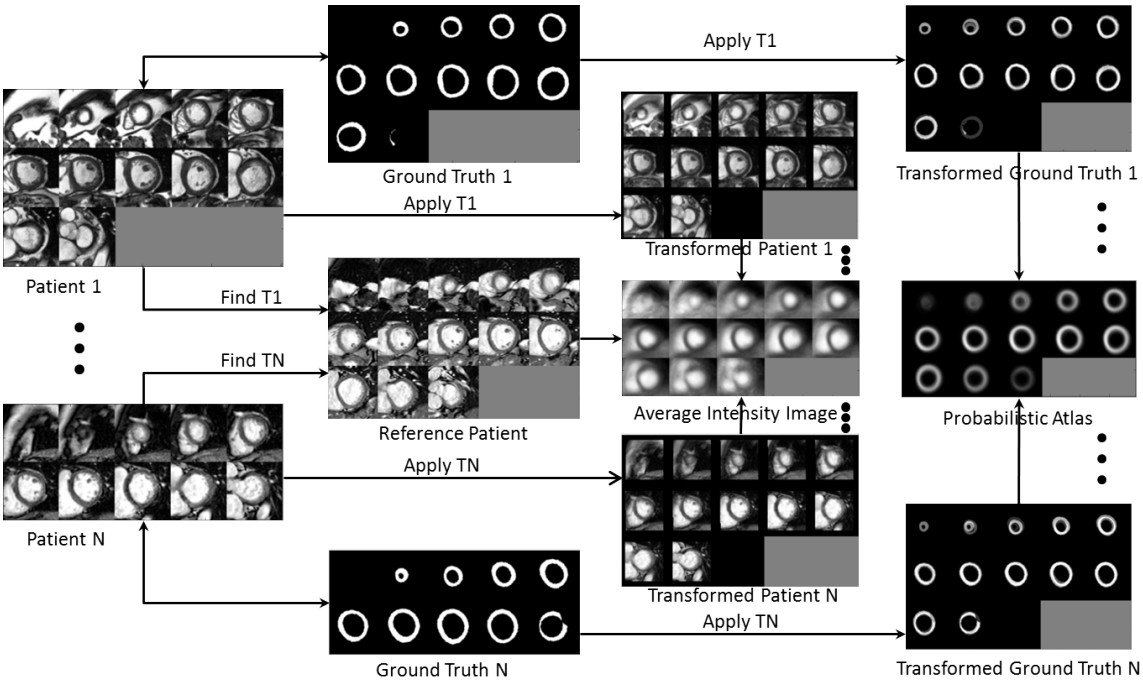


Figure 3.1: All patient images are affinely registered to the reference patient, and the obtained optimum transformation is applied to the corresponding ground truth images. An average intensity image is obtained by averaging the intensities of all transformed patient images, while, the averaging of the transformed ground truth images yields a probabilistic atlas.

LV-RV orientation (as shown in **Fig. 3.1**). All patient volumes are rotated about the z-axis (i.e., slice-encoding direction) to roughly align their orientation with that of the reference patient using the DICOM Image Orientation Patient (IPP) field. The region of interest (ROI) (in the xy-plane) enclosing the left and right ventricles is extracted using the method described in [11] by correlating the 2D motion images generated from the 3D volumes across the cardiac cycle. The only manual input required by our algorithm is the start and end slices of the LV, such that the ROI is restricted in the z-direction, preventing over/under segmentation of slices that do not belong to the desired anatomy. The patient volumes are cropped to the above ROI, and, to compensate for any intensity differences (due to the slice-wise acquisition), each slice is normalized (0.0—1.0) prior to further processing.

3.2.2 Atlas Generation

The cropped 3D volumes (at the end diastole phase) for all patients are first histogram matched and then affinely registered to the reference patient image volume using the intensity based Nelder-Meade downhill simplex algorithm [12] available in SimpleITK. The resulting 3D affine transforms are applied to the respective ground truth segmentations. The transformed volumes and transformed ground truths are then averaged to obtain an average appearance atlas and a probabilistic atlas, respectively (**Fig. 3.1**).

The average appearance atlas is registered to a test volume using intensity-based affine registration. The resulting registration transformation is used to transform the myocardial probabilistic label to the test data, which, in turn, serves as a shape constraint for the graph cut framework.

3.2.3 LV Blood Pool Segmentation using Iterative Graph Cuts

To leverage the 3D LV geometry, we use the blood pool (BP) segmentation of a given slice to help refine the BP ROI in the neighboring slices. As such, we first segment the BP from the mid-slice, followed by its neighboring slices, and proceed accordingly, until the complete volume is segmented.

3.2.3.1 Intensity Distribution Model:

The myocardium probability map for each slice is normalized and inverted to produce the probability map corresponding to the blood pool (BP) and background (BG). The resulting BP/BG probability map is thresholded at 0.5 and the inner connected component is isolated to obtain the high confidence BP ROI. Otsu thresholding [13] is applied within this ROI to obtain the initial BP region. The intensity values within this extracted BP region are then fitted to a Gaussian distribution to generate the

BP intensity model.

A binary mask enclosing the myocardium is obtained by thresholding the myocardial probability map at a very small value (i.e. 0.1). Holes in the binary mask are filled to obtain a ROI enclosing the BP, myocardium, and BG. To generate the BG intensity model, we use the Expectation-Maximization (EM) algorithm available in the scikit-learn library. Specifically, we fit the intensity values within the ROI, excluding the initial BP region, to a Gaussian Mixture Model (GMM) comprising two Gaussians. **Fig. 3.2a** shows the resulting BP log-likelihood map.

Note that we propose the Gaussian distribution for modeling intensity noise in MR images instead of a more appropriate Rician distribution [14]; this simplifies our model and is a good approximation when the signal-to-noise ratio is high.

3.2.3.2 Blood Pool/Background Probabilistic Map:

To obtain a ROI that includes the myocardium and BP, we threshold the myocardial probability map at 0.5, fill in the blood pool, and erode the resulting ROI by 15% (selected empirically) of the radius of its smallest circumscribed circle to obtain the BP ROI. The BP/BG probability map masked by the BP-ROI represents the BP probability map, and its inverse represents the BG probability map. **Fig. 3.2b** shows the BP log-likelihood map.

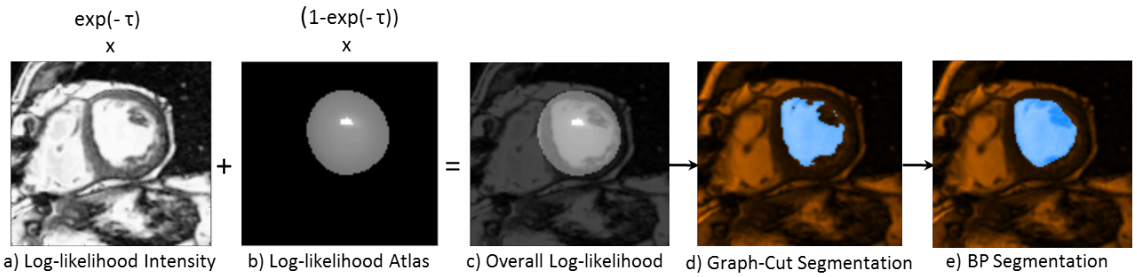


Figure 3.2: Log-likelihood image obtained from: a) the intensity distribution model, b) the BP probabilistic map, c) weighted sum of (a) and (b); d) segmentation obtained from graph cut, e) convex hull of (d) yields the BP segmentation.

3.2.3.3 Graph-Cut Segmentation:

We construct a graph with each node (i.e., pixel) connected to its east, west, north, and south neighbors. Two special terminal nodes representing two classes — the source (blood pool), and the sink (background) — are added to the graph and all other nodes are connected to each terminal node. The segmentation is formulated as an energy minimization problem over the space of optimal labelings f:

$$E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q), \tag{3.1}$$

where the first term represents the data energy that reduces the disagreement between the labeling f_p given the observed data at every pixel $p \in P$, and the second term represents the smoothness energy that forces pixels p and q defined by a set of interacting pair $\mathcal{N}$ (in our case, the neighboring pixels) towards the same label.

The data energy term is represented by the terminal link (t-link) between each node and the source (or sink), which is defined as the weighted sum of the log probabilities of the intensity distribution model and the probabilistic map corresponding to the BP (or BG):

$$D_p(f_p) = exp(\tau) * [-lnPr(I_p|f_p)] + (1 - exp(-\tau)) * [-lnPr(f_p)] \tag{3.2}$$

where, τ is the iteration number, $Pr(I_p|f_p)$ is the likelihood of observing the intensity I_p given that pixel p belongs to class f_p, and $Pr(f_p)$ is the prior probability for class f_p obtained from the BP/BG probability map. The log-likelihood difference between BP and BG labels for $\tau = 1$ is shown in **Fig. 3.2c**. The intensity likelihood term (first term) allows the expansion of the BP region in the first few iterations, whereas the prior probability map (second term) restricts its "spilling" (due to over-segmentation) in subsequent iterations.

The smoothness energy term is computed over the links between neighboring nodes

(n-links), which are weighted based on their intensity similarity:

$$V_{p,q}(f_p, f_q) = \begin{cases} \tau * exp\left(-\frac{|I_p - I_q|}{\tau}\right) & \text{if } f_p \neq f_q \\ 0 & \text{if } f_p = f_q \end{cases} \tag{3.3}$$

where I is the pixel intensity. To avoid the "spilling" of the BP into the myocardium or BG, the smoothness term changes with each iteration, such that, in order for the neighboring pixels to be assigned to the same label during the current iteration, their intensities must be closer than in the previous iteration.

Once weights are assigned to all edges in the graph, the minimum cut equivalent to the maximum flow is identified via the expansion algorithm described in [15]. This approach yields the labeling (graph-cut) that minimizes the global energy of the graph that corresponds to the optimal segmentation (**Fig. 3.2d**). Lastly, the convex hull applied to the graph-cut result constitutes the final BP segmentation, such that, the papillary muscles are included within the BP (**Fig. 3.2e**).

3.2.3.4 Myocardial Probability Map Refinement

The myocardial probability map is thresholded at 0.5, and the inner hollow circular region representing the BP is extracted. The signed distance map corresponding to the boundary of the extracted BP region is affinely registered to the signed distance map generated from the boundary of the graph-cut extracted BP (**3.2.3**) segmentation. The optimum affine transformation that minimizes the sum of squared differences between the two distance maps is applied to the myocardial probability map, such that, it fits the shape of the segmented BP.

3.2.3.5 Iterative Refinement:

The latest BP segmentation obtained from the graph cut is used to update the intensity distribution model. The refined myocardial probability map is used to construct a new BP/BG probability map. The pixels within the latest BP segmentation

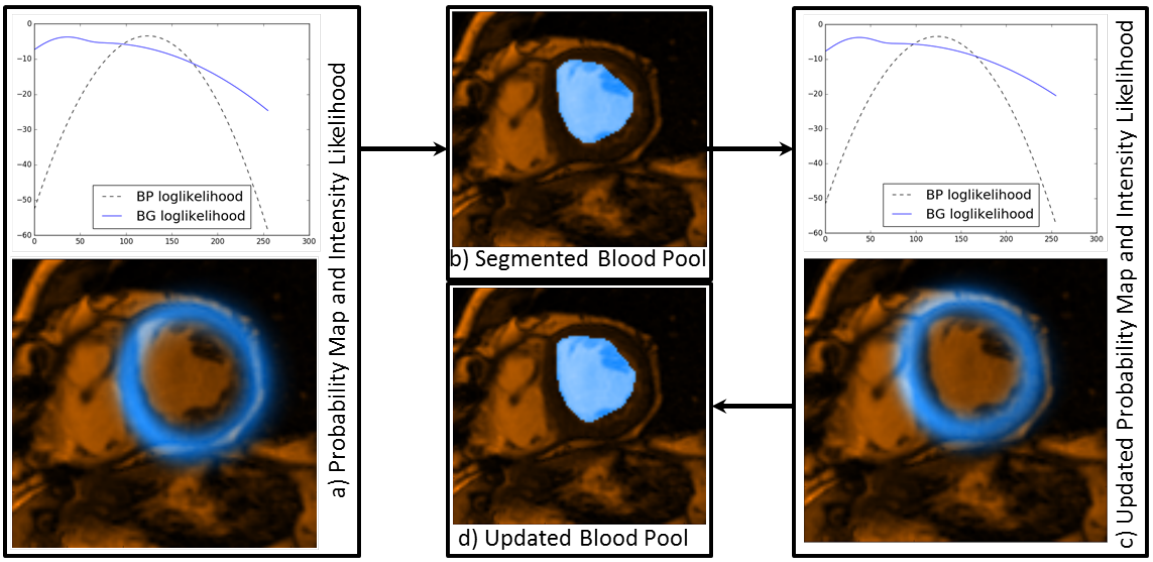


Figure 3.3: a) Probability map and intensity distribution model for current iteration, b) BP segmentation obtained from graph cut using (a), c) updated probability map and intensity distribution model obtained using (b), d) new BP segmentation obtained from graph cut using (c).

are assigned very high likelihood (for belonging to the BP), and hence their labels do not change. An updated BP segmentation is obtained via another graph cut operating on the new graph energy configuration. This iterative process is repeated until the changes in the affine transform parameters for the myocardium probability map are below a predefined threshold; this iterative process usually converges within three iterations. Upon convergence, the convex hull defined by the latest segmentation result constitutes the final BP segmentation. The iterative refinement process takes three iterations in average for convergence and is illustrated in **Fig. 3.3**.

3.2.4 Myocardium Segmentation

The information from the BP segmentation along with the refined myocardial probability map is used to segment the myocardium.

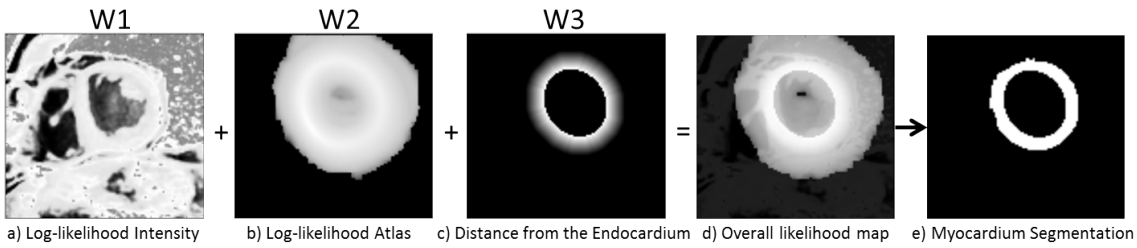


Figure 3.4: Log-likelihood image obtained from: a) the intensity distribution model, b) the refined myocardium probability map, c) distance from the endocardium; d) the weighted sum (w1, w2, and w3) of (a), (b), and (c), respectively; e) final myocardium segmentation obtained from graph cuts

3.2.4.1 Intensity Distribution Model:

We select a ROI in each slice based on the refined probability map, and we match the histogram of the pixel intensities within this ROI to the histogram of the mid-slice. We select the mid-slices (i.e. no apical/basal slices) to obtain a single intensity distribution model for the whole volume. The intensities of the pixels within the refined myocardial mask with probability higher than 0.5 are fitted to a single Gaussian GMM to obtain the myocardium intensity distribution model. Similarly, the intensities of the remaining pixels are fitted to a three Gaussian GMM to obtain the BG intensity distribution model. **Fig. 3.4a** shows the log-likelihood map for the myocardium.

3.2.4.2 Distance from the Endocardial Border:

The endocardial border is obtained from the outer edge of the final BP segmentation (**Fig. 3.2e**). The knowledge that myocardium should be closer to the endocardial border is encoded in the data term represented by the truncated distance map (empirically selected as 10 pixels). This constraint increases the likelihood of pixels near the endocardial border to be labeled as myocardium, while reducing this likelihood for the pixels located further away. Furthermore, to prevent the BP region from being labeled as myocardium, it is assigned the lowest likelihood value (**Fig. 3.4c**).

3.2.4.3 Graph-Cut Segmentation:

A graph is constructed similar to the formulation described in **3.2.3.3**, but this time to classify the myocardial rather than blood pool pixels. The data term is defined as the weighted sum of the intensity distribution model, refined myocardial probability map (as described in **3.2.3.4** and **Fig. 3.4b**), and the distance from endocardial border, with increasing relative influence, respectively. The smoothness term varies spatially according to the intensity difference between the neighboring pixes, as discussed in **3.2.3.3**. The minimum cut in the graph yields the final myocardium segmentation (**Fig. 3.4e**).

3.3 Results

The proposed algorithm was implemented in Python and required 45 seconds on average to segment the BP and myocardium from cine MRI volumes on an Intel® Xenon® 3.60 GHz 32GB RAM PC.

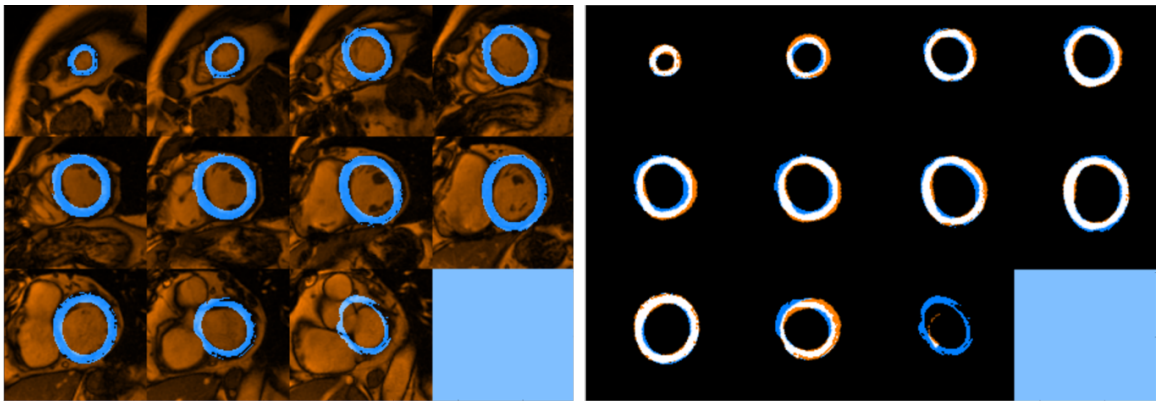
Adhering to the collated results reported for the LV segmentation challenge in [16], we evaluated our segmentation with four-fold cross-validation on 30 patient dataset according to the following metrics: dice index, jaccard index, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) [16]. To maintain approximately equal number of myocardium and non-myocardium pixels for evaluation, such that the NPV conveys some useful information, we dilated each slice of the myocardium region, for the provided gold standard segmentation, by one fourth of the radius of the disk with equivalent area. The segmentation results for a patient dataset are overlaid onto each slice of the patient volume and shown in **Fig. 3.5a**. **Fig. 3.5b** shows a visual comparison of our segmentation results vis-à-vis the provided semi-automated segmentation serving as a gold-standard. The metrics are summarized in **Table 3.1** for all slices together, as well as for the mid-slices and apical/basal slices (first and last two slices, respectively) separately.

Table 3.1: Evaluation of our segmentation results against the provided gold-standard semi-automated segmentation for the mid-slices, apical/basal slices according to Dice Index, Jaccard Index, Sensitivity, Specificity, PPV, and NPV.

Assessment Metric	Mid-Slices	Apical/Basal-Slices	All Slices
Dice Index	0.811 ± 0.068	0.568 ± 0.241	0.740 ± 0.180
Jaccard Index	0.687 ± 0.091	0.433 ± 0.222	0.613 ± 0.183
Sensitivity	0.854 ± 0.104	0.596 ± 0.268	0.783 ± 0.195
Specificity	0.788 ± 0.103	0.725 ± 0.180	0.770 ± 0.134
PPV	0.789 ± 0.079	0.714 ± 0.160	0.767 ± 0.114
NPV	0.866 ± 0.086	0.640 ± 0.224	0.800 ± 0.174

Table 3.2: Evaluation of segmentation results against the provided gold-standard semi-automated segmentation according to Jaccard Index. The values for methods AU, AO, SCR, and INR are obtained from Table 2 in [16] computed for the LVSC validation set against the consensus manual segmentation (CSMAN) as described in [16]. Values are provided as mean ± standard deviation, and in descending order by Jaccard index. SA/FA — Semi/Fully-Automatic

Method	SA/FA	Jaccard Index
AU [10]	SA	0.80 ± 0.19
AO [17]	SA	0.79 ± 0.19
Proposed	FA	0.61 ± 0.18
SCR [18]	FA	0.59 ± 0.19
INR [19]	FA	0.49 ± 0.10



a) Segmentation result superimposed on the original volume

b) Our segmentation result versus the Gold-standard segmentation

Figure 3.5: a) Final myocardium segmentation of all slices of a patient dataset (shown in blue) superimposed with the patient volume (shown in red); b) Final myocardium segmentation assessed against the provided gold-standard semi-automatic segmentation; white regions represent true positives, red regions represent false negatives, and blue regions represent false positives.

3.4 Discussion, Conclusion, and Future Work

Our validation experiments summarized in **Table 3.2** show that our segmentation results are as good as or better than the fully automatic methods, SCR [18] and INR [19], as reported in [16]. Nevertheless, both the semi-automatic methods AU [10] and AO [17] involved substantial human involvement: specifically, AU used guide-point inputs from human observer to fit a finite element cardiac model to the MR data, while AO used a modified optical flow algorithm to track an initial contour manually drawn on the first frame. However, it should be noted that the metrics reported in [16] were evaluated against the consensus manual segmentation (CSMAN) estimated based on three semi-automated methods obtained using STAPLE algorithm on the LV segmentation challenge validation set across the cardiac cycle, whereas ours is compared against the provided semi-automatic gold standard segmentation on the end-diastole phase, when myocardium is thinnest and more prone to segmentation errors, evaluated on 30 training set patients based on four-fold cross-validation. Hence, the metrics provide only an approximate estimate of our algorithm's performance

compared to the ones that participated in the challenge. Moreover, the average segmentation time of 45 secs per volume for an unoptimized code in Python presents a great potential of our algorithm for near real-time clinical applications.

Since the BP region in the mid-slices are better defined than in the apical/basal slices, the segmentation results are consistently better for the mid-slices. We also observed that the slice-wise processing and iterative refinement might compromise the segmentation of the apical/basal slices due to ill-defined BP regions, suggesting the need for special processing for these slices.

As part of our future work, we plan to automate the ROI detection in z-direction to eliminate the manual input required by our algorithm. In addition, instead of using a constant truncating endocardial distance constraint, we plan to use image-derived edge information to enable spatially varying truncating distances to improve the myocardium segmentation. Similarly, we will study the effect of selecting different thresholds for the probability maps, weight variability on the likelihood terms and, in turn, on the final myocardium segmentation. Lastly, we plan to extend the work and evaluate the segmentation performance on all 100 patient datasets and report performance according to the metrics outlined above.

Bibliography

[1] Caroline Petitjean and Jean-Nicolas Dacher. A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis*, 15(2):169 – 184, 2011.

[2] Juan Eugenio Iglesias and Mert R. Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205 – 219, 2015.

[3] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, Feb 2004.

[4] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan. Graph cut segmentation with a statistical shape model in cardiac {MRI}. *Computer Vision and Image Understanding*, 117(9):1027 – 1035, 2013.

[5] Daniel Freedman and Tao Zhang. Interactive graph cut based segmentation with shape priors. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, I:755–762, 2005.

[6] Dwarikanath Mahapatra. Cardiac Image Segmentation from Cine Cardiac MRI Using Graph Cuts and Shape Priors. *Journal of Digital Imaging*, 26(4):721–730, 2013.

[7] Greg Slabaugh and Gozde Unal. Graph cuts segmentation using an elliptical shape prior. *Proceedings - International Conference on Image Processing, ICIP*, 2:1222–1225, 2005.

[8] Nhat Vu and B. S. Manjunath. Shape prior segmentation of multiple objects with graph cuts. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[9] Alan H. Kadish, David Bello, J. Paul Finn, Robert O. Bonow, Andi Schaechter, Haris Subacius, Christine Albert, James P. Daubert, Carissa G. Fonseca, and Jeffrey J. Goldberger. Rationale and design for the defibrillators to reduce risk by magnetic resonance imaging evaluation (determine) trial. *Journal of Cardiovascular Electrophysiology*, 20(9):982–987, 2009.

[10] Bo Li, Yingmin Liu, Christopher J. Occleshaw, Brett R. Cowan, and Alistair A. Young. In-line automated tracking for ventricular function with magnetic resonance imaging. *JACC: Cardiovascular Imaging*, 3(8):860 – 866, 2010.

[11] Yehuda Kfir Ben-Zikri and Cristian A. Linte. A robust automated left ventricle region of interest localization technique using a cardiac cine MRI atlas, 2016.

[12] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

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

[14] HKon Gudbjartsson and Samuel Patz. The rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914, 1995.

[15] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[16] Avan Suinesiaputra, Brett R. Cowan, Ahmed O. Al-Agamy, Mustafa A. Elattar, Nicholas Ayache, Ahmed S. Fahmy, Ayman M. Khalifa, Pau Medrano-Gracia, Marie-Pierre Jolly, Alan H. Kadish, Daniel C. Lee, Jn Margeta, Simon K. Warfield, and Alistair A. Young. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac {MR} images. *Medical Image Analysis*, 18(1):50 – 62, 2014.

[17] Ahmed S. Fahmy, Ahmed O. Al-Agamy, and Ayman Khalifa. Myocardial segmentation using contour-constrained optical flow tracking. In *STACOM*, pages 120–128, 2012.

[18] Marie-Pierre Jolly et al. Automatic segmentation of the myocardium in cine MR images using deformable registration. In *STACOM*, pages 98–108, 2012.

[19] Ján Margeta et al. Layered spatio-temporal forests for left ventricle segmentation from 4D cardiac MRI data. In *STACOM*, pages 109–119, 2012.

Chapter 4

Probabilistic Atlas Prior based Graph Cut Segmentation: Application and Validation on Right Ventricle Slice-wise Segmentation from Cine Cardiac MRI

Right ventricle segmentation helps quantify many functional parameters of the heart and construct anatomical models for intervention planning. Here we propose a fast and accurate graph cut segmentation algorithm to extract the right ventricle from cine cardiac MRI sequences [1]*. A shape prior obtained by propagating the right*

[1]This chapter is adapted from:
Dangi S. et al., "Integrating atlas and graph cut methods for right ventricle blood-pool segmentation from cardiac cine MRI," Proc. SPIE 10135, Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling, 1013519 (3 March 2017)
Dangi S. et al., "Using Atlas Prior with Graph Cut Methods for Right Ventricle Segmentation from Cardiac MRI." In: Pop M., Wright G. (eds) Functional Imaging and Modelling of the Heart. FIMH

ventricle label from an average atlas via affine registration is incorporated into the graph energy. The optimal segmentation obtained from the graph cut is iteratively refined to produce the final right ventricle blood pool segmentation.

4.1 Introduction

The complex motion and anatomy of the RV, indistinct borders due to blood flow and partial volume effect, and the presence of trabeculations (with similar gray level as the surrounding myocardium), makes the segmentation of RV a challenging task [1]. RV segmentation is performed manually in clinical practice, requiring about 15 minutes, and is prone to inter and intra-expert variability [2]. This suggests a need for fast, accurate, and robust semi- or fully-automatic RV segmentation algorithms.

Various techniques have been proposed in the literature [1]. Nevertheless, the focus has primarily been on the left ventricle (LV) segmentation and very limited work has been disseminated on the RV segmentation. Hence, here we adapted the techniques researched and applied to segment the LV from cine cardiac MR images to a different dataset specifically focused on the RV. We exploit the speed of the graph cut framework and improve it by incorporating an average atlas based segmentation (obtained using a single affine registration) as a shape constraint. Subsequently, we iteratively refine the segmentation using techniques similar to those described in [3, 4] to obtain an accurate and robust segmentation of the RV blood pool. The fast and accurate segmentation obtained from the developed method can be used for several time-constraining applications, including peri-operative generation of subjects-specific RV models for surgical planning and guidance applications.

2017. Lecture Notes in Computer Science, vol 10263. Springer, Cham

4.2 Methods

4.2.1 Data Preprocessing

This study employed 16 cardiac cine MRI datasets available through MICCAI 2012 Cardiac MR Right Ventricle Segmentation Challenge (RVSC)[2]. Expert manual segmentation of the RV images were provided with the dataset and served as the gold-standard for evaluating the proposed segmentation technique.

We first select a reference patient volume with good contrast (via visual inspection of the dataset) and LV-RV orientation with RV appearing on the top right position (as shown in **Fig. 4.1**). The DICOM Image Orientation Patient (IPP) field is used to rotate the patient volumes about the z-axis to roughly align the LV-RV orientation of each dataset with that of the selected reference patient. We then find the region of interest (ROI) (in the xy-plane) enclosing the left and right ventricles using the method described in [5]. To restrict the ROI along the z-axis, we require a manual input that indicates the start and end slices of RV anatomy. The image volumes are then cropped according to the above ROI, and the intensity range of each slice is normalized prior to the subsequent steps.

4.2.2 Atlas Generation

We affinely register all cropped 3D volumes to the selected reference patient using the intensity based Nelder-Meade downhill simplex algorithm [6] available in SimpleITK. The resulting affine transforms are applied to the respective ground truth datasets. Averaging the transformed volumes as well as the transformed ground truths yields an average appearance atlas and a blood pool (BP) probabilistic atlas, respectively **Fig. 4.1**.

[2] `http://www.litislab.eu/rvsc`

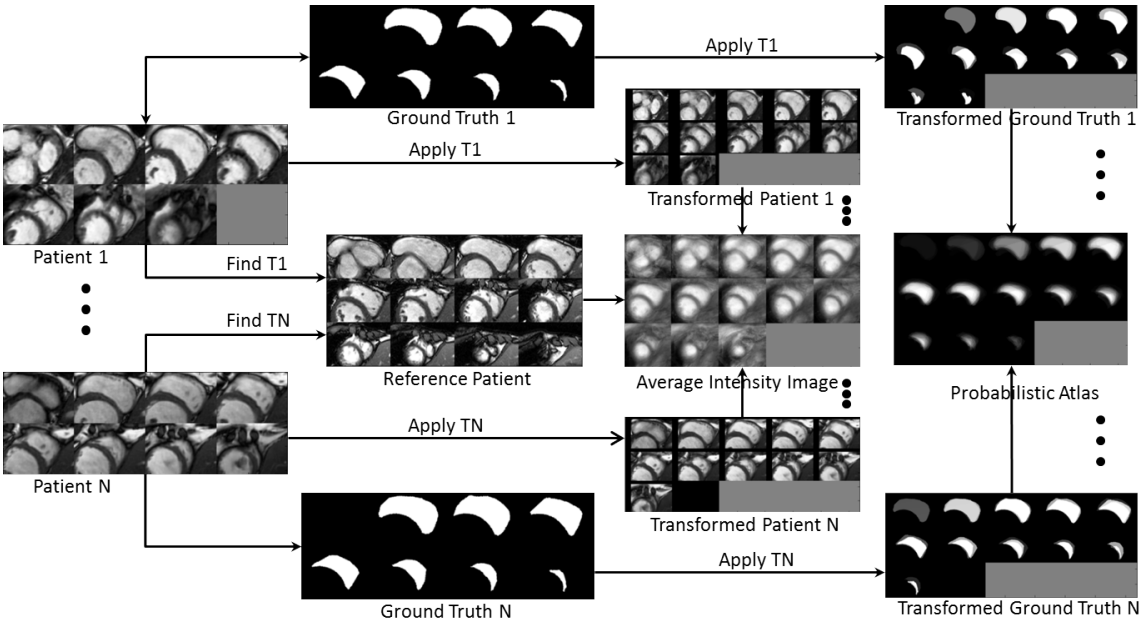


Figure 4.1: All patient images are affinely registered to the reference patient, and the obtained optimum transformation is applied to the corresponding ground truth images. An average intensity image is obtained by averaging the intensities of all transformed patient images, while the averaging of the transformed ground truth images yields a probabilistic atlas.

4.2.3 RV Blood Pool Segmentation using Iterative Graph Cuts

To exploit the 3D cardiac geometry information, we use the BP segmentation of a given slice to initialize the BP ROI in the neighboring slices. As such, we first segment the mid-slice, followed by its neighboring slices, and proceed accordingly, until the complete volume is segmented.

4.2.3.1 Blood Pool Probabilistic Map:

The average appearance atlas is registered to a test volume (**Fig. 4.2a**) using intensity based 3D affine registration. The resulting affine transform is used to transfer the BP probabilistic label to the test data. The obtained label is normalized (0-1) per slice to generate a BP probabilistic map, which helps in BP initialization and serves

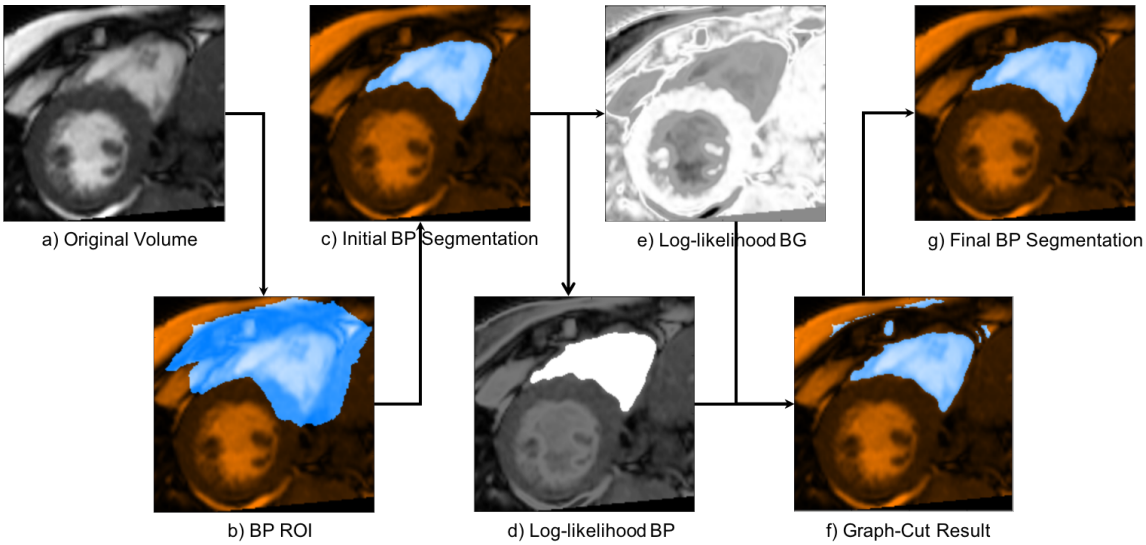


Figure 4.2: BP Segmentation procedure: a) original MR image slice, b) the initial BP ROI obtained by thresholding the BP probability map, c) initial BP segment obtained by otsu thresholding and cleaning; d) BP log-likelihood and e) BG log-likelihood obtained from intensity distribution model; (f) the graph-cut segmentation result; g) final BP segmentation for current iteration obtained after post-processing.

as a shape constraint for subsequent graph cut segmentation.

4.2.3.2 Blood Pool Initialization:

The probabilistic map is thresholded to obtain an initial BP ROI (**Fig. 4.2b**). All the basal slices (including and above the mid-slice) are thresholded at a very small value (0.01) to account for the large BP size, whereas, the threshold is determined automatically, as a mean value for the probabilistic map, for the apical slices to restrict the BP size to a smaller region. Otsu thresholding [7] applied to the obtained BP ROI yields a set of potentially bright BP segments. A single connected component closest to the center of the BP ROI is used as initial BP segmentation shown in **Fig. 4.2c**.

4.2.3.3 Intensity Distribution Model:

The BP intensity model is obtained by fitting the intensity values within the initial BP segmentation to a Gaussian Mixture Model (GMM) comprising one Gaussian.

The remaining intensity values inside the initial BP ROI are fitted to a two-Gaussian GMM to yield the background (BG) intensity model. **Fig. 4.2d** and **Fig. 4.2e** show the resulting BP and BG log-likelihood map, respectively. Although Rician distribution [8] has been found more appropriate for modeling intensity noise in MR images, signal-to-noise ratio for our MR data is high, hence, we rely on a simpler model using Gaussian approximation.

4.2.3.4 Graph-Cut Segmentation:

A four-neighborhood graph is constructed with each pixel representing a node. Two special terminal nodes representing two classes — the source blood pool (BP), and the sink background (BG) — are added to the graph. The segmentation is formulated as an energy minimization problem over the space of optimal labelings f:

$$E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q), \tag{4.1}$$

where the first term represents the data energy that reduces the disagreement between the labeling f_p given the observed data at every pixel $p \in P$, and the second term represents the smoothness energy that forces pixels p and q defined by a set of interacting pair $\mathcal{N}$ (in our case, the neighboring pixels) towards the same label.

The data energy term encoded as terminal link (t-link) between each node to source (or sink) is assigned based on the log likelihood computed from the intensity distribution model:

$$D_p(f_p) = -lnPr(I_p|f_p) \tag{4.2}$$

where $Pr(I_p|f_p)$ is the likelihood of observing the intensity I_p given that pixel p belongs to class f_p. The log-likelihood for BP and BG are shown in (**Fig. 4.2d**) and (**Fig. 4.2e**), respectively.

The smoothness energy term is computed over the links between neighboring nodes (n-links) and is assigned as a weighted sum of intensity similarity between the pixels

and average probability of the pixels belonging to BP based on BP probabilistic atlas:

$$V_{p,q}(f_p, f_q) = \begin{cases} w_I * exp\left(-\frac{|I_p - I_q|}{\tau}\right) + w_A * exp\left(\frac{P_A(p)+P_A(q)}{2}\right) & \text{if } f_p \neq f_q \\ 0 & \text{if } f_p = f_q \end{cases} \tag{4.3}$$

where, τ is the iteration number, w_I and w_A are weights for the intensity similarity term and atlas prior term, respectively, and $P_A(\cdot)$ is the probability of a pixel belonging to BP obtained from the atlas. The intensity similarity term changes during each iteration such that the pixels with higher intensity difference can be assigned the same label to allow the expansion of BP as the iteration proceeds.

After defining the graph energy configuration, the minimum cut equivalent to the maximum flow is identified via the expansion algorithm described in [9]. This approach yields the labeling that minimizes the global energy of the graph and corresponds to the optimal segmentation. **Fig. 4.2f** shows the graph-cut result, which yields the BP segmentation after some post processing **Fig. 4.2g**.

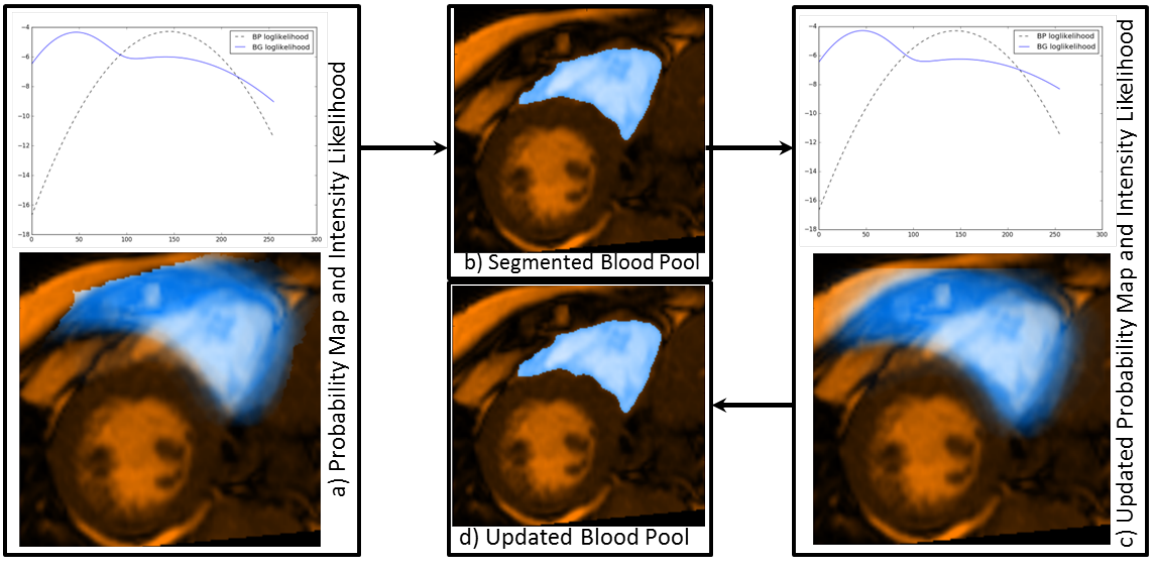


Figure 4.3: a) Probability map and intensity distribution model for current iteration, b) BP segmentation obtained from graph cut using (a), c) updated probability map and intensity distribution model obtained using (b), d) new BP segmentation obtained from graph cut using (c).

4.2.3.5 Blood Pool Probability Map Refinement:

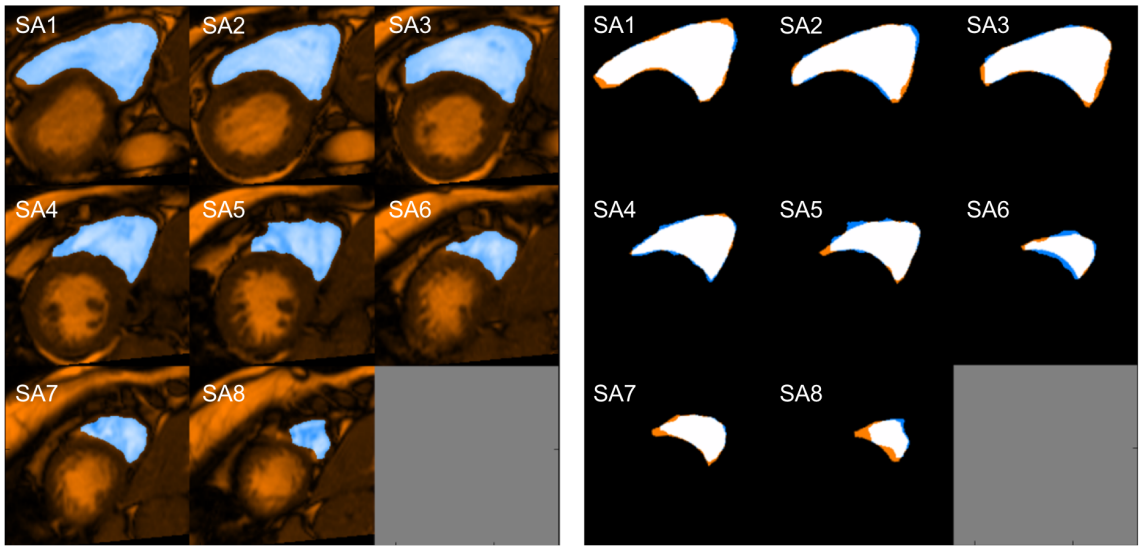
The mean value of the BP probabilistic map is used as a threshold to obtain the BP region. The signed distance map corresponding to the extracted BP region is affinely registered to the signed distance map generated from the boundary of the BP segmentation obtained from graph-cuts, such that, the sum of squared differences between the two distance maps is minimized. The optimum affine transform, when applied to the BP probability map, transforms it according to the latest BP segmentation.

4.2.3.6 Iterative Refinement:

We employed the iterative refinement technique as described in [10]. The latest BP segmentation obtained from the graph cut is used to update the intensity distribution model for BP/BG. Very high intensity likelihood $Pr(I_p|f_p)$ is assigned to the pixels labeled as BP in the current iteration, such that they don't change their labels. The refined BP probability map is used to impose shape constraint into the graph-cut framework in the form of smoothness energy. An updated BP segmentation is obtained via another graph cut operating on the new graph energy configuration. This iterative process is repeated until the dice coefficient between two latest BP segmentations exceed 99%. On average, the process requires about three iterations for convergence, hence the maximum number of iterations is restricted to 10. The latest segmentation result yields the final BP segmentation. **Fig. 4.3** illustrates the iterative refinement process.

4.3 Results

Segmentation results for a patient dataset are overlaid onto each slice of the patient volume and shown in **Fig. 4.4a**. **Fig. 4.4b** shows the visual comparison of our segmentation result against the provided manual segmentation.



a) Segmentation result superimposed on the original volume

b) Our segmentation result versus the Gold-standard segmentation

Figure 4.4: a) Final BP segmentation of all slices of a patient dataset (shown in blue) superimposed with the patient volume (shown in red); b) Final BP segmentation assessed against the provided gold-standard manual segmentation; white regions represent true positives, red regions represent false negatives, and blue regions represent false positives.

Table 4.1: Evaluation of our segmentation results against the provided gold-standard manual segmentation for the basal and apical slices according to Dice Index, Jaccard Index, Mean Absolute Distance, and Hausdorff Distance

Assessment Metric	Basal Slices	Apical Slices	All Slices
Dice Index	0.89 ± 0.12	0.62 ± 0.32	0.83 ± 0.22
Jaccard Index	0.82 ± 0.15	0.52 ± 0.30	0.75 ± 0.23
Mean Absolute Distance (mm)	3.73 ± 4.16	11.11 ± 15.05	5.50 ± 8.69
Hausdorff Distance (mm)	8.25 ± 7.42	15.57 ± 12.57	10.00 ± 9.40

We evaluated the obtained RV blood pool segmentations against the provided expert manual segmentations for 16 datasets according to the following metrics: Dice Metric (DM), Jaccard Metric (JM), Mean Absolute Distance (MAD), and Hausdorff Distance (HD) [1]. The metrics are summarized in **Table 4.1** for all slices together, as well as for the apical-slices (last two slices) and remaining slices separately.

Table 4.2: Evaluation of the segmentation results against the provided gold-standard manual segmentation for the end-diastole phase according to Dice Index and Hausdorff Distance, shown for various methods extracted from the published manuscripts. Values are provided as mean ± standard deviation in descending order by Dice index. SA/FA — Semi/Fully-Automatic. * — cross-validated on training set, † — evaluated on held out test set

Method	SA/FA	Dice Index	Hausdorff Distance
Zuluaga et al.† [11]	FA	0.83 ± 0.17	9.77 ± 7.88
Proposed*	FA	0.83 ± 0.22	10.00 ± 9.40
Ou et al.† [12]	FA	0.71 ± 0.22	14.45 ± 9.34
Bai et al.† [13]	SA	0.86 ± 0.11	7.70 ± 3.74
Grosgeorge et al.† [14]	SA	0.83 ± 0.15	9.48 ± 5.41
Nambakhsh et al.* [15]	SA	0.70 ± 0.21	15.83 ± 10.92

The collated results for the RVSC are reported in [1]. The inter-expert variability study in [1] shows the variability in DM and HD are 0.90 ± 0.10 and 5.02 ± 2.87mm, respectively. As evident from **Table 4.2**, our proposed algorithm performs equivalently to the best fully automatic method and is as competitive as the semi-automatic methods against which it is assessed. However, it should be noted that our method was cross-validated on the training dataset (compared against the provided ground truth segmentation), whereas some of the reported results were obtained by testing on the validation dataset held-out by the challenge organizers. Nevertheless, there is still a room for improvement to reach the performance equivalent to the inter-expert variability.

Axial, Sagittal, Coronal, and the 3D reconstruction of the segmented RV blood pool overlayed onto the MRI volume is shown in **Fig. 4.5**. The signed distance error computed between the 3D models of the segmented RV blood pool and the provided expert manual segmentation is overlaid onto the latter and shown for three different cases in **Fig. 4.6**.

We computed the End Diastole (ED) RV blood pool volume for 16 datasets and

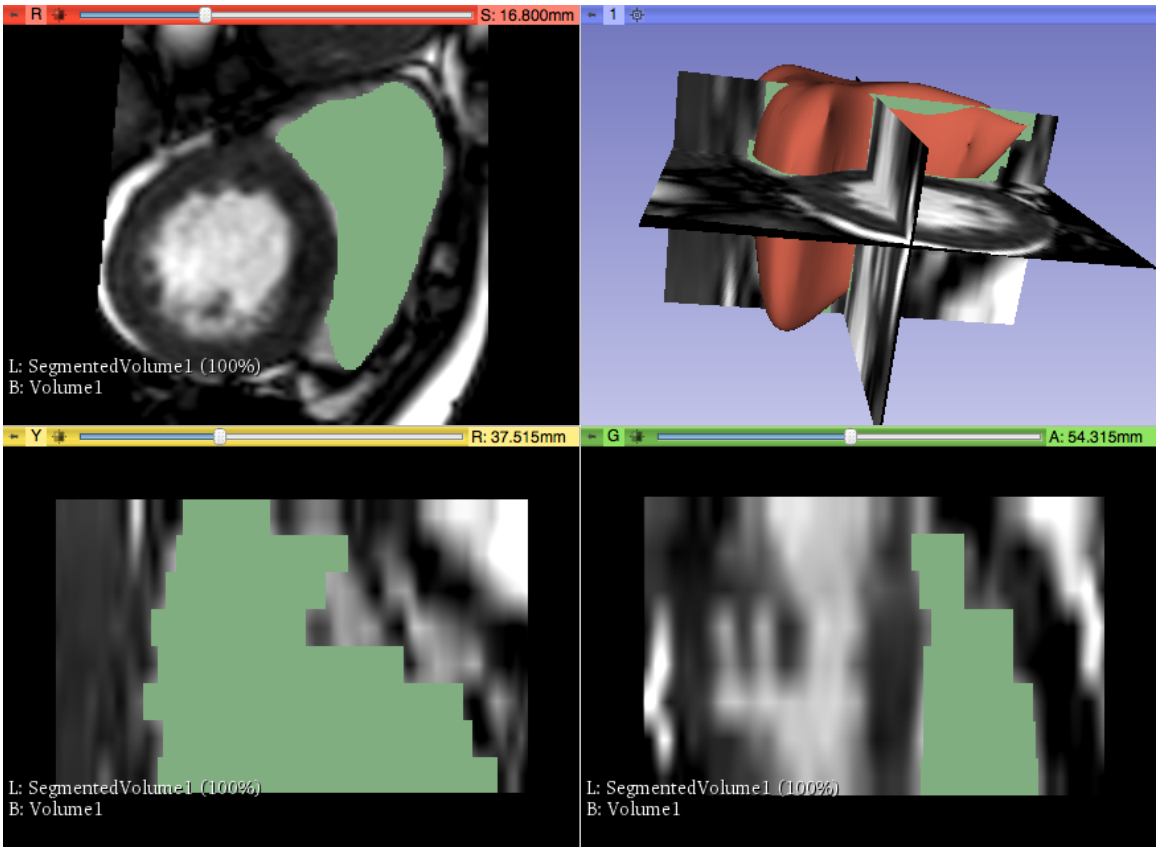


Figure 4.5: Axial, Sagittal, Coronal, and the 3D model (counter-clockwise from top left) of the segmented RV blood pool overlayed onto the MRI image.

Table 4.3: Comparing the end diastole RV blood pool volume estimated by the proposed algorithm against the ground truth volume for 16 cine MRI images

Assessment Metric	Ground Truth	Proposed Algorithm	Difference
ED Volume (ml)	129.7 ± 38.1	122.0 ± 29.0	7.8 ± 16.6

compared it against the volume computed from the provided expert manual segmentation as shown in **Table 4.3**. **Fig. 4.7a** shows the correlation between the RV volume reconstructed from our automated segmentation and the volume reconstructed from the manual expert segmentation. As illustrated, a linear correlation (defined by $y = 1.2x$) exists, where y is the ground truth RV volume and x is the segmented RV

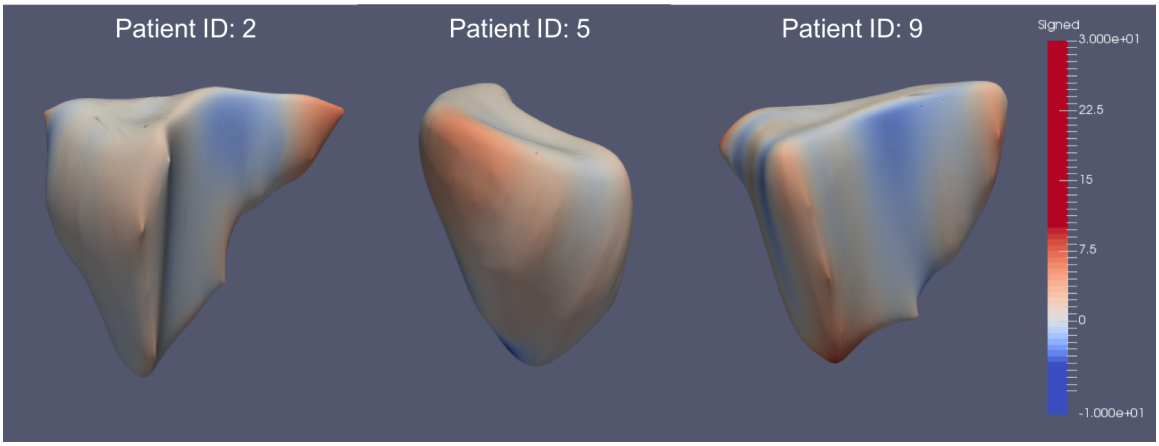


Figure 4.6: Shown for three cases: Signed distance error between the segmented RV blood pool and the corresponding gold standard manual segmentation overlayed onto the 3D model of the latter.

volume in mL, respectively. Moreover, the linear R^2 value is 0.83, showing a strong correlation between the automatically segmented and ground truth RV volume.

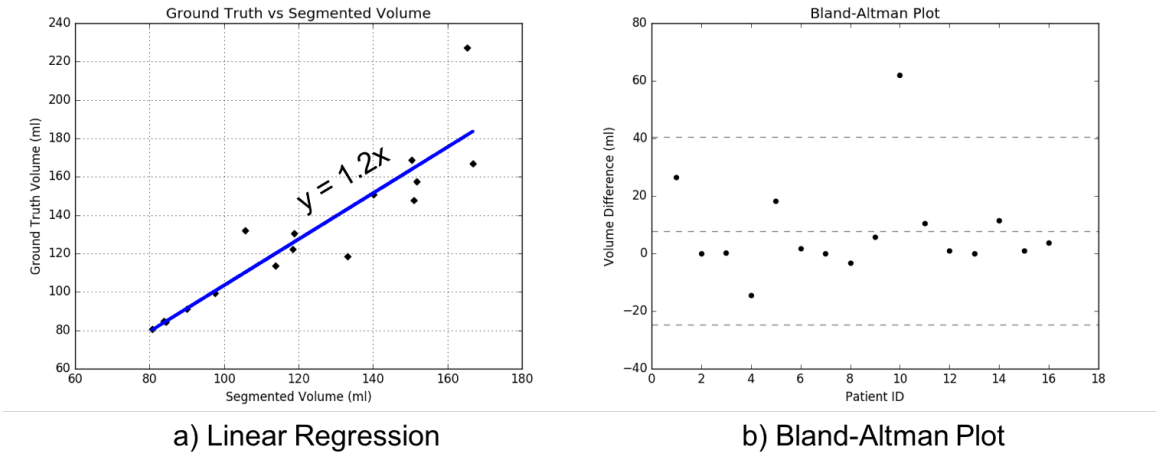


Figure 4.7: a) Linear regression result for ground truth volume against the volume estimated by the algorithm. The best fit line is $y = 1.2x$, with r-squared value of 0.83; b) Difference between the ground truth volume and volume estimated by the algorithm plotted for all 16 cases. The mean difference (7.8 ml) along with the 95% confidence interval ($\pm$1.96SD: -24.8 to 40.4 ml) are shown as the dotted lines.

The Bland-Altman plot in **Fig. 4.7b** shows the difference between the ground truth and automatically segmented volume for all 16 datasets. Although the es-

timated volume is linearly correlated to the ground truth volume, our automatic segmentation algorithm slightly under-estimated the RV volume by 7.8 ± 16.6 mL on average. Specifically, the volume was under-estimated for 5 of the patients (1, 5, 10, 11, 14) and over-estimated it for the remaining 11 patients. Lastly, with the exception of one single dataset (patient 10), who is clearly an outlier as it exhibited a difference between the automatically segmented and ground truth volume of more than 3 SDs from the mean (i.e., a z-score larger than 3), all other cases exhibited a volume difference within 2 SDs from the mean.

The proposed algorithm was implemented in Python and required 76 secs on average to propagate atlas label and segment RV blood pool from cine MRI volumes (12 ± 1 axial slices) on a Intel® Xenon® 3.60 GHz 32GB RAM PC, therefore posing a great potential for peri-operative segmentation of RV without delaying the procedure workflow. The atlas is precomputed, in 450 seconds, by co-registering and averaging the training images.

4.4 Discussion

We observed that the segmentation results for apical slices with smaller BP region are consistently worse than that for the basal slices. Although they have very little impact on volume computation, they could be a limiting factor on other fields such as studies of the fiber structure. Slice-wise refinement process might be hurting the performance of our segmentation algorithm due to the misleading cues on apical slices, which suggests a need for special processing of these slices. An alternative solution for better segmentation of apical slices would be to perform a complete 3D segmentation, which requires a robust algorithm to align the cine MRI slices into a complete 3D volume to correct for stair-step artifacts caused by inherent motion between subsequent slices prior to segmentation.

Since the expert annotated ground truth segmentations of the right ventricle are

only available in end diastole and end systole, only two atlases could be constructed – a diastole and a systole atlas. We have in fact attempted to generate a systole atlas based on the same techniques used to generate the diastole atlas, however the apical slices are compromised due to the wrapping of the right ventricle around the left ventricle, making it difficult to identify the blood pool from the myocardium in this region and hence compromising the accuracy of the atlas and its robustness across all datasets.

Nevertheless, a more feasible approach would be to take advantage of the temporal information and correlation of the subsequent frames in the cardiac cycle and use the frame-to-frame motion to animate the single phase diastolic right ventricle segmentation extracted using the diastole right ventricle atlas faithfully generated as described in this paper. The proposed method will follow the approach described in [16]. Briefly, we will utilize non-rigid image registration to extract the frame-to-frame motion from the sequence of multi-phase cardiac images depicting the right ventricle and then apply the non-rigid displacement field to propagate the segmented right ventricle diastolic surface through the cardiac cycle.

4.5 Conclusion and Future Work

We proposed a fast and automatic segmentation method using atlas prior in the graph cut framework with iterative refinement to segment the RV blood pool from cardiac cine MRI images. Quantitative results of our blood pool segmentation in the ED phase are better than the fully-automatic methods and comparable to the semi-automatic methods reported in the challenge.

We plan to segment the RV myocardium and extend the evaluation of our algorithm to end diastole and end systole phases of all the 48 datasets (16 training, 32 testing). Furthermore, we will be studying the effect of weighting factor variability on different terms of graph energy towards final segmentation result and lastly, we will

also be exploring other options to improve the segmentation results for apical slices.

Bibliography

[1] Caroline Petitjean, Maria A. Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jrme Caudron, Su Ruan, Ismail Ben Ayed, M. Jorge Cardoso, Hsiang-Chou Chen, Daniel Jimenez-Carretero, Maria J. Ledesma-Carbayo, Christos Davatzikos, Jimit Doshi, Guray Erus, Oskar M.O. Maier, Cyrus M.S. Nambakhsh, Yangming Ou, Sbastien Ourselin, Chun-Wei Peng, Nicholas S. Peters, Terry M. Peters, Martin Rajchl, Daniel Rueckert, Andres Santos, Wenzhe Shi, Ching-Wei Wang, Haiyan Wang, and Jing Yuan. Right ventricle segmentation from cardiac MRI: A collation study. *Medical Image Analysis*, 19(1):187 – 202, 2015.

[2] Jrme Caudron, Jeannette Fares, Valentin Lefebvre, Pierre-Hugues Vivier, Caroline Petitjean, and Jean-Nicolas Dacher. Cardiac MRI assessment of right ventricular function in acquired heart disease: Factors of variability. *Academic Radiology*, 19(8):991 – 1002, 2012.

[3] Greg Slabaugh and Gozde Unal. Graph cuts segmentation using an elliptical shape prior. *Proceedings - International Conference on Image Processing, ICIP*, 2:1222–1225, 2005.

[4] Nhat Vu and B. S. Manjunath. Shape prior segmentation of multiple objects with graph cuts. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[5] Yehuda Kfir Ben-Zikri and Cristian A. Linte. A robust automated left ventricle region of interest localization technique using a cardiac cine MRI atlas, 2016.

[6] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

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

[8] HKon Gudbjartsson and Samuel Patz. The rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914, 1995.

[9] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[10] Shusil Dangi, Nathan Cahill, and Cristian A. Linte. *Integrating Atlas and Graph Cut Methods for Left Ventricle Segmentation from Cardiac Cine MRI*, pages 76–86. Springer International Publishing, Cham, 2017.

[11] MA Zuluaga, MJ Cardoso, and Sebastien Ourselin. Automatic right ventricle segmentation using multi-label fusion in cardiac MRI. 2012.

[12] Yangming Ou, Jimit Doshi, Guray Erus, and Christos Davatzikos. Multi-atlas segmentation of the cardiac mr right ventricle. *Proceedings of 3D Cardiovascular Imaging : A MICCAI Segmentation Challenge*, 2012.

[13] Wenjia Bai, Wenzhe Shi, Haiyan Wang, Nicholas S Peters, and Daniel Rueckert. Multiatlas based segmentation with local label fusion for right ventricle MR images. *image*, 6:9, 2012.

[14] Damien Grosgeorge, Caroline Petitjean, Jérôme Caudron, Jeannette Fares, and Jean-Nicolas Dacher. Automatic cardiac ventricle segmentation in MR images: a validation study. *International Journal of Computer Assisted Radiology and Surgery*, 6(5):573–581, 2011.

[15] Cyrus M.S. Nambakhsh, Jing Yuan, Kumaradevan Punithakumar, Aashish Goela, Martin Rajchl, Terry M. Peters, and Ismail Ben Ayed. Left ventricle segmentation in mri via convex relaxed distribution matching. *Medical Image Analysis*, 17(8):1010 – 1024, 2013.

[16] Shusil Dangi, Yehuda K. Ben-Zikri, Yechiel Lamash, Karl Q. Schwarz, and Cristian A. Linte. Automatic LV feature detection and blood-pool tracking from multi-plane TEE time series. *Functional Imaging and Modeling of the Heart: 8th International Conference, FIMH 2015, Lect Notes Comput Sci.*, 9126:29–39, 2015.

Chapter 5

Towards Deep Learning Techniques for Cardiac Cine MRI Slice Misalignment Correction and 3D Hybrid Left Ventricle Segmentation

There exists an inevitable misalignment between the slices in the cine MRI image due to the 2D + time acquisition, rendering 3D segmentation methods ineffective. A large part of published work on cardiac MR image segmentation focuses on 2D segmentation methods that yield good results in mid-slices, however with less accurate results for the apical and basal slices. Here, we propose an algorithm [1] to correct for the slice misalignment using a Convolutional Neural Network (CNN)-based regression, and then perform a 3D graph-cut based segmentation of the LV using atlas shape prior.

[1]This chapter is adapted from: Dangi S. et al., "Cine cardiac MRI slice misalignment correction towards full 3D left ventricle segmentation," Proc. SPIE 10576, Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling, 1057607 (12 March 2018)

5.1 Introduction

Cine cardiac MRI images are acquired one slice at a time throughout the cardiac cycle (2D + time acquisition), during a single breath-hold. Multiple short-axis slices covering the entire heart generates a pseudo-4D volume. Hence, there is an inherent misalignment between the slices due to the patient motion and breathing. A comprehensive review of segmentation techniques applied to short-axis cardiac MR images can be found in [1]. Most of the algorithms operate on the 2D image slices, including our methods presented in **Chapter 3** and **Chapter 4**. The obtained 2D segmentations are then stacked together to reconstruct a 3D volume. However, extensive blurring and interpolation is required to get rid of the stair-step artifacts due to the slice misalignment, as well as the partial volume effects arising from the finite slice thickness. Hence, there is a significant potential of leveraging the 3D context during segmentation, such as the 3D extension of the 2D atlas-prior based graph cut technique presented in earlier chapters. Nevertheless, the full 3D segmentation first requires the slice misalignment correction to generate a contiguous 3D volume.

Several methods for slice misalignment correction have been previously proposed. Elen *et al.*[2] optimized the intensity similarity between the 2D long axis (LA) and short axis (SA) slices along the line of intersection. Similarly, Slomka *et al.*[3] performed 3D registration of LA and SA slices by minimizing the cost function derived from plane intersections for all cine phases. Although helpful, these algorithms feature limited robustness, as they rely on a limited amount of information available at the plane intersections. Therefore, since the alignment of acquired 2D cine MR slices into a full, cohesive 3D volume is challenging and prone to error, most of the segmentation algorithms rely on 2D processing, compromising the segmentation results in the apical and basal regions.

Here we propose a novel CNN architecture that uses multi-resolution features to accurately regress the center of the LV blood-pool from cine MRI slices. We train the

CNN to predict the LV center using a large dataset representative of wide variability in cardiac cine MR short-axis images. For the segmentation of a test image, we predict the LV centers for all the slices using the trained CNN. Subsequently, we correct for the slice misalignment to generate a full 3D test volume. Finally, we leverage the 3D information to segment the LV blood-pool and myocardium using the 3D extension of the atlas and graph-cut based segmentation technique presented in **Chapter 3** and **Chapter 4**.

5.2 Methodology

5.2.1 Cardiac MRI Data

This study employed 97 de-identified cardiac MRI image datasets from patients suffering from myocardial infraction and impaired LV contraction available as a part of the STACOM Cardiac Atlas Segmentation Challenge project [4, 5] database[2]. Cine-MRI images in short-axis and long-axis views are available for each case. The images were acquired using the Steady-State Free Precession (SSFP) MR imaging protocol with the following settings: typical thickness $\leq 10mm$, gap $\leq 2mm$, TR $30-50ms$, TE $1.6ms$, flip angle 60^0, FOV $360mm$ and $256\times256mm$ image matrix using multiple scanners from various manufacturers.

We divided the 97 available dataset into 80% training, 10% validation (to avoid over-fitting during training), and 10% test set, and perform our evaluation on the test set.

5.2.1.1 LV blood-pool ground-truth generation

Ground truth myocardium segmentation generated from expert analyzed 3D surface finite element model is available for all 97 cases. As the slice segmentations are

[2] `http://www.cardiacatlas.org`

obtained from the intersection of image slices with the 3D model, partial myocardium can be observed in some basal slices. To obtain the corresponding blood-pool segmentation, each slice of the provided myocardium segmentation is inverted and a morphological opening operation is performed. If the connected component analysis results in two connected components, the smaller connected region is selected as the blood-pool. Hence, the slices with partial myocardium do not yield any blood-pool region. This has a significant effect in the segmentation evaluation of basal slices.

5.2.2 Data Preparation and Augmentation

The physical pixel spacing in SA images ranged from 0.7031 to 2.0833 mm. We used SimpleITK [6] to resample all images to the most common spacing of 1.5625 mm along both x- and y-axis. The resampled images were center cropped or zero padded to a common resolution of 192 $\times$ 192 pixels. To simulate the real-world scenario, the images were transformed using combinations of 9 translations ($\pm$10%) along x- and y- axis, 7 rotations ($\pm10^0$,$\pm20^0$,$\pm30^0$), and 3 scaling ($\pm$10%). The resulting 189 different combinations were used to train the CNN. The ground truth LV center was computed as the centroid of the LV blood pool obtained from the inner contour of the provided ground truth myocardium segmentation. The LV center from closest mid-slice was propagated to the apical/basal slices with partial/no ground truth myocardium segmentation.

5.2.3 CNN Architecture for LV Center Regression

We propose a novel CNN architecture to predict the LV blood-pool center from SA cine MR images. The proposed approach requires minimal preprocessing (i.e. resampling and crop/pad to fixed resolution) and minimal user input to generate a training set, as the expert needs to select a single LV center point per image.

The CNN network architecture is shown in **Fig. 5.1**. The input image is passed

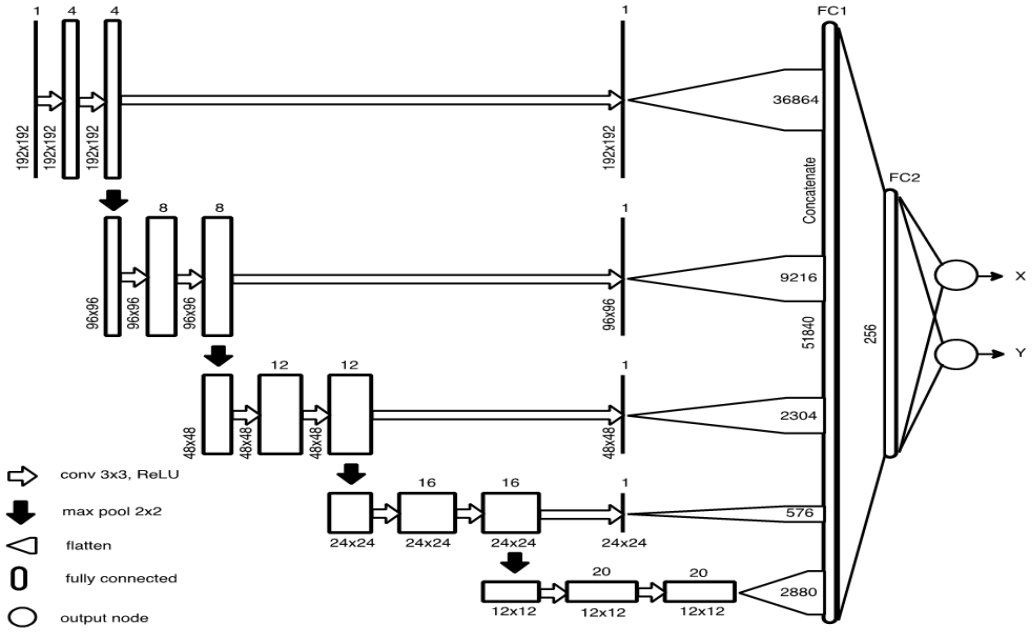
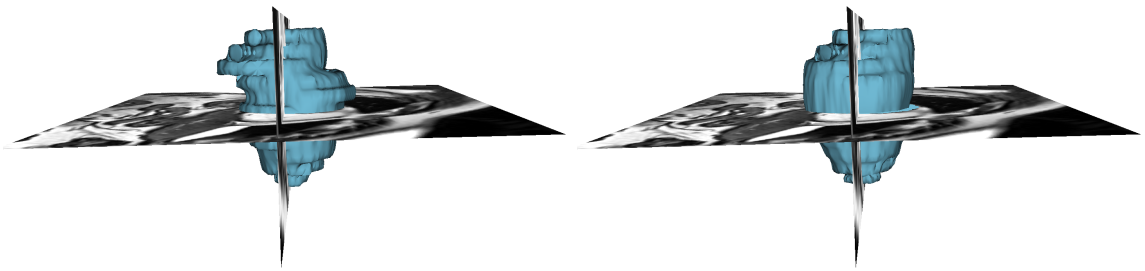


Figure 5.1: CNN Architecture for regression of the LV center from cardiac cine MRI short-axis slices.

through several convolutional layers followed by Rectified Linear Unit (ReLU) non-linearity and four max-pooling layers spread across the network. The global image information obtained in the final convolutional layers are flattened into a single vector to form a fully connected layer. In addition, we generate a single feature map from each resolution and feed it directly to the fully connected layer. These skip connections enable the network to use multi-resolution features to yield a more accurate prediction of the LV center, while maintaining the number of tractable network parameters . The output of the network consists of two values representing the x- and y- coordinates of the LV center.

5.2.4 Slice-misalignment Correction

Assuming there is only translational misalignment between the slices, we translate the predicted LV centers so that they are collinear, resulting in a corrected 3D volume.



(a) Before Misalignment Correction (b) After Misalignment Correction

Figure 5.2: Reconstructed 3D models generated from the ground-truth LV myocardium segmentation before and after slice misalignment correction.

Fig. 5.2 shows the 3D volume reconstructed from the ground-truth LV myocardium segmentation before and after misalignment correction. This step helps restore the 3D connectivity structure of the LV improving the subsequent graph-cut segmentation.

5.2.5 LV Blood-pool Segmentation

We extend the atlas prior based graph-cut segmentation method presented in [7] to 3D for the segmentation of LV blood-pool from the slice-misalignment corrected 3D volume.

5.2.5.1 Atlas Generation

We select a patient volume with an average LV size as a reference. All other training patient volumes are registered to this reference utilizing the ITKv4 registration framework [8] via the SimpleITK interface [6]. We use an affine transformation, with the mutual information similarity metric and the Nelder-Meade optimizer. The registration initialization uses the ground-truth myocardium bounding boxes to obtain initial scaling. The optimum transformation parameters are applied to the corresponding blood-pool segmentations. The registered patient volumes and blood-pool segmentations are averaged to obtain the average intensity atlas and blood-pool prob-

ability map, respectively.

5.2.5.2 Blood-pool Label Transfer

Robust registration of the average intensity atlas to a new test patient volume is crucial for the subsequent graph-cut segmentation. However, due to large variability in LV sizes in the dataset, and the tendency of optimizers to converge to a local minimum, registration results using a single starting point in parameter space are unreliable. We therefore perform the registration from multiple starting points that are selected using an exhaustive search strategy on the scale parameters along each axis and on the translation along the long axis, z direction, of the affine transformation. We evaluate the value of the normalized cross-correlation (NCC) between the atlas and the test volume using multiple scale factors. The parameters corresponding to the top k=5 similarity metric values are used as initial values for the subsequent registrations. Finally, the optimum transformation resulting in the best NCC metric is applied to the blood-pool probability map to transfer the label to the test volume as shown in **Fig. 5.3a**. The test patient image is cropped based on the blood-pool probability map to reduce the computational complexity for subsequent graph-cut segmentation. The algorithm for blood-pool label transfer is shown in **Algorithm 1**.

5.2.5.3 Graph-cut Segmentation

We represent the cropped test volume as a 3D graph with 6-neighbor connectivity. Two special terminal nodes representing two classes — the source background (BG), and the sink blood pool (BP) — are added to the graph. The segmentation is formulated as an energy minimization problem over the space of optimum labelings f:

$$E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q), \tag{5.1}$$

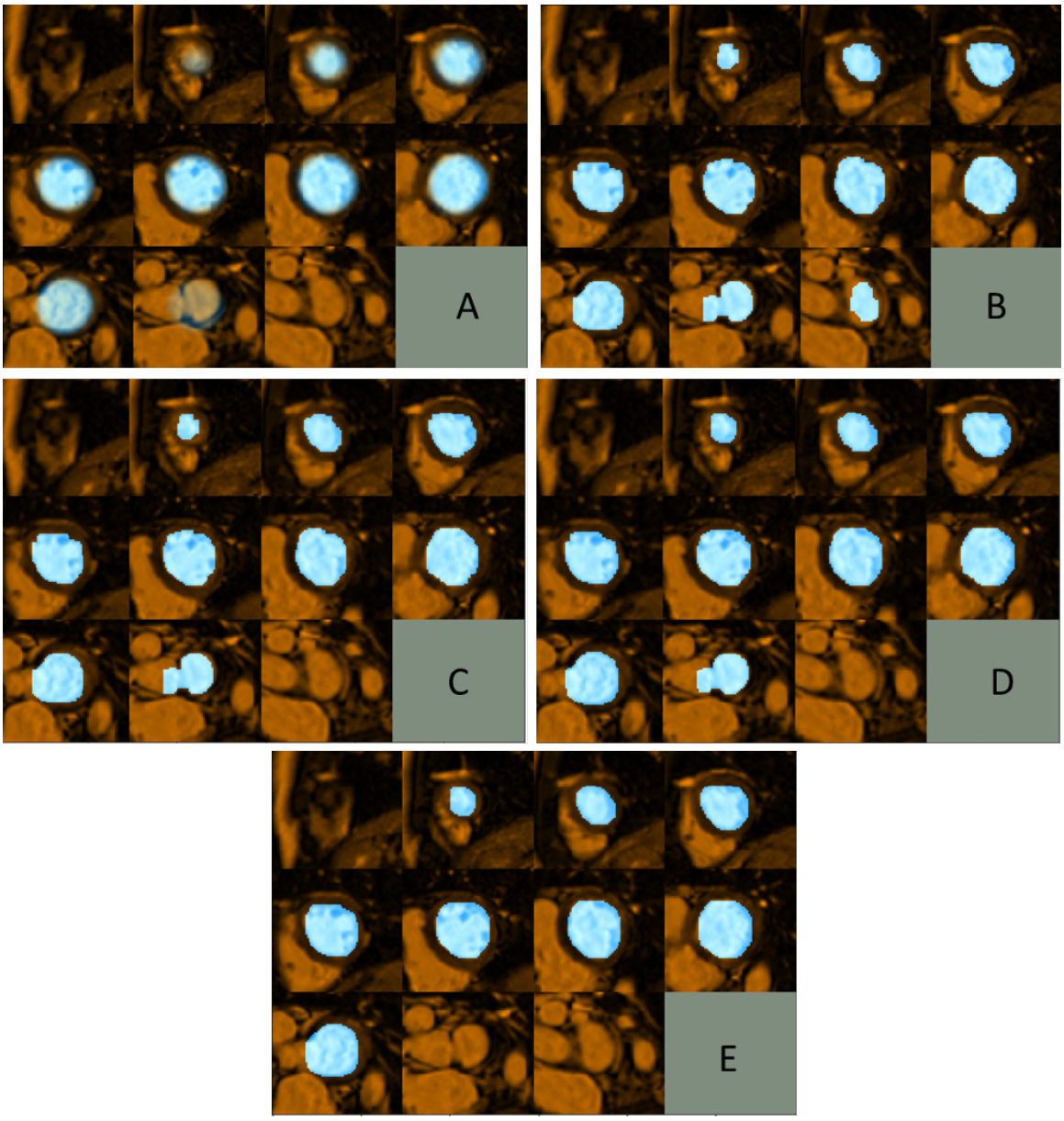


Figure 5.3: Example of left ventricle blood-pool segmentation using the proposed approach: a) Blood-pool probability map transferred to the test volume via registration; b) Initial Graph-cut segmentation (first Iteration); c) Segmentation result after refinement based on Intersection-over-Union, small over-segmented region in aorta similar in intensity with blood-pool has been removed; d) Segmentation result after iterative refinement (converged on third iteration); e) Final blood-pool segmentation after refinement using Stocastic Outlier Selection, over-segmented basal slice with distinct shape statistics compared to other slices has been removed.

Algorithm 1: Blood-pool Label Transfer

Input: Average Intensity Volume, A; BP Probability Map, P; Test Volume, I; Training LV scaling factors, $X_{N\times 3}$

Parameters: Number of scales, s; Number of z-translations, t; Number of transform initializations, k

Output: Optimum affine transformation, T_{opt}

Sample scales from 3-variate Gaussian fitted to Training LV scales;

$M_{1\times 3} \leftarrow \texttt{mean}(X_{N\times 3})$; $C_{3\times 3} \leftarrow (X_{N\times 3}\text{-}M_{1\times 3})^T(X_{N\times 3}\text{-}M_{1\times 3})$;

$C_{3\times 3} \rightarrow L_{3\times 3}{}^T\ L_{3\times 3}$;

S_x,S_y,$S_z \leftarrow \texttt{linspace}(-2,2,\texttt{s})$;

$N_{\texttt{s}^3\times 3} \leftarrow \texttt{Permutation}(S_x,S_y,S_z)$;

$S_{\texttt{s}^3\times 3} \leftarrow M_{1\times 3} + N_{\texttt{s}^3\times 3}\ L_{3\times 3}$;

$Tx_{t\times 3} \leftarrow (0, 0, \texttt{linspace(-Spacing(I)},\texttt{Spacing(I)},t))$;

$E_{\texttt{s}^3 t\times 6} \leftarrow \texttt{Permutation}(S_{\texttt{s}^3\times 3},Tx_{t\times 3})$;

for i **in** $E_{\texttt{s}^3 t\times 6}$ **do**

    Compute Normalized cross-correlation between I *and transformed* A;

    $SM\ [i] \leftarrow \texttt{NCC(I},T_i\texttt{(A))}$;

end

$e_k \leftarrow E_{\texttt{s}^3 t\times 6}\ [\texttt{MaxInd}(SM,\texttt{k})]$ **if** unique S_z;

for j **in** e_k **do**

    Perform registration using each k *initializations*;

    $T_{opt\ j} \leftarrow \underset{T_j}{\text{argmax}}\ \ \texttt{NCC(I},T_j\texttt{(A))}$;

    $SM\ [j] \leftarrow \texttt{NCC(I},T_{opt_j}\texttt{(A))}$;

end

maxInd $\leftarrow \texttt{MaxInd}(SM,1)$; $T_{opt} \leftarrow T_{opt_{\texttt{maxInd}}}$;

where the first term represents the data energy that reduces the disagreement between the labeling f_p given the observed data at every pixel $p \in P$, and the second term represents the smoothness energy that forces pixels p and q defined by a set of interacting pair $\mathcal{N}$ (in our case, the neighboring pixels) towards the same label.

The data energy term encoded as terminal link (t-link) between each node to source (or sink) is assigned as the weighted sum of the log-likelihood computed from the Gaussian Mixture Model (GMM) of intensity distributions, blood-pool probability map, and signed distance map obtained from the thresholded blood-pool probability map:

$$D_p(f_p) = -w_1 ln Pr(I_p|f_p) + \frac{w_2}{1+e^{-itr}} P(f_p) + \frac{w_3}{1+e^{-itr}} Dm(f_p) \tag{5.2}$$

where, $Pr(I_p|f_p)$ is the likelihood of observing the intensity I_p given that pixel p belongs to class f_p. The log-likelihood for BP and BG are obtained by fitting intensity values within the convex-hull of most recent BP segmentation and outside the BP probability map thresholded at Th_{BG} to 1- and 2-Gaussian GMM models, respectively. $P(f_p)$ is the probability of pixel p being class f_p. The probabilities for BP and BG are the BP-probability map and its inverse, respectively. $Dm(f_p)$ is the likelihood of pixel p being class f_p computed from a signed distance map obtained via BP-probability map thresholded at Th_{BP}, with the inside regions being positive and outside regions being negative. This strongly encourages pixels inside and outside the thresholded BP-probability map to be assigned as BP and BG, respectively. itr is the iteration number. The weights are assigned such that the contribution of BP probability map increases with increasing iteration number, reflecting its increasing reliability as the iteration proceeds.

The smoothness energy term is computed over the links between neighboring nodes (n-links) and is assigned as a weighted sum of intensity similarity between the pixels and average probability of the pixels belonging to BP based on the BP probability

map:

$$V_{p,q}(f_p, f_q) = \begin{cases} w_4 * exp\left(-\frac{|I_p - I_q|)}{\sigma}\right) + w_5 * exp\left(\frac{P(p)+P(q)}{2}\right) & \text{if } f_p \neq f_q \\ 0 & \text{if } f_p = f_q \end{cases} \tag{5.3}$$

where, w_4 and w_5 are weights for the intensity similarity term and atlas prior term, respectively, and $P(\cdot)$ is the probability of a pixel belonging to BP obtained from the BP probability map.

After assigning appropriate unary and pairwise potentials to the graph, the minimum cut is identified using the expansion algorithm [9]. The obtained labeling minimizes the global energy of the graph and corresponds to the optimal BP/BG segmentation as shown in **Fig. 5.3b**.

5.2.5.4 Segmentation Refinement using Intersection-over-Union

Due to the intensity similarity between blood-pool and some background regions, the raw graph-cut segmentation is sometimes noisy and requires additional processing. We perform slice-wise refinement via connected-component analysis, such that a single connected region per slice is retained which maximizes the Intersection-over-Union (IoU) metric to the BP of closest mid-slice and to the BP probability map thresholded at Th_{BP}. The refinement starts at mid-slice and then proceeds to the apical/basal slices. To accommodate for small blood-pool regions in apical slices, the IoU value is set to 1.0 if one object is completely inside the other. Further, we only retain the slice segmentations with IoU greater than a predefined threshold Th_{Iou} to filter out small implausible segmentation regions. The BP segmentation result after IoU based refinement is shown in **Fig. 5.3c**. The algorithm for IoU based segmenation refinement is shown in **Algorithm 2**.

Algorithm 2: Refining Graph-cut Segmentation Result using IoU

Input: Noisy graph-cut segmentation, S_N; Thresholded BP probability map, $P_{Th_{BP}}$

Parameters: Intersection-over-Union (IoU) threshold, Th_{IoU}

Output: Refined Segmentation, S_R

Start at mid-slice and proceed to apical/basal slices;
$S_R \leftarrow$ `zeros(size(`S_N`))`;
slices $\leftarrow [midSlice, apicalSlices, basalSlices]$;
for i **in** slices **do**
    $j \leftarrow$ `NearestMidSlice(`i`)`;
    for c **in** `ConnectedComponents(`$S_N\ [i]$`)` **do**
        $IoU_s\ [c] \leftarrow$ `IoU(`$S_N\ [i][c]$,$S_R\ [j]$`)`;
        $IoU_p\ [c] \leftarrow$ `IoU(`$S_N\ [i][c]$,$P_{Th_{BP}}\ [i]$`)`;
        $IoU\ [c] \leftarrow IoU_s\ [c] + IoU_p\ [c]$
    end
    maxInd $\leftarrow$ `MaxInd(`IoU`)`;
    if IoU [maxInd] $> Th_{IoU}$ **then**
        $S_R\ [i] \leftarrow S_N\ [i]$[maxInd]
    end
end

5.2.5.5 Iterative Segmentation Refinement

The initial global registration of the average intensity atlas to a test patient volume might not be accurate, hence producing sub-optimal graph-cut segmentation result. To address this limitation, Otsu thresholding [10] of the region inside the registered BP probability map thresholded at Th_{BP} yields the approximate BP segmentation for the first iteration, which is further refined iteratively.

We compute a slice-wise convex-hull for the recently obtained refined graph-cut segmentation. A 3D thresholded distance map, Dm^{S}_{Th}, is computed from the convex-hull with the regions inside the segmentation assigned a constant value of 0. Similarly, a 3D thresholded distance map, $Dm^{P_{Th_{BP}}}_{Th}$, is computed from the BP probability map thresholded at Th_{BP}, $P_{Th_{BP}}$. The distance map $Dm^{P_{Th_{BP}}}_{Th}$ is registered to Dm^{S}_{Th} using gradient descent optimizer with mean squared difference (MSD) as the similarity metric within a mask defined in the apical regions up until the basal slice with non-zero $P_{Th_{BP}}$. We exclude the basal regions during the registration as they could contain over-segmented aortic valve with intensity similar to the BP, adversely affecting the segmentation refinement. The obtained optimum 3D affine transformation is applied to the BP probability map and hence used to update the graph energy in (5.2) and (5.3).

The intensity values within the convex-hull of the latest refined graph-cut segmentation is used to update the 1-Gaussian GMM model for BP. Similarly, the intensity values outside the transformed BP probability map thresholded at Th_{BG} are used to update the 2-Gaussian GMM model for the BG. The intensity likelihoods obtained from updated BP and BG GMM models are used to update the $Pr(I_p|f_p)$ term of graph energy in (5.2).

The optimal binary labeling of the graph is obtained via minimum-cut using the expansion algorithm. The obtained noisy graph-cut segmentation is "cleaned" using the method described in **Algorithm 2** and hence used to update the BP probability map and the graph energy for the next iteration. This iterative process is

repeated until the IoU of BP segmentations between consecutive iterations is below some threshold, Th_{stop}, or maximum number of iterations, itr_{max}, has been reached. **Fig. 5.3d** shows the BP segmentation result obtained after the iterative refinement process converges in three iterations.

Algorithm 3: Segmentation Refinement using Stocastic Outlier Selection

Input: BP Segmentation after iterative refinement, S_{in}; Thresholded BP Probability Map, $P_{Th_{BP}}$

Parameters: Outlier probability threshold, Th_{SOS}; SOS Perplexity, k_{SOS}

Output: Refined BP Segmentation, S_{out}

```
n — number of segmented slices in S_in;
S_out ← S_in;
X_n×4 ← ShapeFeatures(S_in);
Pr_n×1 ← SOS(X_n×4);
maxInd ← MaxInd(Pr_n×1);
startSlice, endSlice ← StartStop(P_Th_BP);
if Pr_n×1 [maxInd] > Th_SOS then
    if maxInd ≤ startSlice then S_out [ :maxInd ] ← 0;
    if maxInd ≥ endSlice then S_out [maxInd: ] ← 0;
end
```

5.2.5.6 Segmentation Refinement using Stochastic Outlier Selection

The segmentation result obtained after iterative refinement might contain over-segmented regions of the aorta or incorrectly segmented apical slices. Hence, we analyze the shape statistics of the segmented region for all the slices and remove outlier apical/basal slices.

We extract four shape statistics from the segmented region in each slice:

1) Thinness ratio: $\frac{4\pi \times Area}{Perimeter^2}$, measures circularity of the segmented region

2) Eccentricity: Ratio of focal distance over the major axis length of least square fitted ellipse

3) Solidity: Ratio of pixels in the region to pixels in its convex hull

4) Extent: Ratio of pixels in the region to pixels in the total bounding box

The Stochastic Outlier Selection (SOS) algorithm [11] computes the probability of each slice being an outlier based on the affinity matrix obtained from the shape features. The variance of a data point depends on the density of the neighborhood, which is set such that each data point has the same number of neighbors. This number is controlled via the only parameter of SOS, called perplexity (k_{SOS}).

If the slice with maximum probability of being an outlier has a higher probability than a predefined probability threshold, Th_{SOS}, and belongs to apical/basal region (determined by the thresholded BP probability map), slices above/below this slice have a high probability of being an over-segmented aorta/apex region, and hence are removed to obtain the final BP segmentation, as described in **Algorithm 3**. As observed in **Fig. 5.3e**, over-segmented aorta in the basal region has been removed after the SOS refinement to obtain the final BP segmentation.

5.3 Implementation Details

The CNN model was implemented in Python using the Keras application programming interface (API) [12] running on top of TensorFlow [13]. The programming environment was setup as a docker[3] container for portability and reproducibility. The system comprised of Intel(R) Xeon(R) CPU X5650 @ 2.67GHz with 12 cores, 96 GB of system memory, and two 12 GB Nvidia Titan Xp GPUs.

The 4D dataset with 189 different augmentations were saved as 189 shuffled files (2.8 GB), with each shuffled file containing a randomly augmented volume from all 97 datasets. During training, the files were read ahead and pushed to a queue such that

[3] `http://www.docker.com`

the data generator could generate a random batch of training data without significant IO delay. The network weights were initialized using the Xavier uniform initializer, from a uniform distribution within [-L,+L] with $L = \sqrt{\frac{6}{fan_{in}+fan_{out}}}$, where fan_{in} and fan_{out} are the number of input and output units in the weight tensor, respectively.

The 97 datasets were randomly split into 79 training, 9 validation, and 9 test sets. The network was trained for 100 epochs with each epoch comprising of 100 random shuffled files and requiring 46 minutes on average. The sum of squared difference error between predicted and ground truth LV center was used as the loss function. The model yielding the lowest validation loss at the end of an epoch was saved and used to evaluate the results on the test set. The model requires 1.23 seconds on average to predict the LV centers for 9 test datasets (2670 SA slices).

The parameters for LV blood-pool segmentation were empirically tuned based on the validation dataset. For the exhaustive search based image registration, the number of scaling factors, s, per dimension is set to 5, and the number of z-translations, t, is set to 3, such that the NCC similarity measure for $5^3 \times 3 = 425$ different scale and translation combinations have to be computed between the test and transformed average intensity image. We select $k = 5$ best initial transforms for subsequent registration and use the transformation producing best NCC similarity to transfer the BP probability map to a test dataset. The BP and BG thresholds for the BP probability map are set to $Th_{BP} = 0.5$ and $Th_{BG} = 0.0$, respectively. The weights for data energy term in **(5.2)** are set to $w_1 = 10.0$, $w_2 = 2.0$, $w_3 = 15.0$, and the weights for the smoothness energy term in **(5.3)** are set to $w_4 = 50.0$ and $w_5 = 50.0$. Similarly, the intensity spread parameter in the smoothness term **(5.3)** is set to $\sigma = 0.1$, as we rescale the image intensities to a range of 0.0 to 1.0. The threshold for IoU based BP refinement as defined in **Algorithm 2** is set to $Th_{IOU} = 0.6$. The iterative segmentation refinement is stopped when the IoU between two consecutive segmentations is greater than $Th_{stop} = 0.95$ or maximum number of iterations $itr_{max} = 10$ has been reached. Finally, for the segmentation refinement using SOS, the outlier probability

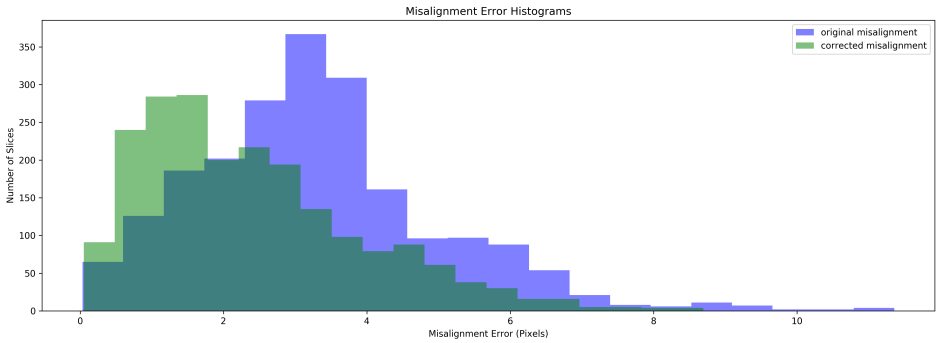
threshold Th_{SOS} is set to 0.6, and the perplexity parameter, k_{SOS}, is set to number of slices with BP segmentation subtracted by 2, allowing only few slices to be considered as an outlier.

5.4 Results

Table 5.1: Mean, Standard Deviation, and Median slice misalignment in pixels before and after the correction.

Method	Before Correction (pixels)	After Correction (pixels)
Mean ± Std	3.30 ± 1.71	2.40 ± 1.54
Median	3.13	2.07

Figure 5.4: Histogram for misalignment errors before and after the correction.



We computed the mean of the LV center across all the slices in the 3D test volume. Assuming the slices need to be translated to the mean center point for slice misalignment correction, we computed the distance of each true LV center point to the mean and designated it as the initial slice misalignment in the test data. Similarly, the Euclidean distance between the predicted and true LV center points yielded the residual slice misalignment after the proposed correction. **Table 5.1** shows the

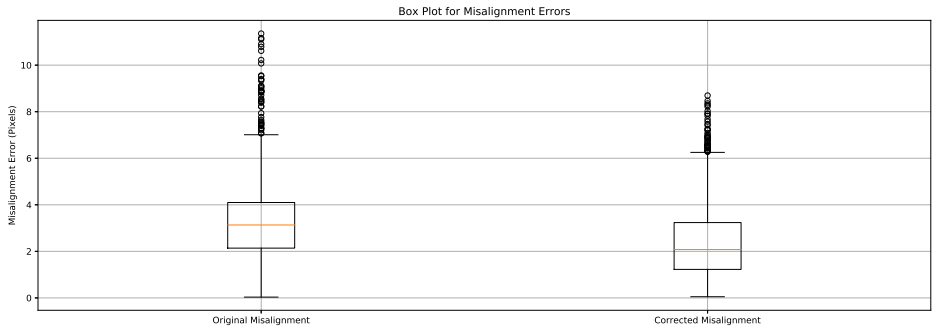


Figure 5.5: Boxplot for misalignment errors before and after the correction. Median (orange line), Interquartile range (box), and outliers (points outside the whiskers) can be observed in the plot. The median misalignment error is reduced from 3.20 to 2.14 pixels.

misalignment statistics, **Fig. 5.2** shows significant reduction of stair-step artifact on the 3D reconstructed LV myocardium, **Fig. 5.4** shows the histogram of misalignment errors, and **Fig. 5.5** shows the box plot for errors, before and after the misalignment correction. Further, the Kolmogorov-Smirnov statistics on the errors before and after misalignment correction was able to reject the null hypothesis that the two error samples come from the same distributions with a p-value of $1.617e^{-76}$. Hence, the proposed CNN regression architecture was successful in reducing the slice misalignment error statistically significantly from median error of 3.13 to 2.07 pixels.

Table 5.2: Evaluation of the end-diastole blood-pool segmentation results against the ground truth blood-pool segmentation using Jaccard, Dice, Hausdorff Distance and Mean Surface Distance measures for the Validation and Test datasets.

Dataset	Jaccard	Dice	Hausdorff Distance (mm)	Mean Surface Distance (mm)
Test Set	0.829 ± 0.077	0.904 ± 0.048	9.446 ± 4.936	0.560 ± 0.566
Validation Set	0.825 ± 0.074	0.902 ± 0.045	9.290 ± 2.088	0.371 ± 0.213

On average, the proposed algorithm converges in 3 iterative refinement steps and requires $\sim$ 2.2 mins to segment a 3D test volume. The obtained segmentation re-

Table 5.3: Evaluation of the end-diastole blood-pool segmentation results against the ground truth blood-pool segmentation using Dice overlap shown for several published methods extracted from the corresponding manuscripts. Note the evaluation is performed on different datasets.

Method	Dice
Queiros et al.[14]	0.93 ± 0.03
Proposed	0.90 ± 0.05
Eslami et al.[15]	0.83 ± 0.03
Grosgeorge et al.[16]	0.82 ± 0.01

sults were validated against the blood-pool segmentation extracted from the provided ground-truth myocardium segmentation as shown in **Fig. 5.6**. We computed the Jaccard and Dice similarity measures along with the mean surface distance and Hausdorff distance for both the validation and the test datasets as shown in **Table 5.2**. Although the parameters for the algorithm are tuned using the validation dataset, the proposed method generalizes well in the test dataset with similar results.

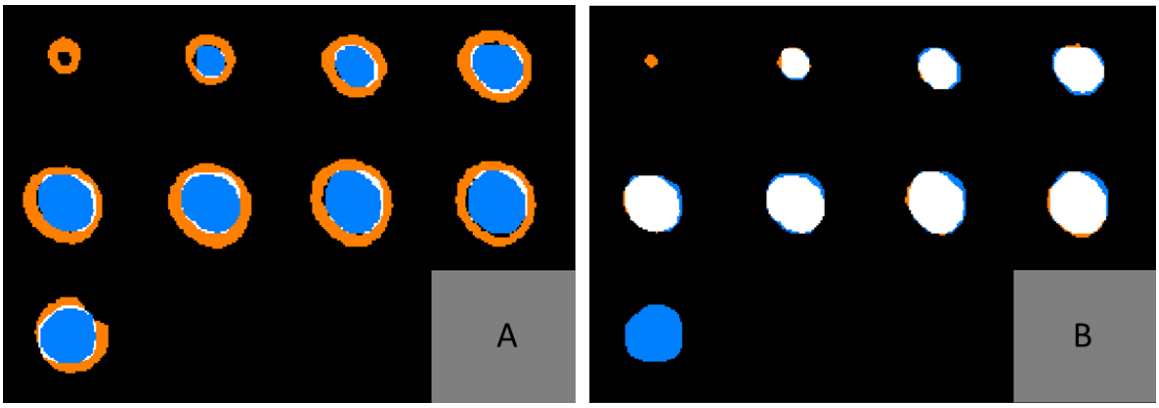


Figure 5.6: Blood-pool segmentation result obtained from the proposed method compared against: a) Ground-truth myocardium segmentation; b) Blood-pool segmentation extracted from the myocardium. The ground-truth, obtained segmentation result, and the intersection regions are shown in red, blue, and white, respectively. Due to the difficulty of obtaining blood-pool segmentation from the partial myocardium slices, these slices are assumed to not contain blood-pool, hence significantly affecting the blood-pool segmentation evaluation.

Furthermore, **Table 5.3** clearly shows that the proposed method features an equivalent performance to the other published methods. Nevertheless, since our proposed method and the previously published methods were evaluated on different datasets, this table provides a holistic evaluation of the described method.

5.5 Discussion, Conclusion, and Future Work

We presented a CNN based regression architecture to predict the LV blood-pool center from the SA cine MR slices. The predicted LV center points for all slices were translated to image center to reduce the median slice-misalignment from 3.13 to 2.07 pixels. As our algorithm in its current form requires minimal preprocessing, specifically resampling the SA images to a spacing of 1.5625×1.5625 mm and center cropping/zero-padding to a common resolution of 192×192 pixels, we obtain a large number of network parameters at the fully connected layers, requiring a large training dataset. We plan to crop an ROI from the original SA images using the Fourier first harmonics information obtained from the image sequence throughout the cardiac cycle, such that the CNN network parameters would be reduced and possibly yield better LV alignment. Furthermore, we will be exploring a segmentation-based approach, where we predict a Gaussian kernel centered at the LV center for misalignment correction.

After obtaining a coherent 3D volume from the slice-misalignment correction, we performed a full 3D segmentation of the LV blood-pool by exploiting the 3D context information using the LV atlas. The 6-neighborhood graph structure ensured smoothness between slice segmentations. Since image registration is highly dependent on a good initialization, we perform an exhaustive search and refinement to obtain the best possible registration result between the average intensity atlas and the test patient volume. Any initial misalignments are further corrected by a graph-cut based iterative refinement process. In addition, due to the intensity similarity of the aortic

region in basal slices to the blood-pool region, basal slices could be over-segmented, and are removed if their segmented shape is significantly different from that of the other slices using the Stochastic Outlier Selection algorithm.

Despite the difficulty of obtaining the gold-standard blood-pool segmentation from the provided ground-truth myocardium segmentation in the partial myocardium regions, our segmentation algorithm is able to obtain a mean dice similarity metric of over 90%, with mean surface distance of ~ 0.5 mm, and Hausdorff distance of ~ 9.4 mm. As there would be no gold-standard blood-pool in the partial myocardium slices, it significantly affects the similarity metrics and surface distance measurements. Furthermore, although the current algorithm parameters are tuned empirically using a validation dataset, we plan to do an extensive study of their impact in the final segmentation result. We also plan to extend our work for myocardium segmentation throughout the cardiac cycle.

Acknowledgements: *This work was supported by the Intramural Research Program of the U.S. National Institutes of Health, National Library of Medicine.*

Bibliography

[1] Caroline Petitjean and Jean-Nicolas Dacher. A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis*, 15(2):169 – 184, 2011.

[2] A. Elen, J. Hermans, J. Ganame, D. Loeckx, J. Bogaert, F. Maes, and P. Suetens. Automatic 3-D breath-hold related motion correction of dynamic multislice MRI. *IEEE Transactions on Medical Imaging*, 29(3):868–878, March 2010.

[3] Piotr J. Slomka, David Fieno, Amit Ramesh, Vaibhav Goyal, Hidetaka Nishina, Louise E.J. Thompson, Rola Saouaf, Daniel S. Berman, and Guido Germano. Patient motion correction for multiplanar, multi-breath-hold cardiac cine MR imaging. *Journal of Magnetic Resonance Imaging*, 25(5):965–973, 2007.

[4] Carissa G. Fonseca, Michael Backhaus, David A. Bluemke, Randall D. Britten, Jae Do Chung, Brett R. Cowan, Ivo D. Dinov, J. Paul Finn, Peter J. Hunter, Alan H. Kadish, Daniel C. Lee, Joao A. C. Lima, Pau MedranoGracia, Kalyanam Shivkumar, Avan Suinesiaputra, Wenchao Tao, and Alistair A. Young. The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*, 27(16):2288–2295, 2011.

[5] Avan Suinesiaputra, Brett R. Cowan, Ahmed O. Al-Agamy, Mustafa A. Elattar, Nicholas Ayache, Ahmed S. Fahmy, Ayman M. Khalifa, Pau Medrano-Gracia, Marie-Pierre Jolly, Alan H. Kadish, Daniel C. Lee, Jn Margeta, Simon K. Warfield, and Alistair A. Young. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac {MR} images. *Medical Image Analysis*, 18(1):50 – 62, 2014.

[6] Ziv Yaniv, Bradley C. Lowekamp, Hans J. Johnson, and Richard Beare. Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research. *Journal of Digital Imaging*, Nov 2017.

[7] Shusil Dangi, Nathan Cahill, and Cristian A. Linte. *Integrating Atlas and Graph Cut Methods for Left Ventricle Segmentation from Cardiac Cine MRI*, pages 76–86. Springer International Publishing, Cham, 2017.

[8] Brian B. Avants, Nicholas J. Tustison, Michael Stauffer, Gang Song, Baohua Wu, and James C. Gee. The insight toolkit image registration framework. *Front Neuroinform.*, 8:1–13, 2014.

[9] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

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

[11] JHM Janssens, Ferenc Huszár, EO Postma, and HJ van den Herik. Stochastic outlier selection. Report, 2012.

[12] François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[13] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J. Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A. Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B. Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *CoRR*, abs/1603.04467, 2016.

[14] Sandro Queirs, Daniel Barbosa, Brecht Heyde, Pedro Morais, Joo L. Vilaa, Denis Friboulet, Olivier Bernard, and Jan Dhooge. Fast automatic myocardial segmentation in 4d cine CMR datasets. *Medical Image Analysis*, 18(7):1115 – 1131, 2014.

[15] Abouzar Eslami, Athanasios Karamalis, Amin Katouzian, and Nassir Navab. Segmentation by retrieval with guided random walks: Application to left ventricle segmentation in mri. *Medical Image Analysis*, 17(2):236 – 253, 2013.

[16] Damien Grosgeorge, Caroline Petitjean, Jérôme Caudron, Jeannette Fares, and Jean-Nicolas Dacher. Automatic cardiac ventricle segmentation in MR images: a validation study. *International Journal of Computer Assisted Radiology and Surgery*, 6(5):573–581, 2011.

Chapter 6

Left Ventricle Segmentation and Quantification from Cardiac Cine MR Images via Multi-task Learning

Segmentation of the left ventricle and quantification of various cardiac contractile functions is crucial for the timely diagnosis and treatment of cardiovascular diseases. Traditionally, the two tasks have been tackled independently. Here we propose [1] *a convolutional neural network based multi-task learning approach to perform both tasks simultaneously, such that, the network learns better representation of the data with improved generalization performance. Probabilistic formulation of the problem enables learning the task uncertainties during the training, which are used to automatically compute the weights for each task.*

[1]This chapter is adapted from:
Dangi S., Yaniv Z., Linte C.A. (2019) Left Ventricle Segmentation and Quantification from Cardiac Cine MR Images via Multi-task Learning. In: Pop M. et al. (eds) Statistical Atlases and Computational Models of the Heart (STACOM). Springer Lect Notes Comput Sci. Vol. 11295. Pp.: 21-31.

6.1 Introduction

Recent success of deep learning techniques [1] in high level computer vision, speech recognition, and natural language processing applications has motivated their use in medical image analysis. Long *et al.* [2] were the first to propose a fully convolutional network (FCN) for semantic image segmentation by adapting the contemporary classification networks fine-tuned for the segmentation task, obtaining state-of-the-art results. Several modifications to the FCN architecture and various post-processing schemes have been proposed to improve the semantic segmentation results as summarized in [3]. Notably, the U-Net architecture [4] with data augmentation has been very successful in medical image segmentation.

While segmentation indirectly enables the computation of various cardiac indices, direct estimation of these quantities from low-dimensional representation of the image have also been proposed in the literature [5–7]. However, these methods are less interpretable and the correctness of the produced output is often unverifiable, potentially limiting their clinical adoption.

Here we propose a CNN based multi-task learning approach to simultaneously perform both the LV segmentation and cardiac index estimation simultaneously, such that these related tasks regularize the network, hence improving the network generalization performance. Furthermore, our method increases the interpretablity of the output cardiac indices, as the clinicians can infer its correctness based on the quality of produced segmentation result.

6.2 Methodology

Traditionally, the segmentation of the LV and quantification of the cardiac indices have been performed independently. However, due to a close relation between the two tasks, we identified that learning a CNN model to perform both tasks simultaneously is beneficial in two ways: 1) it forces the network to learn features important for

both tasks, hence, reducing the chances of over-fitting to a specific task, improving generalization; 2) the segmentation results can be used as a proxy to identify the reliability of the obtained cardiac indices, and also to perform regional cardiac analysis and surgical planning.

6.2.1 Data Preprocessing and Augmentation

This study employed 97 de-identified cardiac MRI image datasets from patients suffering from myocardial infarction and impaired LV contraction available as a part of the STACOM Cardiac Atlas Segmentation Challenge project [8, 9] database[2]. Cine-MRI images in short-axis and long-axis views are available for each case. The semi-automated myocardium segmentation provided with the dataset served as gold-standard for assessing the proposed segmentation technique. The dataset was divided into 80% training and 20% testing for five-fold cross-validation.

The physical pixel spacing in the short-axis images ranged from 0.7031 to 2.0833 mm. We used SimpleITK [10] to resample all images to the most common spacing of 1.5625 mm along both x- and y-axis. The resampled images were center cropped or zero padded to a common resolution of 192×192 pixels. We applied two transformations, obtained from the combination of random rotation and translation (by maximum of half the image size along x- and y-axis), to each training image for data augmentation.

6.2.2 MTL using Uncertainty-based Loss Weighting

Following [11, 12], we model the likelihood for a segmentation task as the squashed and scaled version of the model output through a softmax function: where, σ is a positive scalar, equivalent to the *temperature*, for the defined Gibbs/Boltzmann distribution. The magnitude of σ determines how *uniform* the discrete distribution

[2] `http://www.cardiacatlas.org`

is, and hence relates to the uncertainty of the prediction measured in entropy. The log-likelihood for the segmentation task can be written as:

$$\begin{aligned}
&\log p(\boldsymbol{y} = c|\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x}), \sigma) \\
&\qquad = \frac{1}{\sigma^2} f_c^{\boldsymbol{W}}(\boldsymbol{x}) - \log \sum_{c'} \exp\left(\frac{1}{\sigma^2} f_{c'}^{\boldsymbol{W}}(\boldsymbol{x})\right) \\
&\qquad = \frac{1}{\sigma^2}(f_c^{\boldsymbol{W}}(\boldsymbol{x}) - \log \sum_{c'} \exp\left(f_{c'}^{\boldsymbol{W}}(\boldsymbol{x})\right) - \\
&\qquad\qquad \log \frac{\sum_{c'} \exp\left(\frac{1}{\sigma_2^2} f_{c'}^{\boldsymbol{W}}(\boldsymbol{x})\right)}{\left(\sum_{c'} \exp\left(f_{c'}^{\boldsymbol{W}}(\boldsymbol{x})\right)\right)^{\frac{1}{\sigma_2^2}}} \\
&\qquad \approx \frac{1}{\sigma^2} \log \text{ Softmax}\left(\boldsymbol{y}, \boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})\right) - \log\sigma
\end{aligned} \tag{6.1}$$

where $f_c^{\boldsymbol{W}}(\boldsymbol{x})$ is the c'th element of the vector $\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})$. In the last step, a simplifying assumption $\frac{1}{\sigma} \sum_{c'} \exp\left(\frac{1}{\sigma^2} f_{c'}^{\boldsymbol{W}}(\boldsymbol{x})\right) \approx \left(\sum_{c'} \exp\left(f_{c'}^{\boldsymbol{W}}(\boldsymbol{x})\right)\right)^{\frac{1}{\sigma^2}}$, which becomes an equality when $\sigma \to 1$, has been made, resulting in a simple optimization objective with improved empirical results [11, 12].

Similarly, for the regression task, we define our likelihood as a Lapacian distribution with its mean and scale parameter given by the neural network output:

$$p(\boldsymbol{y}|\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x}), \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|\boldsymbol{y} - \boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})|}{\sigma}\right) \tag{6.2}$$

The log-likelihood for regression task can be written as:

$$\log p(\boldsymbol{y}|\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x}), \sigma) \approx -\frac{1}{\sigma}|\boldsymbol{y} - \boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})| - \log\sigma \tag{6.3}$$

where σ is the neural networks observation noise parameter — capturing the noise in the output. A constant term has been removed for simplicity, as it does not affect the optimization.

For a network with two outputs — continuous output $\boldsymbol{y}_1$ modeled with a Laplacian likelihood, and a discrete output $\boldsymbol{y}_2$ modeled with a softmax likelihood — the joint

loss is:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{W_1}, \boldsymbol{W_2}, \sigma_1, \sigma_2) \\
&= -\log\, p(\boldsymbol{y}_1, \boldsymbol{y}_2 = c | \boldsymbol{f}^{\boldsymbol{W_1}}(\boldsymbol{x}), \boldsymbol{f}^{\boldsymbol{W_2}}(\boldsymbol{x}), \sigma_1, \sigma_2) \\
&= -\log\, (p(\boldsymbol{y}_1 | \boldsymbol{f}^{\boldsymbol{W_1}}(\boldsymbol{x}), \sigma_1) \cdot p(\boldsymbol{y}_2 = c | \boldsymbol{f}^{\boldsymbol{W_2}}(\boldsymbol{x}), \sigma_2)) \\
&\approx \frac{1}{\sigma_1}\mathcal{L}_1(\boldsymbol{W_1}) + \frac{1}{\sigma_2^2}\mathcal{L}_2(\boldsymbol{W_2}) + \log\sigma_1 + \log\sigma_2
\end{aligned} \tag{6.4}$$

where $\mathcal{L}_1(\boldsymbol{W_1}) = |\boldsymbol{y_1} - \boldsymbol{f}^{\boldsymbol{W_1}}(\boldsymbol{x})|$ is the MAD loss of $\boldsymbol{y}_1$ and $\mathcal{L}_2(\boldsymbol{W_2}) = -\log \text{Softmax}(\boldsymbol{y}_2, \boldsymbol{f}^{\boldsymbol{W_2}}(\boldsymbol{x}))$ is the cross-entropy loss of $\boldsymbol{y}_2$. To arrive at Eq. 6.4, the two outputs are assumed independent given the representation learned by the network. During the training, the joint likelihood loss $\mathcal{L}(\boldsymbol{W_1}, \boldsymbol{W_2}, \sigma_1, \sigma_2)$ is optimized with respect to $\boldsymbol{W_1}$, $\boldsymbol{W_2}$ as well as σ_1, and σ_2, yielding satisfactory results in practice. Alternatively, learning the tasks relationship via covariance matrix, as shown by [7] for a fully-connected network, might slightly improve the performance in expense of higher computational cost required to optimize the resulting complex optimization objective.

From Eq. 6.4, we can observe that the losses for individual tasks are weighted by the inverse of their corresponding uncertainties (σ_1, σ_2) learned during the training. Hence, tasks with higher uncertainty will be weighted less, and vice versa. Furthermore, the uncertainties cannot grow too large due to the penalty imposed by the last two terms in (Eq. 6.4). In practice, the network is trained to predict the log variance, $s := \log\sigma$, for numerical stability and avoiding any division by zero, such that, the positive scale parameter, σ, can be computed via exponential mapping $\exp(s)$.

6.2.3 Network Architecture

In this work, we adapt the U-Net architecture [4], highly successful in medical image segmentation, to perform an additional task of myocardium area estimation as shown in **Fig. 6.1**. The segmentation and regression paths are split at the final up-sampling and concatenation layer. The final feature map in the segmentation

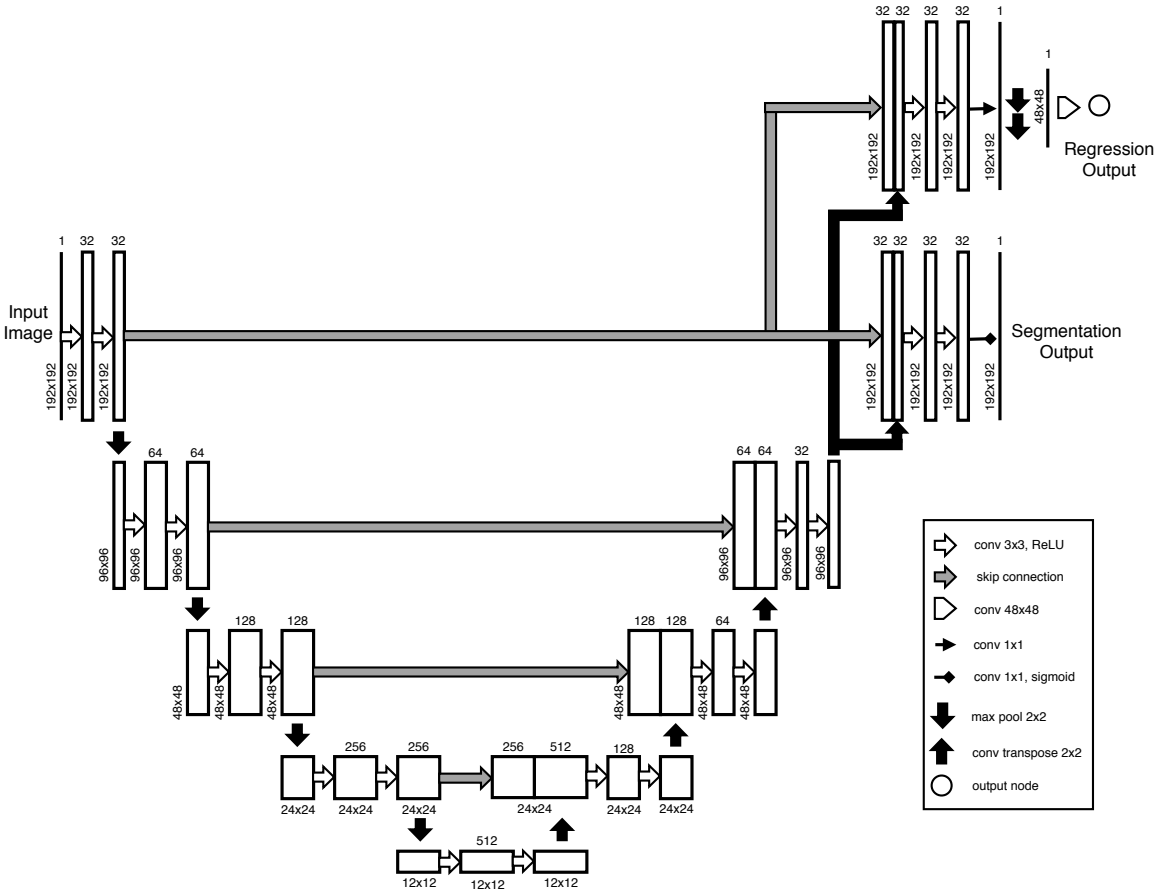


Figure 6.1: Modified U-Net architecture for multi-task learning. The segmentation and regression tasks are split at the final up-sampling and concatenation layer. The final feature map in the segmentation path is passed through a sigmoid layer to obtain a per-pixel image segmentation. Similarly, the final feature map in the regression path is down-sampled (by max-pooling) to $1/4^{th}$ of its size and fed to a fully-connected layer to generate a single regression output. The logarithm of the task uncertainties ($\log\sigma_1$, $\log\sigma_2$) are set as network parameters and are encoded in the loss function **(6.4)**, hence learned during the training.

path is passed through a sigmoid layer to obtain a per-pixel image segmentation. Similarly, the regression output is obtained by down-sampling the final feature map in the regression path by $1/4^{th}$ of its size and passing it through a fully-connected layer. The logarithm of the task uncertainties ($\log\sigma_1$, $\log\sigma_2$) added as the network parameters are used to construct the loss function **(6.4)**, and are learned during the training. Note that we train the network to predict the log uncertainty $s = \log(\sigma)$

due to its numerical stability and the positivity constraint imposed on the computed uncertainty via exponentiation, $\sigma = \exp(s)$.

6.3 Results

Due to the availability of relatively large dataset of cine MR images, it is possible to train our network from random initialization as opposed to initializing it with weights obtained from pre-trained networks. Hence, the network was initialized with the *Kaiming uniform* [13] initializer and trained for 50 epochs using *RMS prop* optimizer with a learning rate of 0.001 (decayed by 0.95 every epoch) in PyTorch[3]. The best performing network, in terms of the Dice overlap between the obtained and gold-standard segmentation, in the test set, was saved and used for evaluation.

The network training required 9 minutes per epoch on average using a 12GB Nvidia Titan Xp GPU. It takes 0.663 milliseconds on average to process a slice during testing. The log uncertainties learned for the segmentation and regression tasks during training are -3.9 and 3.45, respectively, which correspond to weighting the cross-entropy and mean absolute difference (MAD) loss by a ratio of 1556:1. Note that the scale for cross-entropy loss is 10^{-2}, whereas that for MAD loss is 10^{2}.

The 2D segmentation results are stacked to form a 3D volume, and the largest connected component is selected as the final myocardium segmentation. The myocardium segmentation obtained for end-diastole, end-systole, and all cardiac phases from the proposed multi-task network (MTN) and from the baseline U-Net architecture (without the regression task) are both assessed against the gold-standard segmentation provided with the dataset as part of the challenge, using four traditionally employed segmentation metrics — Dice Index, Jaccard Index, Mean surface distance (MSD), and Hausdorff distance (HD) — summarized in **Table 6.1**. Note that the myocardium dice coefficient is higher for end-systole phase where the myocardium is

[3]https://github.com/pytorch/pytorch

Table 6.1: Evaluation of the segmentation results obtained from the baseline U-Net (UNet) architecture and the proposed multi-task network (MTN) against the provided gold-standard myocardium segmentation using — Dice Index, Jaccard Index, Mean Surface Distance, and Hausdorff Distance.

Assessment Metric	**End-Diastole**		**End-Systole**		**All Phases**	
	UNet	**MTN**	UNet	**MTN**	UNet	**MTN**
Dice Index	0.836 ± 0.036	**0.837**± 0.038	**0.850**± 0.033	0.849 ± 0.036	0.847 ± 0.035	**0.849**± 0.036
Jaccard Index	0.719 ± 0.052	**0.721**± 0.054	**0.740**± 0.048	0.739 ± 0.053	0.736 ± 0.050	**0.739**± 0.053
Mean Surface Distance (mm)	0.318 ± 0.089	**0.286**± 0.087	0.299 ± 0.095	**0.274**± 0.090	0.305 ± 0.088	**0.274**± 0.083
Hausdorff Distance (mm)	13.582± 4.337	**13.364**± 4.108	**13.083**± 3.630	13.355± 3.861	**13.211**± 4.212	13.233± 3.810

thickest.

The Kolmogorov-Smirnov test shows that the difference in distributions for Dice, Jaccard and MSD metrics between the proposed multi-task network and baseline U-Net architecture are statistically significant with p-values: $2.156e^{-4}$, $2.156e^{-4}$, and $6.950e^{-34}$, respectively. However, since the segmentation is evaluated on a large sample of 2191 volumes across five-fold cross validation, the p-values quickly go to zero even for slight difference in distributions being compared, representing no practical significance [14]. Hence, we computed the 99% confidence interval for the mean value of each segmentation metric based on 1000 bootstrap re-sampling with replacement, as shown in **Fig. 6.2**. As evident from **Fig. 6.2**, Dice, Jaccard and HD metrics are statistically similar, whereas the reduction in MSD for the proposed multi-task network compared to the baseline U-Net architecture is statistically significant.

In addition to obtaining the myocardium area from the regression path of the proposed network, it can also be computed indirectly from the obtained myocardium segmentation. Hence, we compute and evaluate the myocardium area estimated from

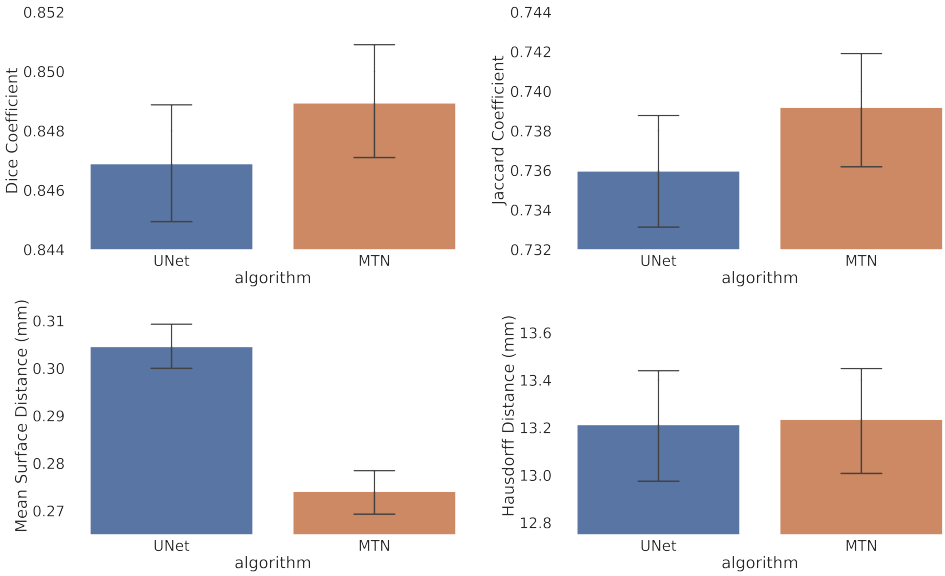


Figure 6.2: Mean and 99% confidence interval for (a) Dice Coefficient, (b) Jaccard Coefficient, (c) Mean Surface Distance (mm), and (d) Hausdorff Distance (mm), for baseline U-Net and the proposed MTN architecture across all cardiac phases. Confidence interval is obtained based on 1000 bootstrap re-sampling with replacement for 2191 test volumes across five-fold cross-validation.

three different sources: (a) regression path of the MTN, (b) segmentation obtained from the MTN, and (c) segmentation obtained from the baseline U-Net model. **Fig. 6.3** shows the myocardium area obtained from these three methods for all phases of the cardiac cycle plotted against the ground-truth myocardium area estimated from the gold-standard myocardium segmentation provided as part of the challenge data. We can observe a linear relationship between the computed and gold-standard myocardium areas, and the corresponding correlation coefficients for the methods (a), (b), and (c) are 0.9466, 0.9565, 0.9518, respectively. Hence the myocardium area obtained from MTN segmentation is closest to the gold-standard segmentation.

Further, we computed the MAD between the ground-truth myocardium area and the area estimated by each of the three methods for end-diastole, end-systole, and all

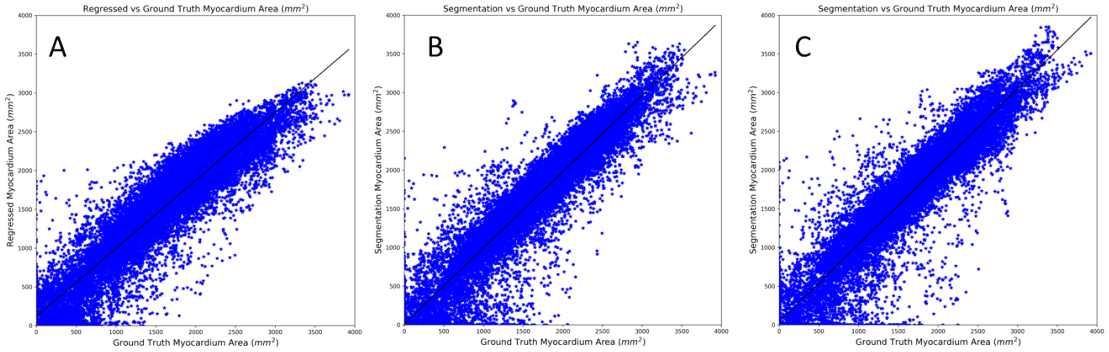


Figure 6.3: The myocardium area computed from A) regression path of the proposed multi-task network, B) segmentation obtained from the proposed multi-task network, C) segmentation obtained from the baseline U-Net model, plotted against the corresponding myocardium area obtained from the provided gold-standard segmentation for all cardiac phases. The best fit line is shown in each plot. The correlation coefficients for A, B, and C are 0.9466, 0.9565, 0.9518, respectively

Table 6.2: Mean absolute difference (MAD), in mm^2, between the myocardium area obtained from the provided gold-standard segmentation and the results computed from: (a) the regression path of the proposed multi-task network, (b) segmentation obtained from the proposed multi-task network, and (c) segmentation obtained from the baseline U-Net model, for end-diastole, end-systole, and all cardiac phases, subdivided into apical, mid, and basal regions of the heart.

Cardiac Regions	**End-Diastole**			**End-Systole**			**All Phases**		
	Reg-MTN	**Seg-MTN**	Seg-UNet	**Reg-MTN**	**Seg-MTN**	Seg-UNet	**Reg-MTN**	**Seg-MTN**	Seg-UNet
All	201 ± 199	**174** ± 209	203 ± 221	211 ± 209	**173** ± 203	187 ± 204	206 ± 198	**170** ± 199	193 ± 208
Apical	**185** ± 180	187 ± 204	194 ± 186	193 ± 199	190 ± 226	**185** ± 185	184 ± 183	**181** ± 210	187 ± 189
Mid	190 ± 172	**141** ± 132	179 ± 142	228 ± 194	**160** ± 151	174 ± 135	212 ± 178	**149** ± 132	176 ± 141
Basal	**250** ± 269	252 ± 331	282 ± 368	193 ± 241	**186** ± 267	216 ± 312	213 ± 248	**210** ± 289	237 ± 319

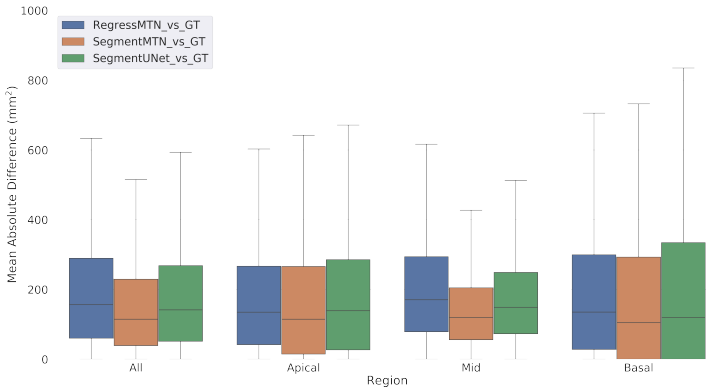
cardiac phases (for 26664 slices across five-fold cross validation). For the regional analysis, slices in the ground-truth segmentation after excluding two apical and two basal

Table 6.3: Mean absolute difference (MAD), in mm^2, between the myocardium area obtained from the provided gold-standard segmentation and the results computed from: (a) Indices-Net [6], (b) Deep Multitask Relationship Learning Network (DMTRL) [7], and (c) segmentation obtained from proposed multi-task network, for mid-slice of the heart. The numbers for (a) and (b) are obtained from Table 1 in [7] evaluated on a different dataset.

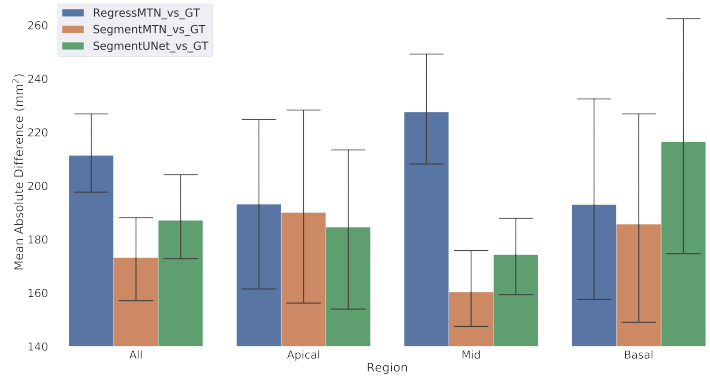
Methods	**MAD mid-slice (mm^2)**
Indices-Net [6]	223 ± 193
DMTRL [7]	189 ± 159
Proposed Seg-MTN	$\mathbf{149 \pm 132}$

slices are considered as mid-slices. **Table 6.2** summarizes the mean and standard deviation for the computed MADs. Box-plots (outliers removed for clarity) comparing the three methods for different regions of the heart throughout the cardiac cycle are shown in **Fig. 6.4a**. The MAD in myocardium area estimation of 206 ± 198 mm^2 obtained from the regression output of the proposed method is similar to the results presented in [6]: 223 ± 193 mm^2. Furthermore, the MAD in mid-slice myocardium area computed from the segmentation obtained from our proposed multi-task network is lower than that reported on [6, 7] as shown in **Table 6.3**, while acknowledging the limitation that the study in [6, 7] was conducted on a different dataset than our study. Moreover, while the regression output of the proposed network yields good estimates of the myocardial area, the box-plot in **Fig. 6.4a** suggests that even further improved myocardial area estimates can be obtained from a segmentation based method, provided that the quality of the segmentation is good.

Lastly, we computed the 99% confidence interval for the mean value of the myocardium area MAD based on 1000 bootstrap re-sampling with replacement, as shown in **Fig. 6.4b**. This confirms that the myocardium area estimated from the segmentation output of the proposed multi-task network is significantly better than that obtained from the regression output; however, there is no statistical significance between other methods. Furthermore, we can observe the variability in MAD is highest



(a) Box-plot for the mean absolute difference (MAD).



(b) Mean and 99% confidence interval for the mean absolute difference (MAD).

Figure 6.4: (a) Box-plot (outliers removed for clarity) and (b) Mean and 99% confidence interval, for the mean absolute difference (MAD) between the myocardium area obtained from the provided gold-standard segmentation and the results obtained from: (1) the regression path of the proposed multi-task network, (2) segmentation obtained from the proposed multi-task network, and (3) segmentation obtained from the baseline U-Net model. Confidence intervals are obtained based on 1000 bootstrap re-sampling with replacement.

for the basal slices, followed by apical, and mid slices.

6.4 Discussion, Conclusion, and Future Work

We presented a multi-task learning approach to simultaneously segment and quantify myocardial area. We adapt the U-Net architecture, highly successful in medical image segmentation, to perform an additional regression task. The best location to incorporate the regression path into the network is a hyper-parameter, tuned empirically. Moreover, we found that adding the regression path in the bottleneck or intermediate decoder layers is detrimental for the segmentation performance of the network, likely due to high influence of the skip connections in the U-Net architecture.

Myocardium area estimates obtained from the regression path of the proposed network are similar to the direct estimation-based results found in the literature. Further, our experiments suggest that segmentation-based myocardium area estimation is superior to that obtained from a direct estimation-based method. However, it should be noted that it is easier to obtain the reference clinical index for direct-estimation, compared to the reference per-pixel segmentation required for segmentation-based method. Lastly, the myocardium segmentation obtained from our method is at least as good as the segmentation obtained from the baseline U-Net model.

To test the generalization performance of the proposed multi-task network, we plan to evaluate the network performance using a lower number of training images. Similarly, we plan to extend this work to segment left ventricle myocardium, blood-pool, and right ventricle, and regress their corresponding areas using the Automated Cardiac Diagnosis Challenge (ACDC)[4] 2017 dataset.

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R35GM128877 and by the Office of Advanced Cyberinfrastructure of the National

[4]https://www.creatis.insa-lyon.fr/Challenge/acdc/

Science Foundation under Award No. 1808530. Ziv Yaniv's work was supported by the Intramural Research Program of the U.S. National Institutes of Health, National Library of Medicine.

Bibliography

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 05 2015.

[2] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[3] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and José García Rodríguez. A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857, 2017.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[5] Xiantong Zhen, Ali Islam, Mousumi Bhaduri, Ian Chan, and Shuo Li. Direct and simultaneous four-chamber volume estimation by multi-output regression. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 669–676, Cham, 2015. Springer International Publishing.

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

[7] Wufeng Xue, Gary Brahm, Sachin Pandey, Stephanie Leung, and Shuo Li. Full left ventricle quantification via deep multitask relationships learning. *Medical Image Analysis*, 43:54 – 65, 2018.

[8] Carissa G. Fonseca, Michael Backhaus, David A. Bluemke, Randall D. Britten, Jae Do Chung, Brett R. Cowan, Ivo D. Dinov, J. Paul Finn, Peter J. Hunter, Alan H. Kadish, Daniel C. Lee, Joao A. C. Lima, Pau MedranoGracia, Kalyanam Shivkumar, Avan Suinesiaputra, Wenchao Tao, and Alistair A. Young. The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*, 27(16):2288–2295, 2011.

[9] Avan Suinesiaputra, Brett R. Cowan, Ahmed O. Al-Agamy, Mustafa A. Elattar, Nicholas Ayache, Ahmed S. Fahmy, Ayman M. Khalifa, Pau Medrano-Gracia, Marie-Pierre Jolly, Alan H. Kadish, Daniel C. Lee, Jn Margeta, Simon K. Warfield, and Alistair A. Young. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac {MR} images. *Medical Image Analysis*, 18(1):50 – 62, 2014.

[10] Ziv Yaniv, Bradley C. Lowekamp, Hans J. Johnson, and Richard Beare. Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research. *Journal of Digital Imaging*, Nov 2017.

[11] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5574–5584. Curran Associates, Inc., 2017.

[12] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[14] Mingfeng Lin, Henry C Lucas Jr, and Galit Shmueli. Research commentary - too big to fail: large samples and the p-value problem. *Information Systems Research*, 24(4):906–917, 2013.

Chapter 7

A Distance Map Regularized CNN for Cardiac Cine MR Image Segmentation

In recent years, convolutional neural networks have been the dominant algorithm for cardiac MR image segmentation. This study [1] *proposes a multi-task learning (MTL)-based regularization of a convolutional neural network (CNN), where the main task of semantic segmentation is accompanied by the simultaneous, auxiliary task of pixel-wise distance map regression. The proposed distance map regularizer is a decoder network added to the bottleneck layer of an existing CNN architecture, facilitating the network to learn robust global features. The regularizer block is removed after training, so that the original number of network parameters does not change. We show that the proposed regularization method improves both binary and multi-class segmentation performance over the corresponding state-of-the-art CNN architectures.*

[1] This chapter is adapted from:
Dangi, S., Linte, C.A. and Yaniv, Z. (2019), A distance map regularized CNN for cardiac cine MR image segmentation. Med. Phys., 46: 5637-5651. doi:10.1002/mp.13853

7.1 Introduction

Fully convolutional networks (FCNs) are currently the dominant approach for image segmentation. Long *et al.* [1] proposed the first fully convolutional network (FCN) for semantic image segmentation, exploiting the capability of Convolutional Neural Networks (CNNs) [2–4] to learn task-specific hierarchical features in an end-to-end manner. However, their initial adoption in the medical domain was challenging, due to the limited availability of medical imaging data and associated costly manual annotation. These challenges were later circumvented by patch-based training, data augmentation, and transfer learning techniques [5, 6].

Specifically, in the context of cardiac image segmentation, Tran [7] adapted a FCN architecture for segmentation of various cardiac structures from short-axis MR images. Similarly, Poudel *et al.* [8] proposed a recurrent FCN architecture to leverage inter-slice spatial dependencies between the 2D cine MR slices. Avendi *et al.* [9] reported improved accuracy and robustness of the LV segmentation by using the output of a FCN to initialize a deformable model. Further, Oktay *et al.* [10] pre-trained an auto-encoder network on ground-truth segmentations and imposed anatomical constraints into a CNN network by adding l_2-loss between the auto-encoder representation of the output and the corresponding ground-truth segmentation. Several modifications to the FCN architecture and various post-processing schemes have been proposed to improve the semantic segmentation results as summarized in [11].

To improve the generalization performance of neural networks, various regularization techniques have been proposed. These include parameter norm penalty (e.g. weight decay [12]), noise injection [13], dropout [14], batch normalization [15], adversarial training [16], and multi-task learning (MTL) [17]. However, norm penalty, noise injection, dropout, and batch normalization are mostly employed to prevent a network from overfitting to the training set. Similarly, adversarial training is focused on improving adversarial robustness of the network to prevent erroneous output

for small perturbations of the input. Hence, to achieve the goal of improving both the network accuracy and its generalization performance, in this work, we focus on MTL-based network regularization.

When a network is trained on multiple related tasks, the inductive bias provided by the auxiliary tasks causes the model to prefer a hypothesis that explains more than one task. This helps the network ignore task-specific noise and hence focus on learning features relevant to multiple tasks, improving the generalization performance [17]. Furthermore, MTL reduces the Rademacher complexity [18] of the model (i.e. its ability to fit random noise), hence reducing the risk of overfitting. An overview of MTL applied to deep neural networks can be found in [19].

MTL has been widely employed in computer vision problems due to the similarity between various tasks being performed. A FCN architecture with a common encoder and task specific decoders was proposed in [20] to perform joint classification, detection, and semantic segmentation, targeting real-time applications such as autonomous driving. A similar single-encoder-multiple-decoder architecture described in [21] performs semantic segmentation, depth regression, and instance segmentation, simultaneously. The architecture was further expanded by [22] to automatically learn the weights for each task based on its uncertainty, obtaining state-of-the-art results.

In the context of medical image analysis, Moeskops *et al.* [23] demonstrated the use of MTL for joint segmentation of six tissue types from brain MRI, the pectoral muscle from breast MRI, and the coronary arteries from cardiac Computed Tomography Angiography (CTA) images, with performance equivalent to networks trained on individual tasks. Similarly, Valindria *et al.* [24] employed a MTL framework to improve the performance for multi-organ segmentation from CT and MR images, exploring various encoder-decoder network architectures. Specific to the cardiac MR applications, Xue *et al.* [25] proposed a network capable of learning multi-task relationship in a Bayesian framework to estimate various local/global LV indices for full quantification of the LV. Similarly, Dangi *et al.* [26] performed joint segmenta-

tion and quantification of the LV myocardium using the learned task uncertainties to weigh the losses, improving upon the state-of-the-art results. Most of these MTL methods in medical image analysis aim to perform various clinically relevant tasks simultaneously. However, the focus of this work is on improving the segmentation performance of various FCN architectures using MTL as a network regularizer.

We propose to use the rich information available in the distance map of the segmentation mask as an auxiliary task for the image segmentation network. Since each pixel in the distance map represents its distance from the closest object boundary, this representation is redundant and robust compared to the per-pixel image label used for semantic segmentation. Furthermore, the distance map represents the shape and boundary information of the object to be segmented. Hence, training the segmentation network on the additional task of predicting the distance map is equivalent to enforcing shape and boundary constraints for the segmentation task.

Related work to ours include [27], which take an image and its semantic segmentation as input and predict the distance transform of the object instances, such that, thresholding the distance map yields the instance segmentation. Similarly, [28] represent the boundary of the object instances using a truncated distance map, which is used to refine the instance segmentation result. However, unlike these methods, our goal is not to perform instance segmentation, but to refine the semantic segmentation result using the distance map as an auxiliary task. The most closely related work to ours is presented in [29] for segmentation of building footprints from satellite images using a MTL framework. In their study, the truncated distance map is predicted at the end of the decoder network and is further used to refine the boundary of the predicted segmentation, resulting in increased model complexity. Unlike that work, we impose a global shape constraint at the bottleneck layer of FCN architectures, using MTL as a network regularizer without increasing the model complexity. The proposed model is customized towards cardiac MRI image segmentation, as we accommodate for slices containing no foreground pixels (in apical and basal regions).

Furthermore, we demonstrate better generalization performance of the proposed network with improved cross-dataset segmentation results.

Contributions: In this work, we propose to impose shape and boundary constraints in a CNN framework to accurately segment the heart chambers from cardiac cine MR images. We impose soft-constraints by including a distance map prediction as an auxiliary task in a MTL framework. We extensively evaluate our proposed model on two publicly available cardiac cine MRI datasets. We demonstrate that the addition of a distance map regularization block improves the segmentation performance of three FCN architectures, without increasing the model complexity and inference time. We employ a task uncertainty-based weighing scheme to automatically learn the weights for the segmentation and distance map regression tasks during training, and show that this method improves segmentation performance over the fixed equal-weighting scheme. Additionally, we show that the proposed regularization technique improves the segmentation performance in the challenging apical and basal slices, as well as across several different pathological heart conditions. This improvement is also reflected on the computed clinical indices important for cardiac health diagnosis. Finally, we demonstrate better generalization ability using the proposed regularization technique with significantly improved cross-dataset segmentation performance, without tuning the network to a new data distribution.

7.2 Methods and Materials

7.2.1 CNN for Semantic Image Segmentation

Let $\boldsymbol{x} = \{x_i \in \mathbb{R}, i \in \mathcal{S}\}$ be the input intensity image and $\boldsymbol{y} = \{y_i \in \mathcal{L}, i \in \mathcal{S}\}$ be the corresponding image segmentation, with $\mathcal{C} = \{0, 1, 2, ..., C-1\}$ representing a set of C class labels, and $\mathcal{S}$ representing the image domain. The task of a CNN based segmentation model, with weights $\boldsymbol{W}$, is to learn a discriminative function

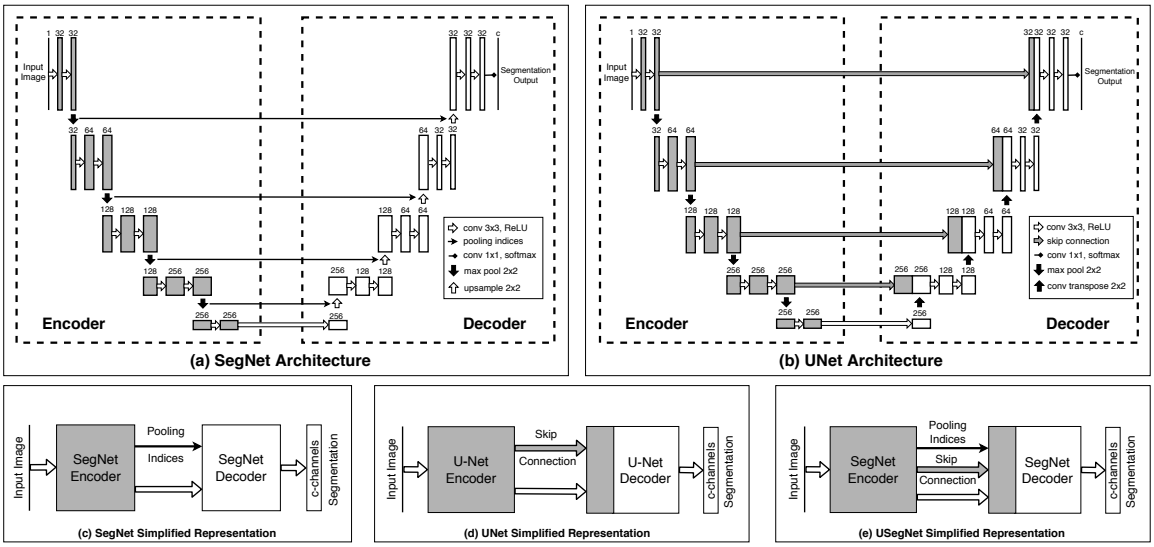


Figure 7.1: Baseline FCN architectures and their simplified block representation. The input image is passed through several convolution, rectified linear unit (ReLU) non-linearity, and down-sampling operations during encoding. This encoded representation is passed through several convolution, ReLU non-linearity, and up-sampling operations during decoding, such that, the final output has the same spatial resolution as the input. (a) SegNet Architecture: max-pooling operation is used for down-sampling, such that the location of the pooled features (i.e. pooling indices) are saved; these pooling-indices are later used to map the features back in their original location during up-sampling; (b) UNet Architecture: skip connections from encoder to decoder layers at different resolutions are added for better flow of information; deconvolution filters are learned for up-sampling the feature maps. Simplified representations of: (c) SegNet Architecture, (d) UNet Architecture, and (e) USegNet Architecture, using both skip-connections as well as the pooling indices for up-sampling.

$\boldsymbol{f}^{\boldsymbol{W}}(\cdot)$ that models the underlying conditional probability distribution $p(\boldsymbol{y}|\boldsymbol{x})$. The output of a CNN model is passed through a softmax function to produce a probability distribution over the class labels, such that, the function $\boldsymbol{f}^{\boldsymbol{W}}(\cdot)$ can be learned by maximizing the likelihood:

$$p(\boldsymbol{y} = c|\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})) = \text{Softmax}(\boldsymbol{f}_c^{\boldsymbol{W}}(\boldsymbol{x})) = \frac{\exp\left(\boldsymbol{f}_c^{\boldsymbol{W}}(\boldsymbol{x})\right)}{\sum_{c' \in \mathcal{L}} \exp\left(\boldsymbol{f}_{c'}^{\boldsymbol{W}}(\boldsymbol{x})\right)} \tag{7.1}$$

where $\boldsymbol{f}_c^{\boldsymbol{W}}(\boldsymbol{x})$ represents the c'th element of the vector $\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})$. In practice, the negative log-likelihood $-\log(p(\boldsymbol{y}|\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})))$ is minimized to learn the optimal CNN model weights, $\boldsymbol{W}$. This is equivalent to minimizing the cross-entropy loss of the

ground-truth segmentation, $\boldsymbol{y}$, with respect to the softmax of the network output, $\boldsymbol{f}^{\boldsymbol{W}}(\boldsymbol{x})$.

A typical FCN architecture (**Fig. 7.1**) for image segmentation consists of an encoder and a decoder network. The encoder network includes multiple pooling (max/average pooling) layers applied after several convolution and non-linear activation layers (e.g. Rectified linear unit (ReLU) [30]). It encodes hierarchical features important for the image segmentation task. To obtain per-pixel image segmentation, the global features obtained at the bottleneck layer need to be up-sampled to the original image resolution using the decoder network. The up-sampling filters can either be fixed (e.g. nearest-neighbor or bilinear upsampling), or can be learned during the training (deconvolutional layer). The final output of a decoder network is passed to a softmax classifier to obtain a per-pixel classification.

In a SegNet [31] (**Fig. 7.1a**) architecture, the location of feature maps during down-sampling (i.e. pooling indices) are saved during encoding, such that the decoder produces sparse feature maps by up-sampling its inputs using these pooling indices. These sparse feature maps are then convolved with a trainable filter bank to obtain dense feature maps, and are finally passed through a softmax classifier to produce per-pixel image segmentation. Since the decoder in the SegNet architecture uses only the global features obtained at the bottleneck layer of the encoder, the high frequency details in the segmentation are lost during the up-sampling process.

The U-Net architecture [32] (**Fig. 7.1b**) introduced skip connections, by concatenating output of encoder layers at different resolutions to the input of the decoder layers at corresponding resolutions, hence preserving the high frequency details important for accurate image segmentation. Furthermore, the skip connections are known to ease the network optimization [33] by introducing multiple paths for backpropagation of the gradients, hence, mitigating the vanishing/exploding gradient problem. Similarly, skip connections also allow the network to learn lower level details in the outer layers and focus on learning the residual global features in the deeper encoder

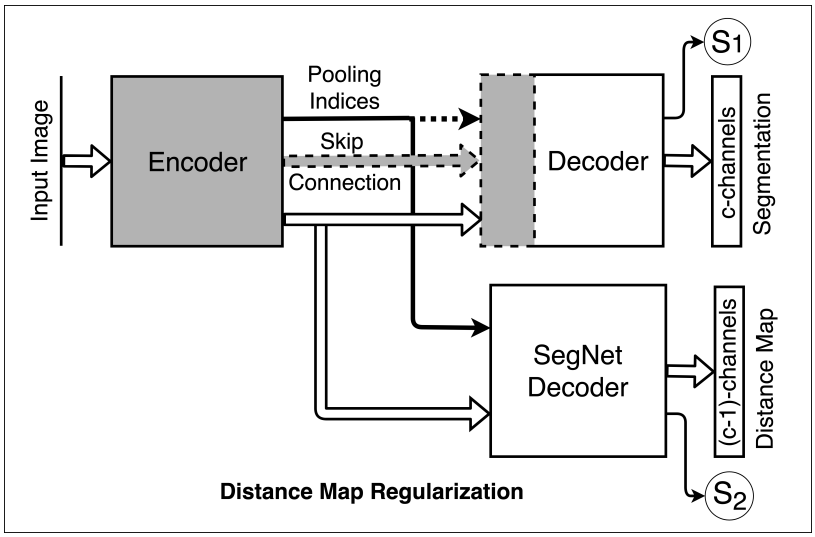


Figure 7.2: Distance map regularizer added to the bottleneck layer. The number of distance map channels is one (1) fewer than the number of classes. Segmentation networks optionally use the pooling indices (yes/no) and skip-connections (yes/no), shown by dashed lines, during decoding: (a) DMR-SegNet: pooling indices (yes), skip connections (no); (b) DMR-USegNet: pooling indices (yes), skip connections (yes); and (c) DMR-UNet: pooling indices (no), skip connections (yes). Uncertainties associated with each task — S_1 corresponding to the semantic segmentation and S_2 corresponding to the pixel-wise distance map regression are also predicted, then subsequently used to scale the corresponding losses during network training.

layers. Hence, the U-Net architecture is able to produce excellent segmentation results using limited training data with augmentation, and has been extensively used in medical image segmentation.

We observed that learned deconvolution filters in the original U-Net architecture can be replaced by a SegNet-like decoder to form a hybrid architecture with reduced network parameters. We refer to this modified architecture as U-SegNet (**Fig 7.1e**) throughout this chapter, and use it as one of the baseline FCN architectures.

7.2.2 Distance Map Regularization Network

The distance map of a binary segmentation mask can be obtained by computing the Euclidean distance of each pixel from the nearest boundary pixel [34]. This

representation provides rich, redundant, and robust information about the boundary, shape, and location of the object to be segmented. For a binary segmentation mask, where $\Omega = \{x_i : y_i = 1, i \in \mathcal{S}\}$ is the set of foreground pixels, $\partial\Omega$ represent the boundary pixels, and $d(\cdot, \cdot)$ is the Euclidean distance between any two pixels, the truncated signed distance map, $D(\boldsymbol{x})$, is computed as:

$$D(x_i) = \begin{cases} d(x_i, \partial\Omega) & \text{if } x_i \in \Omega, \Omega \notin \emptyset \\ -\min(d(x_i, \partial\Omega), T) & \text{if } x_i \notin \Omega, \Omega \notin \emptyset \\ -T & \text{if } \Omega \in \emptyset \end{cases} \tag{7.2}$$

where,

$$d(x_i, \partial\Omega) = \min_{q_i \in \partial\Omega} d(x_i, q_i)$$

is the minimum distance of pixel $x_i \in \boldsymbol{x}$ from the boundary pixels $q_i \in \partial\Omega$. We truncate the signed distance map at a predefined distance threshold, $-T$, hence assigning this maximum negative distance to the slices not containing any foreground pixels (i.e. $\Omega \in \emptyset$), indicating all pixels in the slice are far from the foreground (typically in the apical/basal regions of cardiac cine MR images).

The distance map regularization network is a SegNet-like decoder network, upsampling the feature maps obtained at the bottleneck layer of the encoder to the size of the input image, with the number of output channels equal to the number of foreground classes (i.e. $C - 1$). For example, for a four-class segmentation problem ($C = 4$): background, RV blood-pool, LV myocardium, and LV blood-pool, the regularization network has three output channels, predicting the truncated signed distance maps (Eq. 7.2) computed from the binary masks of the foreground classes: RV bood-pool, LV myocardium, and LV blood-pool.

Fig. 7.2 shows the regularization network added to the bottleneck layer of existing FCN architectures. Network training loss is the weighted sum of the cross-entropy loss for segmentation and the mean absolute difference (MAD) loss between the predicted and the reference distance maps. The network also predicts two scalars, uncertainties

associated with each task, which are subsequently used to weigh the two losses as described in section 7.2.3. Since our goal is to perform semantic segmentation we do not need the distance map prediction at inference time. Therefore, we remove the regularization block after training, such that, the original FCN architecture remains unchanged. Additionally, we found that the quality (mean absolute difference) of the predicted distance maps is insufficient for improving the predicted segmentations from the standard path (see **Fig. 7.5**).

7.2.3 MTL using Uncertainty-based Loss Weighting

Details of Uncertainty-based Loss Weighting can be found in **Chapter 6 subsection 6.2.2**.

7.2.4 Clinical Datasets

7.2.4.1 Left Ventricle Segmentation Challenge (LVSC)

This study employed 200 de-identified cardiac MRI image datasets from patients suffering from myocardial infraction and impaired LV contraction available as a part of the STACOM 2011 Cardiac Atlas Segmentation Challenge project [35, 36] database[2]. Cine-MRI images in short-axis and long-axis views are available for each case. The images were acquired using the Steady-State Free Precession (SSFP) MR imaging protocol with the following settings: typical thickness $\leq 10mm$, gap $\leq 2mm$, TR $30-50ms$, TE $1.6ms$, flip angle 60^0, FOV $360mm$, spatial resolution 0.7031 to 2.0833 $mm^2/pixel$ and $256 \times 256mm$ image matrix using multiple scanners from various manufacturers. Corresponding reference myocardium segmentation generated from expert analyzed 3D surface finite element model are available for 100 training cases throughout the cardiac cycle. The reference segmentation for remaining 100 validation cases are retained by the organizers for an unbiased comparison of segmentation

[2] http://www.cardiacatlas.org/challenges/lv-segmentation-challenge/

results submitted by the challenge participants.

7.2.4.2 Automated Cardiac Diagnosis Challenge (ACDC)

This dataset[3] is composed of short-axis cardiac cine-MR images acquired for 150 patients divided into 5 evenly distributed subgroups: normal, myocardial infarction, dilated cardiomyopathy, hypertropic cardiomyopathy, and abnormal right ventricle, available as a part of the STACOM 2017 ACDC challenge [37]. The acquisitions were obtained over a 6 year period using two MRI scanners of different magnetic strengths (1.5T and 3.0T). The images were acquired using the SSFP sequence with the following settings: thickness $5mm$ (sometimes $8mm$), interslice gap $5mm$, spatial resolution 1.37 to 1.68 $mm^2/pixel$, 28 to 40 frames per cardiac cycle. Corresponding manual segmentations for RV blood-pool, LV myocardium, and LV blood-pool, performed by a clinical expert for the end-systole (ES) and end-diastole (ED) phases are provided for 100 training cases, which we use for our cross-validation experiments. Manual segmentations for the remaining 50 test cases are kept privately by the organizers, such that an unbiased comparison of segmentation results can be performed upon submission.

7.2.5 Data Preprocessing and Augmentation

SimpleITK [38] was used to resample short-axis images to a common resolution of 1.5625 $mm^2/pixel$ and crop/zero-pad to a common size of 192×192 and 256×256 for LVSC and ACDC dataset, respectively. Image intensities were clipped at 99^{th} percentile and normalized to zero mean and unit standard deviation. Each dataset was divided into 80% train, 10% validation, and 10% test set with five non-overlaping folds for cross-validation. Train-validation-test fold was performed randomly over the whole LVSC dataset, whereas it was performed per subgroup (stratified sampling) for the ACDC dataset to maintain even distribution of subgroups over the training,

[3] https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html

validation, and testing sets. The training images were subjected to random similarity transform with: isotropic scaling of 0.8 to 1.2, rotation of 0^o to 360^o, and translation of $-1/8^{th}$ to $+1/8^{th}$ of the image size along both x- and y-axes. The training set for LVSC and ACDC dataset included the original images along with augmentation of two and four randomly transformed versions of each image, respectively. We heavily augment the ACDC dataset, as the labels are available only for the ES and ED phases, whereas, lightly augment the LVSC dataset, as the labels are available throughout the cardiac cycle.

7.2.6 Network Training and Testing Details

Networks implemented in PyTorch[4] were initialized with the *Kaiming uniform* initializer [39] and trained for 30 and 100 epochs for LVSC and ACDC dataset, respectively, with batch size of 15 images. *RMS prop* optimizer [40] with a learning rate of 0.0001 and 0.0005 for single- and multi-task networks, respectively, decayed by 0.99 every epoch was used. We saved the model with best average Dice coefficient on the validation set, and evaluated on the test set.

Networks were trained on the NVIDIA Titan Xp GPU. The distance map threshold was selected empirically and set to a large value of 250 *pixels*, i.e. full distance map. The cross-entropy and the MAD loss were initialized with equal weights of 1.0, such that, the optimal weighting was learned automatically. The auxillary task of distance map regression was removed after the network training. The obtained 2D slice segmentations were rearranged into a 3D volume, and the largest connected component for each heart chamber was retained to yield the final segmentation. Model complexity and average timing requirements for training and testing the models is shown in **Table 7.1**.

[4] `https://github.com/pytorch/pytorch`

Table 7.1: Model complexity, training and testing time. The model size for DMR networks are equivalent to corresponding baseline FCN architectures during test time. The inference time for DMR networks without removing the regularization block are shown in brackets.

	Train Time (min/epoch)		Test Time (ms/volume)		#Parameters ($\times 10^6$)	
	ACDC	LVSC	ACDC	LVSC	Train	Test
SegNet	2.49	14.91	70	67	2.96	2.96
USegNet	2.41	14.49	70	67	3.75	3.75
UNet	2.65	15.50	72	68	4.10	4.10
DMR-SegNet	4.44	20.57	70(157)	63(94)	3.56	2.96
DMR-USegNet	4.84	19.03	73(158)	65(96)	4.35	3.75
DMR-UNet	4.85	21.16	75(160)	67(97)	4.70	4.10

7.2.7 Evaluation Metrics

We use overlap and surface distance measures, as detailed in **Chapter 1 subsection 1.3.4**, to evaluate the segmentation. Additionally, we evaluate the clinical indices associated with the segmentation.

7.3 Results

7.3.1 Segmentation and Clinical Indices Evaluation

The proposed Distance Map Regularized (DMR) SegNet, USegNet, and UNet models along with the baseline models were trained for the joint segmentation of RV blood-pool, LV myocardium, and LV blood-pool from the ACDC challenge dataset. The provided reference segmentation and the corresponding automatic segmentation obtained from the DMR-UNet model for a test patient is shown in **Fig. 7.3**. Similarly, **Fig. 7.5** shows the segmentation results for 2D slices across the heart of a patient, compared against the corresponding reference segmentation, obtained from different networks. Automatic segmentation obtained from all networks, for ED and ES phases,

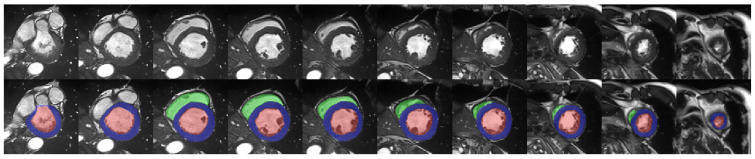
Figure 7.3: Segmentation results for LV blood-pool, LV myocardium, and RV blood-pool. First column shows the short-axis view, second and third columns show orthogonal long-axis views, and the fourth column shows generated three dimensional models. Reference (top row) and segmentation obtained from the DMR-UNet model (bottom row).

are evaluated against the reference segmentation and summarized in **Table 7.2a**; also shown is the evaluation of subsequently computed clinical indices in **Table 7.2b**.

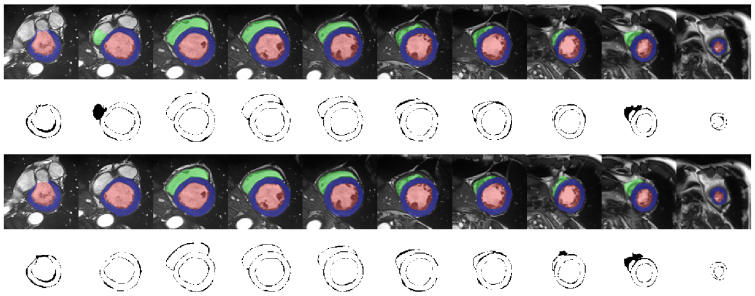
We observe consistent improvement in the average segmentation performance of the models after the DM-Regularization. Specifically, there is statistically significant improvement[5] on several segmentation metrics for all evaluated models. Same results manifest onto the clinical indices with better correlation and LoA on both EF and myocardium mass. Furthermore, the DMR-UNet model outperforms other evaluated networks in many segmentation metrics.

To further analyze the improvement in segmentation performance, we performed a regional analysis by sub-dividing the slices into apical (25% slices in the apical region and beyond), basal (25% slices in the basal region and beyond) and mid-region (remaining 50% mid slices), based on the reference segmentation. From **Fig. 7.6a**, we can observe consistent improvement in segmentation performance at the problematic apical and basal slices [37]; however, due to the small size of these regions, the improvement does not have a large effect on the overall performance, though it

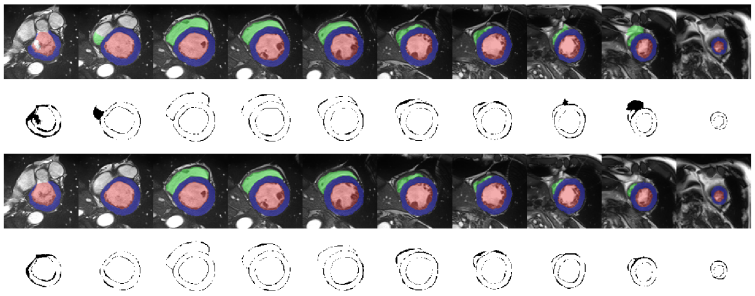
[5]Wilcoxon signed-rank test performed for statistical significance test



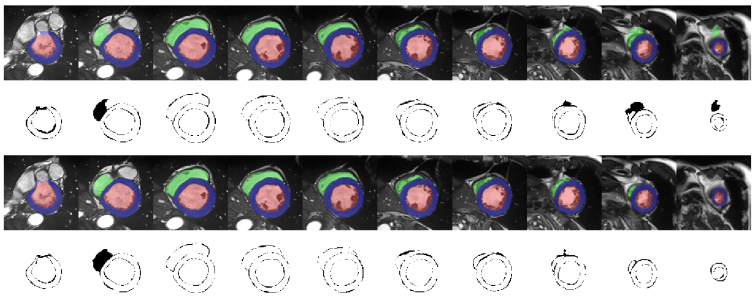
(a) Input volume (top row) with overlaid ground-truth segmentation (bottom row).



(b) Segmentation results for SegNet (top two rows) and DMR-SegNet (bottom two rows).

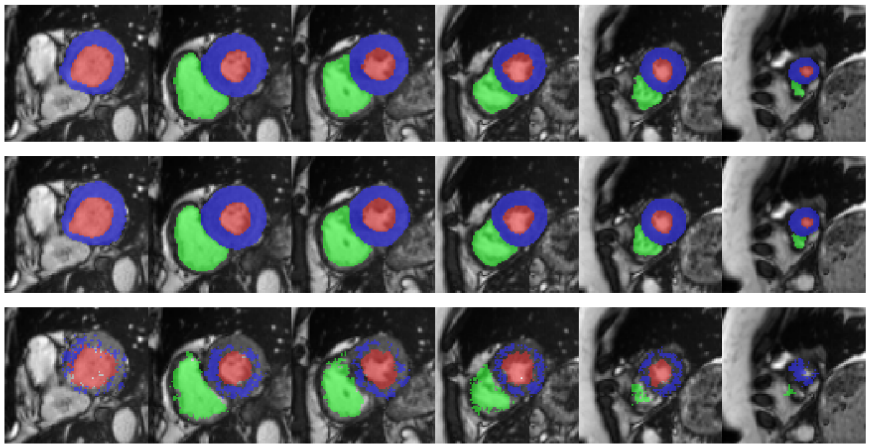


(c) Segmentation results for USegNet (top two rows) and DMR-USegNet (bottom two rows)

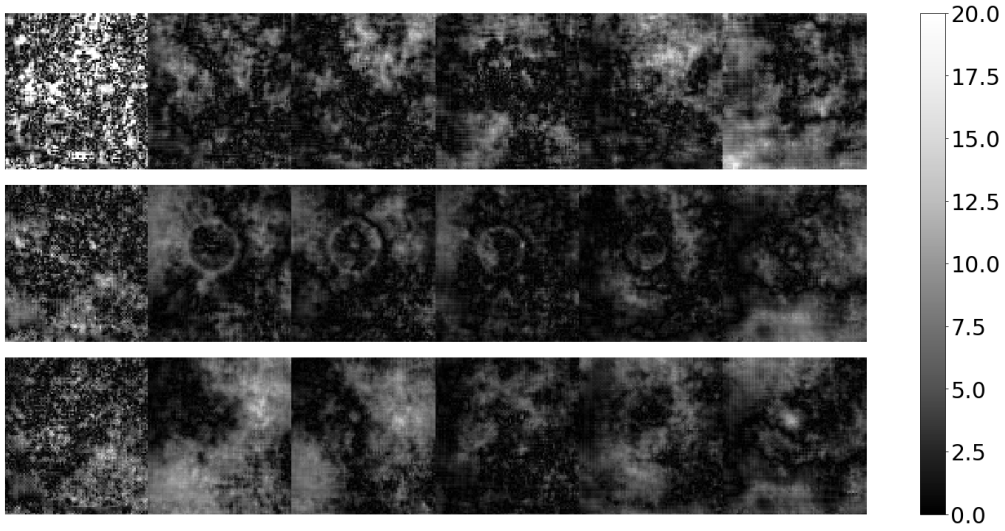


(d) Segmentation results for UNet (top two rows) and DMR-UNet (bottom two rows)

Figure 7.4: Ground-truth and automatic segmentation obtained from all trained models for a test patient. In each sub-figure, the segmentation obtained from the baseline and regularized model are overlaid onto the volume and shown in first and third rows, respectively; corresponding disagreement (in black) between the obtained segmentations and the ground-truth is shown in second and fourth rows, respectively.



(a) Input volume with: (top row) ground-truth segmentation overlaid, (middle row) segmentation obtained from the DMR-UNet model, and (bottom row) segmentation obtained after thresholding the predicted distance map at zero levelset.



(b) Absolute difference between the ground-truth and predicted distance maps. First, second, and third row show the error in RV, LV myocardium, and LV bloodpool, respectively.

Figure 7.5: Visualization of (a) the segmentation obtained by thresholding the predicted distance map and (b) absolute error between the ground-truth and predicted distance maps for all chambers. Shown is only a cropped region around the heart, the error in predicted distance map is higher for the regions farther from the heart.

Table 7.2: Evaluation of the average segmentation results on ACDC dataset for RV blood-pool, LV myocardium, and LV blood-pool (mean value reported), obtained from all networks against the provided reference segmentation. The statistical significance of the results for DM regularized model compared against the baseline model are represented by * and ** for p-values less than 0.05 and 0.005, respectively. Also shown are the clinical indices evaluated for each heart chamber. The best performing model for each metric has been **highlighted**. SN: SegNet, USN: USegNet, UNet: UNet.

(a) Evaluation of Average (across all heart chambers) Segmentation Results

	SN	DMR SN	USN	DMR USN	UNet	DMR UNet
	End Diastole (ED)					
Dice (%)	91.1	91.7**	91.5	92.0**	91.6	**92.2****
Jaccard (%)	84.0	85.1**	84.7	85.5**	85.0	**85.9****
MSD (mm)	0.55	0.53*	0.58	**0.52***	0.54	0.53*
HD (mm)	10.26	9.87	10.26	9.67	10.03	**9.52**
	End Systole (ES)					
Dice (%)	87.3	88.0*	87.7	88.7**	87.2	**88.8***
Jaccard (%)	78.1	79.3*	78.7	80.3**	78.3	**80.4***
MSD (mm)	0.92	0.85	0.92	0.84	1.08	**0.83**
HD (mm)	11.33	**10.31***	11.66	10.91	12.61	10.96*

(b) Evaluation of the Clinical Indices

	SN	DMR SN	USN	DMR USN	UNet	DMR UNet
	Correlation Coefficient					
LV EF	0.939	0.947	0.944	**0.970**	0.962	0.963
RV EF	0.874	0.871	0.866	**0.895**	0.856	0.870
Myo Mass	0.948	0.970	0.958	0.973	0.933	**0.978**
	Bias+LOA					
LV EF	1.00 (13.15)	0.31 (12.44)	0.58 (12.57)	-0.42 **(9.24)**	0.31 (10.41)	0.40 (10.40)
RV EF	1.04 (17.40)	1.77 (17.34)	0.85 (17.40)	0.38 **(15.42)**	0.09 (18.94)	0.29 (18.30)
Myo Mass	3.10 (32.94)	-0.43 (25.17)	0.35 (29.65)	0.21 (23.89)	2.85 (37.39)	0.80 **(21.75)**

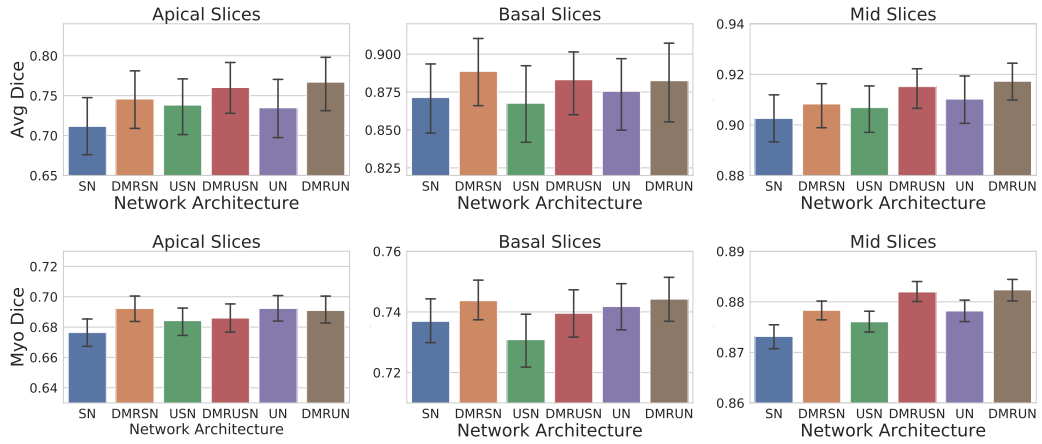


Figure 7.6: Mean and 95% bootstrap confidence interval for average Dice coefficient on apical, basal, and mid slices. Top: Average Dice coefficient for LV blood-pool, LV myocardium, and RV blood-pool segmentation on ACDC dataset (100 volumes). Bottom: Dice coefficient for myocardium segmentation on LVSC dataset (1050 volumes). SegNet: SN, DMR-SegNet: DMRSN, USegNet: USN, DMR-USegNet: DMRUSN, UNet: UN, DMR-UNet: DMRUN.

is of significance when constructing patient specific models of the heart for simulation purposes [41]. We postulate that the additional constraint imposed by a very high negative distance assigned to empty apical/basal slices prevents the network from over-segmenting these regions, hence, improving the regional dice overlap and effectively reducing the overall Hausdorff distance.

To study the effect of the distance map regularization across the five patient sub-groups, we plot the average Dice coefficient for each sub-group computed for all six models in **Fig. 7.7**. As expected, we observe the segmentation performance is better for the normal patients in comparison to the pathological cases. Furthermore, we observe consistent improvement in segmentation performance after the distance map regularization for all patient sub-groups.

We segmented the heart structures from 50 patients ACDC held-out testset and submitted to the challenge organizers. Majority voting prediction of ensemble of DMR-UNet models trained for five-fold cross-validation followed by a 3D connected component analysis yielded the final segmentation. **Table 7.3** shows the comparison

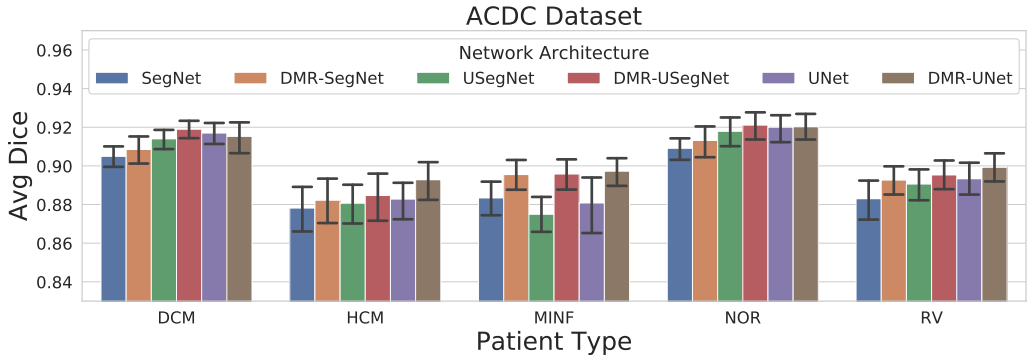


Figure 7.7: Mean and 95% bootstrap confidence interval of average Dice coefficient for segmentation results on ACDC dataset obtained from several architectures divided according to the five sub-groups: DCM — dilated cardiomyopathy, HCM — hypertrophic cardiomyopathy, MINF — previous myocardial infarction, NOR — normal subjects, and RV — abnormal right ventricle.

Table 7.3: Comparison of the segmentation results obtained from the DMR-UNet model against the top three ACDC challenge participants, evaluated on the held-out 50 patient challenge testset. The Dice metric, Hausdorff Distance (HD), and correlation of clinical indices for all three heart chambers is shown.

	End Diastole (ED)							EF	
	LV		RV		Myo			LV	RV
	Dice	HD	Dice	HD	Dice	HD	Corr	Corr	Corr
Baumgartner[42]	0.96	6.53	0.93	12.67	0.89	8.70	0.982	0.988	0.851
Khened[43]	0.96	8.13	0.94	13.99	0.89	9.84	**0.990**	0.989	0.858
Isensee[44]	**0.97**	7.38	**0.95**	10.12	**0.90**	8.72	0.989	**0.991**	**0.901**
DMR-UNet	0.96	**6.05**	0.94	**9.52**	0.89	**7.92**	0.989	0.989	0.851
	End Systole (ES)								
Baumgartner[42]	0.91	9.17	0.88	14.69	0.90	10.64	0.983		
Khened[43]	0.92	8.97	0.88	13.93	0.90	12.58	0.979		
Isensee[44]	**0.93**	**6.91**	**0.90**	**12.14**	**0.92**	8.67	0.985		
DMR-UNet	0.92	8.16	0.88	13.05	0.91	**8.39**	**0.987**		

of our segmentation results against the top three methods submitted to the challenge. Baumgartner *et al.* [42] tested several architectures and found that 2D U-Net with a cross-entropy loss performed the best. Khened *et al.* [43] used a 2D U-Net with dense blocks and an inception first layer to obtain the segmentation. Isensee *et al.*

ensembled 2D and 3D U-Net architectures trained with a Dice loss to obtain the best result in the challenge. Our 2D DMR-UNet model is able to perform as good or better than the other two 2D methods, however, the combination of 2D and 3D context has marginal improvement in the Dice overlap metric. Based on this observation, we believe the ensemble of 2D and 3D DMR-UNet model should be able to perform as good or better than [44], which is not the main objective of this work. Nonetheless, we can observe the constraint imposed by the DM regularization is successful in reducing the errors in apical/basal regions, manifested in the improved Hausdorff distance.

Table 7.4: Evaluation of the segmentation results on LVSC dataset for LV myocardium (mean values reported), obtained from all networks against the provided reference segmentation. The statistical significance of the results for DM regularized model compared against the baseline model are represented by * and ** for p-values less than 0.05 and 0.005, respectively. The best performing model for each metric has been highlighted. SN: SegNet, USN: USegNet, UNet: UNet.

	SN	DMR SN	USN	DMR USN	UNet	DMR UNet
	End Diastole (ED)					
Dice (%)	82.2	83.0*	82.5	83.2**	83.1	**83.6**
Jaccard (%)	70.0	71.1*	70.4	71.5**	71.3	**72.0**
MSD (mm)	0.78	0.74	0.79	0.72*	0.74	**0.70**
HD (mm)	13.20	13.14	13.67	13.12	12.98	**12.80**
Mass (Corr)	0.908	0.937	0.923	**0.938**	0.917	0.936
Mass(gram) (Bias+LOA)	2.56 (35.25)	-0.92 (29.52)	3.91 (32.48)	3.34 **(29.08)**	2.88 (33.75)	0.06 (29.92)
	End Systole (ES)					
Dice (%)	83.5	84.2	83.8	84.3*	84.3	**84.6**
Jaccard (%)	71.9	72.9	72.4	73.0*	73.0	**73.5**
MSD (mm)	0.81	0.77	0.77	0.78	**0.74**	0.75
HD (mm)	12.96	12.96	12.71*	13.71	13.08	**12.51**
Mass (Corr)	0.921	0.935	0.929	0.926	**0.939**	0.922
Mass(gram) (Bias+LOA)	5.48 (32.49)	1.96 (29.58)	5.04 (30.92)	5.49 (31.49)	5.18 **(28.79)**	2.56 (32.18)

Table 7.4 shows the segmentation performance evaluated on the LVSC dataset,

demonstrating superior performance of the DM regularized models over their baseline. Specifically, there is statistically significant improvement on the Dice and Jaccard metric for the ED phase. Furthermore, the correlation and LoA for the myocardial mass improves after network regularization. The improvement in performance is consistent across different heart regions as shown in **Fig. 7.6b**.

Table 7.5: Comparison of the LV myocardium segmentation results on the LVSC validation set against the consensus segmentation (**CS***) as described in [36]. The values for AU, AO, SCR, and INR are obtained from Table 2 in [36], CNR from Table 3 in [45], FCN from Table 3 in [7], and DFCN from Table 12 in [43]. Values are provided as mean (standard deviation), and in descending order by Jaccard index. SA/FA — Semi/Fully-Automatic

Method	SA/FA	Jaccard	Sensitivity	Specificity	PPV	NPV
AU [46]	SA	0.84 (0.17)	0.89 (0.13)	0.96 (0.06)	0.91 (0.13)	0.95 (0.06)
CNR [45]	SA	0.77 (0.11)	0.88 (0.09)	0.95 (0.04)	0.86 (0.11)	0.96 (0.02)
FCN [7]	FA	0.74 (0.13)	0.83 (0.12)	0.96 (0.03)	0.86 (0.10)	0.95 (0.03)
DFCN [43]	FA	0.74 (0.15)	0.84 (0.16)	0.96 (0.03)	0.87 (0.10)	0.95 (0.03)
DMR-UNet	FA	0.74 (0.16)	0.85 (0.16)	0.95 (0.03)	0.86 (0.10)	0.95 (0.03)
AO [47]	SA	0.74 (0.16)	0.88 (0.15)	0.91 (0.06)	0.82 (0.12)	0.94 (0.06)
SCR [48]	FA	0.69 (0.23)	0.74 (0.23)	0.96 (0.05)	0.87 (0.16)	0.89 (0.09)
INR [49]	FA	0.43 (0.10)	0.89 (0.17)	0.56 (0.15)	0.50 (0.10)	0.93 (0.09)

We segmented the myocardium from the LVSC held-out validation set of 100 patients. Majority voting prediction from ensemble of DMR-UNet models trained for five-fold cross-validation followed by a 3D connected-component analysis yielded the final segmentation. **Table 7.5** shows our segmentation results (computed per slice) compared against several other semi-/fully-automatic algorithms. Reported segmentation results are computed against the consensus segmentation (CS*) built

from multiple challenge submissions [36]. Segmentation results for the four challenge participants — AU [46], AO [47], SCR [48], and INR [49], and the details on segmentation evaluation metrics can be found in the challenge summary report [36]. The AU method [46] used the interactive guide-point modeling technique to fit a finite element cardiac model to the CMR data and required expert approval of all slices and all frames. This segmentation was provided as the reference segmentation to the challenge participants. The CNN regression (CNR) method [45] regressed the endo- and epi-cardium contours in polar coordinates, while manually eliminating the problematic slices beyond the apex and base of the heart, hence, obtaining a good segmentation result. The mean (std dev) of Jaccard coefficients computed for our DMR-UNet model in apical, mid, and basal slices are 0.66 (0.18), 0.77 (0.12), and 0.74 (0.17), respectively. Our DMR-UNet model has similar performance to competing fully-automatic segmentation algorithms based on the fully convolutional network (FCN) [7] and the densely connected FCN (DFCN) [43] architectures. The DFCN method involves a computationally expensive region of interest (ROI) identification based on a Fourier transform applied across the cardiac cycle, followed by the circular Hough transform; whereas our method requires minimal pre-processing.

Lastly, the segmentation performance on the LVSC dataset (**Table 7.5**) is significantly lower than ACDC dataset (**Table 7.3**) due to large variability and noise exhibited by the LVSC data as compared to the ACDC dataset.

7.3.2 Cross Dataset Evaluation (Transfer Learning)

To analyze the generalization ability of our proposed distance map regularized networks, we performed a cross-dataset segmentation evaluation. The networks trained on ACDC dataset for five-fold cross-validation were tested on the LVSC dataset, and vice versa; such that, the majority voting scheme produced the final per-pixel segmentation. We observe a significant boost in Dice coefficient of 5% to 12% for distance map regularized networks over their baseline models when trained on ACDC

Table 7.6: Cross-dataset segmentation evaluation for LV myocardium segmentation (mean values reported). The statistical significance of the results for DM regularized model compared against the baseline model are represented by * and ** for p-values less than 0.01 and 0.001, respectively. SN: SegNet, USN: USegNet, UNet: UNet.

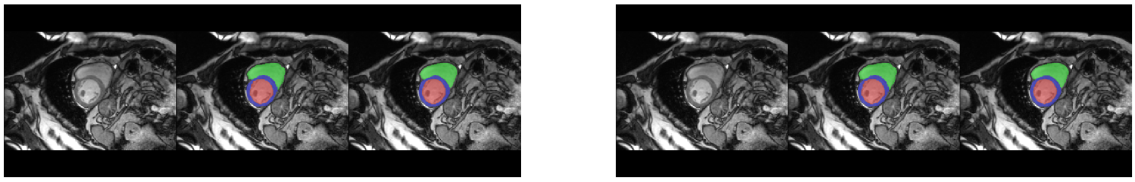
(a) Trained on ACDC and tested on LVSC (194 volumes)

	SN	DMR SN	USN	DMR USN	UNet	DMR UNet
	End Diastole (ED)					
Dice(%)	70.4	73.3**	68.3	76.6**	72.3	**76.7****
Jaccard(%)	55.6	58.9**	53.6	62.9**	58.0	**63.1****
MSD(mm)	2.68	2.07**	3.33	1.80**	2.46	**1.80****
HD(mm)	25.01	22.44**	26.93	20.33**	24.61	**20.16****
	End Systole (ES)					
Dice(%)	68.0	71.9**	65.5	74.9**	69.7	**76.4****
Jaccard(%)	53.3	58.1**	50.8	61.5**	55.5	**63.1****
MSD(mm)	3.56	2.93**	4.19	2.58**	3.49	**2.35****
HD(mm)	25.96	22.62**	27.37	21.67**	25.68	**20.98****

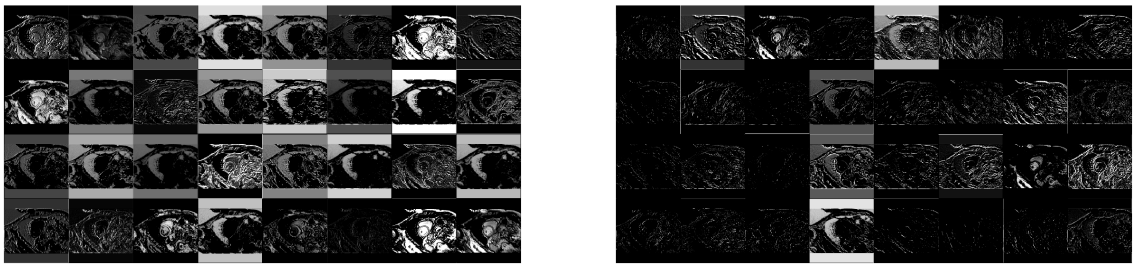
(b) Trained on LVSC and tested on ACDC (200 volumes)

	SN	DMR SN	USN	DMR USN	UNet	DMR UNet
	End Diastole (ED)					
Dice(%)	69.5	78.4**	62.5	80.1**	62.1	**80.2****
Jaccard(%)	56.5	66.3**	49.3	68.2**	49.3	**68.5****
MSD(mm)	4.92	1.77**	6.75	**1.30****	6.29	1.59**
HD(mm)	26.04	17.06**	29.08	**13.93****	29.50	14.16**
	End Systole (ES)					
Dice(%)	57.7	77.6**	51.9	**79.3****	50.3	79.1**
Jaccard(%)	45.4	65.3**	40.1	**67.3****	38.8	67.1**
MSD(mm)	9.59	2.53**	13.27	**2.35****	10.97	2.52**
HD(mm)	35.13	19.25**	39.60	**18.77****	37.44	19.58**

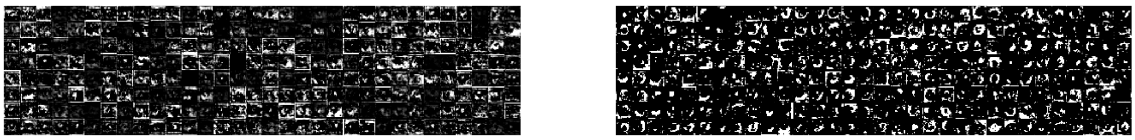
and tested on LVSC dataset (194 ED and ES volumes), as shown in **Table 7.6a**. Similarly, the distance map regularized models significantly outperform the baseline models by 23% to 42% improvement in Dice coefficient, when trained on LVSC and tested on ACDC dataset (200 ED and ES volumes), as shown in **Table 7.6b**. The



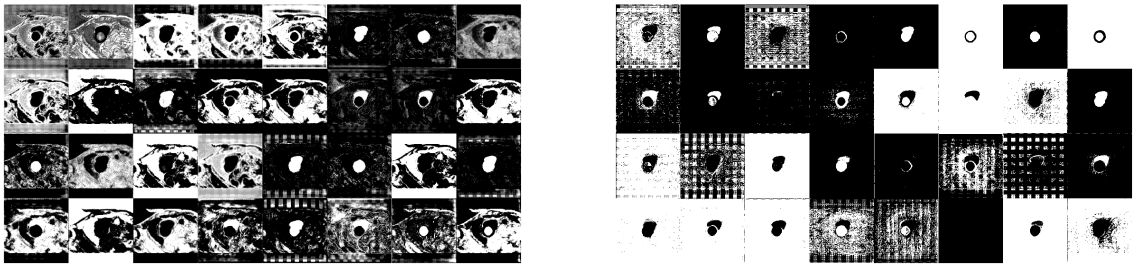
(a) From left to right: input image, ground-truth, and automatic segmentation overlay.



(b) 32 feature maps before first max-pooling operation.



(c) 256 feature maps from the bottle-neck layer.



(d) 32 feature maps before the final 1×1 convolution.

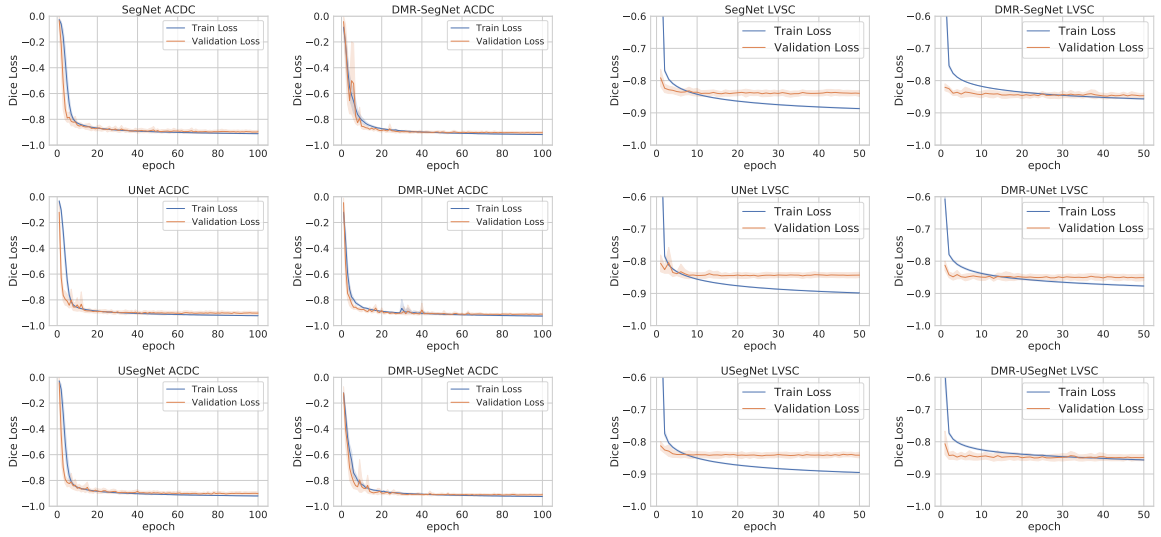
Figure 7.8: Feature maps visualized for the UNet (left column) and DMR-UNet (right column) model. We can observe the UNet model preserves the intensity information and propagates it throughout the network, hence, is more sensitive to the dataset-specific intensity distribution. On the other hand, the DMR-UNet model focuses more on the edges and other discriminative features, producing sparse feature maps, while ignoring dataset-specific intensity distribution. However, the results obtained for intra-dataset segmentation (shown here for ACDC dataset) is similar for both models, whereas, there is a significant improvement in cross-dataset segmentation after distance map regularization.

improvement in generalization performance for the regularization networks trained on LVSC dataset is higher, likely due to the availability of large number of heterogeneous training examples. Similar improvement can be observed in the MSD and HD metric. We want to emphasize that our networks are trained separately on each dataset and are completely unaware of the new data distribution, unlike a typical domain adaptation [50] setting. Nonetheless, the distance map regularized networks are able to generalize better to a new dataset compared to the baseline models.

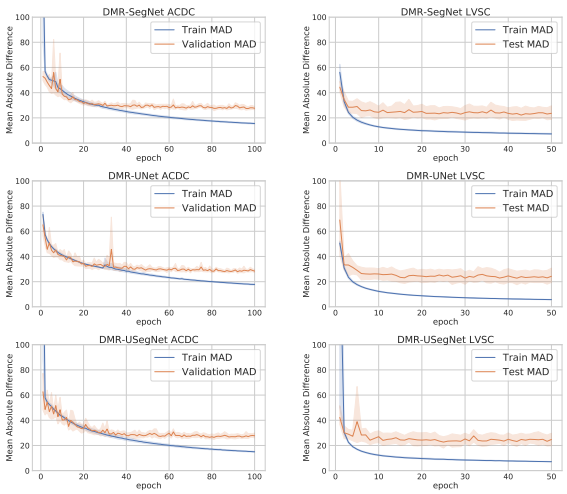
We further analyzed the feature maps across different layers of the baseline and distance map regularized networks, shown in **Fig. 7.8**. We can observe the baseline models preserve the intensity information and propagate it throughout the network, hence, they are more sensitive to the dataset-specific intensity distribution. On the other hand, the multi-task regularized networks focus more on the edges and other discriminative features, producing sparse feature maps, while ignoring dataset-specific intensity distribution. Moreover, from the feature maps at the decoding layers, we observe a clear delineation of several cardiac structures in the regularized network, while those for the baseline models are less discriminative, and contain information about all structures present in the image. Hence, we verify that multi-task learning-based distance map regularization helps the network learn generalizable features important for the segmentation task, demonstrated by their excellent transfer learning capabilities. Furthermore, the network learning curves in **Fig. 7.9** shows small generalization gap for the distance map regularized models, demonstrating their robustness against overfitting.

7.3.3 Comparison with Models Trained on Different Loss Functions

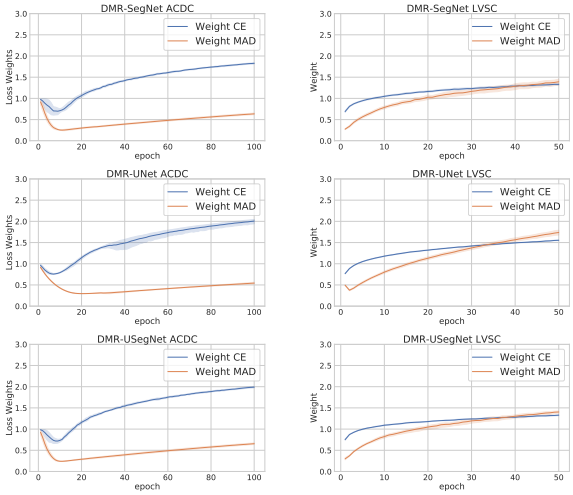
Several modifications to the categorical cross-entropy loss have been proposed to improve segmentation results. A popular variant is weighted categorical cross-entropy,



(a) Training and validation Dice loss for segmentation task. ACDC (left two columns) and LVSC (right two columns).



(b) Training and validation mean absolute difference error for distance map regression task. ACDC (left) and LVSC (right).



(c) Log Weights learned for cross-entropy and mean absolute difference losses. ACDC (left) and LVSC (right).

Figure 7.9: Mean and 95% bootstrap confidence interval for training and validation losses (a and b), and the learned weights for cross-entropy and mean absolute difference losses (c), on ACDC and LVSC dataset across five-fold cross-validation. Since the cross-entropy loss is harder to interpret, we plot the corresponding dice loss computed during training and validation. We can observe lower difference between the training and validation dice loss for the distance map regularized models, demonstrating their ability to prevent overfitting.

where the loss contribution of each class is multiplied by a weight proportional to the inverse frequency of that class in the training set. We compute the weights as $w_c = \frac{\sum_c N_c}{N_c}$, where $c = \{0, 1, ..., C-1\}$ for C classes and N_c is the number of pixels of class c in the training set. The weights w_c are then normalized by their median value during weighted categorical cross-entropy loss computation.

Similarly, Ronneberger *et al.* [32] proposed a spatial weighting scheme, where the pixels closer to segmentation boundaries were assigned higher weights, to incentify the network to produce better segmentation results by avoiding misclassification of boundary pixels. The spatial weight map is computed as:

$$w(\boldsymbol{x}) = w_c(\boldsymbol{x}) + w_0 \cdot \exp\left(-\frac{(d_1(\boldsymbol{x}) + d_2(\boldsymbol{x}))^2}{2\sigma^2}\right) \tag{7.3}$$

where, w_c is the weight map to balance the class frequencies, d_1 and d_2 are the distances to the border of nearest and second nearest object classes. In our experiments we set $w_0 = 1.0$ and $\sigma = 5.0$.

Table 7.7: Evaluation of the segmentation results on ACDC dataset for RV blood-pool, LV myocardium, and LV blood-pool (mean values reported), obtained from different weighting schemes of the categorical cross-entropy loss function. UNet model trained with cross-entropy loss: without any weighting, class frequency weighting, spatial weighting (with uniform class weight), and spatial with class frequency weighting, compared against the proposed DMR-UNet model.

	None	Class	Spatial	Spatial w/Class	DMR UNet
	End Diastole (ED)				
Dice (%)	91.6	89.2	91.7	91.8	**92.2**
Jaccard (%)	85.0	81.2	85.1	85.2	**85.9**
MSD (mm)	0.54	0.71	0.53	**0.51**	0.53
HD (mm)	10.03	10.48	10.06	9.99	**9.52**
	End Systole (ES)				
Dice (%)	87.2	84.7	88.1	87.8	**88.8**
Jaccard (%)	78.3	74.6	79.3	79.0	**80.4**
MSD (mm)	1.08	1.25	0.89	0.95	**0.83**
HD (mm)	12.61	12.60	11.31	12.16	**10.96**

Table 7.7 summarizes the segmentation results obtained on ACDC dataset for UNet models trained on cross-entropy loss with several weighting schemes against the proposed DMR-UNet model. Although class frequency weighted training has been found to improve the performance of a model on limited availability of examples for some classes, in our segmentation problem, we have a large number of examples (pixels) for each class. Furthermore, since the number of background pixels is very high compared to other classes, the weight assigned to background pixels is extremely low, hence discouraging the model to segment ambiguous pixels as a background class, resulting in degraded segmentation performance, as shown in **Table 7.7**. Moreover, while the spatial weighting scheme only provides a sight improvement over the unweighted cross-entropy loss, there is good improvement in the distance metric due to the emphasis on the boundary pixels. Nevertheless, **Table 7.7** clearly shows that our proposed DMR-UNet model significantly outperforms all other weighting schemes, yielding highest overlap and lowest distance metrics.

7.4 Discussion

We performed an extensive study on the effects of hyper-parameters on the performance of the proposed regularization framework. Here we summarize the effects of the learned vs. fixed task weighting, and various choices of the distance map threshold. Furthermore, we analyzed the distribution of network weights before and after regularization.

Task Weighting: At first, we initialized the weights for the cross-entropy and MAD loss equally to 1.0. However, the learned weights for the cross-entropy and MAD loss were around 0.01 and 17, and 0.02 and 13 for ACDC and LVSC dataset, respectively, for the best performing models on the validation set.

To determine the effect of learned task weighting scheme presented in section **7.2.3**, we analyzed the average Dice coefficient of the test set segmentation results

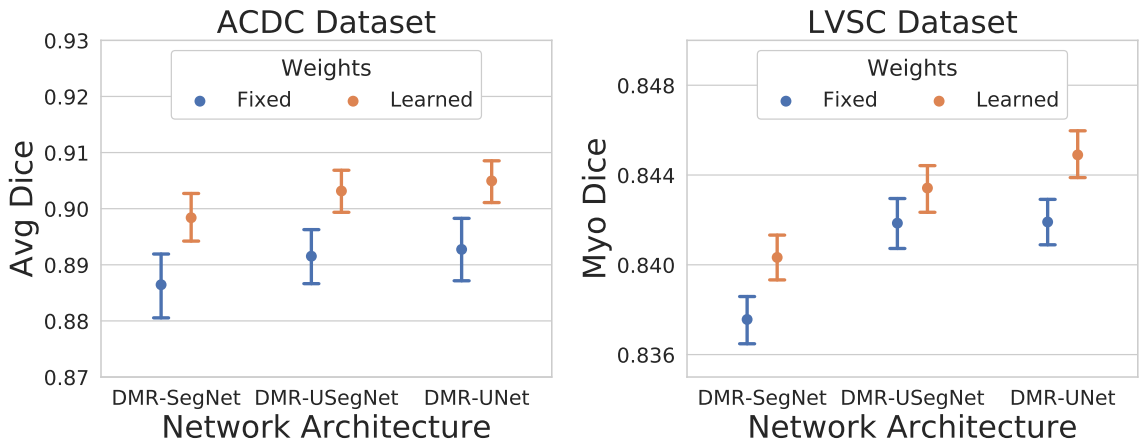


Figure 7.10: Mean and 95% bootstrap confidence interval of average Dice coefficient for Learned vs Fixed equal weighting. Learned task weighting statistically significantly improves the segmentation performance.

for both ACDC (100 volumes) and LVSC (1050 volumes across the full cardiac cycle) datasets with fixed versus learned weighting. From **Fig. 7.10**, we can observe a significant improvement in average Dice coefficient (based on the 95% bootstrap confidence intervals) with learned weights compared to fixed (equal) weighting. Since the scales of the two losses are different, the equal weighting scheme emphasizes the distance map regression task more than it should, hence deteriorating the segmentation performance. On the other hand, the learned task weighting scheme is able to automatically weigh the two losses, bringing them to a similar scale, such that the two tasks are given equal importance, ultimately improving the segmentation performance.

Effect of Distance Map Threshold: We selected three extreme values for the distance map threshold: 5, 60, and 250 *pixels*. The network weights for cross-entropy and MAD loss were equally initialized to $(1, 1)$ and trained with automatically learned task weighting for a fixed number of epochs. The average Dice coefficient on the test-set obtained from the best performing models on the validation-set across five-fold cross-validation is summarized in **Fig. 7.11**. We observe similar performance for different threshold values, demonstrating the low sensitivity of the proposed method

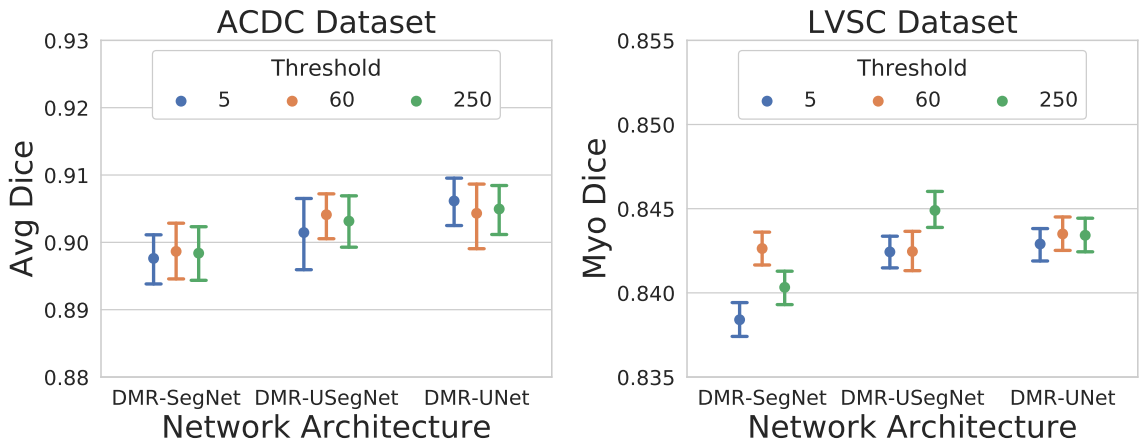
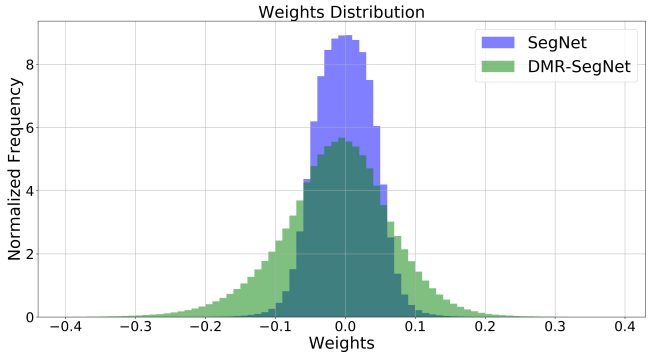


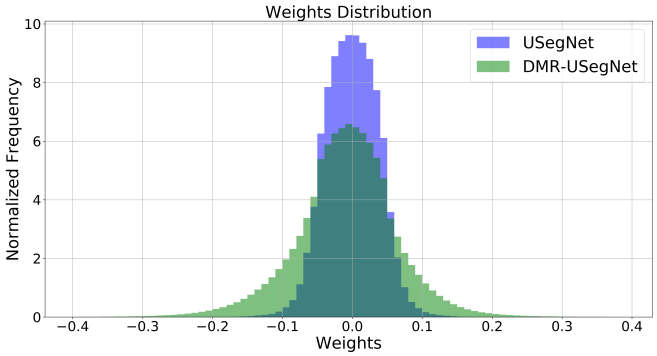
Figure 7.11: Mean and 95% bootstrap confidence interval of average Dice coefficient for a range of distance map thresholds.

to the distance map threshold. Hence, we decided to use a very high threshold of 250 pixels, which is almost equivalent to regressing the full distance map and neglecting this hyper-parameter.

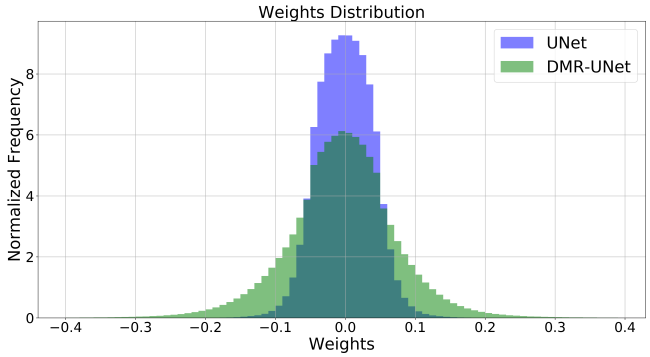
Network Weight Distribution: We also analyzed the weight distribution of the network before and after distance map regularization, as shown in **Fig. 7.12**. We observe the number of non-zero weights increase after the distance map regularization, hence, better utilizing the network capacity. A similar flattening of network weight histogram has been reported for the dropout regularization and Bayesian neural networks [51], both reducing the overfitting and hence improving generalization. Specifically, the network weights are randomly dropped during dropout, forcing the network to use the remaining weights to identify the patterns in data (spreading the weight histogram), hence creating an ensemble effect with reduced over-fitting and improved generalization. We observe a similar pattern in the weight distribution after the distance map regularization.



(a) Weights distribution for SegNet and DMR-SegNet models.



(a) Weights distribution for USegNet and DMR-USegNet models.



(a) Weights distribution for UNet and DMR-UNet models.

Figure 7.12: Weights distribution before and after distance map regularization for models trained across five-fold cross-validation. We can observe the number of non-zero weights increases after the distance map regularization, hence, better utilizing the network capacity.

7.5 Conclusion

In this work we proposed and implemented a multi-task learning-based regularization method for fully convolutional networks for semantic image segmentation and demonstrated its benefits in the context of cardiac MR image segmentation. To implement the proposed method, we appended a decoder network at the bottleneck layer of existing FCN architectures to perform an auxiliary task of distance map prediction, which is removed after training.

We automatically learned the weighting of the tasks based on their uncertainty. As the distance map contains robust information regarding the shape, location, and boundary of the object to be segmented, it facilitates the FCN encoder to learn robust global features important for the segmentation task.

Our experiments verify that introducing the distance map regularization improves the segmentation performance of three FCN architectures for both binary and multi-class segmentation across two publicly available cardiac cine MRI datasets featuring significant patient anatomy and image variability. Specifically, we observed consistent improvement in segmentation performance in the challenging apical and basal slices in response to the soft-constraints imposed by the distance map regularization. We also showed consistent segmentation improvement on all five patient pathology in the ACDC dataset. Furthermore, these improvements were also reflected on the computed clinical indices important for the diagnosis of various heart conditions. Lastly, we demonstrated the proposed regularization significantly improved the generalization ability of the networks on cross-dataset segmentation (transfer learning), without being aware of the new data distribution, with 5% to 42% improvement in average Dice coefficient over the baseline FCN architectures.

Acknowledgment

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R35GM128877 and by the Office of Advanced Cyber infrastructure of the National Science Foundation under Award No. 1808530. Ziv Yaniv's work was supported by the Intramural Research Program of the U.S. National Institutes of Health, National Library of Medicine.

Bibliography

[1] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 05 2015.

[5] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 06 2017.

[6] Geert Litjens et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88, 2017.

[7] Phi Vu Tran. A fully convolutional neural network for cardiac segmentation in short-axis MRI. *CoRR*, abs/1604.00494, 2016.

[8] Rudra P. K. Poudel, Pablo Lamata, and Giovanni Montana. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 83–94, Cham, 2017. Springer International Publishing.

[9] M.R. Avendi, Arash Kheradvar, and Hamid Jafarkhani. 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, 2016.

[10] O. Oktay et al. Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, Feb 2018.

[11] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and José García Rodríguez. A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857, 2017.

[12] Anders Krogh and John A. Hertz. A simple weight decay can improve generalization. NIPS'91, pages 950–957, San Francisco, CA, USA, 1991.

[13] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. ICML '08, pages 1096–1103, New York, NY, USA, 2008. ACM.

[14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ICML'15, pages 448–456. JMLR.org, 2015.

[16] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. ICLR'15, 2015.

[17] Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, Jul 1997.

[18] Peter L. Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *J. Mach. Learn. Res.*, 3:463–482, March 2003.

[19] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

[20] M. Teichmann, M. Weber, M. Zllner, R. Cipolla, and R. Urtasun. MultiNet: Real-time joint semantic reasoning for autonomous driving. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1013–1020, June 2018.

[21] Jonas Uhrig, Marius Cordts, Uwe Franke, and Thomas Brox. Pixel-level encoding and depth layering for instance-level semantic labeling. In *GCPR*, 2016.

[22] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[23] Pim Moeskops et al. Deep learning for multi-task medical image segmentation in multiple modalities. In *MICCAI*, 2016.

[24] V. V. Valindria, N. Pawlowski, M. Rajchl, I. Lavdas, E. O. Aboagye, A. G. Rockall, D. Rueckert, and B. Glocker. Multi-modal learning from unpaired images: Application to multi-organ segmentation in CT and MRI. In *IEEE WACV*, pages 547–556, March 2018.

[25] Wufeng Xue, Gary Brahm, Sachin Pandey, Stephanie Leung, and Shuo Li. Full left ventricle quantification via deep multitask relationships learning. *Medical Image Analysis*, 43:54 – 65, 2018.

[26] Shusil Dangi, Ziv Yaniv, and Cristian A. Linte. Left ventricle segmentation and quantification from cardiac cine MR images via multi-task learning. In *STACOM*, pages 21–31, 2019.

[27] M. Bai and R. Urtasun. Deep watershed transform for instance segmentation. In *IEEE CVPR*, pages 2858–2866, July 2017.

[28] Zeeshan Hayder, Xuming He, and Mathieu Salzmann. Boundary-aware instance segmentation. In *IEEE CVPR*, July 2017.

[29] Benjamin Bischke, Patrick Helber, Joachim Folz, Damian Borth, and Andreas Dengel. Multi-task learning for segmentation of building footprints with deep neural networks. *CoRR*, abs/1709.05932, 2017.

[30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[31] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE CVPR*, June 2016.

[34] Gunilla Borgefors. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing*, 34(3):344 – 371, 1986.

[35] Carissa G. Fonseca, Michael Backhaus, David A. Bluemke, Randall D. Britten, Jae Do Chung, Brett R. Cowan, Ivo D. Dinov, J. Paul Finn, Peter J. Hunter, Alan H. Kadish, Daniel C. Lee, Joao A. C. Lima, Pau MedranoGracia, Kalyanam Shivkumar, Avan Suinesiaputra, Wenchao Tao, and Alistair A. Young. The

cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*, 27(16):2288–2295, 2011.

[36] Avan Suinesiaputra, Brett R. Cowan, Ahmed O. Al-Agamy, Mustafa A. Elattar, Nicholas Ayache, Ahmed S. Fahmy, Ayman M. Khalifa, Pau Medrano-Gracia, Marie-Pierre Jolly, Alan H. Kadish, Daniel C. Lee, Jn Margeta, Simon K. Warfield, and Alistair A. Young. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac {MR} images. *Medical Image Analysis*, 18(1):50 – 62, 2014.

[37] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M. Roh, X. Pennec, M. Sermesant, F. Isensee, P. Jger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Igum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P. Jodoin. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, Nov 2018.

[38] Ziv Yaniv, Bradley C. Lowekamp, Hans J. Johnson, and Richard Beare. SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research. *Journal of Digital Imaging*, 31(3):290–303, 2018.

[39] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[40] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent.

[41] T. Peters, C. Linte, Z. Yaniv, and J. Williams. Mixed and augmented reality in medicine. chapter Chapter 16. Augmented and Virtual Visualization for Image-Guided Cardiac Therapeutics, pages 231–250. CRC Press, 2018.

[42] Christian F. Baumgartner et al. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In *STACOM*, pages 111–119, 2018.

[43] Mahendra Khened, Varghese Alex, and Ganapathy Krishnamurthi. Densely connected fully convolutional network for short-axis cardiac cine MR image segmentation and heart diagnosis using random forest. In *STACOM*, pages 140–151, 2018.

[44] Isensee et al. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In *STACOM*, pages 120–129, 2018.

[45] Li Kuo Tan et al. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Medical Image Analysis*, 39, 2017.

[46] Bo Li, Yingmin Liu, Christopher J. Occleshaw, Brett R. Cowan, and Alistair A. Young. In-line automated tracking for ventricular function with magnetic resonance imaging. *JACC: Cardiovascular Imaging*, 3(8):860 – 866, 2010.

[47] Ahmed S. Fahmy, Ahmed O. Al-Agamy, and Ayman Khalifa. Myocardial segmentation using contour-constrained optical flow tracking. In *STACOM*, pages 120–128, 2012.

[48] Marie-Pierre Jolly et al. Automatic segmentation of the myocardium in cine MR images using deformable registration. In *STACOM*, pages 98–108, 2012.

[49] Ján Margeta et al. Layered spatio-temporal forests for left ventricle segmentation from 4D cardiac MRI data. In *STACOM*, pages 109–119, 2012.

[50] Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135 – 153, 2018.

[51] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. ICML'15, pages 1613–1622. JMLR.org, 2015.

Chapter 8

Discussion, Conclusion, and Future Work

This chapter revisits the challenges posed in the introduction chapter associated with the segmentation of cardiac ultrasound and MRI images and demonstrate how they have been overcome by the methods proposed in this dissertation. Future directions essential for the field is discussed.

8.1 The Big Picture

This dissertation is centered around one of the oldest topics in the computer vision: image segmentation. Specifically, we focus on the segmentation of the heart from ultrasound and MR images to extract clinically relevant diagnostic information. Although several segmentation approaches have been proposed to date, as summarized in **Chapter 1**, segmentation is still an open problem. Furthermore, specific image segmentation methods for different applications (e.g. natural, hyper-spectral, medical images), different modalities (X-ray, Ultrasound, MRI, CT), and different organs (e.g. brain, liver, heart etc.) are tailored specifically for that particular domain to make best use of the available prior knowledge.

The segmentation techniques explored in this dissertation rely on graph-cuts, atlas based, and convolutional neural network based methods. These methods use the global image context with incorporated prior information for segmentation. The segmentation task formulated as an energy minimization problem in a graph can be solved efficiently and quickly using the graph-cut technique [1], within a known factor of the global minimum, providing highly accurate labeling. Hence, most effort in the first half of the thesis focuses on defining the graph energy more accurately to improve the segmentation performance. In the second half, we exploit the performance of the convolutional neural networks to further improve the segmentation of heart chambers, and subsequent clinical indices estimation.

It is very important to tailor the segmentation algorithm towards the imaging modality and the organ under consideration. Hence, to segment the LV from the tri-plane TEE image sequence, we rely on the local phase-based filtering technique [2] and obtain the rough initial estimate of the LV blood-pool. Using this initial estimate, the intensity likelihood (modeled as a Gaussian distribution) for the blood-pool, myocardium, and background can be computed from the image and hence encoded into the graph energy. The minimum energy configuration obtained from the graph-cut labeling yields the final segmentation. Since the LV boundaries are more prominent in the end-diastole phase, we segment the LV from this frame and propagate the segmentation to other cardiac phases via biomechanics-based non-rigid registration algorithm [3]. This method produces better segmentation results compared to segmenting individual frame, as presented in **Chapter 2**.

Although the local neighborhood information is inherently built in the graph formulation, additional prior information of the heart geometry can be encoded into the framework. Some papers [4–6] have explored the use of shape prior information into the graph-cut framework, however, they require the shape prior to be initialized manually. Here we automate the shape prior initialization using a image registration technique. We register training volumes to a reference patient and average the

intensities and labels to obtain the average intensity and probabilistic atlas for the structure of interest, respectively. The average intensity atlas is registered to a test patient to transfer the probabilistic label, which is incorporated into the graph energy to encode the prior shape information. However, since there is a large variability in shape and size of the heart, and image registration produces poor results if not initialized properly, we iteratively refine the segmentation by updating the prior shape as well as the class intensity distributions based on the latest graph-cut segmentation result. Upon convergence, we obtain a robust automatic segmentation of the heart chambers.

We apply the developed probabilistic atlas based iterative graph-cut method for the slice-wise 2D segmentation of the LV (**Chapter 3**) and RV (**Chapter 4**) from two different open-source short-axis cardiac cine MRI datasets [7, 8], with the provided manual labeling for the LV and RV, respectively. However, due to the slice-misalignment in cine MR images, we were not able to leverage the whole 3D context for segmentation. Hence, we train a convolutional neural network to predict the LV center from each 2D short-axis slice, such that, a coherent 3D test volume can be generated by aligning the predicted LV center for each slice in a straight line along the long-axis. This approach allowed us to perform a full 3D segmentation using the developed probabilistic atlas based iterative graph-cut method as presented in **Chapter 5**.

We leverage and further enhance the power of convolutional neural network to accurately segment the heart chambers from cardiac cine MRI datasets in final two chapters. In **Chapter 6**, to compare and contrast segmentation-based and direct-estimation based clinical indices estimation, we train a multi-task learning based fully convolutional network to segment the myocardium and regress its area, simultaneously. The network also predicts the uncertainty of each task. The reciprocal of the uncertainties are used to weigh the two tasks, such that the task with higher uncertainty is weighted less and vice versa. Our evaluation on a 4D cardiac cine MRI

dataset [9] demonstrates that multi-task learning yields better segmentation of the myocardium. Furthermore, we show that segmentation-based estimation of clinical index is better than that obtained from direct-estimation.

In **Chapter 7**, we apply the multi-task uncertainty based weighting scheme to train a fully convolutional network to segment the LV blood-pool, RV blood-pool, and LV myocardium, while simultaneously regressing their corresponding distance maps. The distance map encodes the distance of each pixel from its closest segmented region, hence providing a rich shape and location information of the object to be segmented. We append a distance map regularization block to the bottleneck layer of a fully convolutional network to impose a weak shape and location constraint into the network. The regularizer block is removed after training to reduce the computational cost. We evaluate the proposed method on two open-source 4D cardiac cine MRI datasets [9, 10] and show that it improves the segmentation performance on challenging apical and basal slices with consistent improvement across healthy and pathological cases for both intra- and cross-dataset segmentation of the heart chambers.

8.2 Summary and Contributions

- We proposed and evaluated a LV segmentation method from multi-plane TEE image sequence and performed accurate computation of LV contractile functions based on the 3D reconstruction, as described in **Chapter 2**. Since the traditional intensity based edge-detection methods are highly prone to the speckle noise and low signal-to-noise ratio in ultrasound images, we detect the LV boundaries based on the local-phase asymmetry, which is theoretically invariant to the image contrast. We overcome the challenge of tissue inhomogeneity in ultrasound images by modeling the intensities of the blood-pool, myocardium, and background by a Gaussian distribution based on the image intensities. Similarly, using the smoothness constraint encoded in the graph neighborhood struc-

ture, we are able to overcome the challenges associated with missing anatomical boundaries due to signal dropout and shadow artifacts. Furthermore, to tackle the rapid motion of the LV chamber in the cardiac cycle, we employ a biomechanics-based non-rigid image registration method to propagate the automatic segmentation, obtained from the reliable end-diastole phase, throughout the cardiac cycle.

- We proposed and evaluated an automatic segmentation method using the probabilistic atlas prior into the iterative graph-cut framework. The use of additional prior information, based on the probabilistic atlas, improved the graph-cut segmentation performance. We modeled the intensities of the blood-pool, myocardium, and background classes by a Gaussian mixture model (GMM) to tackle the problem of non-standardized intensity values in MRI and the intensity fuzziness, due to the partial volume effect and blood-flow. As the LV shape is fairly close to the ring shape across the apex-base axis, we were able to obtain a good segmentation results for the LV blood-pool and myocardium, as discussed in **Chapter 3**. However, the segmentation of the RV is more challenging due to its crescent shape with high variability across the apex-base axis, which we were able to segment accurately (as detailed in **Chapter 4**) using the same framework, demonstrating its robustness and versatility.

- Due to the slice-wise acquisition of the cine MR images, the patient motion and breathing introduces slice misalignment. Hence, most of the segmentation methods, including the ones in **Chapter 3** and **4**, have been developed for 2D slice-wise segmentation. We proposed a new method for slice misalignment correction by aligning the LV centers in a straight line along the long-axis. The proposed method assumes that short-axis cine MR acquisitions are perpendicular to the LV long-axis, which is a reasonable assumption, provided the radiologist followed the acquisition protocol correctly. Thus, in **Chapter 5** we

trained a convolutional neural network (CNN) to predict the LV centers for each cine MR slice and generated a coherent 3D volume by correcting the slice misalignment. Furthermore, we extended the probabilistic atlas prior based iterative graph-cut framework for 3D segmentation, to exploit the 3D context and improve the segmentation result.

- We leverage the hierarchical representation power of convolutional neural networks to extract multi-scale features important for the segmentation task from cardiac cine MR images. A fully convolutional network trained on a large set of MR images of healthy and pathological cases, with some artificial data augmentation, has shown good promise in overcoming the challenges of intensity fuzziness, indistinct boundaries, and variability of heart chambers across the patients, leading to good segmentation results from cardiac MR images. In **Chapter 7**, we proposed a multi-task learning framework to perform segmentation and corresponding distance map regression, simultaneously, further improving the robustness and segmentation performance of a fully convolutional network. We show the proposed method is able to improve the segmentation performance on the challenging apical and basal slices, with consistent improvement across healthy and pathological cases. Furthermore, we demonstrate the robustness of the proposed method to extract dataset agnostic features important for segmentation, with significantly improved performance when trained on one cardiac cine MRI dataset and tested on a different dataset.

- We studied the efficacy of the segmentation-based clinical indices estimation in comparison to the direct-estimation method. In **Chapter 6**, we trained a multi-task network to segment the myocardium and regress its area, simultaneously. We demonstrated that segmentation-based area estimation is significantly better than that obtained from direct regression. Furthermore, we advocate the segmentation-based method is more interpretable, and easier to troubleshoot in

case of erroneous results, in comparison to direct-estimation, as the radiologists can visually inspect the segmentation results to infer the reliability of the computed clinical indices. However, it should be noted that it is easier to obtain the reference clinical index for direct-estimation compared to the reference per-pixel segmentation required for segmentation-based method.

- The proposed segmentation techniques for cardiac ultrasound and cine MRI images help obtain a better estimate of the cardiac contractile functions such as blood-pool volume, ejection fraction, and myocardial mass, for the cardiac health diagnosis in a clinic, making the job of a cardiologist faster, easier, and more reproducible. It also enables the production of highly accurate models for LV and RV chambers for personalized pre-operative planning as well as intra-operative guidance.

8.3 Future Directions

Although a lot of research has been conducted to improve the segmentation of heart chambers from cardiac cine MR images, there still exists challenges on correctly segmenting the apical and basal slices, as well as, the segmentation of RV blood-pool and LV myocardium, due to their large variability across patients [10]. Despite the mis-alignment of slices in cine MR images, the temporal information on each slice can help correctly identify and segment the challenging apical and basal slices, indicating a good future direction.

The field can also benefit from accurate and robust algorithms to correct for the slice-misalignment, incorporating both the short- and long-axis MR acquisition information. As such, an open-source cardiac MR image database with true 3D MR image acquisitions would enable researchers to develop machine learning algorithms to correct for slice mis-alignments in cine MR images. Slice-misalignment correction would enable incorporating 3D shape prior information of the heart chambers into the

segmentation framework to overcome the challenges presented by the large variability of heart chambers across the patients.

As can be observed from the progression of this thesis, the convolutional neural network based methods are able to produce significantly better segmentation results compared to the traditional atlas and graph-cut based methods, hence the CNNs deserve more investigation in the future. The convolutional neural network based methods should be equipped with shape prior information of heart-chambers to overcome the challenges of generalization across heterogeneous dataset acquired from different sites. Additionally, a smart combination of convolutional neural networks with traditional segmentation approaches (atlas, level-sets, and graph-cuts), exploiting the benefits of each, could further improve the cardiac MRI segmentation results.

Specifically, in the context of convolutional neural networks, further exploration of multi-task learning could be a potential future direction. One can segment a slice and predict the slice location, simultaneously, to improve the segmentation performance on the apical/basal slices. It would also be interesting to explore other auxiliary tasks that can assist to obtain better segmentation results. Similarly, it will be crucial to improve the robustness of these methods, as such, one can explore data augmentation strategies to better simulate the variability of heart chambers across the patients. Another potential direction is to learn an adversarial loss during network training to better focus on the challenging regions of the images, hence improving the segmentation performance. Similarly, one can employ generative models (variational auto-encoders [11] or Generative Adversarial Networks [12]) to capture the distribution of heart geometry, and hence impose shape constraints into the CNN framework. Finally, it will be critical to quantify the uncertainty [13] of model predictions, to better assist the clinical decision making, suggesting Bayesian neural networks as a good future direction for research.

Lastly, a large dataset with manual segmentation by multiple experts would enable the comparison of semi-/fully-automatic segmentation algorithms with respect

to the inter-observer variability to reliably measure the effectiveness of the developed segmentation algorithm vis-a-vis the gold standard annotations used routinely in the clinical setting.

Bibliography

[1] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[2] M. Mulet-Parada and J. A. Noble. 2D + T acoustic boundary detection in echocardiography. *Med Image Anal.*, 4:21–30, 2000.

[3] Y. Lamash, A. Fischer, S. Carasso, and J. Lessick. Strain analysis from 4D cardiac CT image data. *IEEE Trans Biomed Eng.*, 62:511–21, 2015.

[4] Daniel Freedman and Tao Zhang. Interactive graph cut based segmentation with shape priors. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, I:755–762, 2005.

[5] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan. Graph cut segmentation with a statistical shape model in cardiac {MRI}. *Computer Vision and Image Understanding*, 117(9):1027 – 1035, 2013.

[6] Dwarikanath Mahapatra. Cardiac Image Segmentation from Cine Cardiac MRI Using Graph Cuts and Shape Priors. *Journal of Digital Imaging*, 26(4):721–730, 2013.

[7] Alan H. Kadish, David Bello, J. Paul Finn, Robert O. Bonow, Andi Schaechter, Haris Subacius, Christine Albert, James P. Daubert, Carissa G. Fonseca, and Jeffrey J. Goldberger. Rationale and design for the defibrillators to reduce risk by magnetic resonance imaging evaluation (determine) trial. *Journal of Cardiovascular Electrophysiology*, 20(9):982–987, 2009.

[8] Caroline Petitjean, Maria A. Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jrme Caudron, Su Ruan, Ismail Ben Ayed, M. Jorge Cardoso, Hsiang-Chou Chen, Daniel Jimenez-Carretero, Maria J. Ledesma-Carbayo, Christos Davatzikos, Jimit Doshi, Guray Erus, Oskar M.O. Maier, Cyrus M.S. Nambakhsh, Yangming Ou, Sbastien Ourselin, Chun-Wei Peng, Nicholas S. Peters, Terry M. Peters, Martin Rajchl, Daniel Rueckert, Andres Santos, Wenzhe

Shi, Ching-Wei Wang, Haiyan Wang, and Jing Yuan. Right ventricle segmentation from cardiac MRI: A collation study. *Medical Image Analysis*, 19(1):187 – 202, 2015.

[9] Avan Suinesiaputra, Brett R. Cowan, Ahmed O. Al-Agamy, Mustafa A. Elattar, Nicholas Ayache, Ahmed S. Fahmy, Ayman M. Khalifa, Pau Medrano-Gracia, Marie-Pierre Jolly, Alan H. Kadish, Daniel C. Lee, Jn Margeta, Simon K. Warfield, and Alistair A. Young. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac {MR} images. *Medical Image Analysis*, 18(1):50 – 62, 2014.

[10] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M. Roh, X. Pennec, M. Sermesant, F. Isensee, P. Jger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Igum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P. Jodoin. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, Nov 2018.

[11] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013.

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[13] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5574–5584. Curran Associates, Inc., 2017.