4-5-2023

A Deep Learning-based Pipeline to Generate Patient-specific Anatomical Models of the Heart Using Cardiac MRI

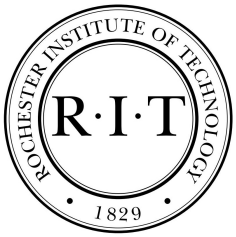
Roshan Reddy Upendra
ru6928@rit.edu

A Deep Learning-based Pipeline to Generate Patient-specific Anatomical Models of the Heart Using Cardiac MRI

Roshan Reddy Upendra

Doctor of Philosophy

in Imaging Science



Chester F. Carlson Center for Imaging Science

College of Science

Rochester Institute of Technology

Rochester, New York

April 5, 2023

Rochester Institute of Technology
Imaging Science

CERTIFICATE OF APPROVAL

The dissertation by

Roshan Reddy Upendra

entitled
A Deep Learning-based Pipeline to Generate Patient-specific Anatomical Models of the Heart Using Cardiac MRI

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

Dissertation Advisor

Dr. Cristian A. Linte
Associate Professor
Center for Imaging Science
Department of Biomedical Engineering
Rochester Institute of Technology, NY

Dissertation Committee Members

Dr. Carl Salvaggio
Professor, Center for Imaging Science
Rochester Institute of Technology, NY

Dr. Suzanne M. Shontz
Associate Dean for Research, School of Engineering
Professor, Electrical Engineering and Computer Science
Bioengineering Program
Information Sciences Institute
University of Kansas, KS

Date

Dr. Satish Kandlikar (Chair)
Gleason Professor
Department of Mechanical Engineering
Rochester Institute of Technology, NY

A Deep Learning-based Pipeline to Generate Patient-specific Anatomical Models of the Heart Using Cardiac MRI

by

Roshan Reddy Upendra

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Imaging Science
Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology
April, 2023

Date _____ _____

Signature of the Author

Date _____ _____

(Certified by)
Ph.D. Program Director

A Deep Learning-based Pipeline to Generate Patient-specific Anatomical Models of the Heart Using Cardiac MRI

by

Roshan Reddy Upendra

Submitted to the

Chester F. Carlson Center for Imaging Science

PhD Program in Imaging Science

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Imaging Science

at the Rochester Institute of Technology

Abstract

Image-based patient-specific anatomical models of the heart have the potential to be used in a variety of clinical scenarios such as diagnosis and prognosis of various cardiovascular diseases (CVDs), including cardiac resynchronization therapy (CRT), ablation therapy, risk stratification, and minimally invasive cardiac interventions. Cardiac magnetic resonance imaging (MRI) provides images with high-resolution and superior soft tissue contrast, rendering it as the gold standard modality for imaging cardiac anatomy.

To obtain meaningful information from such image-based personalized anatomical models of the heart, it is crucial to combine the geometric models of the cardiac chambers extracted from cine cardiac MRI and the scar anatomy from the late gadolinium enhanced (LGE) MRI. There are several challenges to be tackled to generate patient-specific anatomical models of the heart from the cardiac MRI data. Firstly, accurate and robust automated segmentation of the cardiac chambers from the cine cardiac MRI data is essential to estimate cardiac function indices. Secondly, it is important to estimate cardiac motion from 4D cine MRI data to assess

the kinematic and contractile properties of the myocardium. Thirdly, accurate registration of the LGE MRI images with their corresponding cine MRI images is crucial to assess myocardial viability. In addition to the above-mentioned segmentation and registration tasks, it is also crucial to computationally super-resolve the anisotropic (high in-plane and low through-plane resolution) cardiac MRI images, while maintaining the structural integrity of the tissues.

With the advent of deep learning, medical image segmentation and registration have immensely benefited. In this work, we present a deep learning-based framework to generate personalized cardiac anatomical models using cardiac MRI data. Firstly, we segment the cardiac chambers from an open-source cine cardiac MRI data using an adversarial deep learning framework. We evaluate the viability of the proposed adversarial framework by assessing its effect on the clinical cardiac parameters. Secondly, we propose a convolutional neural network (CNN) based 4D deformable registration algorithm for cardiac motion estimation from an open-source 4D cine cardiac MRI dataset. We extend this proposed CNN-based 4D deformable registration algorithm to develop dynamic patient-specific geometric models of the left ventricle (LV) myocardium and right ventricle (RV) endocardium. Thirdly, we present a deep learning framework for registration of cine and LGE MRI images, and assess the registration performance of the proposed method on an open source dataset. Finally, we present a 3D CNN-based framework with structure preserving gradient guidance to generate super-resolution cardiac MRI images, and assess this proposed super-resolution algorithm on an open-source LGE MRI dataset. Furthermore, we investigate the effect of the proposed super-resolution algorithm on downstream segmentation task.

Keywords: Image segmentation; image registration; cine MRI; LGE MRI; deep learning; convolutional neural network; super-resolution; patient-specific modeling; mesh warping; personalized cardiac models

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to each and everyone who supported me throughout my PhD journey.

Firstly, I would like to thank my advisor, Dr. Cristian A. Linte, from the bottom of my heart. He provided me with the freedom to explore the research topics that interest me, while providing challenging opportunities and pushing the boundaries of my knowledge. His expertise, insights, constructive feedback and thoughtful comments has undoubtedly made me a better researcher, writer and presenter. I will always be grateful for his guidance and support, and look forward to continuing to learn from him in the years to come.

I would like to thank my committee members, Dr. Carl Salvaggio, Dr. Suzanne M. Shontz and Dr. Satish Kandlikar for their insightful comments that helped me improve my thesis. Their expertise and guidance played an important role in shaping my thesis.

I would like to extend my gratitude to the faculty and staff of the Chester F. Carlson Center for Imaging Science for the multidisciplinary nature of the courses that were offered, and providing me with a stimulating and nurturing academic environment.

I would also like to acknowledge my internship experiences and my mentors, Bruce Link, Vipul Pai Raikar and Anithapriya Krishnan at Kodak Alaris, Philips and Genentech, respectively. The industrial exposure, practical insights and hands-on experience on real-world problems at these internships have been of immense value in shaping my research approach.

I am very thankful to the fruitful collaborations with all my labmates: Shusil Dangi, Dawei Liu, Kamrul Hasan, Golnaz Jalalahmadi, Kelly Merrell, Zixin Yang and Bidur Khanal. I am very grateful to Richard Simon for his continued guidance, invaluable advise and countless hours of discussion that helped me shape this dissertation. I am also thankful to all the co-authors of my publications related to this dissertation for their valuable contributions.

I have been very fortunate to have met some amazing people at Rochester Institute of Technology (RIT) during the course of my PhD. I would like to thank Harsh, Biswa and Surya for their support in this journey at RIT. They made this journey enjoyable.

I would also like to take this moment to thank each and every friend of mine I met during

my Masters at Drexel University and the friends I met through these friends. I can never thank them enough for their encouragement and support. I would like to express my gratitude to my friends back home, namely Sanketh, Nitish, Nanditha, Aneesh, Shashank, Jayanth and Swathi. Their unconditional love has been a constant source of my energy.

Finally, and most importantly, my deepest and heartfelt gratitude to my beloved family members. In particular, I would like to thank my mother Venna Bharathi, my father C R Upendra, my uncle Srinivas Reddy and my grandmothers V Chandravathi and Sampoornamma, for their unwavering love, belief in me, their unrelenting sacrifices, and encouraging a sense of ambition. They are the foundation upon which I built my academic success. I would like to thank my sister, Vinutha Reddy Upendra for her unconditional love and support. I would also like to thank my aunt Sujatha Thadasina, my uncle Prabhakar Reddy, my cousins Karthik and Deepthi, for their support, love and providing me a second home in this part of the world. Last but not the least, I would like to thank my wife Sneha Devarahally Nanjappa for her constant support, immense love and saint-like patience with me throughout the course of this PhD journey, for reviewing the initial drafts of all my papers, and for being one of the very important pillar of this academic achievement.

To my family

Contents

List of Figures

List of Tables

Chapter 1

Introduction, Background, and Dissertation Overview

This chapter provides a comprehensive review of the clinical background for the proposed work, and presents the idea of patient-specific anatomical models to serve as clinical decision support tools. Additionally, a brief review of medical image segmentation, registration and super-resolution algorithms is provided here, while identifying current clinical challenges. This is followed by an overview of the dissertation and its contributions.

1.1 Cardiovascular System: The Heart

The human heart is a muscular pump, located within the protective thoracic cavity, occupying space between the lungs known as mediastinum. It pumps the oxygenated blood collected from the lungs to all the tissues of the body and pumps the deoxygenated blood collected from the tissues of the body to the lungs. It is important to note that, other than oxygen, the heart also delivers nutrient-rich blood to the tissues and carries away waste from them. A healthy adult human heart beats approximately $100,000$ times a day, pumping around 5.25 liters of blood per minute throughout the body [1, 2]. It is important to understand the anatomy and physiology of the heart to comprehend the functioning of the heart.

1.1.1 The Anatomy and Physiology of the Heart

The human heart, which lies in an oblique position in the thorax, measures approximately 12 cm in length, 8 cm in width and 6 cm in thickness, and weighs 250-350 grams [2]. The size and weight of the heart varies depending on the sex of the individual, exercise habits, cardiac pathology, among other factors. It consists of four chambers: right atrium (RA), left atrium (LA), right ventricle (RV) and left ventricle (LV). The cardiovascular circulation system consists of two circuits: the pulmonary circuit that transports oxygenated blood from the lungs to the heart and delivers deoxygenated blood with carbon dioxide to the lungs, and the systemic circuit that transports oxygenated blood to all the tissues of the body and returns deoxygenated blood with carbon dioxide back to the heart, which will be sent to the pulmonary circuit. The RV pumps deoxygenated blood to the pulmonary arteries to transport the blood to the lungs, where the gas exchange occurs, i.e., carbon dioxide exits and oxygen enters the blood. The oxygenated blood is transported from the lungs to the LA via pulmonary veins, which then pumps the blood to the LV. The LV pumps this oxygenated blood to the aorta and on to the systemic circuit. The deoxygenated blood is returned to the heart through the superior and inferior vena cava to the RA. The heart consists of four valves to regulate the amount of blood that enters and exits different chambers. The atrioventricular valves: mitral valve and tricuspid valve, regulates the blood flow entering the LV and RV, respectively. The semilunar valves: aortic valve and pulmonary valve, regulates the blood flowing to all the tissues of the body and to the lungs, respectively [3]. An illustration of internal anatomy of the heart and the cardiovascular circulation system can be seen in Fig. 1.1.

The heart wall is made up of three layers: the epicardium (outermost layer), the myocardium (middle and thickest layer) and the endocardium (innermost layer) (Fig. 1.2). The epicardium is the outer protective layer of the heart that is primarily made up of connective tissues like elastic and adipose tissue. The myocardium is responsible for the contraction of the heart that pumps the blood. It is made up of cardiac muscle cells called cardiomyocytes. In order to generate high pressure required to pump blood to all the tissues of the body, the LV myocardium is thicker than the RV myocardium. The RV requires relatively lower pres-

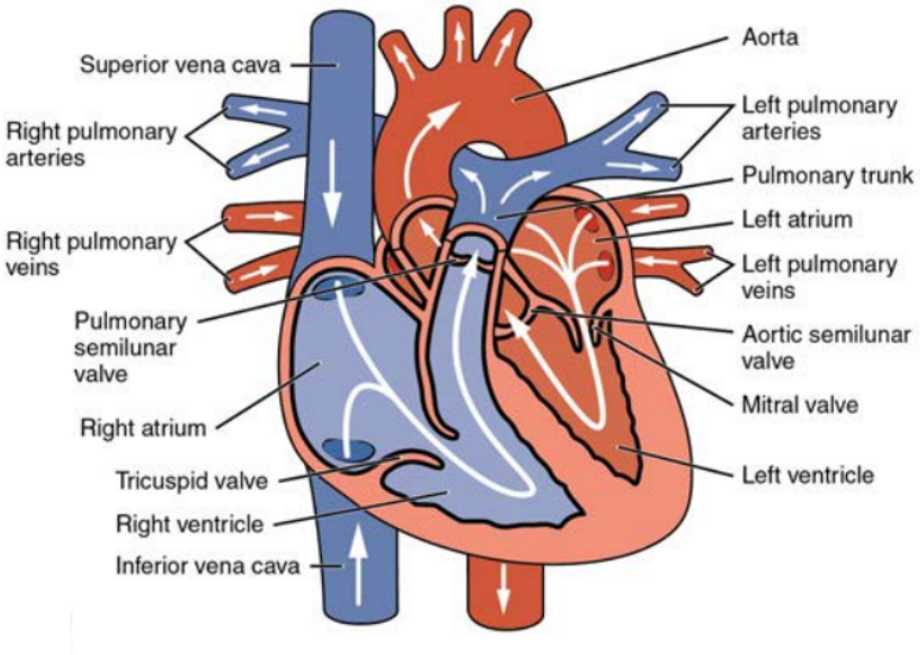


Figure 1.1: Illustration of internal anatomy of the heart and the cardiovascular circulation system [2].

sure to circulate the blood in the pulmonary circuit. The endocardium is the thinnest layer of the heart wall and is made of endothelial cells. It protects heart valves, plays active role in regulating the myocardial muscle contraction and acts as a boundary between blood-pool and myocardium [2].

1.1.2 The Cardiac Cycle

The main purpose of the heart is to pump blood through the body. It does so by coordinated contraction and relaxation of the heart muscles by electrical signals called the cardiac cycle. The cardiac cycle refers to the series of events that occur from the beginning of the atrial contraction to the end of the ventricular relaxation. It includes two phases: diastole and systole. The diastole refers to the period of relaxation and the systole refers to the period of contraction. It is paramount that the diastolic and systolic events occur in a regulated and coordinated manner in the atria and the ventricles to efficiently pump the blood. An

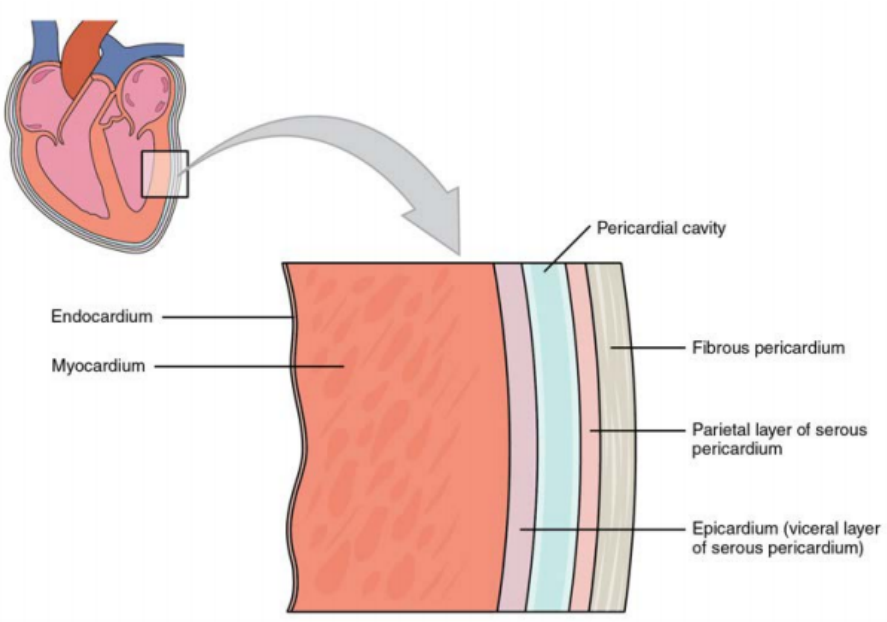


Figure 1.2: Layers of the heart wall [2].

illustration of the cardiac cycle is shown in Fig. 1.3.

The atria and the ventricles are in the diastole phase at the beginning of the cardiac cycle. Therefore, the blood flows into the left and right atria. Also, the tricuspid and mitral valve are open, resulting in an unobstructed flow from the atria to the ventricles. The pulmonary and the aortic valves are closed and approximately 70-80 percent of ventricles are filled with blood. Next, atrial contraction occurs resulting in filling the remaining 20-30 percent of the ventricles. The atrial systole lasts for around 100 ms and is followed by ventricular systole, as the atria returns to diastole. The ventricular systole is divided into two phases: isovolumetric contraction phase and ventricular ejection phase, lasting a total of around 270 ms. During the isovolumetric contraction phase, the blood pressure in the chamber is not high enough to open the pulmonary and aortic valves, however it increase above the pressure of the atria, which are now in the diastolic phase. This results in the closing of the tricuspid and mitral valves by the blood that flows back toward the atria. In the ventricular ejection phase, the blood pressure in the ventricles increase further and pushes open the pulmonary and aortic valves. This is followed by ventricular diastole, which is again divided into two phases: isovolumetric

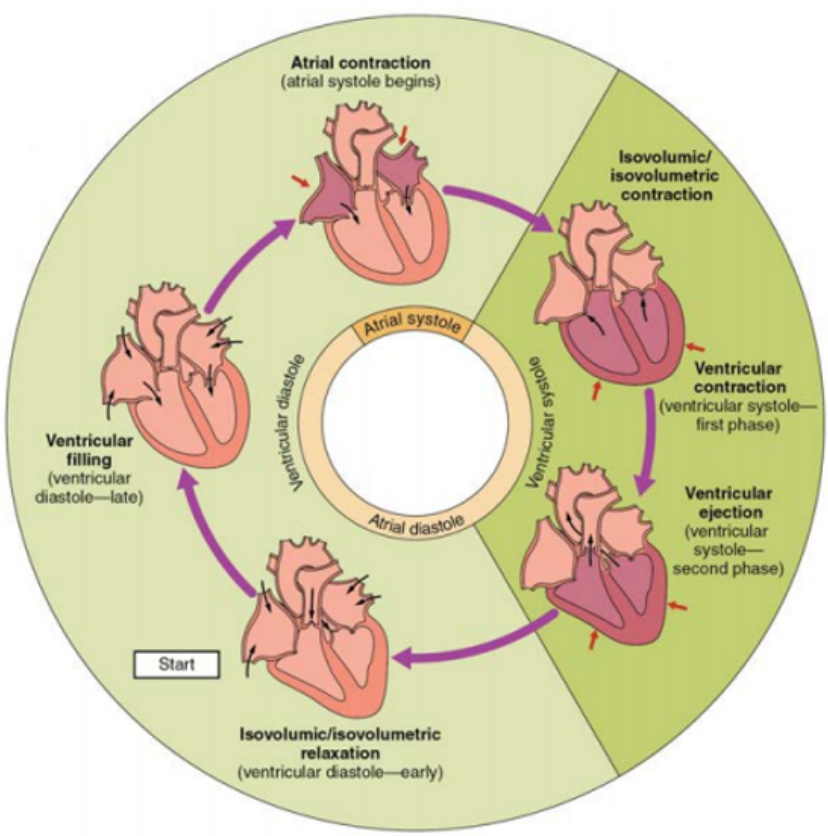


Figure 1.3: Illustration of the cardiac cycle [2].

ventricular relaxation phase and late ventricular diastole phase, lasting a total of approximately 430 ms. In the isovolumetric relaxation phase, as the ventricular muscle relaxes, the pulmonary and aortic valves close due to the backflow of blood from the arteries to the heart. This prevents further backflow of the blood to the heart. In the late ventricular diastole phase, the ventricular muscle continues to relax and eventually the blood pressure in the ventricles drop below that of the atria, pushing open the tricuspid and mitral valves, ensuing blood flow from the major veins to the atria and the ventricle. This results in the atria and the ventricles in the diastolic phase, with the atrioventricular valves open, completing one cardiac cycle.

1.1.3 Cardiovascular Diseases

As described in the previous sections, the heart is a complex and intricate organ that requires proper functioning of different systems in unison at various levels. As such, problems affecting the functioning of the heart at any level can result in cardiac health issues. The most common cardiovascular diseases (CVDs) can be divided into two types: ischemic and non-ischemic heart diseases.

Ischemic heart diseases, also known as coronary heart diseases (CHD), are the primary cause of morbidity and mortality worldwide [3]. It is primarily caused due to the decreased blood flow in the coronary arteries, which occurs because of the accumulation of cholesterol particles on the walls of these arteries. This accumulation of the cholesterol particles narrows the coronary arteries, resulting in myocardial ischemia, wherein the heart muscle does not receive adequate amount of oxygen to function properly, and can ultimately lead to myocardial infarction (MI). The prolonged myocardial ischemia can lead to myocardial tissue becoming necrotic, and this myocardial injury can become irreversible [4]. It is important to understand that the coronary artery occlusion need not necessarily lead to MI. In many cases, the ischemia leads to dysfunctional, but viable myocardial tissues. Myocardial revascularization therapy is a procedure that restores blood supply to the viable ischemic myocardial tissues. To this end, it is important to identify the viable but dysfunctional myocardium and differentiate it from the necrotic tissue [3, 4].

Non-ischemic heart diseases, also known as non-ischemic cardiomyopathies, includes a range of myocardial disorders such as, hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), restrictive cardiomyopathy (RCM) and arrhythmogenic cardiomyopathy (ACM). HCM is mostly a genetic condition that causes thickening of the septum, which is the muscular wall that divides the left and the right chambers of the heart. This thickening narrows the LV and affects the blood flow to the aorta, reducing the volume of the blood pumped to the body. DCM, the most common type of cardiomyopathy, is a condition in which the LV is enlarged and weakened, affecting the ability of the heart to pump blood. RCM is the rarest form of cardiomyopathy that affects the diastolic function of the ventricles as they become

abnormally rigid, impairing ventricular filling. ACM, another rare form of cardiomyopathy, affects the RV muscle. In this case, the heart muscle of the RV is substituted by fat and/or fibrous tissue, affecting the ability of the RV to expand and contract effectively [5, 6]. These cardiomyopathies may develop to other heart complications, such as arrhythmia and heart failure. In order to diagnose, treat and prognosticate these cardiac diseases, it is crucial to analyze the ventricular structure and function.

1.1.4 Cardiac Function

Cardiac diseases have major impact on the cardiac motion and its pump activity. Therefore, it is essential to assess the cardiac function in an accurate and reproducible way to diagnose and plan the treatment. The cardiac function can be broadly classified into two types: global and regional function parameters.

The assessment of global ventricular function includes myocardial mass, end-diastolic volume (EDV), end-systolic volume (ESV), stroke volume (SV), ejection fraction (EF), heart rate (HR) and cardiac output (CO). Myocardial mass is computed by multiplying the total volume of the ventricle with the assumed average density of the myocardium (1.05 g/ml). EDV is the total amount of blood in the ventricles at end-diastole and ESV is the remaining amount of blood in the ventricles after ejection. These two parameters can be used to deduce SV, EF and CO:

$$SV = EDV - ESV, \tag{1.1}$$

$$EF = \frac{SV}{EDV}, \tag{1.2}$$

$$CO = SV \times HR. \tag{1.3}$$

In clinical cardiology, the analysis of these global ventricular functional parameters is crucial for diagnosis of various CVDs, planning therapeutic procedures and prognosis. For example, LV ESV is one of the major determinant of survival post-MI [7].

In some cases, global ventricular functional parameters such as EF and CO may not correlate with myocardial contractility. For instance, a small imbalance in the oxygen supply as

a result of narrowing of coronary artery in ischemic heart diseases can result in contractile dysfunction of the specific area of myocardial tissue [8]. Therefore, the assessment of regional myocardial function characterized by myocardial wall thickness, myocardial motion, strain and torsion can help better understand these CVDs. An instance of the clinical application of wall motion and thickness quantification is the assessment of dysfunctional yet viable myocardium in ischemic heart diseases. To this end, accurate and consistent computation of these cardiac function parameters, both global and regional, are of utmost importance.

1.2 Cardiac Imaging

Cardiac imaging enables to capture these global and regional function parameters non-invasively. In this section, we describe the most common non-invasive cardiac imaging modalities, i.e., ultrasound (US) imaging, computed tomography (CT) imaging, and magnetic resonance imaging (MRI). We briefly outline the advantages and limitations of these clinical imaging modalities, and explain the reason to focus our research on MRI.

1.2.1 Ultrasound Imaging

Ultrasound (US) imaging is a non-invasive technique that is based on the acoustic pulse-echo measurement. The US probes, called transducers, transmit sound waves with frequencies above the human hearing threshold ($>$ 20 KHz), i.e., the US pulse. However, most of the US transducers operate in the range of $1 - 15$ MHz, and they also receive the echo signals to generate images [9]. These US transducers are made up of special ceramic crystal materials called piezoelectrics. They produce US waves by converting the electric field applied to them, and conversely produce electric signals when they receive the reflected acoustic signal. The US waves are partially reflected back to the transducer by the surfaces where the density of the matter changes (e.g. the boundary between soft tissue and bone) [10].

In clinical cardiology, US imaging has been used to assess both global and regional function parameters, such as systolic and diastolic ventricular function, myocardial velocities during systole and diastole, ventricular filling pressure, ventricular dyssynchrony, myocardial defor-

mation, as well as isovolumetric contraction and relaxation peaks. These parameters can provide prognostic markers for a number of CVDs. These assessments are enabled by a variety of US imaging techniques like tissue doppler imaging (TDI), strain imaging, contrast echocardiography and 3D transesophageal echocardiography (TEE) [11, 12].

Some of the advantages associated with US imaging are real-time imaging, low-cost, easily portable and free of ionizing-radiation. In most cases, it tends to be the initial choice of imaging modality due to its high accessibility. However, they do suffer from poor image quality compared to CT and MRI, and trade-off between spatial and temporal resolution. Additionally, a great deal of expertise is required to capture images of diagnostic significance and to interpret them.

1.2.2 Computed Tomography Imaging

Computed tomography (CT) imaging produces cross-sectional images of the body using the established X-ray technology. During a CT scan, the subject lies on a bed that moves axially through the ring-shaped structure called a gantry, a motorized X-ray source moves around the gantry, emanating narrow beams of X-rays through the body. The digital X-ray detectors located opposite to the source picks up the X-ray leaving the patient and transmits it to the computer. A 2D image slice is reconstructed after each rotation of the X-ray source, and multiple such 2D slices are stacked to generate a 3D image of the organs [13].

The high spatial and temporal resolution of the CT imaging enables visualization of the beating heart and in-turn, comprehensively assess the cardiovascular anatomy. Cardiac CT is routinely used in clinical cardiology to detect coronary calcification from atherosclerosis, evaluation of coronary stents, calcium scoring, detection of coronary artery stenosis, evaluate cardiac masses, and other coronary anomalies [14]. Cardiac CT also provides accurate measurements of ventricular volumes, regional wall motion and wall thickening [15]. The primary disadvantage of CT imaging is the use of ionizing radiation, continuous exposure of which can have harmful effects on the body [16]. In addition to the ionizing radiation, intravenous contrast agents are injected into the body to visualize soft tissues better and this can cause

allergic reactions and/or kidney failure in some cases [13].

1.2.3 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive, ionizing radiation-free imaging technology that produces high-resolution, dynamic 3D images of the human anatomy, with superior soft tissue details. The MRI technology is based on the excitation and detection of the change in the direction of the rotational axis of the protons in the hydrogen atoms found in the fat and water that make up most of the human body [17]. The strong magnetic field (1.5 T or 3 T for clinical use) produced by the powerful magnets in the MRI machines forces the protons in the human body to align with respect to the applied field, which is followed by pulsing a radio frequency (RF) current through the area of the body that we want to examine. This stimulates the protons and causes it to spin out of equilibrium, which is the "resonance" part of the MRI. The protons are forced to spin at a particular frequency called the Larmour frequency and in a particular direction against the magnetic field. When the RF pulse is turned off, the protons release the energy absorbed from the RF pulses as they realign with the magnetic field, and this energy is picked up by the MRI sensors. This is recorded as k-space data and converted to images using inverse Fourier transform. The amount of energy released by the protons and the time it takes to realign with the magnetic field depends on the chemical nature of the molecules, and these properties are used to distinguish various types of tissues [17, 18].

1.2.3.1 Cardiac MRI

In clinical cardiology, cardiac MRI is the current gold standard to depict cardiac structure, cardiac function and myocardial viability, due to its high sensitivity to soft tissues [19, 20, 21, 22, 23, 24]. Despite certain shortcomings such as high equipment cost, large equipment size, operational complexity, slower acquisition speed, patient discomfort during acquisition (patient must hold their breath for a specified time during image acquisition) and incapability to image patients with metallic implants, MRI is the preferred modality for imaging the cardiac structure due to its unparalleled high soft tissue contrast. In particular, the cine cardiac MRI is

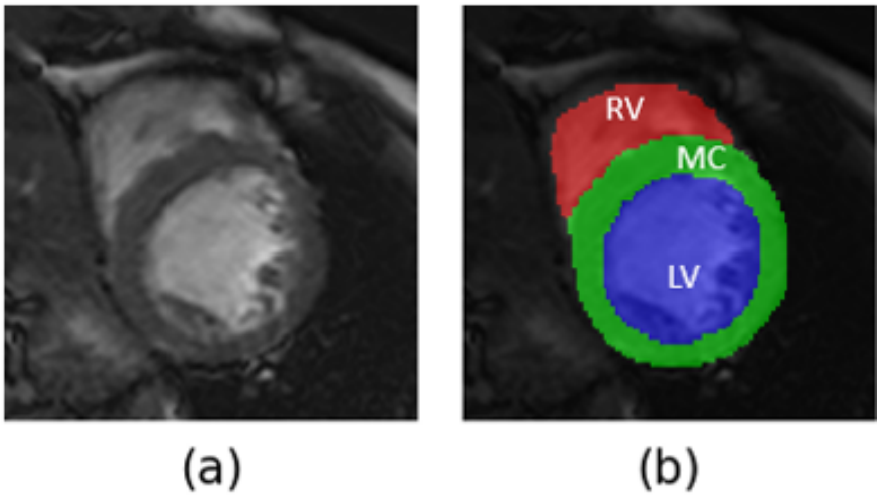


Figure 1.4: Example of (a) cine MRI image with its (b) manual annotations - LV blood-pool (LV) in blue, LV myocardium (MC) in green and RV blood-pool (RV) in red overlaid on it.

the primary technique used for cardiac function analysis and ventricular volume quantification [24], the tagged cine MRI is used to evaluate dynamic deformations of the myocardium [25], and the late gadolinium enhancement (LGE) cardiac MRI is the benchmark for assessing myocardial viability by detection and quantification of myocardial scar tissues [19]. Therefore, we utilize cardiac MRI datasets, in particular cine MRI and LGE MRI datasets, made available publicly, for the research work in this dissertation.

Cine MRI The balanced steady-state free precession (bSSFP) pulse sequence, coupled with retrospective electrocardiogram (ECG) gating enables the breath-hold cine cardiac MRI acquisition, wherein a sequence of short-axis image slices of the heart throughout the cardiac cycle are captured. A stack of such short-axis cine cardiac image slices are stacked together to generate 4D image of the cardiac chambers. Due to the high signal-to-noise ratio (SNR), contrast-to-noise ratio and acquisition speed enabled by the bSSFP pulse sequence, excellent contrast between the myocardium and blood-pool is achieved by the cine acquisition protocols (Fig. 1.4). In a typical cine cardiac MRI acquisition protocol, 20-40 frames of each short-axis slices are captured, covering the entire cardiac cycle. A typical short-axis slice consists of an in-plane resolution of 1-2 mm and a through-plane resolution of 5-10 mm [24]. The excellent

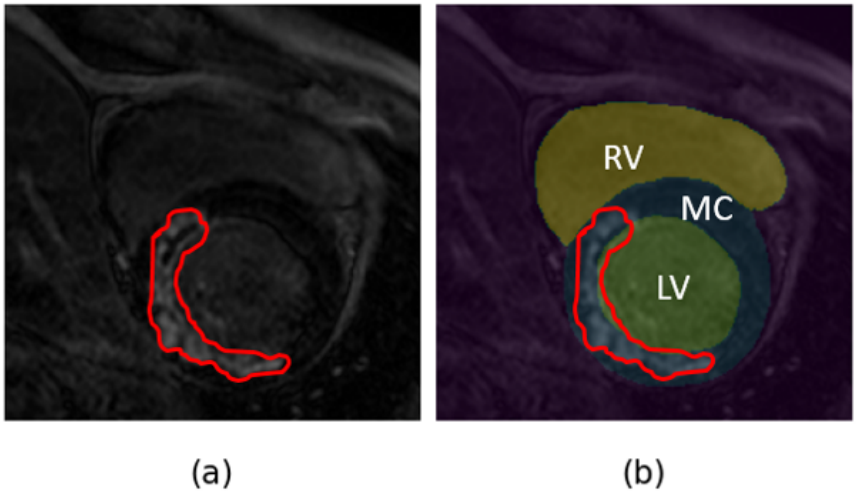


Figure 1.5: Example of (a) LGE MRI image and associated hyper-enhanced regions marked by red contour with (b) overlaid manual annotations - LV blood-pool (LV), LV myocardium (MC) and RV blood-pool (RV).

contrast between the myocardium and the blood-pool, and the 4D nature of the acquired cine cardiac MRI images enable the quantification of ventricular volumes, myocardial wall thickness, myocardial motion irregularities, and other cardiac function indices.

Late Gadolinium Enhancement MRI The introduction of gadolinium-based contrast agents in cardiac MRI established late gadolinium enhancement (LGE) cardiac MRI as the standard clinical practice for assessment of myocardial tissue infarction [19, 26]. LGE is based on the regional differences in extracellular space of the myocardium, i.e., it is based on the concept of delayed wash in and wash out of the contrast agent in the infarcted areas of the myocardial tissue, causing it to appear brighter than the surrounding non-infarcted regions (Fig. 1.5). In a typical clinical set-up, the gadolinium-based contrast agent is administered to a patient 10-20 minutes prior to the MRI image acquisition. In addition to assessing the transmural extent of the infarct to predict the success of recovery of cardiac revascularization therapy (CRT), LGE MRI can also be used in the diagnosis of myocarditis, cardiac sarcoidosis, cardiac amyloidosis, DCM and HCM [27].

1.3 Towards Personalized Models of Cardiac Anatomy

Cardiovascular diseases (CVD) are the most common non-communicable diseases worldwide, and are associated with substantial morbidity and mortality [28]. According to World Health Organization (WHO), the number of CVD-related deaths was 17.9 million in 2019, comprising 32% of the total number of deaths, and is expected to reach 22.2 million by 2030. In order to tackle the adverse effects of CVD, early detection and prediction of the disease progression is of utmost importance.

The clinical assessment for diagnosis of CVD includes examination of information from medical history, physical tests, laboratory tests and cardiac imaging. Cardiac imaging, in particular cardiac MRI, is used to assess the cardiac structure, as well as the global and regional function indices of the heart, for diagnosis of ischemic and non-ischemic heart diseases [29, 30]. Cardiac MRI is also routinely used to accurately depict the transmural extent of the myocardial infarction to guide revascularization therapy [31].

In clinical cardiology, the excellent soft tissue contrast of cine MRI enables accurate and reproducible delineation of the blood-pool and myocardium. Therefore, cine MRI is the benchmark for serial assessments of ventricular function and LV mass in patients who have undergone various cardiac therapeutic interventions. Cine MRI can also be used to accurately determine the regional myocardial wall thickness, which can be used in conjunction with LGE MRI to differentiate between viable and non-viable myocardium in order to salvage the viable myocardium [29].

In traditional clinical practice, crude averages and other statistics of population-based metrics are used to diagnose and recommend therapies, serving the "average patient". Although such approaches have been useful to understand the general behavior of pathologies, substantial inter-subject variation prevails, which can significantly affect the prospect of benefiting from a therapy or being harmed by the therapy [32]. For instance, it is a common practice in clinical cardiology to make therapeutic decisions based on surrogate biomarkers, notably left ventricular ejection fraction (LVEF) [33]. While reduced LVEF identifies patients with high risk of mortality, a number of patients who die prematurely as a result of ventricular

tachyarrhythmias after previous myocardial infarction have indicated normal or only mildly reduced LVEF [34]. This is due to the fact that LVEF reflects changes in LV dimensions during the cardiac cycle, and not the changes in contractility. With the progress in medical imaging and computational power in the recent years, personalized image-based computational models have shown immense clinical potential to provide a more accurate diagnosis, prognosis prediction and therapy planning tailored to the "individual patient" [35, 36, 37].

To be able to use patient-specific computational models of the heart for clinical decision support, the model should ideally be a combination of cardiac anatomy, electrophysiology, biomechanics, and hemodynamics [35, 38, 39, 40]. The personalized anatomical model of the heart usually includes geometric model of the cardiac chambers extracted from cine MRI, scar anatomy extracted from LGE MRI and the fiber architecture from diffusion tensor (DT) MRI [35]. In this dissertation, we focus on developing patient-specific bi-ventricular geometric models from cine cardiac MRI images, and integrating the scar anatomy from LGE MRI images with the myocardial anatomy from cine MRI images. To this end, accurate and robust segmentation of the cardiac chambers from cine MRI images to estimate cardiac function indices, accurate cardiac motion estimation from the cine MRI images to assess the kinematic and contractile properties of the myocardium and accurate registration of LGE MRI images with their corresponding cine MRI images to assess myocardial viability are crucial. Furthermore, computationally enhancing the image resolution of anisotropic 3D cardiac MRI images alleviates the challenges imposed on the segmentation and registration tasks due to the low through-plane resolution of these cardiac MRI volumes.

1.3.1 Cardiac MRI Segmentation

In order to compute the cardiac structural and functional indices from the short-axis cine MRI images, it is essential to delineate the boundaries of cardiac chambers from the MRI data. Although the manual segmentation is the current gold-standard, it can be a very laborious and time intensive task. Furthermore, manual segmentation is subject to significant intra- and inter-observer variability. For example, in the inter-observer variability study done by

some researchers, the limit of agreement for ejection fraction computation obtained by manual segmentation was in the range of [9.9%-10.3%] [41], [6%-22%] [42] and [10.3%-11.3%] [43]. Similarly, the intra-observer variability in these studies were in the range of 7% [43] and 20% [44], for left ventricle myocardial mass computation. Hence, there remains an uncertainty in these manual segmentations, as the computed clinical cardiac function indices from these segmentations could differ from the clinical fact due to intra- and inter-observer variability. Therefore, accurate, robust and consistent segmentation of cardiac chambers from cardiac MRI segmentation is crucial to improve the precision of computation of the clinical cardiac function indices. To this end, a number of semi-automatic and fully automatic segmentation techniques have been developed to aid the cardiologists in clinical practice. The popular cardiac MRI segmentation methodologies can be classified into three categories: (i) segmentation algorithms with no prior or weak prior, (ii) segmentation algorithms with strong prior, and (iii) deep learning-based segmentation algorithms [45, 46, 47, 48, 49].

1.3.1.1 Segmentation Algorithms with No Prior or Weak Prior

These segmentation algorithms use little or no prior information about the input image, instead they rely on the intensity differences between the blood-pool, myocardium and the surrounding tissues in the cine MRI data. Typical segmentation algorithms with weak or no prior knowledge include thresholding, region growing, pixel or voxel classification and active contours.

Thresholding and Region Growing Thresholding algorithms usually analyze the intensity histograms of the input MRI image to determine a threshold value and use it to localize the region of interest (RoI), i.e., the blood-pool and/or myocardium [50, 51, 52]. In region growing algorithms, one or more seed points are selected in the MRI images (for instance, one seed point each in blood-pool and myocardium), and the neighboring pixels with similar features (intensity, texture) are appended to these seed points, growing the region. The region growing stops when no more neighboring pixels meet the inclusion criteria, resulting in a segmented image [53, 54]. Often, thresholding is used to distinguish the blood pool, the myocardium and the surrounding regions, and then followed by region growing to refine

the segmentation [55, 56, 57]. While these approaches require no prior knowledge and work fairly well to segment the cardiac chambers in mid-ventricle slices, they often fail in basal and apical slices, and are affected by obstacles such as papillary muscles. Also, they are prone to inter-user variability as they require considerable user-intervention.

Pixel or Voxel Classification Here, each pixel or voxel is classified into a particular class depending on their features (intensity, color, and texture) using unsupervised or supervised techniques. In cardiac MRI segmentation, unsupervised methods typically include k-means clustering [58, 59] and Gaussian Mixture Model (GMM) with Expectation-Minimization (EM) [60]. In k-means clustering, k initial centroids (number of clusters) are chosen and each pixel is assigned to a cluster according to their distances to the centroids. This is followed by computation of new centroids and the process is repeated until no change occurs. In cardiac MRI segmentation, the LV blood-pool is identified by computing the distance to a circle [61]. Fuzzy c-means algorithm [62], a generalization of k-means allows for soft clustering instead of the hard clustering performed by k-means algorithm, can be used to segment the LV region [63]. GMM is another generalization of k-means algorithm and is estimated using the EM algorithm. The EM algorithm is used to find the maximum likelihood estimates of parameters of a statistical model. This method can model soft, non-spherical clusters to segment the cardiac chambers like LV blood-pool as well as RV blood-pool [64, 65, 66]. These methods require some prior knowledge like the geometric assumptions, as they have to compensate for the minimal spatial information.

In contrast to the clustering methods, supervised classifiers like the support vector machine (SVM) [67], random forest [68], k-nearest neighbour [69] and neural network [70] need labelled training data. In cardiac MRI data, each pixel is usually labelled as LV blood-pool, LV myocardium, RV blood-pool and background. The training data and their corresponding labels are used to train these classifiers by learning to minimize a cost-function that punishes misclassification of the training labels [71, 72, 73]. In general, the supervised techniques provide more accurate segmentation than the unsupervised ones as it uses annotation from the experts to train the classifier models. However, these supervised classifiers require annotations for

these training data, does not generalize well if the testing data statistically deviates from the training data and often ignore the spatial dependencies of the local features [46].

Active Contours or Snakes Active contours [74], or deformable models, have been widely used in cardiac image segmentation [75, 76, 77, 78]. This method involves iteratively deforming curves that search for cardiac chamber walls with weak prior knowledge, instead of directly classifying the pixel/voxel regions. Here, a parameterized spline curve $C(s) = (x(s), y(s))$, where s is a free parameter, is deformed locally towards the target boundaries based on the predefined internal spline energy E_{int}, and the external image forces E_{ext}. The internal spline energy aims to maintain the topology and smoothness of the parameterized spline curve and the external image force pushes the spline curve to the target boundary. The total energy function is given by:

$$E = \int E_{int}(C(s)) + E_{ext}(C(s)) + E_c(C(s))ds, \tag{1.4}$$

where E_c denotes additional constraints that focuses on enhancing convergence or penalising unwanted shape irregularities.

Some of the limitations of the active contour-based methods include user interaction to initialize the contours, collapse of the snake due to internal spline energy, poor convergence of boundary concavities, often gets stuck in local minima states, computational complexity, and longer computation times. To mitigate some of these limitations, many researchers have proposed various modifications to the energy functions, such as including pressure force ("balloons") [79] normal to the direction of the curve [80], gradient vector flow (GVF) [81] as the external energy term in the snake formulation [76, 82], and the level set framework [83], where the curve is implicitly defined as the zero level of a higher dimensional function to handle large topological changes, i.e., when the the curve has to be morphed significantly [84, 85].

1.3.1.2 Segmentation Algorithms with Strong Prior

While the segmentation algorithms with weak or no prior knowledge are computationally efficient, they produce inaccurate segmentation results in the ill-defined regions and require significant user-interaction to correct them. On the other hand, methods that employ strong priors, such as graph-cut algorithms, statistical shape models and atlas-based models can increase the segmentation accuracy and robustness in the ill-defined regions, however, they require manual building of training set that is representative of the population.

Graph-cut Algorithms In a two-class graph-cut segmentation algorithm [86, 87], a graph $G = (V, E)$ is defined such that the vertices V represent a set of pixels/voxels/super-pixels and E represents a set of edges connecting a pair of neighboring vertices. In addition, there are two special (terminal) vertices, source vertex (representing foreground object in the image) and sink vertex (representing background object in the image). Also, there are two types of edges, terminal edges (connects terminal and non-terminal vertices) and non-terminal edges (connects the non-terminal vertices only). The graph is a flow network, for which there are efficient algorithms to compute a minimal cut that separates the two terminal vertices, and this minimal cut defines a segmentation. The idea of the graph-cut algorithm is that that the minimal cut solution will keep the pixels/voxels with high probabilities (foreground) to belong to the side of the source vertex and similarly the background pixels/voxels on the other side of the cut near the sink vertex. Other than the minimal cut criterion as cost function for image segmentation [88], researchers have proposed normalized cut [89], region cut [90], mean cut [91] and ratio cut [92]. The graph-cut based algorithms for segmentation of LV and RV from cardiac MRI usually requires shape priors [93, 94, 95, 96] or context-based information [97, 98]. These graph-cut algorithms produce accurate segmentation results in the mid-slices of the 3D MRI images, however, the segmentation is often compromised in the apical/basal slices as well as in complex cardiac structures, such as papillary and trabecular muscles.

Active Shape and Appearance Models One of the most popular segmentation algorithm used in medical imaging, active shape model (ASM) [99] learns the pattern of shape variability

as in the training set of labelled examples, but constrained by the point distribution model, and iteratively deformed to fit the target shape in a new image. Here, we need the models of the image appearance around each model point, for instance, a simple model assumes that the points lie on strong edges. The initial estimate of the pose and shape parameters are iteratively updated until convergence by repeating these two steps: calculating the adjustments for each model point (landmark) by looking along normal vector to the surface to find the best local match for the model of the image appearance for the landmark in question and updating the model parameters to best fit the model instance by minimizing the squared distances to the found best positions. The ASM segmentation method exploits the strong prior knowledge like the specific shape variability of the cardiac chambers to produce accurate and robust segmentation of these chambers from the cardiac MRI images [100, 101].

An extension of ASM, active appearance model (AAM) [102] includes gray level modeling and learns both the shape and texture variability from the annotated training set. This results in a more robust and realistic statistical model. AAM is applied extensively in cardiac MRI segmentation, and demonstrates clinical potential in quantifying the cardiac function indices such as ventricular volumes and myocardial mass [103, 104, 105]. However, the AAM segmentation methods use global contour optimization instead of the local structures, as in ASM. This has led to hybrid models that combine the strengths of both ASM and AAM methods such that the AAM is used to model the cardiac chambers and the ASM allows for position refinement [106, 107, 108].

Regardless, the ASM and AAM segmentation methods are not applicable for widespread clinical use as they demand annotation of the training sets, the model fitting process can be computationally expensive and slow, prone to local minima and not generalizable.

Atlas-based Models In the context of cardiac MRI segmentation, an atlas is a reference cardiac MRI image that describes the different cardiac structures (LV and RV) that needs to be segmented along with its ground truth segmentation. Given this atlas, a cardiac MRI image can be segmented via image registration. The atlas-based segmentation technique involves obtaining an optimum registration transform to register the atlas image to the test image,

and using this obtained registration transform to deform the ground truth segmentation map of the atlas to segment the test image. The cardiac atlas can be constructed either using a single segmented image (single atlas model) [109] or an average obtained from multiple atlases (probabilistic atlas model) [110, 111] on the same reference coordinate system. While both these methods require only one registration step, the single atlas model-based segmentation method produces poor segmentation results due to the large variability in the test images. This can be overcome using the probabilistic model by incorporating the variability in the multiple training atlases. Alternatively, multiple atlases can be registered to the test image and a label fusion strategy can be used to obtain a segmentation map of the test image (multi-atlas model) [112, 113]. This method is relatively more robust than the single and probabilistic atlas models, at the cost of increased computational time and resources.

1.3.1.3 Deep Learning-based Segmentation Algorithms

Over the last few years, deep learning [114] algorithms have demonstrated substantial improvement in computer vision tasks, including medical image segmentation. The availability of large number of open-source cardiac MRI datasets with ground truth segmentation maps [115, 116, 117] and advanced computer hardware for training the deep learning algorithms has benefited cardiac MRI segmentation immensely, outperforming the previous state-of-the-art non-deep learning segmentation methods [115, 49, 118, 119].

A standard convolutional neural network (CNN) takes an input image, learns hierarchical features by passing the image through a stack of convolutional filters, followed by normalization layer and non-linear activation function to extract feature maps. The extracted feature maps are downsampled using pooling layers. These downsampled spatial feature maps are then passed through fully connected layers to further reduce the dimension of the features and find the most relevant features for inference. To perform inference, the CNN must be trained to minimize a cost function and to update the model parameters. This cost function accounts for the error between the CNN prediction and the ground truth labels during training, and provides information for the optimizer to update the CNN parameters through backpropaga-

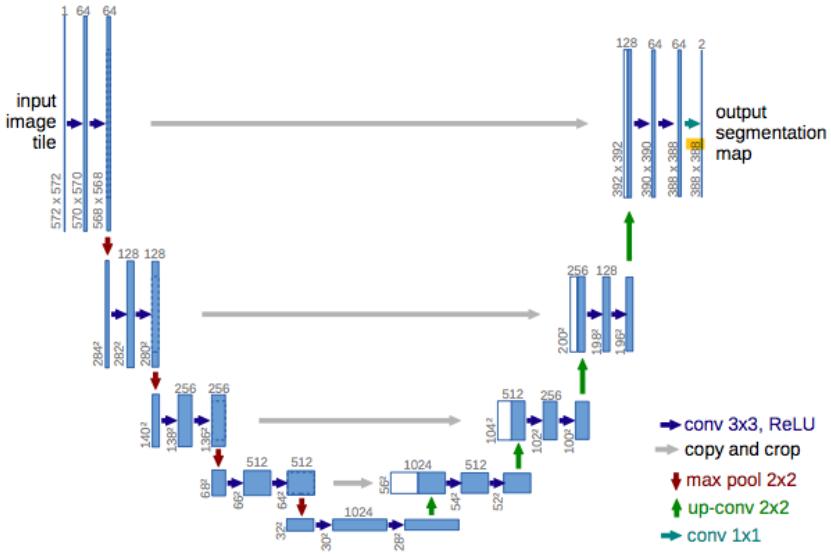


Figure 1.6: U-Net architecture [120]

tion [121], thereby predicting outputs that is as close as possible to the ground truth. In order to obtain pixel-wise segmentation of the cardiac chambers from cine cardiac MRI datasets, fully convolutional neural networks (FCN) [122] and its variants are more commonly used [123, 124, 125]. FCNs are CNNs that do not have any fully connected layers, instead they have an encoder-decoder framework that transforms the input image into high-level feature representation (encoding) and recovers spatial information back to the image space to produce pixel-wise segmentation maps. A number of variants of FCN have been proposed for medical image segmentation, the most popular of which is U-Net [120]. The U-Net model consists of a contracting path, a bottleneck and an expansive path that give it the U-shaped architecture (Fig. 1.6). Additionally, it employs skip connections between the contracting and expansive path to allow the network to propagate spatial context information to higher resolution layers. The U-Net model and its variants have been widely adopted for the segmentation of cardiac chambers from cardiac MRI datasets with various loss functions [126, 127], as 3D network [128, 129], as multi-task learning network [130, 131], etc. A more comprehensive review of deep learning-based algorithms for cardiac MRI segmentation can be found in [49].

1.3.2 Multimodal Image Registration: Cine MRI and LGE MRI

As mentioned in the Section 1.2.3.1, LGE MRI enhances the infarcted myocardium such that it appears distinctively brighter than the surrounding healthy myocardial tissue. The precise localization and quantification of the infarcted regions from these LGE MRI images is critical for diagnosis and planning therapy. In a clinical set-up, the viability of the myocardium is visually assessed based on these LGE MRI images, however, manual delineation of the compromised myocardial tissue is time-consuming and subject to intra- and inter-observer variations. While automatic segmentation of the scar tissue from LGE MRI [132, 133, 134] is desired, direct segmentation of these tissues is very challenging due the following reasons: (i) overlap in the intensity range of different tissues in the LGE MRI images (for example, the scar tissue can appear identical to the LV blood-pool, and the myocardium can have similar intensity range as adjacent lung/liver tissues); (ii) heterogeneous intensity of the myocardium due to the different pathologies associated with the heart; and (iii) inter-subject variations in the enhancement patterns of the scar tissue such as location, size and shape. These issues make it difficult to directly segment the infarcted regions from the LGE MRI data without any prior knowledge [135, 136].

To accurately delineate the scar tissue from the LGE MRI images, integration of prior shape information of the myocardium would be very useful. Although LGE MRI is useful to identify scarred myocardium regions, it does not allow for high contrast between the blood-pool and the myocardium. Therefore, a number of reported methods use the excellent contrast between myocardium and blood-pool provided by the cine MRI, acquired in the same session as the LGE MRI, as *a priori* knowledge [135, 136]. Therefore, accurate registration of the LGE MRI images with their corresponding cine MRI images enables visualization of all desired features, i.e., blood pool, myocardium, and scarred regions. The methods in the literature for multimodal image registration of cine MRI and LGE MRI can be broadly classified into: (i) rigid registration, (ii) affine registration, and (iii) deformable registration.

1.3.2.1 Rigid Registration

In rigid image registration, the cine and LGE MRI images are aligned by simply translating and rotating with respect to each other to achieve correspondence, thus preserving the internal cardiac structure. The automated rigid registration of cine and LGE MRI can be 2D rigid registration based on a shift window [137], 3D rigid registration using mutual information as the similarity measure [138], or rigid registration of the 3D cine volume to each of the 2D LGE slices to propagate 3D cine contours onto the 2D LGE slices [139]. It is common for the shape and local deformations to differ in the cine and LGE MRI datasets, which is not accounted for in the rigid registration methods, and this can have adverse impact on the localization of the scarred regions in LGE MRI. However, these rigid registration methods can be useful for initial alignment of the cine and LGE MRI images [140].

1.3.2.2 Affine Registration

In addition to translation and rotation, the affine registration of cine and LGE MRI images can be used to correct for scaling and shearing differences. A number of affine registration methods have been proposed to register the LGE MRI images with their corresponding cine MRI images by using similarity measures based on cross-correlation [141], mutual information [142] and/or pattern intensity [143]. While the affine registration methods account for the scaling and shearing differences, it cannot sufficiently correct for the local deformations. However, the affine registration methods can provide better initial alignment than rigid registration [142].

1.3.2.3 Deformable Registration

Deformable image registration involves warping an image (moving image) to align to the other image (target image) via a deformation field such that it can account for the differential local deformations between the two images. The deformable registration step is usually followed post the initial alignment by rigid or affine registration to fine-tune the registration using transformation models such as B-spline based free form deformation, thin-plate splines, Demons algorithm, or finite element method-based linear elastic [140, 142, 144, 145]. To summarize,

these methods employ traditional approaches to iteratively optimize the registration objective function for a given image pair. An overview of the transformation models and the objective functions for deformable image registration can be found in [146] and an overview of multi-modal image registration methods with transformation models and objective functions specific to cardiac diagnosis and treatment can be found in [147].

With the advent of deep learning and introduction of deep learning-based image registration algorithm like spatial transformer networks (STN) [148], several researchers proposed the training of neural network-based image registration algorithms to optimize the registration cost function for a given pair multimodal medical images [149, 150, 151, 152]. A detailed review of deep learning-based medical image registration, including multimodal registration algorithms can be found in [153].

1.3.3 Cardiac Motion Estimation

Cardiac motion estimation is crucial to assess regional cardiac function such as myocardial wall deformation, strain, torsion and thickness from cardiac MRI images. While tagged MRI (tMRI) is the current reference modality to obtain regional information on myocardial deformation, a number of researchers showed that it is possible to obtain accurate cardiac motion estimation from untagged cine cardiac MRI [154, 155, 156].

1.3.3.1 Cardiac Motion Estimation from Tagged MRI

As surgically implanting invasive markers into the myocardium tends to influence the regional motion pattern of the myocardial muscle and are impractical for clinical applications, MRI tagging [157] was developed to provide non-invasive mathematical markers inside the myocardium. MRI tagging involves superimposing grid-like structures on the myocardium by applying special RF pulses called spatial modulation of magnetization (SPAMM). The grid-like structures (tags) on the images are deformed along with the myocardium, allowing quantification of deformation and evaluation of the myocardial motion [25] (Fig. 1.7).

Cardiac motion tracking from 3D tMRI sequences can be achieved using harmonic phase

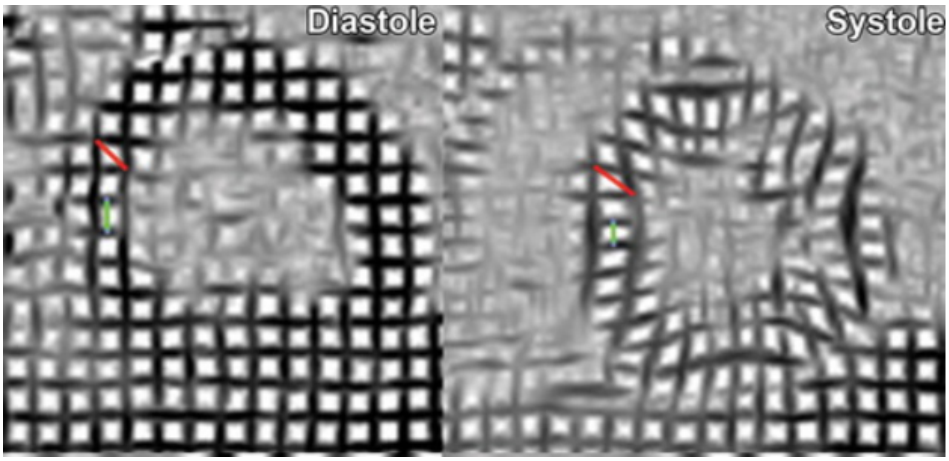


Figure 1.7: Example of tagged MRI images during diastole (left) and systole (right), with red lines representing radial strain measurements and green lines representing circumferential strain measurements [25]

(HARP) technique [158], local sine wave modeling [159], Gabor filter banks [160], deformable models [161], optical flow methods [162] or non-rigid image registration methods [163]. A comparative study of the myocardial function quantification techniques using tMRI can be found in [164]. Although MRI tagging allows for non-invasive mathematical landmarks for tracking the myocardial wall deformation, it has considerable limitations. Some of the common issues with tMRI are the fading of the tags through the cardiac cycle, low temporal resolution, masking the cardiac anatomy, and additional acquisition time. Therefore, cardiac motion estimation from untagged cine cardiac MRI is gaining popularity as it alleviates most of the limitations of tMRI.

1.3.3.2 Cardiac Motion Estimation from Untagged Cine MRI

Due to the high soft tissue contrast and high temporal resolution of cine cardiac MRI images, they can be used for cardiac motion estimation. Cardiac motion estimation methods from cine cardiac MRI can be mainly classified into (i) incompressible deformable models, (ii) physiome models, (iii) feature-based tracking methods, and (iv) registration-based methods [155].

The incompressible deformable model strategy is based on a 3-D deformable model that is incompressible. It is based on the assumptions that myocardium is almost incompress-

ible and that there is no transmural bending during myocardial movement [165, 166]. The physiome model involves combining electrophysiology, kinematics, and/or mechanics of the heart by merging the electrical propagation model, electromechanical coupling model, and/or biomechanical model, respectively. The myocardial deformation can be recovered by coupling of these models [167, 168]. The relatively popular feature-based tracking methods involves segmentation of the myocardial wall from the cine cardiac MRI data, followed by geometrical and mechanical modeling to extract the displacement field in order to perform cardiac motion analysis [169, 170, 171]. The registration-based method for cardiac motion estimation from cine cardiac MRI can be energy-based warping or optical flow techniques.

A number of 4D deformable registration methods based on voxel similarity measures have been proposed for cardiac motion estimation from 4D cine MRI data [172, 154, 173]. The main idea behind these methods involves establishing correspondence in both the temporal and the spatial domains between two 3D cardiac MRI images by employing methods like 4D spatio-temporal B-spline models [172, 174]. More specifically, registering 3D images acquired at different phases of the cardiac cycle to each other can estimate cardiac motion. A more detailed review of non-deep learning cardiac image registration methods can be found in [175].

With increase in popularity of deep learning algorithms for image registration, a number of deep learning-based 4D deformable registration methods have been proposed for cardiac motion estimation from cine cardiac MRI [131, 176, 177].

1.3.4 Cardiac MRI Super-Resolution

In clinical cardiac MRI, the 3D volumes obtained usually have high in-plane resolution, but low through-plane resolution (slice thickness). For example, in a typical cine and LGE cardiac MRI, the anisotropic cardiac volumes have an in-plane resolution of 1 to 1.5 mm and through-plane resolution of 5 to 10 mm. This is due to the inherent trade-off in the MRI imaging protocol between the signal-to-noise ratio, spatial and temporal resolution, and acquisition time. Therefore, the obtained anisotropic cardiac MRI may impose challenges in downstream segmentation and registration tasks. In order to overcome this limitation, researchers have

proposed a number of super-resolution methods to computationally enhance the resolution of the image [178, 179, 180, 181].

1.3.4.1 Image Interpolation and Non-Deep Learning Super-Resolution Methods

The aim of image interpolation is to improve image resolution by upsampling. Image interpolation methods can be broadly classified into: (i) polynomial-based interpolation and (ii) edge-directed interpolation methods. A comprehensive review of the both polynomial-based and edge-directed interpolation methods can be found in [178]. The image interpolation methods involve only upsampling the low-resolution (LR) image and assume the LR images are aliased as they are a direct downsampled version of the high-resolution (HR) images. During the upsampling process of the image interpolation methods, they often exploit this aliasing property and perform dealiasing of the LR image [182]. However, this leads to a blurry HR image, as the upsampling does not usually recover high-frequency semantic and structural information, which is crucial for cardiac MRI images.

In order to address the limitation, researchers proposed super-resolution algorithms as they involves upsampling, deblurring and denoising. A detailed review of super-resolution algorithms, prior to deep learning-based methods are described in [178].

1.3.4.2 Deep Learning-based Super-Resolution Methods

In recent years, a number of research efforts proposed deep learning-based super-resolution methods to computationally enhance the resolution of the image. These deep learning-based super-resolution algorithms can be classified into recursive learning, residual learning, dense connection-based learning, multi-scale learning, advanced convolution-based learning and attention-based learning. A comprehensive review of these deep learning architectures for image super-resolution can be found in [180].

With the advent of deep learning-based super-resolution methods for natural images, several researchers proposed these methods for medical images, especially brain MRI images [181]. However, relatively limited efforts have been made to improve the through-plane resolution of

cardiac MRI images.

1.3.5 Evaluation Metrics

The following metrics are used to evaluate the various algorithms presented in this manuscript.

1.3.5.1 Intersection of Union and Dice

The intersection of union (IoU) and Dice of two binary segmentation masks, X and Y, are defined as:

$$\text{IoU} = \frac{|X \cap Y|}{|X \cup Y|}, \qquad \text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1.5}$$

where, $|\cdot|$ represents the cardinality of each set. Both, the Dice and IoU metrics are defined such that, the values 1.0 and 0.0 indicate 100% and 0% overlap between the two segmentation masks, respectively.

1.3.5.2 Average Surface Distance and Hausdorff Distance

The average surface distance (ASD) computes the average distance between the two surfaces and the Hausdorff distance (HD) computes the largest distance between the two surfaces. They are defined as:

$$\text{ASD} = \frac{1}{2}\left(\frac{1}{N_X}\sum_{p \in S_X} d(p, S_Y) + \frac{1}{N_Y}\sum_{q \in S_Y} d(q, S_X)\right) \tag{1.6}$$

$$\text{HD} = \max\left(\max_{p \in S_X} d(p, S_Y) \max_{q \in S_Y} d(q, S_X)\right), \tag{1.7}$$

where, S_X and S_Y (with N_X and N_Y points, respectively) are surfaces corresponding to the two binary segmentation masks, X and Y, respectively, and $d(p, S)$ is the minimum Euclidean distance of the point p from the points $q \in S$.

1.3.5.3 Peak Signal-to-Noise Ratio and Structural Similarity Index

Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [183] are the two most widely used evaluation metrics for image quality assessment.

The PSNR can be defined using the mean squared error (MSE) such that, for image I and its noisy approximation J of size mxn with c number of channels, MSE and PSNR are expressed as:

$$\text{MSE} = \frac{1}{i * j * c} \sum (I - J)^2 \tag{1.8}$$

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \tag{1.9}$$

where, MAX_I is the maximum possible value for a pixel in the image. The PSNR values typically range between 15 and 30, where higher value represents closer to approximation of the image J to image I.

The SSIM metric is based on the comparison between the luminance, contrast and structure between the two images I and J. It is defined as:

$$\text{SSIM(I, J)} = \frac{(2\mu_I\mu_J + c_1)(2\sigma_{IJ} + c_2)}{(\mu_I^2 + \mu_J^2 + c_1)(\sigma_I^2 + \sigma_J^2 + c_2)} \tag{1.10}$$

where, $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, μ represents the average, σ^2 represents the variance, σ_{IJ} represents the covariance between I and J. Here, L represents the dynamic range of the pixel values, with k_1 and k_2 as constants. The SSIM index ranges from 0 to 1, with 0 indicating no similarity between the two images and 1 indicating perfect similarity.

1.4 Dissertation Contributions

Contributions to Cardiac MRI Segmentation

In our efforts to develop a deep learning pipeline to generate patient-specific computational models of the heart for clinical decision support, the first specific aim was to obtain accurate

and robust automated segmentation of the cardiac chambers from cine MRI to estimate cardiac function indices. In the 2017 automated cardiac diagnosis challenge (ACDC) [115], U-Net inspired architectures proved to be the state-of-the-art for the segmentation of cardiac chambers from the cine MRI data. In order to obtain improved segmentation results, we investigated the integration of these U-Net inspired architectures into an adversarial framework to segment LV blood-pool from the ACDC dataset. Subsequently, we evaluated the viability of the proposed adversarial framework for multi-class segmentation of LV blood-pool, LV myocardium and RV blood-pool. Furthermore, we show its effect on the clinical cardiac parameters, specifically, stroke volume, ejection fraction and myocardial mass.

Contributions to Cardiac Motion Estimation

In order to asses the regional heart function, accurate cardiac motion estimation from cine MRI data is important. Cardiac motion estimation from 4D cine MRI dataset involves finding an optical flow representation between the consecutive 3D cine cardiac frames. To this end, we propose a CNN-based 4D deformable registration technique for consistent cardiac motion estimation from an open-source 4D cine cardiac MRI dataset.

Furthermore, we extend the proposed preliminary, proof of concept, CNN-based 4D deformable registration method to develop dynamic patient-specific geometric models of the LV myocardium across subjects with different pathologies, namely normal, DCM, HCM and subjects with prior myocardial infarctions. We also extend the proposed cardiac motion extraction method to generate dynamic, deformable models of the RV blood-pool across subjects with normal and abnormal RV.

Contributions to Cine MRI and LGE MRI Registration

To accurately localize and quantify compromised myocardium, it is crucial to co-register the cine and LGE MRI images. Several researchers proposed CNN-based unsupervised registration algorithms using similar cost functions as the traditional unsupervised registration algorithms to register the cine and LGE MRI data. These unsupervised deep learning-based registration

methods do not necessarily improve registration accuracy beyond that achieved using traditional unsupervised approaches as the cost functions used to optimize them are similar. To overcome these challenges, we propose a regions of interest (RoI) guided registration technique to improve the registration accuracy beyond the aforementioned unsupervised techniques. This method relies on the annotations of cardiac structures obtained by manual annotation or by previously validated automatic segmentation techniques.

While the proposed RoI-guided CNN architecture can be used to reliably register cine and LGE MRI images, it requires annotations of cardiac structures for large number of training data. Therefore, we propose a joint deep learning framework for registration of cine and LGE MRI images, and the segmentation of cardiac chambers from both cine and LGE MRI data. The aim of this coupling of the segmentation and the registration tasks is to improve the registration accuracy by sharing the weights learned from the segmentation models, thereby, using fewer training datasets and reducing the need for large number of manual annotations.

Contributions to Cardiac MRI Super-Resolution

In a typical clinical cardiac MRI acquisition, multiple high resolution short-axis MRI slices are acquired resulting in anisotropic 3D volumes of the heart that have high in-plane resolution (1 to 2 mm) and low through-plane resolution (5 to 10 mm). The anisotropic 3D cardiac MRI images result in low resolution representation of the cardiac anatomy, which may impose challenges in previously mentioned segmentation and registration techniques. To address this limitation, we propose a self-supervised 2D deep learning framework to compute super-resolution cardiac MRI images. Additionally, we also propose a 3D CNN-based architecture with gradient guidance to generate super-resolution cardiac MRI images. The aim of the gradient guidance is to "pay more attention" to the 3D structure of the tissues in the cardiac MRI images. We assess the performance of the proposed method on an open source high-resolution LGE MRI dataset. In addition to training and testing the proposed method on the open source LGE MRI dataset, we also evaluate the generalization ability of the the trained models on a completely different dataset. Furthermore, we investigate the effect of the

proposed super-resolution method on the downsampling segmentation task.

1.5 Dissertation Overview

The dissertation chapters provide a detailed description of the proposed deep learning-based methods to overcome certain challenges of cardiac MRI segmentation and registration. Chapter 1 of this dissertation serves as an introduction to the cardiovascular system, cardiac imaging, the steps towards personalized models of the cardiac anatomy, and their challenges. The subsequent chapters describe the proposed deep learning-based methods for the development of patient-specific models of cardiac anatomy.

Chapter 2

This chapter deals with segmentation of cardiac chambers from cardiac MRI data. It provides a detailed description of an adversarial network architecture to segment LV blood-pool, LV myocardium and RV blood-pool from an open-source cine cardiac MRI dataset. Also, the effect of the proposed segmentation method on the clinical cardiac parameters is investigated. The materials presented in this chapter are adapted from the manuscripts published in 2019 Function Imaging and Modeling of the Heart (FIMH) conference in Springer's Lecture Notes in Computer Science series and 2020 SPIE Medical Imaging conference.

Chapter 3

This chapter deals with motion estimation from 4D cine MRI data. It provides a detailed description of the proposed CNN-based 4D deformable registration technique to estimate motion from an open-source 4D cine cardiac MRI dataset. Furthermore, the chapter demonstrates the use of proposed cardiac motion estimation method to build dynamic patient-specific LV myocardial models across subjects with different pathologies. Additionally, the proposed CNN-based motion estimation method is applied for motion extraction of the RV from the open-source 4D cine cardiac MRI dataset. The materials presented in this chapter are adapted from the manuscripts published in 2020 Computing in Cardiology (CinC) conference, 2021 FIMH

conference in Springer's Lecture Notes in Computer Science series and 2021 International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

Chapter 4

This chapter deals with registration of cine MRI data with their corresponding LGE MRI data. It provides a detailed description of the proposed CNN-based supervised image registration method to register cine MRI images to its corresponding LGE MRI images. In addition, an extension of the proposed method, a joint deep learning framework that enables a multi-task training of segmentation and registration is described in detail. The materials presented in this chapter are adapted from the manuscripts published in 2020 Medical Image Understanding and Analysis (MIUA) conference in Springer's Lecture Notes in Computer Science series and 2021 SPIE Medical Imaging conference.

Chapter 5

This chapter provides a detailed description of a self supervised 2D CNN-based framework and a 3D CNN framework with gradient guidance to compute super-resolution LGE cardiac MRI images. The proposed method is assessed on an open-source high-resolution LGE MRI dataset. Additionally, the effect of the proposed super-resolution method on downstream segmentation task is investigated. The materials presented in this chapter are adapted from the manuscripts published in 2021 CinC conference and 2022 International Conference of the IEEE EMBC.

Chapter 6

This chapter summarizes the contributions of this dissertation with respect to patient-specific models of the cardiac anatomy and provides some potential future research directions.

Bibliography

[1] Anthony J Weinhaus and Kenneth P Roberts. Anatomy of the human heart. In *Handbook of cardiac anatomy, physiology, and devices*, pages 51–79. Springer, 2005. 1.1

[2] J Gordon Betts, P Desaix, E Johnson, JE Johnson, O Koral, D Kruse, B Poe, JA Wise, M Womble, and KA Young. Anatomy & physiology.(2013). *Open Stax College*. 1.1, 1.1.1, 1.1, 1.1.1, 1.2, 1.3

[3] Jan Bogaert, Steven Dymarkowski, Andrew M Taylor, and Vivek Muthurangu. *Clinical cardiac MRI*. Springer Science & Business Media, 2012. 1.1.1, 1.1.3

[4] Martina Perazzolo Marra, João AC Lima, and Sabino Iliceto. MRI in acute myocardial infarction. *European heart journal*, 32(3):284–293, 2011. 1.1.3

[5] Prabhakar Rajiah, Jacobo Kirsch, Michael A Bolen, Juan C Batlle, Richard KJ Brown, Christopher J Francois, Mauricio S Galizia, Kate Hanneman, Joao R Inacio, Amer M Johri, et al. ACR appropriateness criteria® nonischemic myocardial disease with clinical manifestations (ischemic cardiomyopathy already excluded). *Journal of the American College of Radiology*, 18(5):S83–S105, 2021. 1.1.3

[6] Cleveland Clinic. Cardiomyopathy: What is it, types, causes, symptoms and treatment. Accessed Nov. 30, 2021 [Online]. 1.1.3

[7] Harvey D White, Robin M Norris, Michael A Brown, Peter W Brandt, RM Whitlock, and Christopher J Wild. Left ventricular end-systolic volume as the major determinant

of survival after recovery from myocardial infarction. *Circulation*, 76(1):44–51, 1987. 1.1.4

[8] Ernesto Castillo, João AC Lima, and David A Bluemke. Regional myocardial function: advances in MR imaging and analysis. *Radiographics*, 23(suppl_1):S127–S140, 2003. 1.1.4

[9] Thomas L Szabo. *Diagnostic ultrasound imaging: inside out*. Academic press, 2004. 1.2.1

[10] National Institute of Health. Ultrasound. Accessed Dec. 01, 2021 [Online]. https://www.nibib.nih.gov/science-education/science-topics/ultrasound. 1.2.1

[11] Maja Cikes, Ling Tong, George R Sutherland, and Jan D'hooge. Ultrafast cardiac ultrasound imaging: technical principles, applications, and clinical benefits. *JACC: Cardiovascular Imaging*, 7(8):812–823, 2014. 1.2.1

[12] 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. 1.2.1

[13] National Institute of Health. Computed Tomography (CT). Accessed Dec. 01, 2021 [Online]. https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct. 1.2.2

[14] Susanna Prat-Gonzalez, Javier Sanz, and Mario J Garcia. Cardiac CT: indications and limitations. *Journal of nuclear medicine technology*, 36(1):18–24, 2008. 1.2.2

[15] Kevin Kalisz and Prabhakar Rajiah. Computed tomography of cardiomyopathies. *Cardiovascular diagnosis and therapy*, 7(5):539, 2017. 1.2.2

[16] 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. 1.2.2

[17] National Institute of Health. Magnetic Resonance Imaging (MRI). Accessed Dec. 01, 2021 [Online]. https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri. 1.2.3

[18] Joseph P Hornak. The basics of MRI. *http://www.cis.rit.edu/htbooks/mri*, 2006. 1.2.3

[19] James P Earls, Vincent B Ho, Thomas K Foo, Ernesto Castillo, and Scott D Flamm. Cardiac MRI: recent progress and continued challenges. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 16(2):111–127, 2002. 1.2.3.1, 1.2.3.1

[20] Anil K Attili, Andreas Schuster, Eike Nagel, Johan HC Reiber, and Rob J Van der Geest. Quantification in cardiac MRI: advances in image acquisition and processing. *The international journal of cardiovascular imaging*, 26(1):27–40, 2010. 1.2.3.1

[21] Gabriella Captur, Charlotte Manisty, and James C Moon. Cardiac MRI evaluation of myocardial disease. *Heart*, 102(18):1429–1435, 2016. 1.2.3.1

[22] Masaki Ishida, Shingo Kato, and Hajime Sakuma. Cardiac MRI in ischemic heart disease. *Circulation Journal*, pages 0908060454–0908060454, 2009. 1.2.3.1

[23] Michael Salerno, Behzad Sharif, Håkan Arheden, Andreas Kumar, Leon Axel, Debiao Li, and Stefan Neubauer. Recent advances in cardiovascular magnetic resonance: techniques and applications. *Circulation: Cardiovascular Imaging*, 10(6):e003951, 2017. 1.2.3.1

[24] Mushabbar A Syed, Subha V Raman, and Orlando P Simonetti. *Basic Principles of Cardiovascular MRI: Physics and Imaging Techniques*. Springer, 2015. 1.2.3.1, 1.2.3.1

[25] Mi-Young Jeung, Philippe Germain, Pierre Croisille, Soraya El ghannudi, Catherine Roy, and Afshin Gangi. Myocardial tagging with MR imaging: overview of normal and pathologic findings. *Radiographics*, 32(5):1381–1398, 2012. 1.2.3.1, 1.3.3.1, 1.7

[26] Adelina Doltra, Brage Hoyem Amundsen, Rolf Gebker, Eckart Fleck, and Sebastian Kelle. Emerging concepts for myocardial late gadolinium enhancement MRI. *Current cardiology reviews*, 9(3):185–190, 2013. 1.2.3.1

[27] Dr. Joachim Feger. Late gadolinium enhancement. Accessed Dec. 01, 2021 [Online]. https://radiopaedia.org/articles/late-gadolinium-enhancement-2?lang=us. 1.2.3.1

[28] Bayram Şahin and Gülnur İlgün. Risk factors of deaths related to cardiovascular diseases in world health organization (WHO) member countries. *Health & Social Care in the Community*, 2020. 1.3

[29] Hajime Sakuma. Magnetic resonance imaging for ischemic heart disease. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 26(1):3–13, 2007. 1.3

[30] Oliver Wieben, Christopher Francois, and Scott B Reeder. Cardiac MRI of ischemic heart disease at 3T: potential and challenges. *European journal of radiology*, 65(1):15–28, 2008. 1.3

[31] Raymond J Kim, Edwin Wu, Allen Rafael, Enn-Ling Chen, Michele A Parker, Orlando Simonetti, Francis J Klocke, Robert O Bonow, and Robert M Judd. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. *New England Journal of Medicine*, 343(20):1445–1453, 2000. 1.3

[32] David M Kent and Rodney A Hayward. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *Jama*, 298(10):1209–1212, 2007. 1.3

[33] Patrick T O'gara, Frederick G Kushner, Deborah D Ascheim, Donald E Casey, Mina K Chung, James A De Lemos, Steven M Ettinger, James C Fang, Francis M Fesmire, Barry A Franklin, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Journal of the American college of cardiology*, 61(4):e78–e140, 2013. 1.3

[34] Nikolaos Dagres and Gerhard Hindricks. Risk stratification after myocardial infarction: is left ventricular ejection fraction enough to prevent sudden cardiac death? *European heart journal*, 34(26):1964–1971, 2013. 1.3

[35] Maxime Sermesant and Reza Razavi. Personalized computational models of the heart for cardiac resynchronization therapy. In *Patient-specific modeling of the cardiovascular system*, pages 167–182. Springer, 2010. 1.3

[36] Nic Smith, Adelaide de Vecchi, Matthew McCormick, David Nordsletten, Oscar Camara, Alejandro F Frangi, Hervé Delingette, Maxime Sermesant, Jatin Relan, Nicholas Ayache, et al. euHeart: personalized and integrated cardiac care using patient-specific cardiovascular modelling. *Interface focus*, 1(3):349–364, 2011. 1.3, 3.1

[37] Richard A Gray and Pras Pathmanathan. Patient-specific cardiovascular computational modeling: diversity of personalization and challenges. *Journal of cardiovascular translational research*, 11(2):80–88, 2018. 1.3, 3.1

[38] Elham Kayvanpour, Tommaso Mansi, Farbod Sedaghat-Hamedani, Ali Amr, Dominik Neumann, Bogdan Georgescu, Philipp Seegerer, Ali Kamen, Jan Haas, Karen S Frese, et al. Towards personalized cardiology: multi-scale modeling of the failing heart. *PLoS One*, 10(7):e0134869, 2015. 1.3

[39] Maxime Sermesant, Jean-Marc Peyrat, Phani Chinchapatnam, Florence Billet, Tommaso Mansi, Kawal Rhode, Herve Delingette, Reza Razavi, and Nicholas Ayache. Toward patient-specific myocardial models of the heart. *Heart failure clinics*, 4(3):289–301, 2008. 1.3

[40] Jazmin Aguado-Sierra, Adarsh Krishnamurthy, Christopher Villongco, Joyce Chuang, Elliot Howard, Matthew J Gonzales, Jeff Omens, David E Krummen, Sanjiv Narayan, Roy CP Kerckhoffs, et al. Patient-specific modeling of dyssynchronous heart failure: a case study. *Progress in biophysics and molecular biology*, 107(1):147–155, 2011. 1.3

[41] Laurent Bonnemains, Damien Mandry, Pierre-Yves Marie, Emilien Micard, Bailiang Chen, and Pierre-André Vuissoz. Assessment of right ventricle volumes and function by cardiac mri: quantification of the regional and global interobserver variability. *Magnetic Resonance in Medicine*, 67(6):1740–1746, 2012. 1.3.1

[42] Oronzo Catalano, Serena Antonaci, Cristina Opasich, Guido Moro, Maria Mussida, Mariarosa Perotti, Giuseppe Calsamiglia, Mauro Frascaroli, Maurizia Baldi, and Franco Cobelli. Intra-observer and interobserver reproducibility of right ventricle volumes, function and mass by cardiac magnetic resonance. *Journal of cardiovascular medicine*, 8(10):807–814, 2007. 1.3.1

[43] Philipp Beerbaum, Peter Barth, Siegfried Kropf, Samir Sarikouch, Andrea Kelter-Kloepping, Diana Franke, Matthias Gutberlet, and Titus Kuehne. Cardiac function by mri in congenital heart disease: impact of consensus training on interinstitutional variance. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 30(5):956–966, 2009. 1.3.1

[44] Peter MT Pattynama, Hildo J Lamb, Edo A Van der Velde, Rob J Van der Geest, Ernst E Van der Wall, and Albert De Roos. Reproducibility of mri-derived measurements of right ventricular volumes and myocardial mass. *Magnetic resonance imaging*, 13(1):53–63, 1995. 1.3.1

[45] 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. 1.3.1, 2.1

[46] 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, 2016. 1.3.1, 1.3.1.1, 2.1

[47] Amruta K Savaashe and Nagaraj V Dharwadkar. A review on cardiac image segmentation. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pages 545–550. IEEE, 2019. 1.3.1

[48] Leon Lenchik, Laura Heacock, Ashley A Weaver, Robert D Boutin, Tessa S Cook, Jason Itri, Christopher G Filippi, Rao P Gullapalli, James Lee, Marianna Zagurovskaya, et al. Automated segmentation of tissues using CT and MRI: a systematic review. *Academic radiology*, 26(12):1695–1706, 2019. 1.3.1

[49] Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, and Daniel Rueckert. Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine*, 7:25, 2020. 1.3.1, 1.3.1.3, 1.3.1.3

[50] Ardeshir Goshtasby and David A Turner. Segmentation of cardiac cine MR images for extraction of right and left ventricular chambers. *IEEE transactions on medical imaging*, 14(1):56–64, 1995. 1.3.1.1

[51] Juyang Weng, Ajit Singh, and MY Chiu. Learning-based ventricle detection from cardiac MR and CT images. *IEEE transactions on medical imaging*, 16(4):378–391, 1997. 1.3.1.1

[52] Ehud Nachtomy, Rafael Cooperstein, Mordechy Vaturi, Elyakim Bosak, Zvi Vered, and Solange Akselrod. Automatic assessment of cardiac function from short-axis MRI: procedure and clinical evaluation. *Magnetic resonance imaging*, 16(4):365–376, 1998. 1.3.1.1

[53] Noel CF Codella, Jonathan W Weinsaft, Matthew D Cham, Matthew Janik, Martin R Prince, and Yi Wang. Left ventricle: automated segmentation by using myocardial effusion threshold reduction and intravoxel computation at MR imaging. *Radiology*, 248(3):1004–1012, 2008. 1.3.1.1

[54] Hae-Yeoun Lee, Noel CF Codella, Matthew D Cham, Jonathan W Weinsaft, and Yi Wang. Automatic left ventricle segmentation using iterative thresholding and an active contour model with adaptation on short-axis cardiac MRI. *IEEE Transactions on Biomedical Engineering*, 57(4):905–913, 2009. 1.3.1.1

[55] Mohammed Ammar, Saïd Mahmoudi, Mohammed Amine Chikh, and Amine Abbou. Endocardial border detection in cardiac magnetic resonance images using level set method. *Journal of digital imaging*, 25(2):294–306, 2012. 1.3.1.1

[56] Ying-Li Lu, Kim A Connelly, Alexander J Dick, Graham A Wright, and Perry E Radau. Automatic functional analysis of left ventricle in cardiac cine MRI. *Quantitative imaging in medicine and surgery*, 3(4):200, 2013. 1.3.1.1

[57] Sandro Queirós, Daniel Barbosa, Brecht Heyde, Pedro Morais, João L Vilaça, Denis Friboulet, Olivier Bernard, and Jan D'hooge. Fast automatic myocardial segmentation in 4d cine CMR datasets. *Medical image analysis*, 18(7):1115–1131, 2014. 1.3.1.1

[58] Stuart Lloyd. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2):129–137, 1982. 1.3.1.1

[59] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967. 1.3.1.1

[60] 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)*, 39(1):1–22, 1977. 1.3.1.1

[61] Michael Lynch, Ovidiu Ghita, and Paul F Whelan. Automatic segmentation of the left ventricle cavity and myocardium in MRI data. *Computers in biology and medicine*, 36(4):389–407, 2006. 1.3.1.1

[62] James C Bezdek, Robert Ehrlich, and William Full. FCM: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203, 1984. 1.3.1.1

[63] Abd-El-Ouahab Boudraa. Automated detection of the left ventricular region in magnetic resonance images by fuzzy c-means model. *The International Journal of Cardiac Imaging*, 13:347–355, 1997. 1.3.1.1

[64] Marie-Pierre Jolly. Automatic segmentation of the left ventricle in cardiac MR and CT images. *International Journal of Computer Vision*, 70(2):151–163, 2006. 1.3.1.1

[65] Huaifei Hu, Haihua Liu, Zhiyong Gao, and Lu Huang. Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming. *Magnetic resonance imaging*, 31(4):575–584, 2013. 1.3.1.1

[66] Amol Pednekar, Uday Kurkure, Raja Muthupillai, Scott Flamm, and Ioannis A Kakadiaris. Automated left ventricular segmentation in cardiac MRI. *IEEE Transactions on Biomedical Engineering*, 53(7):1425–1428, 2006. 1.3.1.1

[67] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. 1.3.1.1

[68] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. 1.3.1.1

[69] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967. 1.3.1.1

[70] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995. 1.3.1.1

[71] George Stalidis, Nikolaos Maglaveras, Serafim N Efstratiadis, Athanasios S Dimitriadis, and Costas Pappas. Model-based processing scheme for quantitative 4-D cardiac MRI analysis. *IEEE transactions on information technology in biomedicine*, 6(1):59–72, 2002. 1.3.1.1

[72] Jenny Folkesson, Eigil Samset, Raymond Y Kwong, and Carl-Fredrik Westin. Unifying statistical classification and geodesic active regions for segmentation of cardiac MRI. *IEEE Transactions on Information Technology in Biomedicine*, 12(3):328–334, 2008. 1.3.1.1

[73] Dwarikanath Mahapatra. Analyzing training information from random forests for improved image segmentation. *IEEE Transactions on Image Processing*, 23(4):1504–1512, 2014. 1.3.1.1

[74] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988. 1.3.1.1

[75] Nikos Paragios. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *International Journal of Computer Vision*, 50(3):345–362, 2002. 1.3.1.1

[76] MF Santarelli, V Positano, C Michelassi, M Lombardi, and Luigi Landini. Automated cardiac MR image segmentation: theory and measurement evaluation. *Medical engineering & physics*, 25(2):149–159, 2003. 1.3.1.1, 1.3.1.1

[77] Einar Heiberg, L Wigstrom, Marcus Carlsson, AF Bolger, and M Karlsson. Time resolved three-dimensional automated segmentation of the left ventricle. In *Computers in Cardiology, 2005*, pages 599–602. IEEE, 2005. 1.3.1.1

[78] Yuwei Wu, Yuanquan Wang, and Yunde Jia. Segmentation of the left ventricle in cardiac cine MRI using a shape-constrained snake model. *Computer Vision and Image Understanding*, 117(9):990–1003, 2013. 1.3.1.1

[79] Laurent D Cohen. On active contour models and balloons. *CVGIP: Image understanding*, 53(2):211–218, 1991. 1.3.1.1

[80] Piotr Makowski, Thomas Sangild Sørensen, Søren Vorre Therkildsen, Andrzej Materka, Hans Stødkilde-Jørgensen, and Erik Morre Pedersen. Two-phase active contour method for semiautomatic segmentation of the heart and blood vessels from MRI images for 3D visualization. *Computerized Medical Imaging and Graphics*, 26(1):9–17, 2002. 1.3.1.1

[81] Chenyang Xu and Jerry L Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing*, 7(3):359–369, 1998. 1.3.1.1

[82] Yuanquan Wang and Yunde Jia. Segmentation of the left ventricle from cardiac MR images based on degenerated minimal surface diffusion and shape priors. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 4, pages 671–674. IEEE, 2006. 1.3.1.1

[83] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988. 1.3.1.1

[84] Ismail Ben Ayed, Shuo Li, and Ian Ross. Embedding overlap priors in variational left ventricle tracking. *IEEE Transactions on Medical Imaging*, 28(12):1902–1913, 2009. 1.3.1.1

[85] Michael Lynch, Ovidiu Ghita, and Paul F Whelan. Segmentation of the left ventricle of the heart in 3-D+ t MRI data using an optimized nonrigid temporal model. *IEEE Transactions on Medical Imaging*, 27(2):195–203, 2008. 1.3.1.1

[86] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001. 1.3.1.2

[87] Yuri Y Boykov and M-P Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, volume 1, pages 105–112. IEEE, 2001. 1.3.1.2

[88] Zhenyu Wu and Richard Leahy. Tissue classification in MR images using hierarchical segmentation. In *1990 IEEE Nuclear Science Symposium Conference Record*, pages 1410–1414. IEEE, 1990. 1.3.1.2

[89] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000. 1.3.1.2

[90] Ingemar J Cox, Satish B Rao, and Yu Zhong. "Ratio regions": a technique for image segmentation. In *Proceedings of 13th International Conference on Pattern Recognition*, volume 2, pages 557–564. IEEE, 1996. 1.3.1.2

[91] Song Wang and Jeffrey Mark Siskind. Image segmentation with minimum mean cut. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 1, pages 517–524. IEEE, 2001. 1.3.1.2

[92] Song Wang and Jeffrey Mark Siskind. Image segmentation with ratio cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):675–690, 2003. 1.3.1.2

[93] Jie Zhu-Jacquot and Ramin Zabih. Segmentation of the left ventricle in cardiac MR images using graph cuts with parametric shape priors. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 521–524. IEEE, 2008. 1.3.1.2

[94] Shusil Dangi, Nathan Cahill, and Cristian A Linte. Integrating atlas and graph cut methods for left ventricle segmentation from cardiac cine MRI. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 76–86. Springer, 2016. 1.3.1.2

[95] Dwarikanath Mahapatra. Cardiac image segmentation from cine cardiac MRI using graph cuts and shape priors. *Journal of digital imaging*, 26(4):721–730, 2013. 1.3.1.2

[96] Damien Grosgeorge, Caroline Petitjean, J-N Dacher, and Su Ruan. Graph cut segmentation with a statistical shape model in cardiac MRI. *Computer Vision and Image Understanding*, 117(9):1027–1035, 2013. 1.3.1.2

[97] Dwarikanath Mahapatra. Cardiac MRI segmentation using mutual context information from left and right ventricle. *Journal of digital imaging*, 26(5):898–908, 2013. 1.3.1.2

[98] Ismail Ben Ayed, Kumaradevan Punithakumar, Shuo Li, Ali Islam, and Jaron Chong. Left ventricle segmentation via graph cut distribution matching. In *International conference on medical image computing and computer-assisted intervention*, pages 901–909. Springer, 2009. 1.3.1.2

[99] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 1.3.1.2

[100] Hans C Van Assen, Mikhail G Danilouchkine, Alejandro F Frangi, Sebastián Ordás, Jos JM Westenberg, Johan HC Reiber, and Boudewijn PF Lelieveldt. SPASM: a 3D-

ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical image analysis*, 10(2):286–303, 2006. 1.3.1.2

[101] Karim Lekadir, Robert Merrifield, and Guang-Zhong Yang. Outlier detection and handling for robust 3-D active shape models search. *IEEE Transactions on Medical Imaging*, 26(2):212–222, 2007. 1.3.1.2

[102] Gareth J Edwards, Christopher J Taylor, and Timothy F Cootes. Interpreting face images using active appearance models. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 300–305. IEEE, 1998. 1.3.1.2

[103] Steven C Mitchell, Boudewijn PF Lelieveldt, Rob J van der Geest, Jorrit Schaap, Johan HC Reiber, and Milan Sonka. Segmentation of cardiac MR images: An active appearance model approach. In *Medical Imaging 2000: Image Processing*, volume 3979, pages 224–234. International Society for Optics and Photonics, 2000. 1.3.1.2

[104] Mikkel Bille Stegmann, Jens Chr Nilsson, and Bjørn A Grønning. Automated segmentation of cardiac magnetic resonance images. *Proc. International Society of Magnetic Resonance In Medicine-ISMRM 2001*, 2:827, 2001. 1.3.1.2

[105] Steven C Mitchell, Johan G Bosch, Boudewijn PF Lelieveldt, Rob J Van der Geest, Johan HC Reiber, and Milan Sonka. 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *IEEE transactions on medical imaging*, 21(9):1167–1178, 2002. 1.3.1.2

[106] Steven C Mitchell, Boudewijn PF Lelieveldt, Rob J Van Der Geest, Hans G Bosch, JHC Reiver, and Milan Sonka. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. *IEEE Transactions on medical imaging*, 20(5):415–423, 2001. 1.3.1.2

[107] Sebastian Zambal, Jiří Hladvka, and Katja Bühler. Improving segmentation of the left ventricle using a two-component statistical model. In *International Conference on*

Medical Image Computing and Computer-Assisted Intervention, pages 151–158. Springer, 2006. 1.3.1.2

[108] Honghai Zhang, Andreas Wahle, Ryan K Johnson, Thomas D Scholz, and Milan Sonka. 4-D cardiac MR image analysis: left and right ventricular morphology and function. *IEEE transactions on medical imaging*, 29(2):350–364, 2009. 1.3.1.2

[109] Maria Lorenzo-Valdés, Gerardo I Sanchez-Ortiz, Raad Mohiaddin, and Daniel Rueckert. Atlas-based segmentation and tracking of 3D cardiac MR images using non-rigid registration. In *International conference on medical image computing and computer-assisted intervention*, pages 642–650. Springer, 2002. 1.3.1.2

[110] Maria Lorenzo-Valdés, Gerardo I Sanchez-Ortiz, Andrew G Elkington, Raad H Mohiaddin, and Daniel Rueckert. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis*, 8(3):255–265, 2004. 1.3.1.2

[111] Jyrki Lötjönen, Sari Kivistö, Juha Koikkalainen, Daniel Smutek, and Kirsi Lauerma. Statistical shape model of atria, ventricles and epicardium from short-and long-axis MR images. *Medical image analysis*, 8(3):371–386, 2004. 1.3.1.2

[112] Wenjia Bai, Wenzhe Shi, Declan P O'regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters, and Daniel 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, 2013. 1.3.1.2

[113] Wenjia Bai, Wenzhe Shi, Christian Ledig, and Daniel Rueckert. Multi-atlas segmentation with augmented features for cardiac MR images. *Medical image analysis*, 19(1):98–109, 2015. 1.3.1.2

[114] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 1.3.1.3

[115] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez

Ballester, et al. 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, 2018. 1.3.1.3, 1.4, 2.2.5, 3.1

[116] Víctor M Campello, Polyxeni Gkontra, Cristian Izquierdo, Carlos Martín-Isla, Alireza Sojoudi, Peter M Full, Klaus Maier-Hein, Yao Zhang, Zhiqiang He, Jun Ma, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. *IEEE Transactions on Medical Imaging*, 40(12):3543–3554, 2021. 1.3.1.3

[117] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015. 1.3.1.3

[118] Surayya Ado Bala and Shri Kant. Deep learning-based model architectures for cardiac MRI segmentation: A survey. *International Journal of Innovative Science, Engineering & Technology*, pages 129–135, 2020. 1.3.1.3

[119] Qian Tao, Boudewijn PF Lelieveldt, and Rob J van der Geest. Deep learning for quantitative cardiac MRI. *American Journal of Roentgenology*, 214(3):529–535, 2020. 1.3.1.3

[120] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1.6, 1.3.1.3, 2.1, 2.2.4.2, 3.2.2, 5.2.2, 5.2.3.4

[121] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986. 1.3.1.3

[122] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1.3.1.3

[123] Phi Vu Tran. A fully convolutional neural network for cardiac segmentation in short-axis MRI. *arXiv preprint arXiv:1604.00494*, 2016. 1.3.1.3

[124] Mahendra Khened, Varghese Alex Kollerathu, and Ganapathy Krishnamurthi. Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical image analysis*, 51:21–45, 2019. 1.3.1.3

[125] Ahmed S Fahmy, Hossam El-Rewaidy, Maryam Nezafat, Shiro Nakamori, and Reza Nezafat. Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks. *Journal of Cardiovascular Magnetic Resonance*, 21(1):1–12, 2019. 1.3.1.3

[126] Mingqiang Chen, Lin Fang, and Huafeng Liu. FR-NET: Focal loss constrained deep residual networks for segmentation of cardiac MRI. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 764–767. IEEE, 2019. 1.3.1.3

[127] Yeonggul Jang, Yoonmi Hong, Seongmin Ha, Sekeun Kim, and Hyuk-Jae Chang. Automatic segmentation of LV and RV in cardiac MRI. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 161–169. Springer, 2017. 1.3.1.3

[128] Fabian Isensee, Paul F Jaeger, Peter M Full, Ivo Wolf, Sandy Engelhardt, and Klaus H Maier-Hein. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In *International workshop on statistical atlases and computational models of the heart*, pages 120–129. Springer, 2017. 1.3.1.3, 2.2.3, 2.2.4.1, 3.2.2, 3.3.0.1, 3.3.0.1

[129] Christian F Baumgartner, Lisa M Koch, Marc Pollefeys, and Ender Konukoglu. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 111–119. Springer, 2017. 1.3.1.3

[130] Shusil Dangi, Ziv Yaniv, and Cristian A Linte. Left ventricle segmentation and quantification from cardiac cine MR images via multi-task learning. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 21–31. Springer, 2018. 1.3.1.3, 4.1, 4.4

[131] Chen Qin, Wenjia Bai, Jo Schlemper, Steffen E Petersen, Stefan K Piechnik, Stefan Neubauer, and Daniel Rueckert. Joint motion estimation and segmentation from undersampled cardiac MR image. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 55–63. Springer, 2018. 1.3.1.3, 1.3.3.2, 3.1, 4.1, 4.2.4

[132] Arunark Kolipaka, George P Chatzimavroudis, Richard D White, Thomas P O'Donnell, and Randolph M Setser. Segmentation of non-viable myocardium in delayed enhancement magnetic resonance images. *The international journal of cardiovascular imaging*, 21(2):303–311, 2005. 1.3.2

[133] Andrew S Flett, Jonathan Hasleton, Christopher Cook, Derek Hausenloy, Giovanni Quarta, Cono Ariti, Vivek Muthurangu, and James C Moon. Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. *JACC: cardiovascular imaging*, 4(2):150–156, 2011. 1.3.2

[134] Martin Rajchl, Jing Yuan, James A White, Eranga Ukwatta, John Stirrat, Cyrus MS Nambakhsh, Feng P Li, and Terry M Peters. Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac MR images. *IEEE transactions on medical imaging*, 33(1):159–172, 2013. 1.3.2

[135] Xiahai Zhuang. Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 581–588. Springer, 2016. 1.3.2, 4.1, 4.2.1

[136] Xiahai Zhuang. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):2933–2946, 2018. 1.3.2, 4.1, 4.2.1

[137] R El Berbari, N Kachenoura, F Frouin, A Herment, E Mousseaux, and I Bloch. An automated quantification of the transmural myocardial infarct extent using cardiac DE-MR images. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4403–4406. IEEE, 2009. 1.3.2.1

[138] Yasmina Chenoune, Constantin Constantinides, Racha El Berbari, Elodie Roullot, Frédérique Frouin, Alain Herment, and Elie Mousseaux. Rigid registration of delayed-enhancement and cine cardiac MR images using 3D normalized mutual information. In *2010 Computing in Cardiology*, pages 161–164. IEEE, 2010. 1.3.2.1, 4.1

[139] Robert S Xu, Prashant Athavale, Yingli Lu, Perry Radau, and Graham A Wright. Myocardial segmentation in late-enhancement MR images via registration and propagation of cine contours. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 856–859. IEEE, 2013. 1.3.2.1

[140] Fumin Guo, Mengyuan Li, Matthew Ng, Graham Wright, and Mihaela Pop. Cine and multicontrast late enhanced MRI registration for 3D heart model construction. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 49–57. Springer, 2018. 1.3.2.1, 1.3.2.3, 4.1

[141] Julian Andrés Betancur Acevedo. *Multimodal image registration for the characterization of the hypertrophic cardiomyopathy and the cardiac asynchronism*. PhD thesis, Université Rennes 1, 2014. 1.3.2.2

[142] Qian Tao, Sebastiaan RD Piers, Hildo J Lamb, and Rob J van der Geest. Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment. *Journal of Magnetic Resonance Imaging*, 42(2):390–399, 2015. 1.3.2.2, 1.3.2.3

[143] Dong Wei, Ying Sun, Ping Chai, Adrian Low, and Sim Heng Ong. Myocardial segmentation of late gadolinium enhanced MR images by propagation of contours from cine

MR images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 428–435. Springer, 2011. 1.3.2.2, 4.1, 4.1

[144] Cybele Ciofolo, Maxim Fradkin, Benoit Mory, Gilion Hautvast, and Marcel Breeuwer. Automatic myocardium segmentation in late-enhancement MRI. In *2008 5th IEEE International Symposium on Biomedical Imaging: from nano to macro*, pages 225–228. IEEE, 2008. 1.3.2.3

[145] Fumin Guo, Philippa RP Krahn, Terenz Escartin, Idan Roifman, and Graham Wright. Cine and late gadolinium enhancement MRI registration and automated myocardial infarct heterogeneity quantification. *Magnetic Resonance in Medicine*, 85(5):2842–2855, 2021. 1.3.2.3

[146] Seungjong Oh and Siyong Kim. Deformable image registration in radiation therapy. *Radiation oncology journal*, 35(2):101, 2017. 1.3.2.3

[147] Azira Khalil, Siew-Cheok Ng, Yih Miin Liew, and Khin Wee Lai. An overview on image registration techniques for cardiac diagnosis and treatment. *Cardiology research and practice*, 2018, 2018. 1.3.2.3, 4.1

[148] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28:2017–2025, 2015. 1.3.2.3, 3.2.3.1, 4.2.2

[149] Xi Cheng, Li Zhang, and Yefeng Zheng. Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):248–252, 2018. 1.3.2.3

[150] Martin Simonovsky, Benjamín Gutiérrez-Becker, Diana Mateus, Nassir Navab, and Nikos Komodakis. A deep metric for multimodal registration. In *International conference on medical image computing and computer-assisted intervention*, pages 10–18. Springer, 2016. 1.3.2.3

[151] Grant Haskins, Jochen Kruecker, Uwe Kruger, Sheng Xu, Peter A Pinto, Brad J Wood, and Pingkun Yan. Learning deep similarity metric for 3D MR–TRUS image registration. *International journal of computer assisted radiology and surgery*, 14(3):417–425, 2019. 1.3.2.3

[152] Robert Wright, Bishesh Khanal, Alberto Gomez, Emily Skelton, Jacqueline Matthew, Jo V Hajnal, Daniel Rueckert, and Julia A Schnabel. LSTM spatial co-transformer networks for registration of 3D fetal US and MR brain images. In *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis*, pages 149–159. Springer, 2018. 1.3.2.3

[153] Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1):1–18, 2020. 1.3.2.3, 6.2

[154] Dinggang Shen, Hari Sundar, Zhong Xue, Yong Fan, and Harold Litt. Consistent estimation of cardiac motions by 4D image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 902–910. Springer, 2005. 1.3.3, 1.3.3.2

[155] Hui Wang and Amir A Amini. Cardiac motion and deformation recovery from MRI: a review. *IEEE transactions on medical imaging*, 31(2):487–503, 2011. 1.3.3, 1.3.3.2

[156] Wenzhe Shi, Xiahai Zhuang, Haiyan Wang, Simon Duckett, Duy VN Luong, Catalina Tobon-Gomez, KaiPin Tung, Philip J Edwards, Kawal S Rhode, Reza S Razavi, et al. A comprehensive cardiac motion estimation framework using both untagged and 3-D tagged MR images based on nonrigid registration. *IEEE transactions on medical imaging*, 31(6):1263–1275, 2012. 1.3.3

[157] Elias A Zerhouni, David M Parish, Walter J Rogers, Andrew Yang, and Edward P Shapiro. Human heart: tagging with MR imaging–a method for noninvasive assessment of myocardial motion. *Radiology*, 169(1):59–63, 1988. 1.3.3.1

[158] Nael F Osman, William S Kerwin, Elliot R McVeigh, and Jerry L Prince. Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(6):1048–1060, 1999. 1.3.3.1

[159] Theo Arts, Frits W Prinzen, Tammo Delhaas, Julien R Milles, Alessandro C Rossi, and Patrick Clarysse. Mapping displacement and deformation of the heart with local sine-wave modeling. *IEEE Transactions on medical imaging*, 29(5):1114–1123, 2010. 1.3.3.1

[160] Ting Chen and Leon Axel. Using Gabor filter banks and temporal-spatial constraints to compute 3D myocardium strain. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4755–4758. IEEE, 2006. 1.3.3.1

[161] Jinah Park, Dimitri Metaxas, Alistair A Young, and Leon Axel. Deformable models with parameter functions for cardiac motion analysis from tagged MRI data. *IEEE Transactions on Medical Imaging*, 15(3):278–289, 1996. 1.3.3.1

[162] Lawrence Dougherty, Jane C Asmuth, Aaron S Blom, Leon Axel, and Rakesh Kumar. Validation of an optical flow method for tag displacement estimation. *IEEE Transactions on Medical imaging*, 18(4):359–363, 1999. 1.3.3.1

[163] Raghavendra Chandrashekara, Raad H Mohiaddin, and Daniel Rueckert. Analysis of 3-D myocardial motion in tagged MR images using nonrigid image registration. *IEEE Transactions on Medical Imaging*, 23(10):1245–1250, 2004. 1.3.3.1

[164] Caroline Petitjean, Nicolas Rougon, and Philippe Cluzel. Assessment of myocardial function: a review of quantification methods and results using tagged MRI. *Journal of Cardiovascular Magnetic Resonance*, 7(2):501–516, 2005. 1.3.3.1

[165] Arnaud Bistoquet, John Oshinski, and Oskar Skrinjar. Left ventricular deformation recovery from cine MRI using an incompressible model. *IEEE transactions on medical imaging*, 26(9):1136–1153, 2007. 1.3.3.2

[166] Arnaud Bistoquet, John Oshinski, and Oskar Škrinjar. Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model. *Medical image analysis*, 12(1):69–85, 2008. 1.3.3.2

[167] Ken CL Wong, Linwei Wang, Heye Zhang, Huafeng Liu, and Pengcheng Shi. Physiological fusion of functional and structural images for cardiac deformation recovery. *IEEE Transactions on Medical Imaging*, 30(4):990–1000, 2011. 1.3.3.2

[168] Herve Delingette, Florence Billet, Ken CL Wong, Maxime Sermesant, Kawal Rhode, Matthew Ginks, C Aldo Rinaldi, Reza Razavi, and Nicholas Ayache. Personalization of cardiac motion and contractility from images using variational data assimilation. *IEEE transactions on biomedical engineering*, 59(1):20–24, 2011. 1.3.3.2

[169] Pengcheng Shi, Albert J Sinusas, R Todd Constable, and James S Duncan. Volumetric deformation analysis using mechanics-based data fusion: Applications in cardiac motion recovery. *International Journal of Computer Vision*, 35(1):87–107, 1999. 1.3.3.2

[170] Xenophon Papademetris, Albert J Sinusas, Donald P Dione, and James S Duncan. Estimation of 3D left ventricular deformation from echocardiography. *Medical image analysis*, 5(1):17–28, 2001. 1.3.3.2

[171] Y Chenoune, E Deléchelle, E Petit, T Goissen, J Garot, and A Rahmouni. Segmentation of cardiac cine-MR images and myocardial deformation assessment using level set methods. *Computerized Medical Imaging and Graphics*, 29(8):607–616, 2005. 1.3.3.2

[172] Dimitrios Perperidis, Raad H Mohiaddin, and Daniel Rueckert. Spatio-temporal free-form registration of cardiac MR image sequences. *Medical image analysis*, 9(5):441–456, 2005. 1.3.3.2

[173] Hari Sundar, Harold Litt, and Dinggang Shen. Estimating myocardial motion by 4D image warping. *Pattern recognition*, 42(11):2514–2526, 2009. 1.3.3.2

[174] Ghalib A Bello, Timothy JW Dawes, Jinming Duan, Carlo Biffi, Antonio De Marvao, Luke SGE Howard, J Simon R Gibbs, Martin R Wilkins, Stuart A Cook, Daniel Rueck-

ert, et al. Deep-learning cardiac motion analysis for human survival prediction. *Nature machine intelligence*, 1(2):95–104, 2019. 1.3.3.2, 3.1

[175] Timo Makela, Patrick Clarysse, Outi Sipila, Nicoleta Pauna, Quoc Cuong Pham, Toivo Katila, and Isabelle E Magnin. A review of cardiac image registration methods. *IEEE Transactions on medical imaging*, 21(9):1011–1021, 2002. 1.3.3.2

[176] Huaqi Qiu, Chen Qin, Loic Le Folgoc, Benjamin Hou, Jo Schlemper, and Daniel Rueckert. Deep learning for cardiac motion estimation: supervised vs. unsupervised training. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 186–194. Springer, 2019. 1.3.3.2, 3.1

[177] Manuel A Morales, David Izquierdo-Garcia, Iman Aganj, Jayashree Kalpathy-Cramer, Bruce R Rosen, and Ciprian Catana. Implementation and validation of a three-dimensional cardiac motion estimation network. *Radiology: Artificial Intelligence*, 1(4):e180080, 2019. 1.3.3.2, 3.1

[178] Wan-Chi Siu and Kwok-Wai Hung. Review of image interpolation and super-resolution. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–10. IEEE, 2012. 1.3.4, 1.3.4.1

[179] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12):3106–3121, 2019. 1.3.4

[180] Syed Muhammad Arsalan Bashir, Yi Wang, Mahrukh Khan, and Yilong Niu. A comprehensive review of deep learning-based single image super-resolution. *PeerJ Computer Science*, 7:e621, 2021. 1.3.4, 1.3.4.2

[181] Y Li, Bruno Sixou, and F Peyrin. A review of the deep learning methods for medical images super resolution problems. *Irbm*, 42(2):120–133, 2021. 1.3.4, 1.3.4.2, 5.1

[182] Hussein A Aly and Eric Dubois. Image up-sampling using total-variation regularization with a new observation model. *IEEE Transactions on Image Processing*, 14(10):1647–1659, 2005. 1.3.4.1

[183] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 1.3.5.3, 5.4

Chapter 2

Automated Segmentation of Cardiac Chambers from Cine Cardiac MRI Using an Adversarial Network Architecture

Cine cardiac magnetic resonance imaging , the current gold standard for cardiac function analysis, provides images with high spatio-temporal resolution. Computing clinical cardiac parameters like ventricular blood-pool volumes, ejection fraction and myocardial mass from these high resolution images is an important step in cardiac disease diagnosis, therapy planning and monitoring cardiac health. An accurate segmentation of left ventricle blood-pool, myocardium and right ventricle blood-pool is crucial for computing these clinical cardiac parameters[1]*. U-Net inspired models are the current state-of-the-art for medical image segmentation. SegAN, a*

[1]This chapter is adapted from:
Upendra R.R. et al., "An Adversarial Network Architecture Using 2D U-Net Models for Segmentation of Left Ventricle from Cine Cardiac MRI." In: Coudière Y., Ozenne V., Vigmond E., Zemzemi N. (eds) Functional Imaging and Modeling of the Heart. FIMH 2019. Lecture Notes in Computer Science, vol 11504. Springer, Cham.
Upendra R.R. et al., "Automated segmentation of cardiac chambers from cine cardiac MRI using an adversarial network architecture.", Proc. SPIE 11315, Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, 113152Y (16 March 2020).

novel adversarial network architecture with multi-scale loss function, has shown superior segmentation performance over U-Net models with single-scale loss function. Here, we compare the performance of stand-alone U-Net models and U-Net models in SegAN framework for segmentation of left ventricle blood-pool, myocardium and right ventricle blood-pool from the 2017 automated cardiac diagnosis challenge (ACDC) dataset.

2.1 Introduction

Cardiac magnetic resonance imaging (MRI), a non-invasive and non-ionizing radiation imaging modality, provides high resolution 3D images (parallel short axis slices stacked together) of the cardiac anatomy with superior soft tissue details. This makes cardiac MRI the current gold standard for cardiac function analysis [1, 2]. The analysis of the ventricular structure and function is an important step in cardiac disease diagnosis, treatment and prognosis. Cardiac function indices like stroke volume, ejection fraction, cardiac output, myocardium thickness and strain analysis play a crucial role in predicting and planning therapy for diseases like myocardial infarction, ischemia, arrhythmogenic right ventricular cardiomyopathy, pulmonary hypertension, dilated and hypertrophic cardiomyopathy [2]. The calculation of these cardiac function indices requires accurate delineation of the left ventricle (LV) blood-pool, the LV myocardium and the right ventricle (RV) blood-pool. Therefore, accurate and robust segmentation of these cardiac chambers from the MRI data plays an important role in a large number of cardiac problems.

Manual segmentation can be a very laborious task prone to significant user variability. Therefore, semi-automatic or fully automated segmentation methods would be very useful to cardiologists in the decision making process [3]. The automated segmentation of the cardiac chambers is challenging due to the fuzzy boundaries of the ventricular cavities, motion artifacts, banding artifacts, presence of trabeculae and papillary muscles, and shape variation across phases and pathologies.

Prior to deep learning, a number of automated segmentation algorithms for segmentation of the cardiac chambers from cine MRI images have been proposed [4, 2]. Traditional algorithms

such as thresholding, edge detection, region growing, clustering, etc., were proposed initially [3, 5]. These algorithms work decently for mid-ventricle slices, but often fail in the basal and apical slices. Also, they require considerable user-intervention. In graph based segmentation algorithms [6], graphs are created and a cost is assigned to each pixel or node. A minimum cost path is found using a graph searching algorithm to segment the left ventricle. These methods fail in complex cardiac structures, like papillary and trabecular muscles (PTMs). Active shape models (ASM) [7, 8] were proposed to segment the left ventricles using the energy minimization of rigidity and elasticity internally, and edges externally. ASM-based segmentation methods require extensive computational time and perform poorly in segmentation of cardiac chambers from low contrast images. To summarize, these non-deep learning algorithms require considerable manual or semi-manual interactions, fail to accurately delineate ventricles in basal and apical slices, and require extensive computational time. The inadequacies of these segmentation methods render them unsuitable for clinical applications.

In recent years, deep learning techniques have shown exceptional performance in image segmentation. With the availability of large number of medical images for supervised training, convolutional neural networks (CNN) significantly improved the medical image segmentation performance. The availability of large number of cardiac MRI images enabled the use of deep learning for segmentation of cardiac chambers. Several international challenges have been organized in the past few years to develop and evaluate segmentation algorithms for both the ventricles [9], [10], [11].

The introduction of U-Net by Ronneberger *et al.*[12], a fully convolutional network with a downsampling network that captures context information and an upsampling network that enables accurate localization of the annotated objects is currently the most popular method used for biomedical image segmentation. A majority of the medical image segmentation algorithms introduced in the past few years are variants of the U-Net model.

Generative adversarial networks (GAN)[13] are a type of adversarial networks in which two neural networks compete against each other in a min-max game to generate new image which is as close as possible to the original training image. This inspired algorithms like cycle-GAN

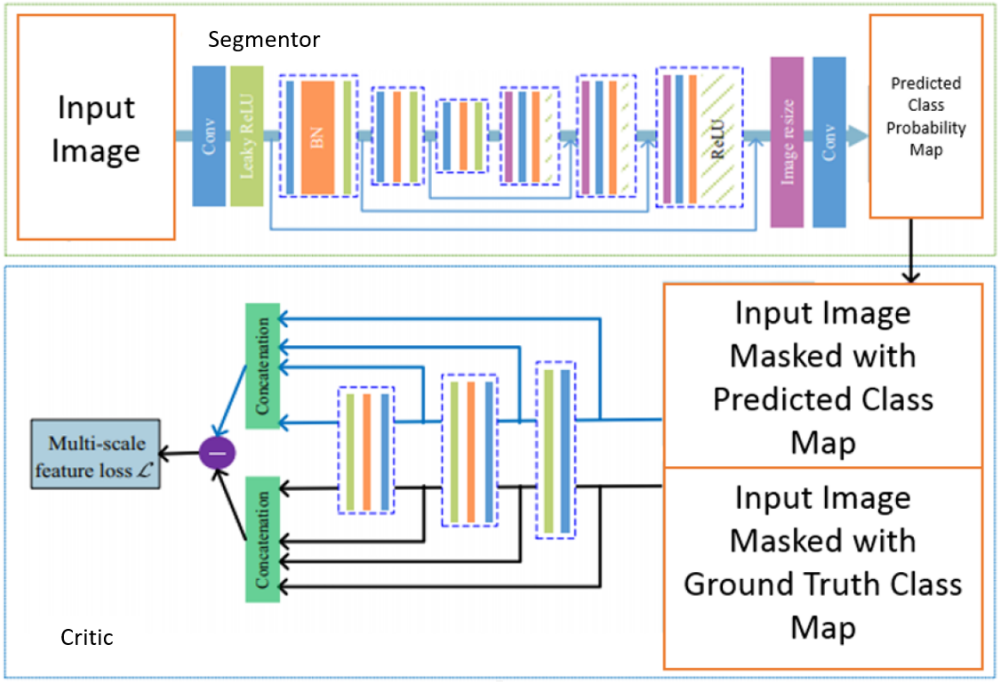


Figure 2.1: SegAN Architecture Inspired from GAN [15]

for automated segmentation of epithelial tissue from microscopic Drosophilia embryos images which outperformed the U-Net models [14]. Xue *et al.* [15] proposed SegAN, an end-to-end adversarial network architecture that achieved better Dice score than the U-Net models in the MICCAI BRATS (2013 and 2015) brain tumor segmentation challenge dataset.

Here, we combine U-Net models and its variants with SegAN adversarial architecture to segment the LV blood-pool on the 2017 ACDC segmentation challenge dataset. We then extend the work to investigate the viability of SegAN framework for multi-class segmentation of LV blood-pool, LV myocardium and RV blood-pool simultaneously and show its effect on clinical cardiac parameters like stroke volume, ejection fraction and myocardial mass.

2.2 Methodology

Inspired by GAN, Xue et al. have come up with SegAN, an adversarial network that has two networks, segmentor and critic, analagous to generator and discriminator in GAN, respectively. The segmentor, a fully convolutional neural network, takes in raw images as input and outputs

a probability label map. The critic network, which is the encoder part of the fully convolutional neural network needs two inputs - the masked image by the ground truth labels and the masked image by predicted labels obtained from the segmentor. The aim of the segmentor network is to minimize the L_1 loss function and the aim of critic network is to maximize the L_1 loss function [15].

2.2.1 Conventional GAN Models

In GANs, the loss function is defined as -

$$\min_{\theta_G} \max_{\theta_D} L(\theta_G, \theta_D) = E_{x \sim P_{data}}[logD(x)] + E_{z \sim P_z}[log(1 - D(G(z)))]. \tag{2.1}$$

In the above equation, θ_G and θ_D are the parameters of generator G and discriminator D, respectively. x and z are real image from unknown distribution P_{data} and random input for G from probability distribution P_z, respectively. The generator G outputs a high dimensional vector which is the input to the discriminator D. The discriminator D is trained to maximize the probability of assigning the correct label to the training data and the data generated from G. The generator G is simultaneously trained to minimize the objective function $log(1 - D(G(z)))$ to generate images that are difficult to differentiate for D [13]. The aim of the generator is to produce images that are as similar as possible to the real image and the aim of the discriminator is to successfully distinguish between the real image and the fake image produced by the generator.

2.2.2 Loss Function in SegAN

In SegAN, the aim is to solve the mapping between input images and their segmentation masks. The loss function L for SegAN is given by -

$$\min_{\theta_S} \max_{\theta_C} L(\theta_S, \theta_C) = \frac{1}{N} \sum_{n=1}^{N} l_{mae}(f_C(x_n \circ S(x_n)), f_C(x_n \circ y_n)). \tag{2.2}$$

In this equation, θ_S and θ_C are the parameters of segmentor S and critic C, respectively and N represents the number of training images. $(x_n \circ S(x_n))$ and $(x_n \circ y_n)$ are input images masked with segmentor predicted label map and ground truth, respectively. $f_c(x)$ are the features extracted from image x by critic and l_{mae} is the mean absolute error (MAE) given by

-

$$l_{mae}(f_C(x), f_C(x')) = \frac{1}{L}\sum_{i=1}^{L} ||f_C^i(x) - f_C^i(x')||_1, \tag{2.3}$$

with L representing the number of layers in the critic network [15].

The segmentor and critic networks are trained alternatively, just like GAN. The difference between GAN and SegAN is that GAN has two seperate losses for generator and discriminator, while, the SegAN has only one multi-scale L_1 loss funtion for both segmentor and critic.

2.2.3 Segmentor and Critic

We use three different segmentor networks to predict the segmented mask and compare their results. The first one is the original U-Net [16]. The second one is a U-Net architecture with skip connection used in [15] (U-Net A). The third segmentor used is a modified version of the U-Net architecture inspired from [17] (U-Net B). The input to all these three networks are raw images and the output is a predicted mask.

For the critic network, we used a similar structure to the downsampling part of the corresponding segmentor network to extract hierarchical features from multiple layers of the network. We then concatenated all these features extracted across multiple layers and computed the overall L_1 loss using the concatenated feature vector [15]. The input to the critic network are two images - input image masked with predicted class map and input image masked with the ground truth class map; and output is a feature vector.

2.2.4 Experiments and Implementation Details

2.2.4.1 Binary Segmentation

The focus of our experiment is to compare the results of a stand-alone 2D U-Net architecture with a SegAN architecture. For example, we obtain segmentation results using U-Net [16] with cross entropy loss as cost function. Then, we use this U-Net [16] as segmentor and the downsampling part of the U-Net as critic in the SegAN architecture with multi-scale L_1 loss as cost function. The results of these two networks are compared to determine if the SegAN architecture improves the segmentation results of the U-Net model. Experiments are performed with three variants of 2D U-Net architectures for the segmentation of LV blood-pool from the cardiac MRI data - the original ***U-Net*** from [16], the encoder-decoder network used as segmentor in [15] (***U-Net A***), and a modified U-Net inspired from [17] (***U-Net B***), the current state-of-the-art for left ventricle segmentation in the ACDC 2017 dataset.

The segmentor and the critic network are trained alternately using back-propagation and the loss function. First, the segmentor outputs a predicted class map. Then, the segmentor is fixed and the critic is trained in the next step using gradients calculated from the loss function. After that, the critic is fixed and the segmentor is trained using gradients from the loss function passed to the segmentor from the critic [15]. As explained in GANs, this process resembles a min-max game, where the segmentor aims to minimize the loss and the critic tries to maximize it. Provided additional data and more epochs, the segmentor will produce segmented masks i.e. labelled maps that are similar to the ground truth. For each U-Net model we use as segmentor, we use the encoder part of that particular U-Net model as critic.

We train the U-Net and SegAN models by resizing each slice to a 224x224 image and feeding it into the network with a learning rate of 0.0008, a batch size of 8, a decay of 0.5, a beta value of 0.5, one GPU and 50 epochs.

2.2.4.2 Multi-class Segmentation

The focus of this experiment is to compare the results of a stand-alone 2D U-Net architecture with a SegAN architecture for multi-class segmentation of LV blood-pool, LV myocardium

and RV blood-pool. Here, we perform our experiments on two stand-alone U-Net models - the original ***U-Net*** [12] and a modified U-Net (***U-Net A***) [15].

For stand-alone 2D U-Net model training, the images are resized to 224x224 and fed into the network in batches of 10. The U-Net model is trained using Adam optimizer with a learning rate of 0.0001 for 100 epochs.

In SegAN architecture, the input to the segmentor network is a 224x224x1 resized CMR image and the output is a 224x224x3 predicted class probability map, where the three layers correspond to the three different segmented classes - left ventricle blood-pool (LV), myocardium (MC) and right ventricle blood-pool (RV). Here, we train three different critic networks, one for each label class. The segmentor network and the three critic networks are trained using the average loss computed from the three different critic networks. Since we are training four networks (one segmentor network and three critic networks) per epoch, the images to SegAN network are fed in batches of two to avoid memory allocation issues. The SegAN network is trained using Adam optimizer with a learning rate of 0.00001 for 50 epochs. These experiments were performed on a machine equipped with NVIDIA RTX 2080 Ti GPU with 11GB of memory.

2.2.5 Dataset

The **A**utomated **C**ardiac **D**iagnosis **C**hallenge (ACDC) dataset was released during the MICCAI 2017 conference in conjunction with the STACOM workshop. The images were acquired using two different MRI scanners with different magnetic strength - 1.5 T and 3.0 T. The short axis slices cover the left ventricle from base to apex such that we get one image every 5 mm to 10 mm. A complete cardiac cycle is usually covered by 28 to 40 images. Their spatial resolution is 1.37 to 1.68 $mm^2/pixel$ [18]. The training dataset is composed of 100 subjects and the test dataset is composed of 50 subjects.

The image dataset corresponding to each subject consists of two image volumes, one at end-diastole and one at end-systole, with each containing 10 slices, therefore leading to a total of $1,902$ images. Since we do not have the ground truth for the 50 test subjects, we divide the

training dataset into 80 subjects for training and 20 subjects for validation. The evaluation metrics in this paper are the result of 5-fold cross validation of the training dataset.

2.3 Results

Table 2.1 summarizes the segmentation performance of the investigated frameworks with and without the SegAN integration for binary segmentation (LV blood-pool only). We can observe that the mean Dice scores and mean IoU values of the three SegAN architectures are higher than their corresponding U-Net models. To compare the performance the three stand-alone U-Net models with their SegAN frameworks, we conducted a statistical significance (T-test) test. The mean Dice score showed significant improvement ($p < 0.05$) from 93.41% (***U-Net***) to 94.71% (***SegAN + U-Net***), ($p < 0.1$) from 92.62% (***U-Net A***) to 93.88% (***SegAN + U-Net A***) and ($p < 0.05$) from 94.91% (***U-Net B***) to 95.87% (***SegAN + U-Net B***) in end diastole, and ($p < 0.1$) from 90.30% (***U-Net A***) to 91.10% (***SegAN + U-Net A***) and ($p < 0.1$) from 92.72% (***U-Net B***) to 93.14% (***SegAN + U-Net B***) in end systole. The mean IoU values showed significant improvement ($p < 0.05$) from 87.25% (***U-Net***) to 89.55% (***SegAN + U-Net***) and ($p < 0.1$) from 91.55% (***U-Net B***) to 92.94% (***SegAN + U-Net B***) in end diastole.

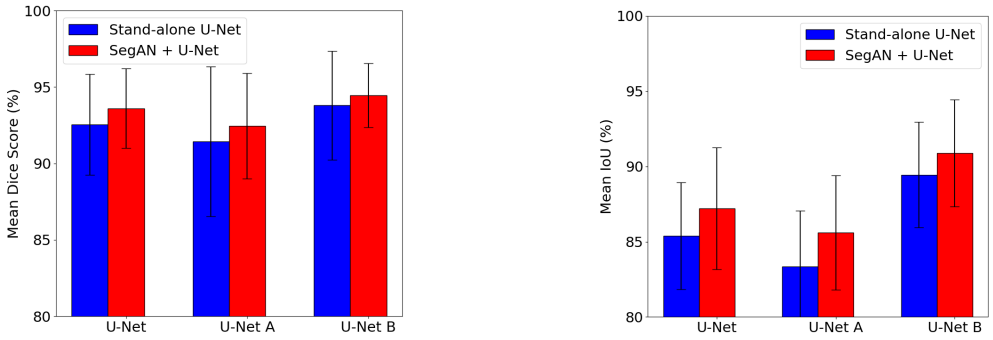
The highest mean Dice score and mean IoU in our experiments are obtained using the SegAN architecture with U-Net B as its segmentor network and the U-Net B's encoder as the critic network. The ***SegAN + U-Net B*** outperforms ***U-Net*** by 2.46% (Dice) in end diastole and 1.40% in end systole.

In Fig. 2.2, it can be observed that the segmentation performance of SegAN frameworks (shown in red) is better than the performance of the corresponding stand-alone U-Net architectures (shown in blue). When we use these U-Net models as segmentor, we see significant improvement in both Dice score and IoU, for ED and ES.

Fig. 2.3 shows examples of mid, apical and basal slices of the heart and the corresponding segmented masks using the six architectures. The white regions represent the overlap between the ground truth mask and the tested mask. The red and blue regions represent the false

Table 2.1: Segmentation evaluation, mean (std-dev) for end diastole (ED) and end systole (ES) left ventricle segmentation in the 2017 ACDC dataset. Statistical significance (T-test) of the results of SegAn architecture compared against U-Net models are represented by * for $p < 0.1$ and ** for $p < 0.05$. The best Dice values achieved are labeled in **bold**.

	Dice (ED) (%)	IoU (ED) (%)	Dice (ES) (%)	IoU (ES) (%)
U-Net	93.41 (4.23)	87.25 (3.12)	91.75 (2.26)	83.64 (4.01)
SegAN + U-Net	*94.71 (1.24)***	*89.55 (2.46)***	*92.54 (3.89)*	*84.91 (5.75)*
U-Net A	92.62 (2.75)	85.27 (1.81)	90.30 (7.11)	81.58 (5.60)
SegAN + U-Net A	*93.88 (2.86)**	*88.54 (1.12)*	*91.10 (4.15)**	*82.74 (5.71)*
U-Net B	94.91 (2.40)	91.55 (3.23)	92.72 (4.71)	87.44 (3.81)
SegAN + U-Net B	***95.87 (1.71)****	***92.94 (3.27)****	***93.14 (2.56)****	***88.94 (3.92)***



(a) (b)

Figure 2.2: Comparison of (a) mean Dice scores and (b) mean IoU values of U-Net models and its corresponding SegAN architecture, for left ventricle segmentation

positive (pixels predicted as left ventricle by the tested algorithm, but not annotated in the ground truth), and false negative (pixels not predicted as left ventricle by the tested algorithm, but annotated in the ground truth) regions, respectively.

In Table 2.2, we summarize the segmentation performance of U-Net models with and without SegAN integration for multi-class segmentation. We can observe that the SegAN architecture, when integrated with both the variants of U-Net models, achieves better Dice score for LV blood-pool and RV blood-pool segmentation in both end diastole and end systole phases. However, in case of LV myocardium segmentation, the Dice scores achieved in the stand-alone U-Net models training are better or similar to the U-Net models trained in SegAN

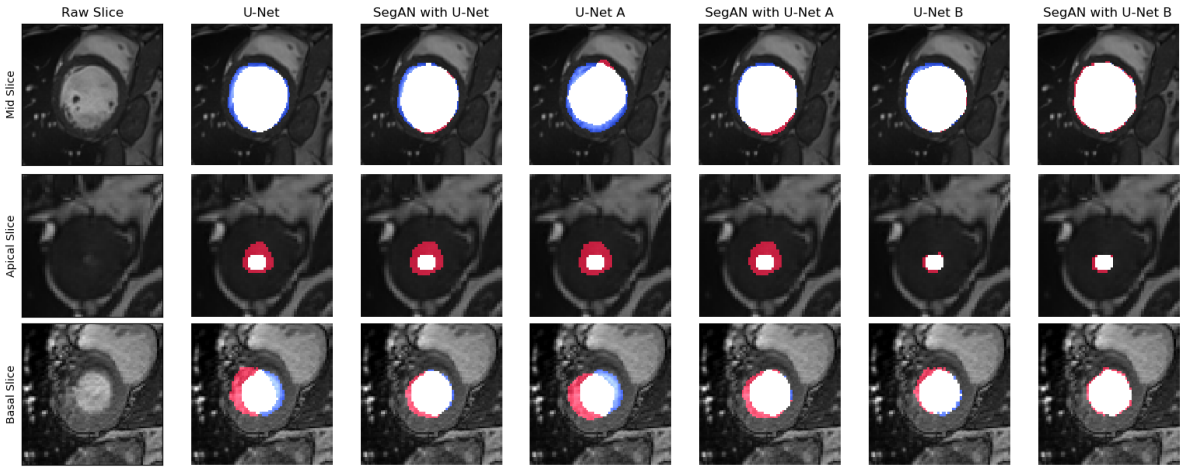


Figure 2.3: Examples of segmentation of the left ventricle in mid, apical and basal slice (top to bottom). The white, red and blue regions represent true positives, false positives and false negatives, respectively.

framework. Fig. 2.4 shows examples of the segmented LV blood-pool, LV myocardium and RV blood-pool in mid, apical and basal slices for the two variants of U-Net models, with and without SegAN integration.

We also investigate the viability of SegAN framework for multi-class segmentation by evaluating the clinical cardiac parameters like LV stroke volume, LV ejection fraction, RV stroke volume, RV ejection fraction and myocardial mass. In Table 2.3, we show the correlation coefficient of these clinical cardiac parameters calculated using the segmentation results obtained from the above mentioned U-Net models and its SegAN integrated counterparts with the clinical cardiac parameters calculated using the ground truth. The correlation coefficient values of ventricular stroke volume and the ejection fraction computed from the segmentation results of SegAN framework are higher than the correlation coefficient values of ventricular stroke volume and the ejection fraction computed from the U-Net models. However, the correlation coefficient values of myocardial mass computed from the segmentation results of SegAN framework are lower than the correlation coefficient values of myocardial mass computed from the U-Net models. These observations are in agreement with the Dice score results shown in Table 2.2, where the segmentation of the myocardial tissue using SegAN integration is not superior to the segmentation using stand-alone U-Net models, i.e., the myocardial mass esti-

Table 2.2: Segmentation evaluation, mean Dice score (std-dev) for end diastole (ED) and end systole (ES) for left ventricle blood-pool (LV), myocardium (MC) and right ventricle blood-pool (RV) segmentation in the 2017 ACDC dataset. Statistical significance (T-test) of the results of SegAN architecture compared against U-Net models are represented by * for $p < 0.05$ and ** for $p < 0.005$. The best Dice scores achieved are labeled in **bold**.

	LV Dice (ED) (%)	LV Dice (ES) (%)	MC Dice (ED) (%)	MC Dice (ES) (%)	RV Dice (ED) (%)	RV Dice (ES) (%)
U-Net	89.04 (1.97)	88.65 (2.03)	90.62 (2.72)	88.17 (3.21)	88.09 (2.05)	88.80 1.92
SegAN + U-Net	91.69 (1.49)**	90.29 (1.61)**	88.19 (4.21)	88.42 (5.93)	91.55 (4.19)	90.06 (1.74)**
U-Net A	90.14 (1.78)	88.29 (1.55)	**91.61 (2.14)**	**88.89 (2.32)****	90.55 (4.41)	87.41 (4.78)
SegAN + U-Net A	**92.19 (1.67)***	**91.09 (2.08)***	91.18 (3.50)	87.95 (7.97)	**92.06 (3.41)**	**90.05 (2.02)***

Table 2.3: Evaluation of clinical indices - LV stroke volume (SV) correlation coefficient, LV ejection fraction (EF) correlation coefficient, myocardium mass correlation coefficient, RV stroke volume correlation coefficient and RV ejection fraction correlation coefficient.

	LV SV Correlation	LV EF Correlation	MC Mass Correlation	RV SV Correlation	RV EF Correlation
U-Net	0.923	0.904	0.955	0.899	0.841
SegAN + U-Net	0.941	0.918	0.931	**0.944**	0.901
U-Net A	0.937	0.957	**0.956**	0.908	0.882
SegAN + U-Net A	**0.974**	**0.963**	0.937	0.935	**0.921**

mates are affected since they are directly related to uncertainties present in the myocardium segmentation.

2.4 Discussion

In this work, the integration of U-Net models into the SegAN framework is evaluated on the 2017 ACDC segmentation challenge dataset. Our goal was to investigate if the SegAN framework improves the segmentation performance of U-Net models. Our experiments reveal that U-Net models, when trained in the SegAN framework, produces significantly better seg-

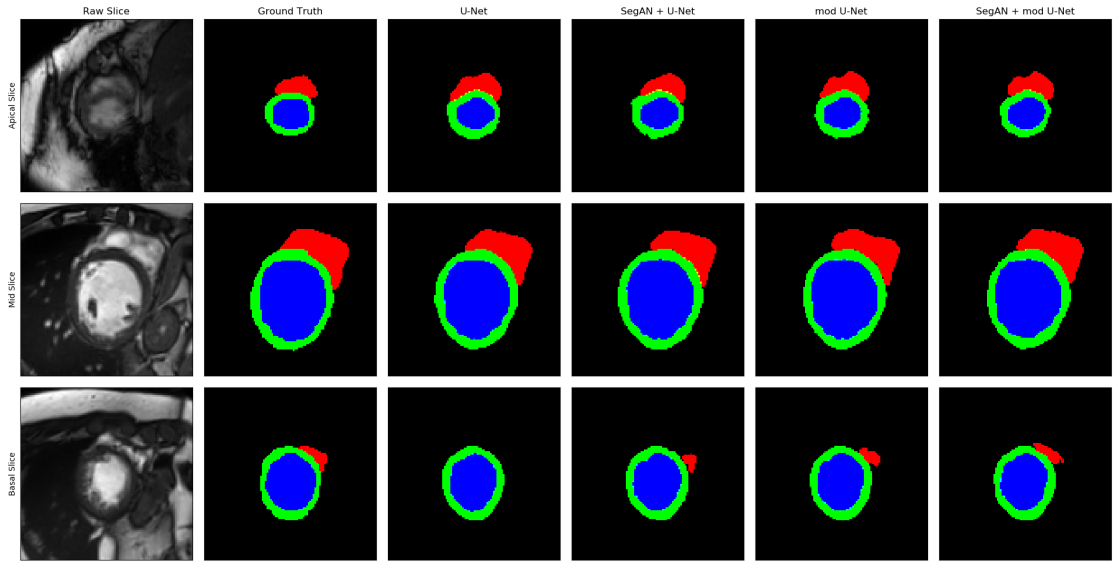


Figure 2.4: Examples of segmentation of the left ventricle blood-pool (blue), myocardium (green) and right ventricle blood-pool (red) in mid, apical and basal slices (top to bottom)

mentation results than when trained stand-alone, consistently. The features extracted across multiple layers of the critic network and concatenated into the feature vector used to compute the multi-scale L_1 loss captures pixel-, low-, mid- and high-level features. This multi-resolution approach to feature extraction enables the SegAN model to learn the dissimilarities between the generated and the ground truth segmentation maps across the multiple layers of the critic network.

We use cross entropy loss as cost function for training U-Net models. We also experimented with training all the U-Net variants with a Dice loss cost function, however the results indicated a consistently lower performance than that achieved using cross entropy loss. We also experimented with multi-scale L_2 loss as cost function for training the SegAN models, however, the results were not very consistent. Further investigation with multi-scale L_2 loss as cost function will be conducted, to determine if it can outperform multi-scale L_1 loss as cost function.

To evaluate our method, we used a 5-fold cross-validation strategy, in which we employed five different combinations of 80 training and 20 testing datasets from the available

100 datasets. This is a common approach used to validate novel deep learning techniques, as it enables testing the robustness of the method across different training datasets, while also removing the bias associated with a single 80 training - 20 testing data split.

The segmentation results of multi-class segmentation are in compliance with the results of the brain tumor segmentation in BRATS 2015 dataset[15]. The SegAN architecture obtained better Dice score in segmenting the whole brain tumor region, but had some drawbacks in segmenting the tumor core and Gd-enhanced tumor core. This is attributed to the fact that the SegAN architecture extracts features from multiple layers of the critic network and the segmentation of regions of smaller areas, like the left ventricle myocardium, may require more concentration at pixel-level features. Therefore, U-Net model with cross entropy loss (pixel-level loss) could have better segmentation performance than SegAN architecture with multi-scale L_1 loss for segmentation of the left ventricle myocardium.

The major drawback of the SegAN architecture in multi-class segmentation is the computational time and the memory required to train one segmentor network and three critic networks simultaneously. The computational time required for one epoch for an U-Net model is around 225 seconds, whereas the U-Net model in SegAN framework requires around 900 seconds (for multi-class segmentation).

2.5 Conclusion and Future Work

In this paper, we demonstrate the use of an adversarial architecture, SegAN with multi-scale L_1 loss function, to segment the LV blood-pool from cine cardiac MR images. Encouraged by the results, we extended the SegAN framework for multi-class segmentation of LV blood-pool, LV myocardium and RV blood-pool. This multi-scale L_1 loss function captures features at multiple levels - pixel-level, superpixel-level and patches-level. Our experiments reveal that this integration of U-Net models in the SegAN framework leads to significant improvement of LV blood-pool and RV blood-pool segmentation in the 2017 ACDC segmentation challenge dataset. The adversarial nature of the architecture and the multi-resolution approach enables the SegAN model to accurately segment the above mentioned heart chambers, which in turn,

enables accurate computation of critical clinical parameters like ventricular stroke volumes and ejection fraction.

We observed that the segmentation result of LV myocardium did not improve with the SegAN integration. An alternative solution to this could be an integration of a weighted cross entropy loss as cost function along with the multi-scale L_1 loss function.

Bibliography

[1] A Gerche La, Guido Claessen, A de Bruaene Van, Nele Pattyn, J Cleemput Van, Marc Gewillig, Jan Bogaert, Steven Dymarkowski, Piet Claus, and Hein Heidbuchel. Cardiac MRI: a new gold standard for ventricular volume quantification during high-intensity exercise. *Circulation. Cardiovascular imaging*, 6(2):329–338, 2013. 2.1

[2] 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, 2016. 1.3.1, 1.3.1.1, 2.1

[3] Mina Nasr-Esfahani, Majid Mohrekesh, Mojtaba Akbari, SM Soroushmehr, Ebrahim Nasr-Esfahani, Nader Karimi, Shadrokh Samavi, and Kayvan Najarian. Left ventricle segmentation in cardiac MR images using fully convolutional network. *arXiv preprint arXiv:1802.07778*, 2018. 2.1

[4] 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. 1.3.1, 2.1

[5] Mustafa A Alattar, Nael F Osman, and Ahmed S Fahmy. Myocardial segmentation using constrained multi-seeded region growing. In *International Conference Image Analysis and Recognition*, pages 89–98. Springer, 2010. 2.1

[6] Xiang Lin, Brett Cowan, and Alistair Young. Model-based graph cut method for segmentation of the left ventricle. In *Engineering in Medicine and Biology Society, 2005.*

IEEE-EMBS 2005. 27th Annual International Conference of the, pages 3059–3062. IEEE, 2006. 2.1

[7] Carlos Santiago, Jacinto C Nascimento, and Jorge S Marques. A new ASM framework for left ventricle segmentation exploring slice variability in cardiac MRI volumes. *Neural Computing and Applications*, 28(9):2489–2500, 2017. 2.1

[8] Georgios Tzimiropoulos and Maja Pantic. Optimization problems for fast AAM fitting in-the-wild. In *Proceedings of the IEEE international conference on computer vision*, pages 593–600, 2013. 2.1

[9] Avan Suinesiaputra, Brett R Cowan, J Paul Finn, Carissa G Fonseca, Alan H Kadish, Daniel C Lee, Pau Medrano-Gracia, Simon K Warfield, Wenchao Tao, and Alistair A Young. Left ventricular segmentation challenge from cardiac MRI: a collation study. In *International Workshop on Statistical Atlases and Computational Models of the Heart. Lecture Notes in Computer Science*, volume 7085, pages 88–97, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. 2.1

[10] 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, et al. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Medical image analysis*, 18(1):50–62, 2014. 2.1

[11] Caroline Petitjean, Maria A Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jérôme Caudron, Su Ruan, Ismail Ben Ayed, M Jorge Cardoso, Hsiang-Chou Chen, et al. Right ventricle segmentation from cardiac MRI: a collation study. *Medical image analysis*, 19(1):187–202, 2015. 2.1

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image com-*

puting and computer-assisted intervention, pages 234–241. Springer, 2015. 1.6, 1.3.1.3, 2.1, 2.2.4.2, 3.2.2, 5.2.2, 5.2.3.4

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 2.1, 2.2.1

[14] Matthias Haering, Joerg Grosshans, Fred Wolf, and Stephan Eule. Automated segmentation of epithelial tissue using cycle-consistent generative adversarial networks. *bioRxiv*, page 311373, 2018. 2.1

[15] Yuan Xue, Tao Xu, Han Zhang, L Rodney Long, and Xiaolei Huang. Segan: Adversarial network with multi-scale L_1 loss for medical image segmentation. *Neuroinformatics*, 16(3-4):383–392, 2018. 2.1, 2.1, 2.2, 2.2.2, 2.2.3, 2.2.4.1, 2.2.4.2, 2.4

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention. Lecture Notes in Computer Science*, volume 9351, pages 234–241. Springer, 2015. 2.2.3, 2.2.4.1, 4.1

[17] Fabian Isensee, Paul F Jaeger, Peter M Full, Ivo Wolf, Sandy Engelhardt, and Klaus H Maier-Hein. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In *International Workshop on Statistical Atlases and Computational Models of the Heart. Lecture notes in Computer Science*, volume 10663, pages 120–129. Springer, 2017. 1.3.1.3, 2.2.3, 2.2.4.1, 3.2.2, 3.3.0.1, 3.3.0.1

[18] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. 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, 2018. 1.3.1.3, 1.4, 2.2.5, 3.1

Chapter 3

CNN-Based Cardiac Motion Extraction to Generate Deformable Geometric Ventricular Models from Cine MRI

Patient-specific left ventricle (LV) myocardial models have the potential to be used in a variety of clinical scenarios for improved diagnosis and treatment plans. Cine cardiac magnetic resonance (MR) imaging provides high resolution images to reconstruct patient-specific geometric models of the LV myocardium.[1] *Here, we propose a deep leaning-based framework for the development of patient-specific geometric models of LV myocardium from cine cardiac MR images, using the Automated Cardiac Diagnosis Challenge (ACDC) dataset. We use the deformation field estimated from the VoxelMorph-based convolutional neural network (CNN) to propagate*

[1]This chapter is adapted from:
Upendra R.R. et al., "A convolutional neural network-based deformable image registration method for cardiac motion estimation from cine cardiac MR images." In 2020 Computing in Cardiology (pp. 1-4). IEEE.
Upendra R.R. et al., "CNN-Based Cardiac Motion Extraction to Generate Deformable Geometric Left Ventricle Myocardial Models from Cine MRI.", In: Ennis D.B., Perotti L.E., Wang V.Y. (eds) Functional Imaging and Modeling of the Heart. FIMH 2021. Lecture Notes in Computer Science, vol 12738. Springer, Cham.
Upendra R.R. et al., "Motion Extraction of the Right Ventricle from 4D Cardiac Cine MRI Using a Deep Learning-Based Deformable Registration Framework." In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC) (pp. 3795-3799). IEEE.

the isosurface mesh and volume mesh of the end-diastole (ED) frame to the subsequent frames of the cardiac cycle. We assess the CNN-based propagated models against segmented models at each cardiac phase, as well as models propagated using another traditional nonrigid image registration technique. Additionally, we generate dynamic LV myocardial volume meshes at all phases of the cardiac cycle using the log barrier-based mesh warping (LBWARP) method and compare them with the CNN-propagated volume meshes.

Furthermore, we describe the development of dynamic patient-specific right ventricle (RV) models associated with normal subjects and abnormal RV patients to be subsequently used to assess RV function based on motion and kinematic analysis. To this end, we first constructed static RV models using segmentation masks of cardiac chambers generated from our accurate, memory-efficient deep neural architecture – CondenseUNet – featuring both a learned group structure and a regularized weight-pruner to estimate the motion of the RV. We then use the deformation field estimated from the proposed deep learning-based deformable network to propagate the RV isosurface mesh of the ED frame to the subsequent frames of the cardiac cycle.

3.1 Introduction

To reduce the morbidity and mortality associated with cardiovascular diseases (CVDs) [1], and to improve their treatment, it is crucial to detect and predict the progression of the diseases at an early stage. In a clinical set-up, population-based metrics, including measurements of cardiac wall motion, ventricular volumes, cardiac chamber flow patterns, etc., derived from cardiac imaging are used for diagnosis, prognosis and therapy planning.

In recent years, image-based computational models have been increasingly used to study ventricular mechanics associated with various cardiac conditions. A comprehensive review of patient-specific cardiovascular modeling and its applications is described in [2]. Cardiovascular patient-specific modeling includes a geometric representation of some or all cardiac chambers of the patient's anatomy and is derived from different imaging modalities [3].

The construction of patient-specific geometric models entails several steps: clinical imaging, segmentation and geometry reconstruction, and spatial discretization (i.e., mesh generation)

[4]. For example, Bello *et al.* [5] presented a deep learning based framework for human survival prediction for patients diagnosed with pulmonary hypertension using cine cardiac MR images. Here, the authors employ a 4D spatio-temporal B-spline image registration method to estimate the deformation field at each voxel and at each timeframe. The estimated deformation field was used to propagate the ED surface mesh of the right ventricle (RV), reconstructed from the segmentation map, to the rest of the timeframes of a particular subject. Qin *et al.* [6] proposed a joint deep learning network for cardiac motion estimation and segmentation of 2D cine cardiac MR images. Qiu *et al.* [7] compared the performance of supervised and unsupervised training strategies for cardiac motion estimation using convolutional neural networks (CNN), performed in the 2D plane. Morales *et al.* [8] proposed an unsupervised CNN-based 3D deformable registration method for cardiac motion estimation; however, they do not account for the out-of-plane motion of the two-dimensional stack of the CMR images that leads to slice misalignment.

In this work, we propose a deep learning-based pipeline to develop patient-specific geometric models of the LV myocardium from cine cardiac MR images (Fig. 3.1). These models may be used to conduct various simulations, such as assessing myocardial viability. We introduce a preliminary, proof of concept, CNN-based 4D deformable registration method for cardiac motion estimation from cine cardiac MR images, using the Automated Cardiac Diagnosis Challenge (ACDC) dataset [9]. We then demonstrate the use of the CNN-based 4D deformable registration technique to build dynamic patient-specific LV myocardial models across subjects with different pathologies, namely normal, dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and subjects with prior myocardial infarctions (MINF). Following segmentation of the ED cardiac frame, we generate both isosurface and volume LV meshes, which we then propagate through the cardiac cycle using the CNN-based registration fields. In addition, we demonstrate the generation of dynamic LV volume meshes depicting the heart at various cardiac phases by warping a patient-specific ED volume mesh based on the registration-based propagated surface meshes, using the LBWARP method [10]. Lastly, we compare these meshes to those obtained by directly propagating the ED volume mesh using

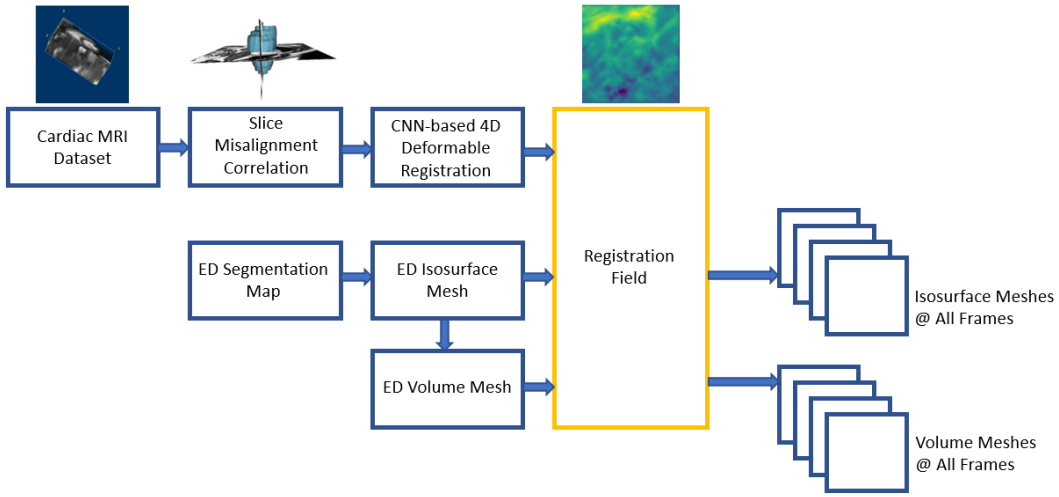


Figure 3.1: Overview of the proposed CNN-based workflow to generate patient-specific LV myocardial geometric model.

the CNN-based deformation fields.

Although cardiac cine MRI has provided a non-invasive method for studying global and regional function of the heart, most of these studies have been centered on the LV. In light of the thin wall structure of the RV and its asymmetric geometry, there have only been very few research endeavors exploring the kinematics of RV, including the extraction of the RV motion and generation of patient-specific RV anatomical models. Therefore, we extend our proposed approach for extracting the RV motion from cine cardiac MR image sequences and generate deformable endocardial RV models that can be later used to study RV kinematics as a biomarker for studying RV-related cardiac disease (Fig. 3.2).

Hence, we propose the deep learning-based approach for extracting the frame-to-frame RV motion from cine cardiac images, and using this motion, along with segmented isosurface meshes at ED, to generate dynamic, deformable models of the RV. Here, we illustrate the potential of the CNN-based 4D deformable registration technique to build dynamic patient-specific RV models across subjects with normal and abnormal RVs. We used the segmented mask of the RV endocardium at all cardiac frames generated via CondenseUNet [11], which substitutes the concept of both standard convolution and group convolution (G-Conv) with

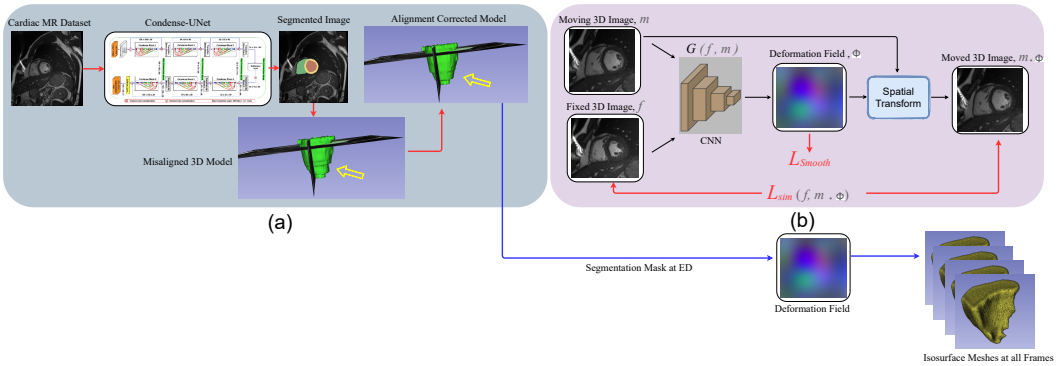


Figure 3.2: Image segmentation and deformable registration pipeline: a) ED frame segmentation and slice misalignment correction; b) deep learning registration framework. The CNN $G(f, m)$ learns to predict the deformation field and register the moving 3D image to the fixed 3D image to generate the transformed image using the spatial transformation function.

learned group-convolution (LG-Conv). Following segmentation of the ED cardiac frame, we generate isosurface meshes, which we then propagate through the cardiac cycle using the CNN-based registration fields. Lastly, we compare these propagated isosurface meshes to those generated directly from the segmentation masks obtained from CondenseUNet [11].

3.2 Methodology

3.2.1 Cardiac MRI Data

We use the 2017 ACDC dataset that was acquired from real clinical exams. The dataset is composed of cine cardiac MR images from 150 subjects, divided into five equally-distributed subgroups: normal, MINF, DCM, HCM and abnormal RV. The MR image acquisitions were obtained using two different MR scanners of 1.5 T and 3.0 T magnetic strength. These series of short axis slices cover the LV from base to apex such that one image is captured every 5 mm to 10 mm with a spatial resolution of 1.37 mm^2/pixel to 1.68 mm^2/pixel.

3.2.2 Image Preprocessing

We first correct for the inherent slice misalignments that occur during the cine cardiac MR image acquisition by leveraging the slice misalignment correction method presented by Dangi

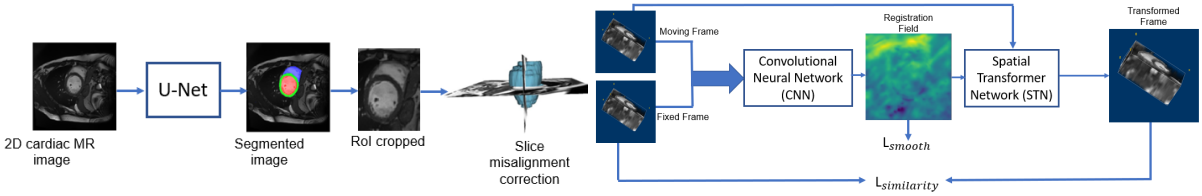


Figure 3.3: (a) Slice misalignment correction and (b) 4D deformable registration workflow.

et al. [12]. We train a modified version of the U-Net model [13] inspired from Isensee *et al.* [14], to segment the cardiac chambers (LV blood-pool, LV myocardium and RV blood-pool) from 2D cine cardiac MR slices. We use these predicted segmentation maps to crop the regions of interest (RoI) and to identify the centers of the LV blood-pool. The 2D slices are stacked such that the LV blood-pool centers are collinear, resulting in a slice misalignment corrected 3D CMR image (Fig. 3.3a). The U-Net model is trained on the ED and end-systole (ES) frames of the CMR data of 80 subjects and validated on 20 subjects.

3.2.3 Deformable Image Registration

3.2.3.1 CNN-based Image Registration.

We employ the VoxelMorph [15] framework to find an optical flow representation between a sequence of 3D image pairs $\{(I_{ED}, I_{ED+t})\}_{t=1,2,3,\ldots,N_T-1}$ where N_T is the total number of frames, and at each iteration, an image pair (I_{ED}, I_{ED+t}) is input to the CNN and a registration field ϕ is output. The registration field is fed to a spatial transformer network (STN) [16] along with the ED frame, I_{ED}, to produce a warped image, $I_{ED} \circ \phi$ (Fig. 3.3b).

To train the CNN, a loss function consisting of two components is used to optimize the network:

$$L = L_{\text{similarity}} + \lambda L_{\text{smooth}}, \tag{3.1}$$

where $L_{\text{similarity}}$ is the mean squared error (MSE) between the target frame I_{ED+t} and the

warped ED frame $I_{ED} \circ \phi$:

$$MSE = \frac{1}{|\Omega|} \sum_{i \in \Omega} \left(I_{ED+t}(i) - [I_{ED} \circ \phi](i) \right)^2, \tag{3.2}$$

where Ω is the spatial domain of the images and $i \in \mathbb{R}^2$ is the position of a point on the frame. The second term in the loss function (eq. (5.1)) is a smoothing loss function L_{smooth} that spatially smoothes the registration field ϕ and λ is the regularization parameter. In general, a diffusion regularizer on the spatial gradients of the registration field is used as the smoothing loss function and is given

$$L_{smooth} = \sum_{i \in \Omega} ||\nabla \phi(\mathrm{i})||^2. \tag{3.3}$$

Here we experiment with a Laplacian operator in the smoothing loss function, inspired from Zhu *et al.* [17]:

$$L_{smooth} = \sum_{i \in \Omega} ||\Delta \phi(\mathrm{i})||^2. \tag{3.4}$$

Unlike the gradient operator, which only considers the local properties of the objective function $y = x^2$, the Laplacian operator considers the global properties of the function $y = x^2$, i.e., it considers the slope magnitude and its trends when choosing a direction. [17].

We divide the total 150 MRI dataset into 110 for training, 10 for validation and 30 for testing. All the cropped input cine MRI frames are resampled to 96 $\times$ 96 $\times$ 16 voxels with 1.5 mm isotropic resolution. We train the VoxelMorph CNN using the Adam optimizer with a learning rate of 10^{-4}, halved at every 10^{th} epoch for 50 epochs on a machine equipped with a NVIDIA RTX 2080 Ti GPU with 11 GB of memory; the regularization parameter λ is set to 10^{-3}.

3.2.4 Mesh Generation and Propagation

3.2.4.1 Left Ventricle Myocardium

We use the manual segmentation map of the ED frame to generate isosurface meshes. The slice thickness of each MRI image slice is 5 mm to 10 mm; however, in order to obtain good quality meshes, the segmentation maps were resampled to a slice thickness of 1 mm. We use the Lewiner marching cubes [18] algorithm to generate the meshes from the resampled segmentation maps of the ED frames, on an Intel(R) Core(TM) i9-9900K CPU, and then simplification techniques, such as r-refinement and edge collapse, were performed using MeshLab 2020.07 [19]. The simplification techniques are repeated multiple times to reduce the number of vertices until the mesh has been fully decimated while preserving the anatomical integrity and aspect ratio of the isosurface meshes.

Volume meshes of the initial surface meshes at the ED phases for four patients with various heart conditions were generated based on the decimated patient-specific surface meshes using Tetgen 1.6 [20]. In particular, a constrained Delaunay mesh generation algorithm was used to generate tetrahedral meshes based on the triangulated surface meshes. Steiner points were added within the boundary of the surface mesh so that the tetrahedra maintained a radius-edge ratio of 1.01 and a maximum volume of 9 mm^3 as needed for generation of valid meshes [20]. Volume mesh quality improvement was performed using the feasible Newton method in Mesquite [21]. This method iteratively minimizes the quadratic approximation of a nonlinear function and converges linearly toward a local minimum while performing an Armijo line search to ensure feasibility of the elements; feasibility in this case refers to a valid, non-inverted element. The volume mesh converged to the highest quality indicated by the minimum average scaled Jacobian of the elements in the mesh.

To demonstrate the VoxelMorph-based motion extraction and propagation to build patient-specific LV myocardial models, we generate two sets of volume meshes at each cardiac frame for each patient in each pathology group (Fig. 3.4).

The first set is produced by propagating the volume meshes at the ED frame to all the subsequent frames of the cardiac cycle using the deformation field estimated by the VoxelMorph-

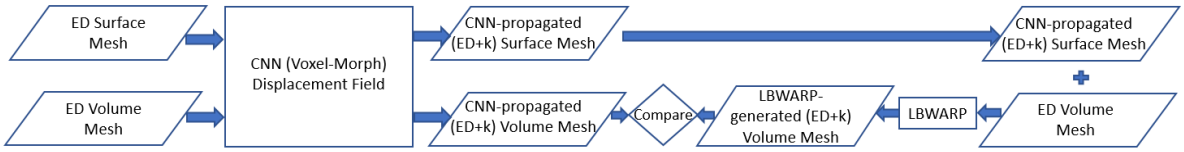


Figure 3.4: Pipeline to generate dynamic volume meshes (at cardiac frames (ED + k)) by direct CNN-based propagation, as well as volume mesh warping based on dynamic boundary meshes.

based registration method. For the second set, the ED volume mesh generated with Tetgen was used to generate the volume meshes corresponding to the other cardiac phases. We employed the LBWARP method [10] to deform the ED volume mesh onto the target surface mesh for the new cardiac phase (Fig. 3.5). The method computes new positions for the interior vertices in the ED volume mesh, while maintaining the mesh topology and point-to-point correspondence [10]. The simplification of the ED isosurface meshes, generation of the ED volume meshes and generation of the volume meshes corresponding to the other cardiac phases using the LBWARP method were performed on a machine equipped with AMD FX(tm)-6300 Six-Cores processor and a NVIDIA GeForce GTX 1050 Ti graphics card.

Briefly, LBWARP first calculates a set of local weights for each interior vertex in the initial ED volume mesh based on the relative inverse distances from each of its neighbors. The projected Newton method is used to solve the strictly convex optimization problems. For each set of local weights, a sparse system of linear equations is formed specifying the representation of each interior vertex in terms of its neighbors. Next, the boundary vertices in the ED surface mesh are mapped onto the new surface boundary. Finally, the interior vertices in the ED volume mesh are repositioned by solving the original system of linear equations with new right-hand side vectors to reflect the updated positions of the boundary nodes, while maintaining edge connectivity and point-to-point correspondence, and ultimately yielding the volume meshes that correspond to each new cardiac phase.

Baseline Comparisons We compare the performance of the VoxelMorph framework with that of the B-spline free form deformation (FFD) non-rigid image registration algorithm [22].

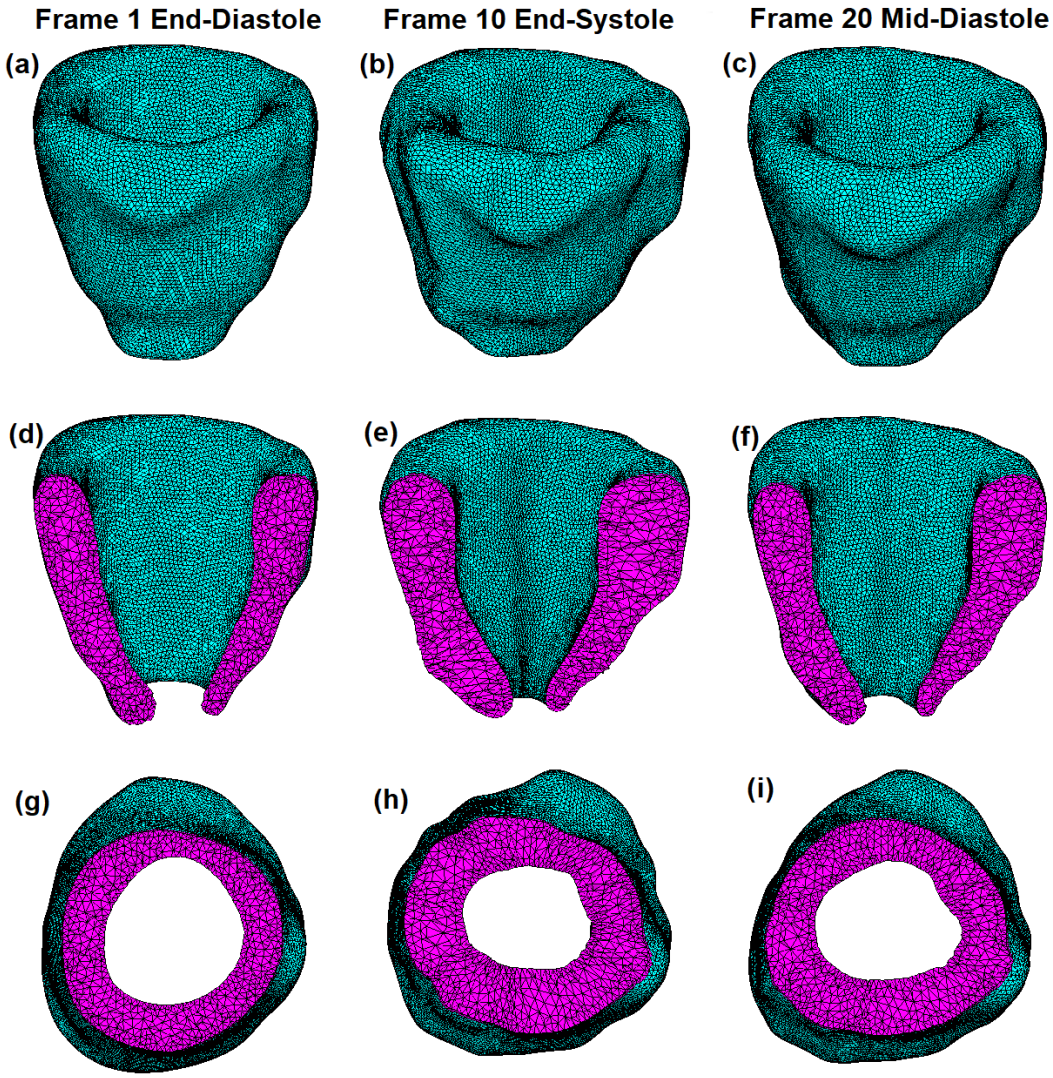


Figure 3.5: Warped volume meshes for a patient with a healthy heart generated using LB-WARP at three cardiac phases (a) end-diastole; (b) end-systole; and (c) mid-diastole; (d-f) Long axis cutaway view of volume meshes at the three cardiac phases, respectively; (g-i) short-axis cutaway view of volume meshes at the three cardiac phases, respectively.

This iterative intensity-based image registration method was implemented using SimpleElastix [23, 24], which enables a variety of image registration algorithms in different programming languages. The FFD algorithm was set to use the adaptive stochastic gradient descent method as the optimizer, MSE as the similarity measure and binding energy as the regularization function. The FFD-based image registration was optimized in 500 iterations, while sampling

2048 random points per iteration, on an Intel(R) Core(TM) i9-9900K CPU.

3.2.4.2 Right Ventricle Blood-Pool

To extract RV anatomical information for incorporation into geometric models, we used *CondenseUNet* [11] framework, which substitutes the concept of both standard convolution and group convolution (G-Conv) with learned group convolution (LG-Conv). Our network learns the group convolution automatically during training through a multi-stage scheme. The capability of our network to learn the group structure allows multiple groups to re-use the same features via condensed connectivity. Moreover, the efficient weight-pruning methods lead to high computational savings without compromising segmentation accuracy [25].

The surface mesh generation pipeline contains two main tasks: surface mesh generation and smoothing. The predominant algorithm for isosurface extraction from original 3D data is marching cubes [18], which produces a triangulation within each cube to approximate the isosurface by using a look-up table of edge intersections. For this purpose, we used the segmentation map of all the frames in a cardiac cycle generated by our *CondenseUNet* model. Since the slice thickness was large and ranged from 5 mm to 10 mm, we re-sampled the dataset to achieve a 1 mm consistent slice thickness. After extracting the isosurface models using the Lewiner marching cubes [18] algorithm implemented using the scikit-image library [26] in the Python programming language, our next task was to remove the surface noise by applying smoothing operations. In order to smooth the isosurface meshes, we used the joint smoothing technique in 3D Slicer 4.10.2 [27], with the smoothing factor in the range of 0.15 to 0.2. This mesh smoothing operation significantly improves mesh appearance as well as shape, by moving mesh vertices without modifying topology.

Besides the RV isosurface meshes generated from the individual cardiac image frame segmentations following marching cubes and smoothing, which served as ground truth, we generated three additional sets of meshes by propagating the isosurface mesh at the ED phase to all the subsequent cardiac frames using the registration field estimated using the proposed VoxelMorph registration, as well as two traditional nonrigid image registration methods: the

B-spline free form deformation (FFD) [22] algorithm and the fast symmetric force Demon's algorithm [28, 29].

Baseline Comparisons The results obtained using the proposed deep learning registration framework were compared to those obtained using traditional iterative image registration methods, including the FFD [22] algorithm and the fast symmetric force Demon's algorithm [29]. The FFD registration method was implemented in SimpleElastix [23]. The FFD algorithm was set to use the adaptive stochastic gradient descent method as the optimizer, MSE as the similarity measure, binding energy as the regularization function, and was optimized in 500 iterations. The Demon's algorithm was implemented in SimpleITK [30]. The standard deviations for the Gaussian smoothing of the total displacement field was set to 1 and optimized in 500 iterations. These algorithms are trained using manually tuned parameters on an Intel(R) Core(TM) i9-9900K CPU.

3.3 Results

3.3.0.1 Left Ventricle Myocardium

The manual segmentation labels of the LV blood-pool, LV myocardium and RV blood-pool for ED and ES frames of 100 subjects are provided in the ACDC challenge dataset. To evaluate the performance of the CNN-based deformable registration algorithm, we warp the segmentation map of the ED frame to ES frame using the estimated registration field, and compute the Dice score and Hausdorff distance (HD) between the segmentation map of ES frame and the warped segmentation map of ED frame. We refer to this as the "gold" standard comparison, as the segmentation maps used for comparison are manually annotated by experts. We also warp the segmentation map of the ED frame to all subsequent cardiac frames, and compute the Dice score and HD between the warped segmentation map of ED frame and the segmentation maps predicted by the modified U-Net model [14]. We refer to this as the "silver" standard comparison, as the segmentation masks used as reference were not annotated by experts, but rather were generated using techniques previously validated against expert annotations.

In Table 3.1, we show the mean Dice score and mean HD for LV blood-pool, LV myocardium and RV blood-pool before registration (post misalignment correction) and after registration on the test dataset, for both “gold" and “silver" standard comparisons. We also compare the effect of the gradient-based operator and the Laplacian-based operator on the VoxelMorph-based deformable registration method.

Table 3.1: Mean Dice score and Hausdorff distance (HD) for LV blood-pool (LV), LV myocardium (MC) and RV blood-pool (RV), for both “gold" and “silver" standard comparisons, for unregistered frames and post registration using VoxelMorph (VM) framework. Statistically significant differences between the registration metrics *before* and *after* registration were evaluated using the Student t-test and are reported using * for $p < 0.05$ and ** for $p < 0.005$. The best evaluation metrics achieved are labeled in **bold**.

		Dice (%)			HD (mm)		
		LV	MC	RV	LV	MC	RV
ED to ES (Gold std.)	Unregistered	87.30	69.15	70.18	7.22	8.93	11.85
	VM (gradient)	92.17**	79.39**	77.58*	5.59*	8.05	11.75
	VM (Laplacian)	**93.73****	**80.59****	**79.63***	**5.11***	**7.98***	**11.62**
ED to all (Silver std.)	Before registration	81.29	80.15	77.32	3.13	6.08	**8.61**
	VM (gradient)	94.67**	84.08**	82.73*	**2.51**	6.07	8.96
	VM (Laplacian)	**94.84****	**85.22****	**84.36****	2.74	**5.88***	9.04

Our proposed method achieves a 83.04% Dice score and 8.46 mm HD for all cardiac chambers following registration using the gradient-based smoothing loss function, and a 84.65% Dice score and 8.23 mm HD following registration using the Laplacian-based smoothing loss function, for our “gold” standard comparison evaluated at ES frames. Similarly, for our “silver” standard comparison, conducted across all frames, we report a 87.16% Dice score and 5.84 mm HD following registration using the gradient-based smoothing loss function and a 88.14% Dice score and 5.88 mm HD following registration using the Laplacian-based smoothing loss function. Fig. 3.6 shows the cardiac chamber contours propagated using our registration from ED frame to the other cardiac frames.

Furthermore, the LV isosurface (generated from the ED image segmentation map) is propagated to all the subsequent cardiac frames using the deformation field estimated by FFD and VoxelMorph. We then compare these isosurfaces to those directly generated by segmenting

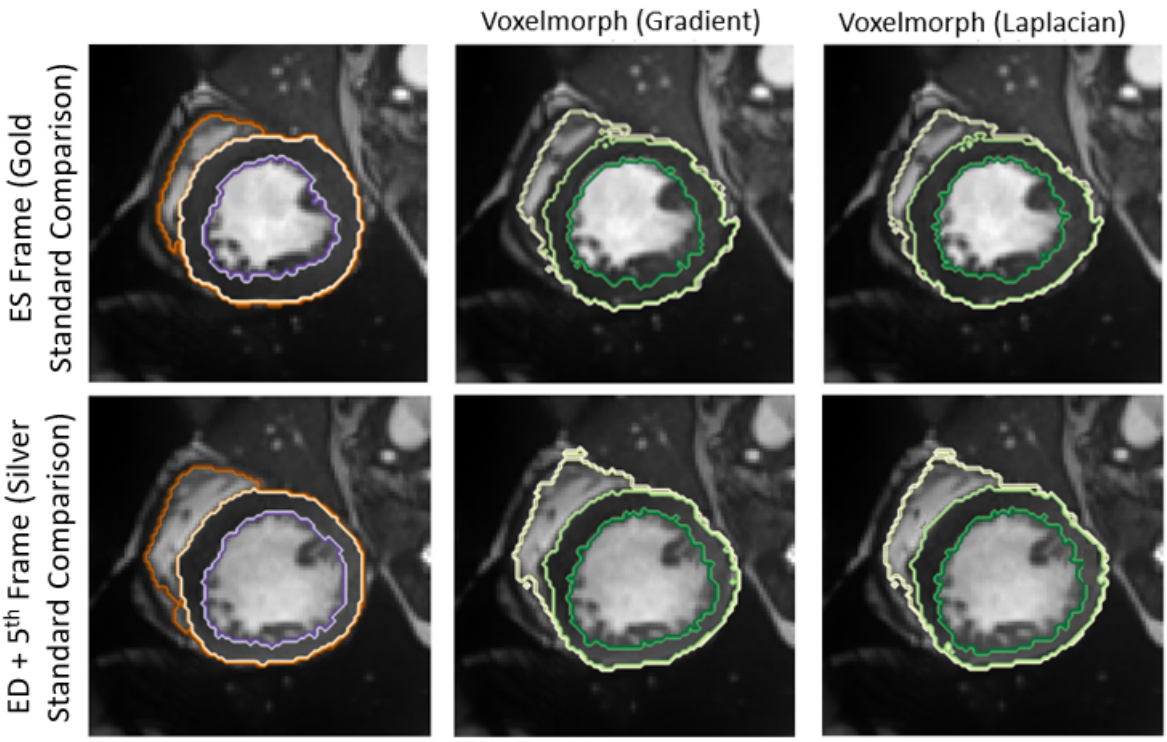


Figure 3.6: Panel 1-1: ES CMR slice with manually annotated segmentation contours of cardiac chambers overlaid on the slice; Panel 1-2: post registration contours using gradient-based operator as smoothing loss with segmentation contours of warped ED frame overlaid on ES frame (Dice: 81.32%, HD: 3.64 mm); Panel 1-3: post registration contours using Laplacian-based operator as smoothing loss (Dice: 83.56%, HD: 3.48 mm). Panel 2-1: ED + 5^{th} frame CMR slice with segmentation contours obtained from U-Net model; Panel 2-2: post registration contours using gradient-based operator as smoothing loss (Dice: 92.36%, HD: 4.12 mm); Panel 1-3: post registration contours using Laplacian-based operator as smoothing loss (Dice: 92.42%, HD: 4.12 mm).

all cardiac image frames using a modified U-Net model [14], which we refer to as the “silver standard".

Table 3.2 summarizes the performance of the FFD and VoxelMorph registration by assessing the Dice score and mean absolute distance (MAD) between the propagated and directly segmented (i.e., “silver standard”) isosurfaces.

Fig. 3.7 illustrates the distance between the three sets of isosurfaces (segmented, CNN-propagated and FFD-propagated) for one patient randomly selected from each pathology. The MAD between the surfaces is less than 2 mm at all frames, with the CNN-propagated

Table 3.2: Mean Dice score (%) and mean absolute distance (MAD) (mm) between FFD and segmentation (FFD-SEG), CNN and segmentation (CNN-SEG), and FFD and CNN (FFD-CNN) results. Statistically significant differences were evaluated using the t-test (* for $p < 0.1$ and ** for $p < 0.05$).

	Normal		MINF		DCM		HCM	
	Dice	MAD	Dice	MAD	Dice	MAD	Dice	MAD
FFD-Segmentation	74.80	1.53	77.69	1.09	80.41	0.91	77.39	1.97
CNN-Segmentation	80.41**	1.15	81.21*	0.87	83.39*	0.91	82.46*	1.09
FFD-CNN	77.81	1.13	82.12	0.75	81.67	0.97	77.34	1.77

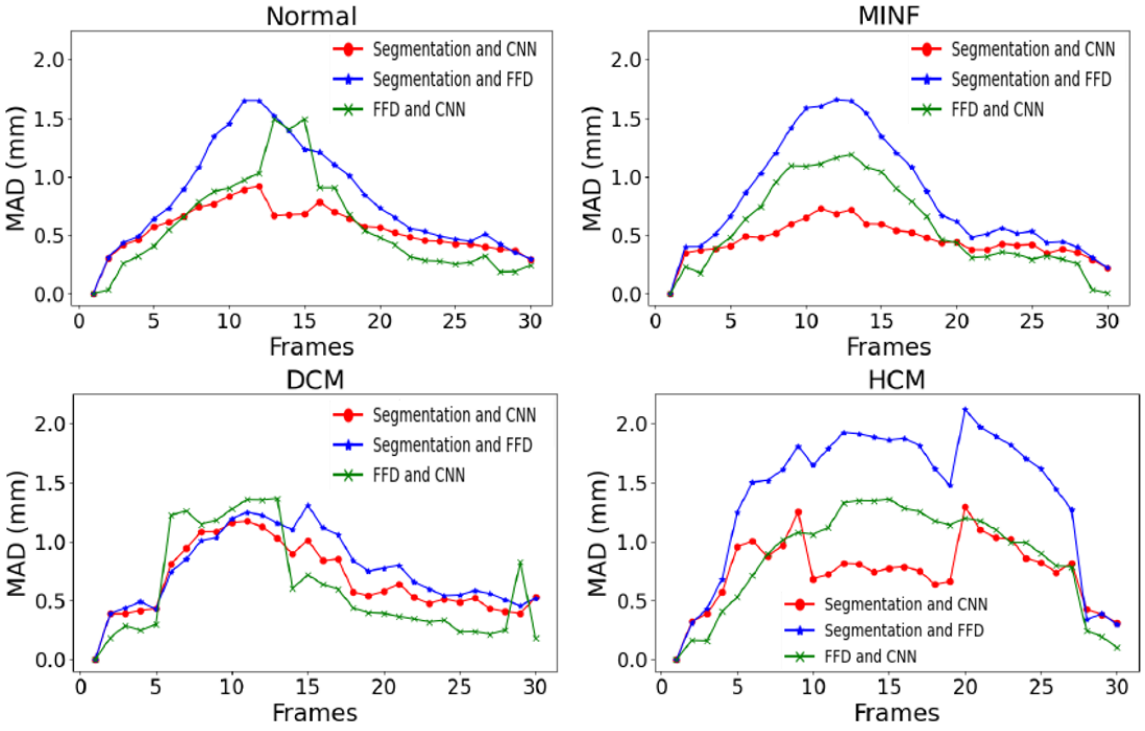


Figure 3.7: MAD between FFD- and CNN-propagated, and segmented (i.e., "silver standard") isosurfaces at all cardiac frames for all patient pathologies.

isosurfaces being closest to the "silver standard" segmented surfaces. Fig. 3.8 illustrates the model-to-model distance between the FFD-propagated and CNN-propagated isosurface meshes at end-systole (ES) and mid-diastole frames for subjects from all four pathologies.

Since the CNN-propagated isosurfaces are in closer agreement to the "silver standard" segmented surfaces compared to the FFD-propagated isosurfaces, we use the CNN-propagation

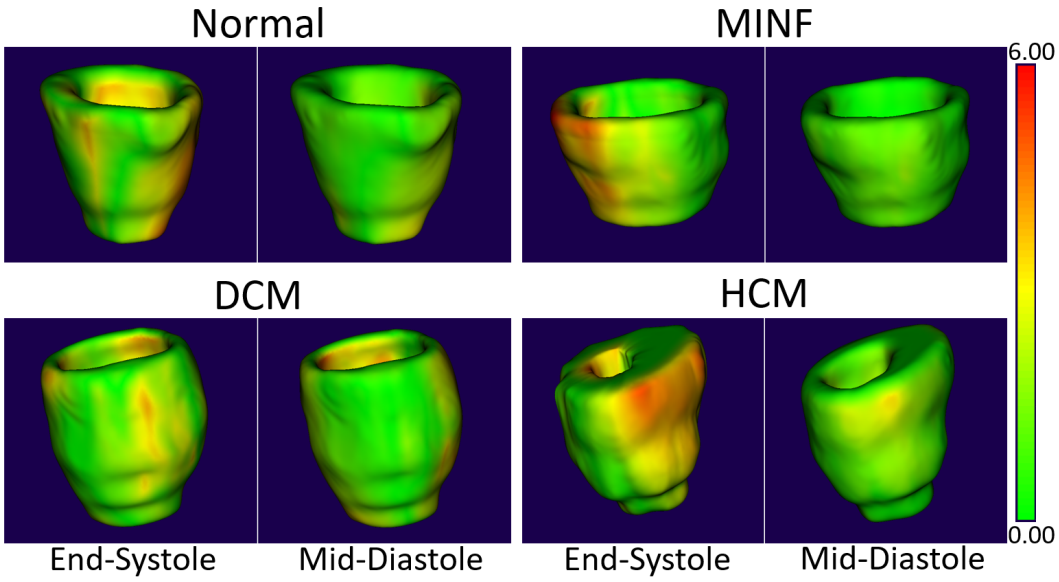


Figure 3.8: [
Model-to-model distance between the left ventricle myocardium isosurface meshes generated from FFD- and the CNN-propagation method at end-systole and mid-diastole frames]Model-to-model distance between the isosurface meshes generated from FFD- and the CNN-propagation method for all patient pathologies at end-systole and mid-diastole frames.

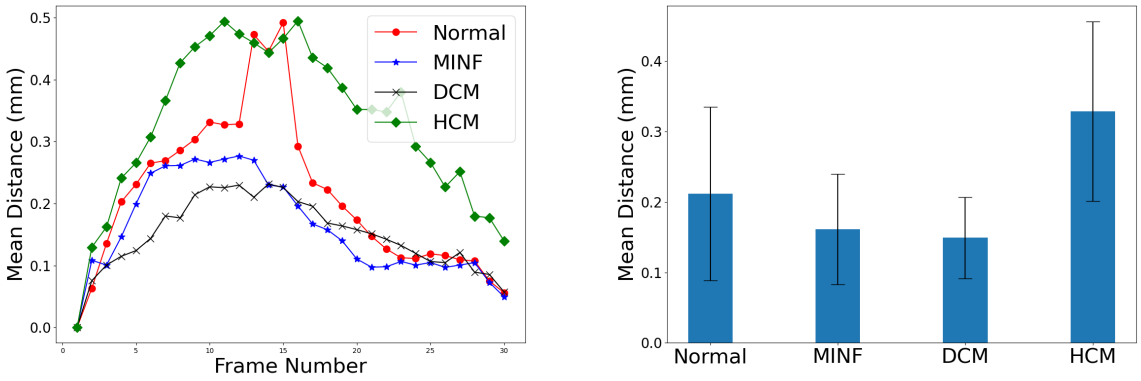


Figure 3.9: Mean node-to-node distance at each cardiac frame between the CNN-propagated and LBWARP-generated volume meshes (left); mean (std-dev) node distance across all frames for each patient pathology (right).

method to generate the volume meshes at each phase of the cardiac cycle. As mentioned in Section 3.2.4.1 and shown in Fig. 3.4, we generate two sets of volume meshes at each frame of the cardiac cycle. Fig. 3.9 shows the mean node distance between the two sets of volume

meshes across all cardiac frames for one subject in each of the four pathologies. Fig. 3.9 also shows the mean node distance between the two sets of volume meshes at each frame of the cardiac cycle for the four subjects. It can be observed that the two sets of volume meshes are in close agreement with each other, exhibiting a mesh-to-mesh distance within 0.5 mm.

3.3.0.2 Right Ventricle Blood-Pool

To evaluate the registration performance of the FFD, Demon's and VoxelMorph methods with respect to RV, the isosurface of the RV generated from the segmentation map in the ED frame is propagated to all the subsequent cardiac frames using the registration field. We then compare the registration accuracy by measuring the overlap between the isosurfaces directly generated by segmenting all cardiac image frames using our CondenseUNet model [11] (i.e., "silver standard") and those propagated by FFD, Demon's and VoxelMorph using Dice score and mean absolute distance (MAD).

Table 3.3 summarizes the registration performance between these propagated and "silver standard" isosurfaces, for both normal and abnormal RV. Fig. 3.10 illustrates the MAD between the propagated and segmented isosurfaces for one patient each with normal and abnormal RV. It can be observed that the CNN-propagated isosurfaces are closer to the segmented isosurfaces than the FFD-propagated isosurfaces; they are comparable to the Demon's-

Table 3.3: RV Endocardium Mean (std-dev) Dice score (%) and mean absolute distance (MAD) between FFD and segmentation (FFD-SEG), Demon's and segmentation (Dem-SEG), CNN and segmentation (CNN-SEG), FFD and CNN (FFD-CNN), and Demon's and CNN (Dem-CNN) results. Statistically significant differences were confirmed via t-test between FFD-SEG and Dem-SEG, and FFD-SEG and CNN-SEG (* $p < 0.1$ and ** $p < 0.05$).

	Normal RV		Abnormal RV	
Methods	Dice	MAD	Dice	MAD
FFD-SEG	75.47 (5.71)	4.37 (1.23)	81.72 (3.32)	2.39 (0.62)
Dem-SEG	79.49 (4.77)**	3.52 (0.93)	84.54 (4.75)**	2.14 (0.46)
CNN-SEG	79.51 (4.93)**	3.34 (0.82)*	83.61 (4.96)**	2.44 (0.63)
FFD-CNN	80.15 (5.86)	1.69 (1.02)	87.31 (3.45)	1.03 (0.56)
Dem-CNN	84.91 (5.58)	1.08 (0.91)	90.64 (2.55)	0.78 (0.31)

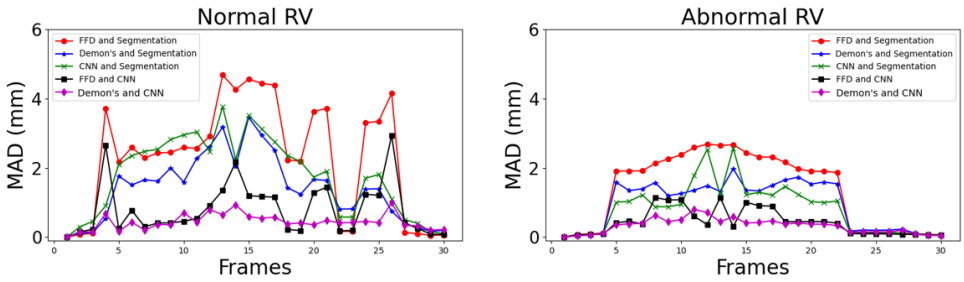


Figure 3.10: Mean absolute distance (MAD) between FFD-, Demon's- and CNN-propagated and segmented (i.e., "silver standard") masks at all cardiac frames for patients with normal and abnormal RVs.

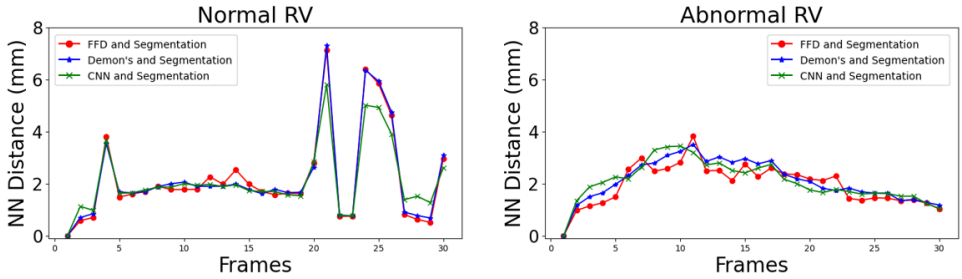


Figure 3.11: (NN) distance between FFD-, Demon's- and CNN-propagated and segmented (i.e., "silver standard") isosurface meshes at all cardiac frames for patients with normal and abnormal RVs.

propagated isosurfaces.

As mentioned in Section 3.2.4.2, we generate four sets of isosurface meshes at each frame of the cardiac cycle for one patient with a normal RV and one patient with an abnormal RV. Fig. 3.11 shows the mean nearest neighbor (NN) distance between the three sets of the registration-propagated isosurface meshes and the isosurface meshes generated directly from the segmented masks at each frame of the cardiac cycle for both the normal and abnormal RV subjects. It can be observed that the isosurface meshes are in close agreement with one another in the subjects with both a normal and an abnormal RV. Fig. 3.12 illustrates the model-to-model distance at the end-systole (ES) frame between the three registration-propagated isosurface meshes and the isosurface meshes generated directly from the segmented masks for both the normal and abnormal RV subjects.

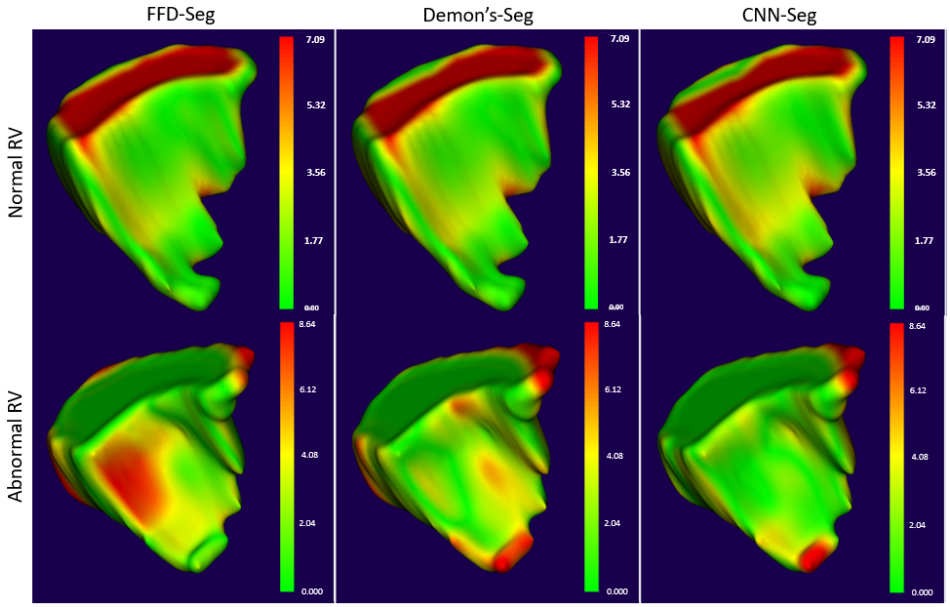


Figure 3.12: Model-to-model distance between the isosurface mesh at end-systole (ES) frame generated from segmentation and propagated using FFD, Demon's and CNN-based deformable registration methods (left to right) for a patient with normal RV (top) and a patient with abnormal RV (bottom).

3.4 Discussion

We present a deep learning-based 4D deformable registration method for cardiac motion estimation from 3D cine CMR images. The workflow also includes a slice misalignment correction step that alleviates the challenges associated with out-of-plane motion in the slice stack that would otherwise impact frame-to-frame image registration and motion extraction. In addition, we evaluate and compare the effect of the gradient-based operator and the Laplacian-based operator for smoothing the registration field on the performance of VoxelMorph-based registration network for cardiac motion estimation. We observe that the Laplacian-based smoothing loss function regularizes better than the gradient-based smoothing loss function. This can be attributed to the fact that the gradient operator only considers the local properties of the objective function $y = x^2$ and the Laplacian operator considers global properties more than the gradient. We also show our intended application of cardiac motion estimation, wherein the registration field obtained from our CNN-based 4D deformable registration is used to propagate

the patient-specific anatomy information from the ED frame to its subsequent frames.

We also briefly investigated the effect of using initial-to-final frame vs. adjacent frame-to-frame registration to extract the cardiac motion throughout the cycle. Although the sequential registration method estimates smaller deformation between two consecutive, adjacent image frames compared to the larger deformations estimated by the initial-to-final frame registration, their concatenation across several frames accumulates considerable registration errors. As such, when using these concatenated registration-predicted deformation fields to propagate the ED isosurfaces and volume meshes to the subsequent cardiac phases, the Dice score and MAD between the propagated and segmented geometries rapidly deteriorate, along with the quality of the propagated surface and volume meshes.

Following the generation of the dynamic, multi-phase meshes, we also assessed the quality of the ES meshes. One set of ES meshes was generated by propagating the ED mesh using the CNN-based extracted motion, while the other set of ES meshes was generated by warping the ED volume mesh based on the dynamic boundary meshes via the LBWARP approach. Unlike the starting ED phase meshes, the ES phase meshes contain some lower quality elements indicated by the lower minimum scaled Jacobian values, but are still suitable for use in simulations.

One of the major advantages of the proposed CNN-based framework over the traditional nonrigid image registration techniques is the significantly faster computing time. For example, it takes around 40 seconds to propagate the RV isosurface mesh at the ED frame to the other frames of the cardiac cycle using a trained VoxelMorph model, compared to 135 and 160 seconds using the FFD and Demon's registration methods, respectively. Similarly, the advantage of using mesh propagation rather than direct mesh generation from individual cardiac image frame segmentation is point correspondence across meshes at different frames, as well as an overall smoother mesh animation over sequential frames, since the individual frame segmentation is accompanied by inherent uncertainty.

Moreover, although the proposed VoxelMorph-based cardiac motion extraction method can capture the frame-to-frame motion with sufficient accuracy, as shown in this work, our ongoing

and future efforts are focused on further improving the algorithm by imposing diffeomorphic deformations [31]. This improvement will help maintain high mesh quality and prevent mesh tangling and element degeneration, especially for the systolic phases.

3.5 Conclusion

In this work, we show that the proposed deep learning framework can be used to build LV myocardial geometric models. The proposed framework is not limited to any pathology and can be extended to LV and RV blood-pool geometry. Ultimately, we intend to use this technique to build dynamic patient-specific myocardial models with associated fiber architecture for biomechanical cardiac simulations.

Bibliography

[1] Emelia J Benjamin, Michael J Blaha, Stephanie E Chiuve, Mary Cushman, Sandeep R Das, Rajat Deo, J Floyd, M Fornage, C Gillespie, CR Isasi, et al. Heart disease and stroke statistics-2017 update: A report from the American Heart Association. *Circulation*, 135(10):e146–e603, 2017. 3.1, 4.1

[2] Nic Smith, Adelaide de Vecchi, Matthew McCormick, David Nordsletten, Oscar Camara, Alejandro F Frangi, Hervé Delingette, Maxime Sermesant, Jatin Relan, Nicholas Ayache, et al. Euheart: personalized and integrated cardiac care using patient-specific cardiovascular modelling. *Interface Focus*, 1(3):349–364, 2011. 1.3, 3.1

[3] Richard A Gray and Pras Pathmanathan. Patient-specific cardiovascular computational modeling: Diversity of personalization and challenges. *Journal of Cardiovascular Translational Research*, 11(2):80–88, 2018. 1.3, 3.1

[4] Paul D Morris, Andrew Narracott, Hendrik von Tengg-Kobligk, Daniel Alejandro Silva Soto, Sarah Hsiao, Angela Lungu, Paul Evans, Neil W Bressloff, Patricia V Lawford, D Rodney Hose, et al. Computational fluid dynamics modelling in cardiovascular medicine. *Heart*, 102(1):18–28, 2016. 3.1

[5] Ghalib A Bello, Timothy JW Dawes, Jinming Duan, Carlo Biffi, Antonio de Marvao, Luke SGE Howard, J Simon R Gibbs, Martin R Wilkins, Stuart A Cook, Daniel Rueckert, et al. Deep-learning cardiac motion analysis for human survival prediction. *Nature Machine Intelligence*, 1(2):95–104, 2019. 1.3.3.2, 3.1

[6] Chen Qin, Wenjia Bai, Jo Schlemper, Steffen E Petersen, Stefan K Piechnik, Stefan Neubauer, and Daniel Rueckert. Joint learning of motion estimation and segmentation for cardiac MR image sequences. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 472–480. Springer, 2018. 1.3.1.3, 1.3.3.2, 3.1, 4.1, 4.2.4

[7] Huaqi Qiu, Chen Qin, Loic Le Folgoc, Benjamin Hou, Jo Schlemper, and Daniel Rueckert. Deep learning for cardiac motion estimation: supervised vs. unsupervised training. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 186–194. Springer, 2019. 1.3.3.2, 3.1

[8] Manuel A Morales, David Izquierdo-Garcia, Iman Aganj, Jayashree Kalpathy-Cramer, Bruce R Rosen, and Ciprian Catana. Implementation and validation of a three-dimensional cardiac motion estimation network. *Radiology: Artificial Intelligence*, 1(4):e180080, 2019. 1.3.3.2, 3.1

[9] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. 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, 2018. 1.3.1.3, 1.4, 2.2.5, 3.1

[10] Suzanne M Shontz and Stephen A Vavasis. A mesh warping algorithm based on weighted Laplacian smoothing. In *12^{th} International Meshing Roundtable*, pages 147–158, 2003. 3.1, 3.2.4.1

[11] SM Kamrul Hasan and Cristian A Linte. CondenseUNet: A memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation. In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11315, page 113151J. International Society for Optics and Photonics, 2020. 3.1, 3.2.4.2, 3.3.0.2, 4.1, 4.4

[12] Shusil Dangi, Cristian A Linte, and Ziv Yaniv. Cine cardiac MRI slice misalignment correction towards full 3D left ventricle segmentation. In *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10576, page 1057607. International Society for Optics and Photonics, 2018. 3.2.2, 4.4

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 1.6, 1.3.1.3, 2.1, 2.2.4.2, 3.2.2, 5.2.2, 5.2.3.4

[14] Fabian Isensee, Paul F Jaeger, Peter M Full, Ivo Wolf, Sandy Engelhardt, and Klaus H Maier-Hein. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In *International Workshop on Statistical Atlases and Computational Models of the Heart. Lecture Notes in Computer Science*, volume 10663, pages 120–129. Springer, 2017. 1.3.1.3, 2.2.3, 2.2.4.1, 3.2.2, 3.3.0.1, 3.3.0.1

[15] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019. 3.2.3.1, 6.2

[16] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015. 1.3.2.3, 3.2.3.1, 4.2.2

[17] Yongpei Zhu, Zicong Zhou Sr., Guojun Liao Sr., and Kehong Yuan. New loss functions for medical image registration based on Voxelmorph. In *Medical Imaging 2020: Image Processing*, volume 11313, page 113132E. International Society for Optics and Photonics, 2020. 3.2.3.1, 3.2.3.1

[18] Thomas Lewiner, Hélio Lopes, Antônio Wilson Vieira, and Geovan Tavares. Efficient implementation of marching cubes' cases with topological guarantees. *Journal of Graphics Tools*, 8(2):1–15, 2003. 3.2.4.1, 3.2.4.2

[19] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. Meshlab: An open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, volume 2008, pages 129–136. Salerno, Italy, 2008. 3.2.4.1

[20] Hang Si. Tetgen, a Delaunay-based quality tetrahedral mesh generator. *ACM Transactions on Mathematical Software (TOMS)*, 41(2):1–36, 2015. 3.2.4.1

[21] The Trilinos Project Team. The Trilinos Project Website, https://trilinos.github.io. 2020; Last accessed 12 Nov. 2020. 3.2.4.1

[22] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999. 3.2.4.1, 3.2.4.2, 3.2.4.2

[23] Kasper Marstal, Floris Berendsen, Marius Staring, and Stefan Klein. SimpleElastix: A user-friendly, multi-lingual library for medical image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 134–142, 2016. 3.2.4.1, 3.2.4.2

[24] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2009. 3.2.4.1

[25] SM Kamrul Hasan and Cristian A Linte. L-CO-Net: Learned condensation-optimization network for segmentation and clinical parameter estimation from cardiac cine MRI. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1217–1220. IEEE, 2020. 3.2.4.2

[26] Stéfan van der Walt *et al.* Scikit-image: Image processing in Python. *PeerJ*, 2:e453, 6 2014. 3.2.4.2

[27] Fedorov *et al.* 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341, 2012. 3.2.4.2

[28] Xavier Pennec, Pascal Cachier, and Nicholas Ayache. Understanding the "Demon's algorithm": 3D non-rigid registration by gradient descent. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 597–605. Springer, 1999. 3.2.4.2

[29] Florence Dru and Tom Vercauteren. An ITK implementation of the symmetric log-domain diffeomorphic Demons algorithm. 2009. 3.2.4.2, 3.2.4.2

[30] Yaniv *et al.* SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research. *Journal of Digital Imaging*, 31(3):290–303, 2018. 3.2.4.2

[31] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical Image Analysis*, 57:226–236, 2019. 3.4

Chapter 4

An Image Registration Approach for Late Gadolinium Enhanced MRI and Cine Cardiac MRI Using Convolutional Neural Networks

Late gadolinium enhanced (LGE) cardiac magnetic resonance (CMR) imaging, the current benchmark for assessment of myocardium viability, enables the identification and quantification of the compromised myocardial tissue regions, as they appear hyper-enhanced compared to the surrounding, healthy myocardium. However, in LGE CMR images, the reduced contrast between the left ventricle (LV) myocardium and LV blood-pool hampers the accurate delineation of the LV myocardium. On the other hand, the balanced-Steady State Free Precession (bSSFP) cine CMR imaging provides high resolution images ideal for accurate segmentation of the cardiac chambers.[1] *In the interest of generating patient-specific hybrid 3D and 4D anatom-*

[1]This chapter is adapted from:
Upendra R.R. et al., "A supervised image registration approach for late gadolinium enhanced MRI and cine cardiac MRI using convolutional neural networks." In: Papież B., Namburete A., Yaqub M., Noble J. (eds) Medical Image Understanding and Analysis. MIUA 2020. Communications in Computer and Information Science, vol 1248. Springer, Cham.
Upendra R.R. et al., "Joint deep learning framework for image registration and segmentation of late gadolinium enhanced MRI and cine cardiac MRI.", Proc. SPIE 11598, Medical Imaging 2021: Image-Guided Procedures,

ical models of the heart, to identify and quantify the compromised myocardial tissue regions for revascularization therapy planning, we present a spatial transformer network (STN) based convolutional neural network (CNN) architecture for registration of LGE and bSSFP cine CMR image datasets made available through the 2019 Multi-Sequence Cardiac Magnetic Resonance Segmentation Challenge (MS-CMRSeg). We perform a supervised registration by leveraging the region of interest (RoI) information using the manual annotations of the LV blood-pool, LV myocardium and right ventricle (RV) blood-pool provided for both the LGE and the bSSFP cine CMR images. Furthermore, in order to reduce the reliance on the number of manual annotations for training such network, we extend the proposed architecture to a joint deep learning framework consisting of three branches: a STN based RoI guided CNN for registration of LGE and bSSFP cine CMR images, an U-Net model for segmentation of bSSFP cine CMR images, and an U-Net model for segmentation of LGE CMR images.

4.1 Introduction

Myocardial infarction, cardiomyopathy and myocarditis represent common cardiac conditions associated with significant morbidity and mortality worldwide [1]. The assessment of myocardium viability for patients who experienced any of these diseases is critical for diagnosis and planning of optimal therapies. Accurate quantification of the compromised myocardium is a crucial step in determining the part of the heart that may benefit from therapy [2]. LGE CMR imaging is the most widely used technique to detect, localize and quantify the diseased myocardial tissue, also referred to as scar tissue. During the typical LGE CMR image acquisition protocol, a gadolinium-based contrast agent is injected into a patient, and the MR images are acquired 15-20 minutes post-injection. In the LGE CMR images, the compromised LV myocardial regions appear much brighter than healthy tissue, due to the trapping and delayed wash-out of the contrast agent from the diseased tissue regions. As a concrete example, in case of myocardial infarction, LGE CMR imaging helps assess the transmural extent of the infarct, which helps predict the success of recovery following revascularization therapy

Robotic Interventions, and Modeling, 115980F (15 February 2021).

and also provides additional insights about other potential complications associated with the disease [3].

In a clinical set-up, radiologists and cardiologists visually assess the viability of the myocardium based on the LGE CMR images. However, the gadolinium-based contrast agent reduces the contrast between the myocardium and the LV blood-pool. Although useful to identify scarred myocardium regions, LGE CMR images do not allow accurate delineation between the LV blood-pool and the LV myocardium. On the other hand, the bSSFP cine CMR images provide excellent contrast between myocardium and blood-pool (Fig. 4.1), and can be successfully employed to identify the myocardium and blood-pool, but they cannot show the scarred regions. Therefore, the LGE and bSFFP CMR images show complementary information pertaining to the heart, but neither image type, on its own, enables the extraction, quantification and global visualization of all desired features: LV blood pool, LV myocardium, and scarred regions.

In the recent 2019 MS-CMRSeg challenge [4, 5], participants were provided with the segmentation labels for LV blood-pool, LV myocardium, and RV blood-pool available for the bSSFP cine CMR images to segment the same cardiac chambers from the LGE CMR images of the same patients. Several studies proposed the generation of synthetic LGE CMR images from the bSSFP cine CMR images using cycleGAN [6], histogram matching [7], shape transfer GAN [8] or style transfer networks like MUNIT [9]. They use these synthetic LGE CMR images and the annotations provided for the bSSFP cine CMR images to train various U-Net architectures to segment LV blood-pool, LV myocardium and RV blood-pool from the actual LGE CMR images. These methods result in good segmentation performance, however, they are time consuming, since they rely on a two-step process: training the adversarial networks to generate synthetic LGE CMR images from the bSSFP cine CMR images, followed by the training of the U-Net architectures (or its variants) on these synthetic LGE CMR images to segment the cardiac chambers from the original LGE CMR images.

An alternative approach to "learning" features from one image type and using them to segment the other image type is to segment the complementary features from the cine MRI

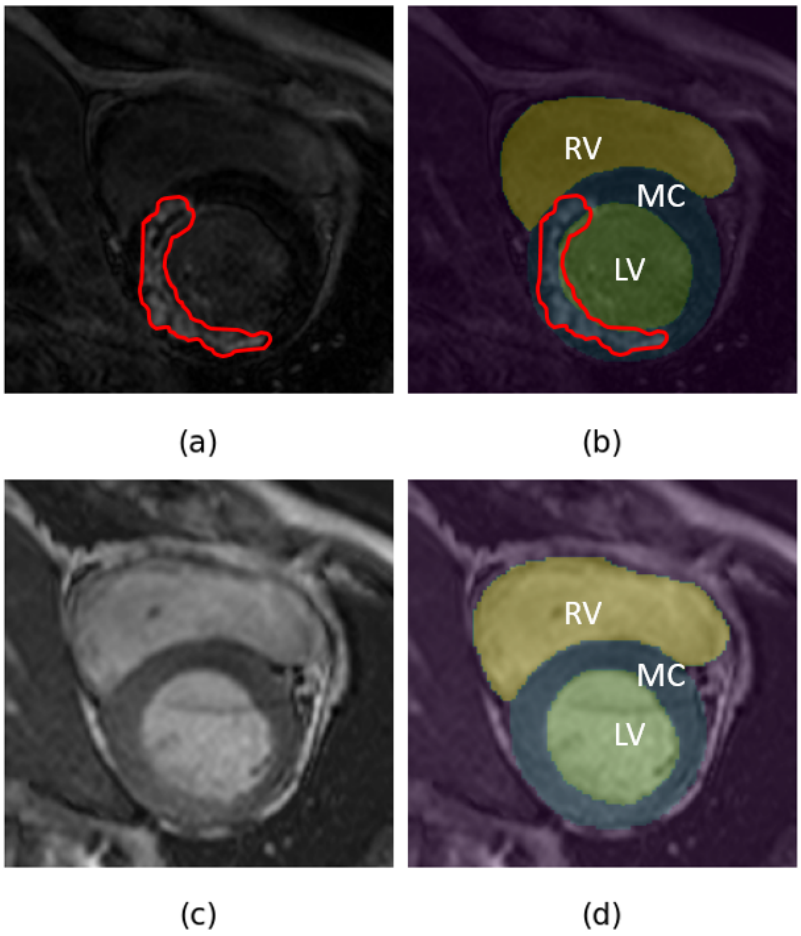


Figure 4.1: Example of (a) bSSFP cine CMR image with its (b) manual annotations - LV blood-pool (LV) in blue, LV myocardium (MC) in green and RV blood-pool (RV) in red overlayed on it and (c) LGE CMR image with its (d) manual annotations overlayed on it.

and LGE MRI images, then co-register the images and use the appropriate registration transformation to propagate the segmentation labels from the cine MRI into the LGE MRI space or vice versa. Chenoune *et al.*[10] rigidly register 3D delay-enhanced images with the cine MRI images using mutual information as the similarity measure. Wei *et al.* [2] use pattern intensity as similarity measure leading to accurate affine registration of cine and LGE MRI images. More recently, Guo *et al.* [11] proposed employing rigid registration to initially align the cine CMR images with the multi-contrast late enhanced CMR images, followed by deformable registration to further refine the alignment. In summary, all these works employ traditional

approaches to iteratively optimize the registration cost function for a given image pair.

With the advent of deep learning, several groups proposed the utilization of neural networks to train image registration algorithms using similarity measures like normalized mutual information (NMI), normalized cross correlation (NCC), local Pearson correlation coefficient (LPC), sum of squared intensity difference (SSD) and sum of absolute difference (SAD) [12]. While the cost functions can be optimized by training large datasets using neural networks, these unsupervised registration methods do not perform significantly better than the traditional approaches, as the similarity measures used are the same [13]. Hence, while these unsupervised machine learning-based registration help in speeding up the registration process compared to the traditional unsupervised registration algorithms, they do not necessarily improve registration accuracy beyond that achieved using traditional unsupervised approaches.

In this paper, we propose a supervised deep learning-based registration approach to register bSSFP cine CMR images to its corresponding LGE CMR images using a STN-inspired CNN. Some literature suggests that supervised registration techniques entail the use of the displacement field for training [14, 15]. Our proposed method, on the other hand, only uses several segmentation labels to guide the registration. Hence, here we refer to it as a supervised registration, although, according to the literature nomenclature mentioned above, it could also be classified as a segmentation-guided registration.

We train the network on the 2019 MS-CMRSeg challenge dataset using the provided manual annotations (required only during training) of the LV blood-pool, LV myocardium, and RV blood-pool and compute a dual-loss cost function that combines the benefits of both the Dice loss and cross-entropy loss. We compare the accuracy of our proposed rigid and affine supervised deep learning-based registration to the accuracy of previously disseminated unsupervised deep learning-based rigid and affine registrations.

Our proposed method aims to address the limitations associated with the aforementioned methods as follows: (i) we fully exploit the information pertaining to the various regions of interest (RoIs) of the cardiac anatomy (i.e., LV blood-pool, LV myocardium and RV blood-pool) and devise a robust ROI-guided registration technique that improved registration accuracy be-

yond the previous unsupervised techniques; (ii) our method requires minimal preprocessing, specifically it only relies on several segmentation labels that could be obtained using manual annotation or using available and previously validated accurate, automatic segmentation techniques [16, 17, 18, 19, 20]; and (iii) does not require the need to train additional adversarial networks to generate synthetic LGE-MRI images [6, 7, 8, 9], therefore reducing network training time without compromising registration accuracy.

In addition to the proposed STN-based RoI-guided CNN, we propose a joint deep learning framework that consists of three branches - a STN-inspired CNN for supervised registration of bSSFP cine CMR images and LGE CMR images, an U-Net model [21] for segmentation of bSSFP cine CMR images and an U-Net model for segmentation of LGE CMR images. Inspired by Qin *et al.* [22], we optimize a composite loss function by training all three networks simultaneously. The aim of the proposed joint deep learning model is to further improve registration accuracy by sharing the weights learned from the segmentation models and to reduce the reliance on the number of manual annotations.

4.2 Methodology

4.2.1 Dataset

The dataset used in this paper was made available through the 2019 MS-CMRSeg challenge [4, 5]. The available data consisted of LGE, T2-weighted, and bSSFP cine MRI images acquired from 45 patients who had been diagnosed with cardiomyopathy. In this work, we utilize the LGE and cine MRI images for registration. The manual annotations of the LV blood-pool, LV myocardium and RV blood-pool were performed by trained personnel and corroborated by expert cardiologists. Both the cine and LGE MRI images were acquired at end-diastole. The cine MRI images were acquired using a TR and TE of 2.7 ms and 1.4 ms, respectively, and consisted of 8-12 slices with an in-plane resolution of 1.25 mm $\times$ 1.25 mm and a slice thickness of 8-13 mm. The LGE MRI images were acquired using a T1-weighted inversion-recovery gradient-echo pulse sequences with a TR and TE of 3.6 ms and 1.8 ms, respectively,

and consisted of 10-18 slices featuring an in-plane resolution of 0.75 $mm \times 0.75$ mm and a 5 mm slice thickness.

To account for the differences in slice thickness, image sizes and in-plane image resolution between the bSSFP cine CMR and LGE CMR images, all the images are resampled to a slice thickness of 5 mm (using spline interpolation), in-plane image resolution of 0.75 $mm \times 0.75$ mm and then resized to 224×224 pixels.

4.2.2 Spatial Transformer Network (STN) Architecture

The STN consists of three parts - a localisation network, a parameterised sampling grid (grid generator) and a differentiable image sampler. The localisation network function $f_{loc}()$ can be any fully connected network or convolutional neural network that takes in an input feature map $I \in R^{W \times H \times C}$ with width W height H and channels C through a number of hidden layers, and outputs θ, where $\theta = f_{loc}(I)$, and contains the parameters of the transformation T_θ.

In an effort to explain the proposed algorithm, let us consider a regular grid G consisting of a set of points with source coordinates (x_i^s, y_i^s). This grid G acts as input to the grid generator and the transformation T_θ is applied on it i.e. $T_\theta(G)$. This operation results in a set of points with target coordinates (x_i^t, y_i^t) which is altered to translate, scale, rotate, skew etc. the input image depending on the values of the θ. Depending on the target coordinates (x_i^t, y_i^t), the differentiable image sampler generates a transformed output feature map $O \in R^{W \times H \times C^{'}}$ [23].

4.2.3 STN-based Registration of bSSFP Cine and LGE MRI Images

In our experiments, we concatenate the bSSFP cine CMR image (224×224) with its corresponding LGE CMR image (224×224) and input the resulting $224 \times 224 \times 2$ tensor into a CNN which is analogous to the localisation network. For a 2D affine registration transformation, the output θ of the CNN is a six-dimensional vector that results in the transformation matrix T_θ,

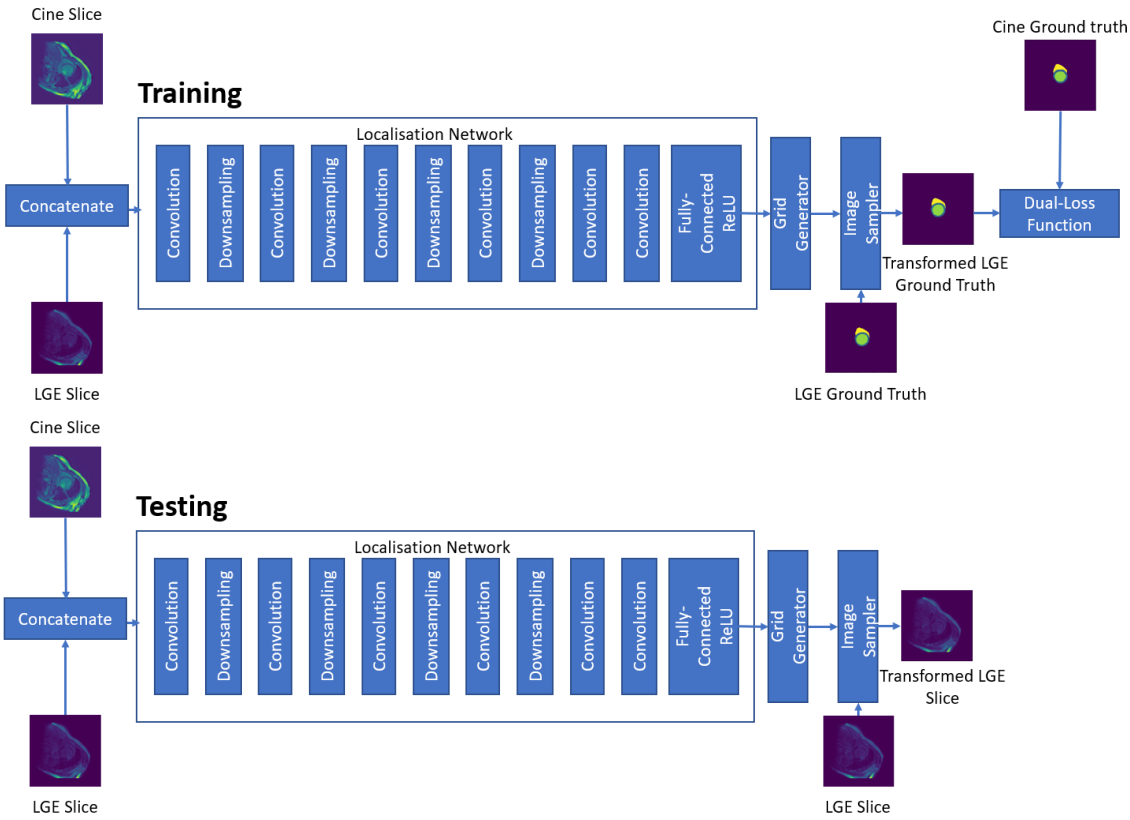


Figure 4.2: Overview of supervised registration of bSSFP cine CMR images and LGE CMR image using STN. In the training network, the GT LGE CMR image is fed into the image sampler and the dual-loss function is computed using the transformed GT features from the LGE and bSSFP cine CMR images. In the testing network, the LGE CMR image slice is fed into the image sampler and the output consists of a spatially transformed LGE CMR image slice.

$$T_\theta = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \tag{4.1}$$

For a rigid registration transformation, the output θ of the CNN is three-dimensional i.e. $\theta = [x, y, z]$ where z is the rotation parameter and (x, y) are the translation parameters. This results in the transformation matrix T_θ,

$$T_\theta = \begin{bmatrix} cos(z) & sin(z) & x \\ -sin(z) & cos(z) & y \end{bmatrix} \tag{4.2}$$

The grid generator outputs a sampling grid $T_\theta(G)$ and the differentiable image sampler transforms the ground truth (GT) map of the LGE CMR image. The Dice loss and cross-entropy loss are computed using the GT map of the cine CMR image and the transformed GT map of the LGE CMR image. The computed loss is then used to back-propagate the CNN (Fig. 4.2).

4.2.4 Joint Deep Learning Model for Registration and Segmentation

As shown in earlier works [24, 25, 22], image registration and segmentation tasks are closely related and it has been shown that learning features from one task can benefit the other task. In this work, we explore a joint deep learning model for registration of bSSFP cine CMR and LGE CMR images, and segmentation of cardiac chambers (LV blood-pool, LV myocardium and RV blood-pool) from the bSSFP cine CMR and LGE CMR images (Fig. 4.3). The coupling of these registration and segmentation tasks result in sharing of the weights learnt from the segmentation task with the registration branch of the network, improving the registration accuracy.

4.2.5 Experiments and Implementation Details

4.2.5.1 Suervised RoI-based Registration Network

Firstly, we compare the registration accuracy of four different registration methods - unsupervised rigid, unsupervised affine, supervised rigid and supervised affine.

The unsupervised registration methods are trained using the CNN shown in Fig. 4.2, using NMI as a cost function. Our proposed supervised registration methods, which could also be classified as segmentation-guided image registration techniques, are trained using the following dual-loss function:

$$\mathcal{L}_{dual-loss} = \alpha.\mathcal{L}_{cross-entropy} + (1-\alpha).\mathcal{L}_{Dice-loss} \tag{4.3}$$

where $\mathcal{L}_{cross-entropy}$ is the cross-entropy loss, $\mathcal{L}_{dice-loss}$ is the Dice loss and α allows us to modulate the effect of the Dice loss and cross-entropy loss on the overall dual-loss function.

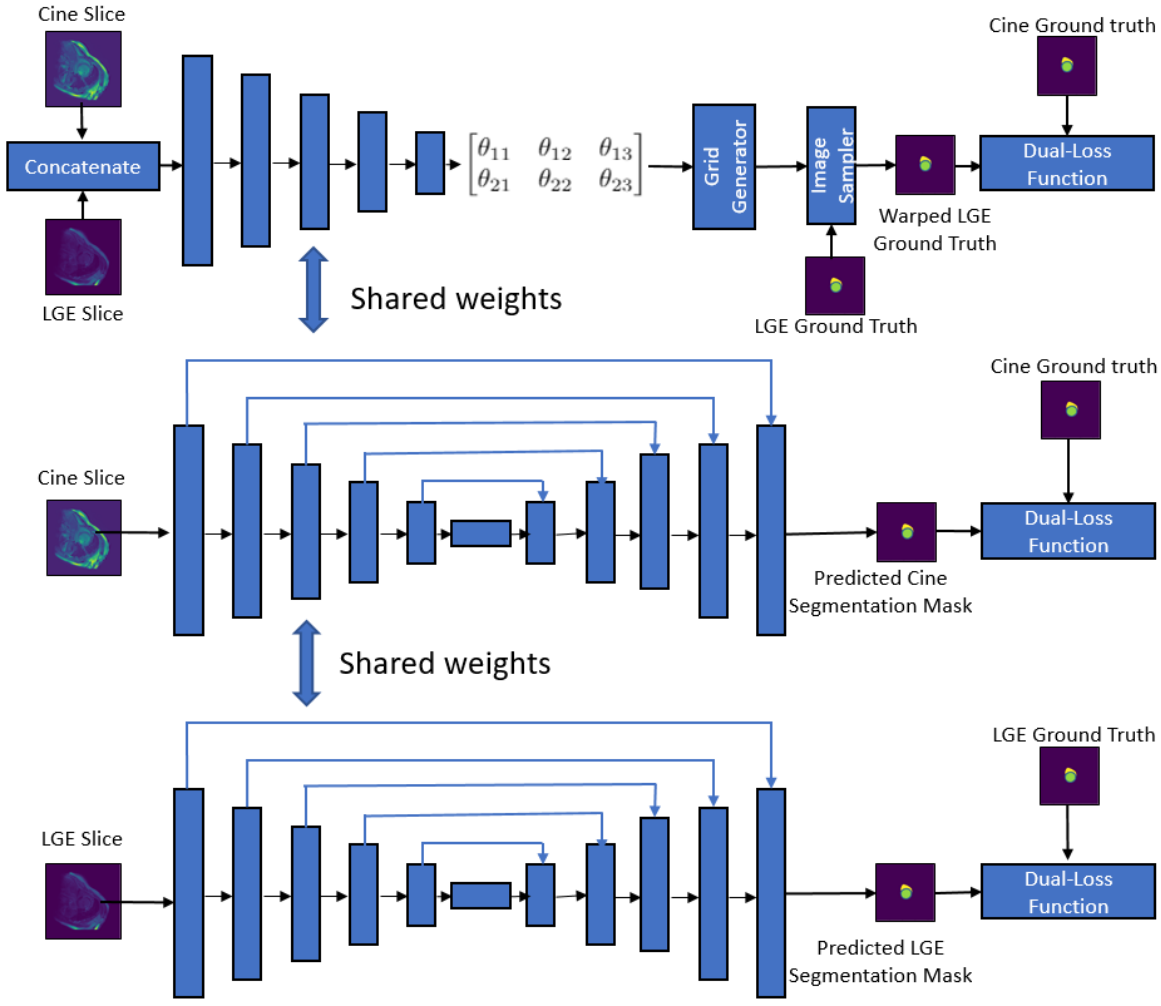


Figure 4.3: Schematic architecture of the proposed joint deep learning framework consisting of three branches - a STN based CNN for registration of bSSFP cine CMR and LGE CMR images, an U-Net model for segmentation of bSSFP cine CMR images and an U-Net model for segmentation of LGE CMR images.

In our experiments, of the total 45 bSSFP and LGE MRI datasets, we use 35 datasets for training, leaving 5 datasets for validation and the remaining 5 datasets for testing. We train these networks using the Adam optimizer with a learning rate of *1e-5* and a gamma decay of 0.99 every alternate epoch for fine-tuning for 100 epochs on a machine equipped with NVIDIA RTX 2080 Ti GPU with 11GB of memory.

4.2.5.2 Joint Deep Learning Network

Secondly, we focus on comparing the registration results of the proposed joint deep learning model with the stand-alone STN based model [26]. We also compare the results obtained by training our networks by splitting the available 45 bSSFP cine and LGE MRI datasets to 35 for training, 5 for validation and 5 for testing, and the results obtained by training our networks by splitting to 25 for training, 15 for validation and 5 for testing.

The three branches of our joint deep learning model are trained using the above-mentioned dual-loss function (Equation 4.3). The loss function is calculated using the predicted segmentation maps and their corresponding GT maps for the bSSFP cine and LGE CMR segmentation networks ($\mathcal{L}_{cine-seg}$ and $\mathcal{L}_{lge-seg}$, respectively). For the supervised RoI-guided registration network, the dual-loss function is computed using the transformed GT map of the LGE CMR images and the GT of cine CMR images ($\mathcal{L}_{reg}$). Therefore, the resulting composite loss function is given by

$$\mathcal{L} = \lambda_1.\mathcal{L}_{reg} + \lambda_2.\mathcal{L}_{cine-seg} + \lambda_3.\mathcal{L}_{lge-seg} \tag{4.4}$$

where λ_1, λ_2 and λ_3 are the trade-off parameters for the three branches of the joint deep learning model.

We train all our networks by randomly augmenting both the bSSFP cine CMR and the LGE CMR images on-the-fly using a series of translation, rotation and gamma correction operations. In these experiments, the networks are trained using the Adam optimizer with a learning rate of 10^{-4} and a gamma decay of 0.99 every alternate epoch for fine-tuning for 100 epochs on a machine equipped with NVIDIA RTX 2080 Ti GPU with 11GB of memory.

4.3 Results

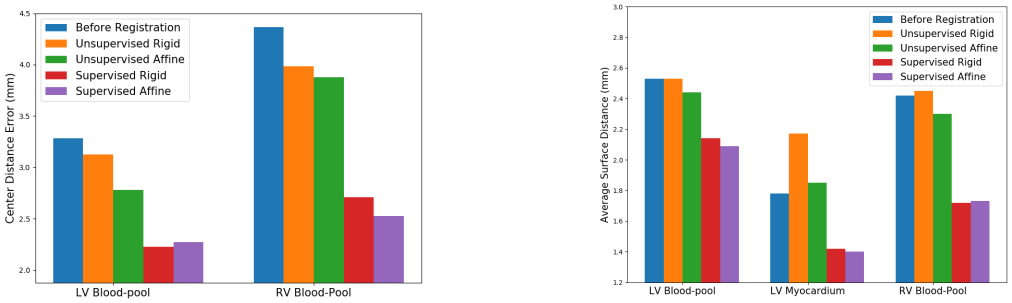
To evaluate our registration, we identify the LV and RV blood-pool centres as the centroid of the segmentation masks of the LV and RV blood-pool corresponding to both the bSSFP cine CMR and LGE CMR images. The Euclidean distance between the blood-pool centers i.e.

center distance (CD) from these two images is compared to the blood-pool CD of the bSSFP cine CMR image and its corresponding transformed LGE CMR image. We also quantify our registration accuracy using average surface distance (ASD), a popular evaluation metric for registration, between the LV blood-pool, LV myocardium and RV blood-pool masks of bSSFP cine CMR image and its corresponding LGE CMR image, before and after registration.

In Table 4.1, we show the mean CD and mean ASD before registration, after unsupervised rigid registration, unsupervised affine registration, supervised rigid registration and supervised affine registration of the test dataset. Fig. 4.4 shows the comparison of CD and ASD of all the four above-mentioned registration approaches. We can observe that the CD is significantly reduced in both the supervised registration algorithms (rigid and affine) for LV blood-pool (p-value < 0.005) and RV blood-pool (Herp-value < 0.05). We can also observe that the ASD is significantly reduced for the LV blood-pool (p-value < 0.05), LV myocardium (p-value < 0.05) and RV blood-pool (p-value < 0.005) for both the rigid and affine supervised registration methods. However, the changes in the CD and ASD after unsupervised registration (both rigid and affine) is not very significant, compared to before registration.

Table 4.1: Summary of registration evaluation for RoI-based supervised registration algorithm. Mean (std-dev) center-to-center distance (CD) and average surface distance (ASD) for LV blood-pool (LV), LV myocardium (MC) and RV blood-pool (RV). Statistically significant differences between the registration metrics *before* and *after* registration were evaluated using the Student t-test and are reported using * for $p < 0.05$ and ** for $p < 0.005$. The best evaluation metrics achieved are labeled in **bold**.

	LV CD (mm)	LV ASD (mm)	MC ASD (mm)	RV CD (mm)	RV ASD (mm)
Before Registration	3.28 (1.83)	2.53 (1.23)	1.78 (0.78)	4.36 (3.79)	2.42 (1.20)
Unsupervised Rigid	3.12 (1.79)	2.53 (1.13)	2.17 (1.58)	2.48 (3.78)	2.45 (1.26)
Unsupervised Affine	2.78 (1.65)	2.44 (1.58)	1.85 (1.56)	3.87 (3.75)	2.30 (1.23)
Supervised Rigid	**2.22 (1.08)****	2.14 (1.20)*	1.42 (1.48)*	2.69 (2.51)*	**1.72 (1.06)****
Supervised Affine	2.27 (1.38)**	**2.09 (1.14)***	**1.40 (1.12)***	**2.52 (2.66)***	1.73 (1.02)**



(a) (b)

Figure 4.4: Comparison of (a) mean CD and (b) mean ASD values before registration, after unsupervised rigid registration, unsupervised affine registration, supervised rigid registration and supervised affine registration

Fig. 4.5 shows an example of the manual annotations of LV blood-pool (green), LV myocardium (blue) and RV blood-pool (yellow) of a bSSFP cine CMR image overlayed on its corresponding LGE CMR image before registration and on the corresponding transformed LGE CMR image after unsupervised rigid, unsupervised affine, supervised rigid and supervised affine registration. Fig. 4.5 also shows that when the LGE CMR image and its associated hyper-enhanced regions (marked by the enclosed pink contour) is overlaid onto the bSSFP cine CMR image and its associated labels, the hyper-enhanced regions erroneously appear as part of the LV blood-pool, instead of the LV myocardium. Nevertheless, following supervised registration, the hyper-enhanced regions correctly align with the LV myocardium, where they truly belong. Lastly, Fig. 4.5 also helps the reader visually appreciate the performance of each registration algorithm by showing the LV and RV blood-pool center-to-center distance before and after each registration algorithm is applied.

In Table 4.2, we summarize the registration performance of the stand-alone STN-based RoI-guided CNN and the proposed joint deep learning model. We compare the mean CD and mean ASD achieved by both the stand-alone STN based CNN and the joint deep learning model by training 25 of the 45 available image datasets and by training 35 of the 45 available image datasets. We can observe that the registration performance of the joint deep learning model achieved by training only 25 image datasets is comparable to that of the stand-alone STN based registration when trained using 35 image datasets and significantly better than

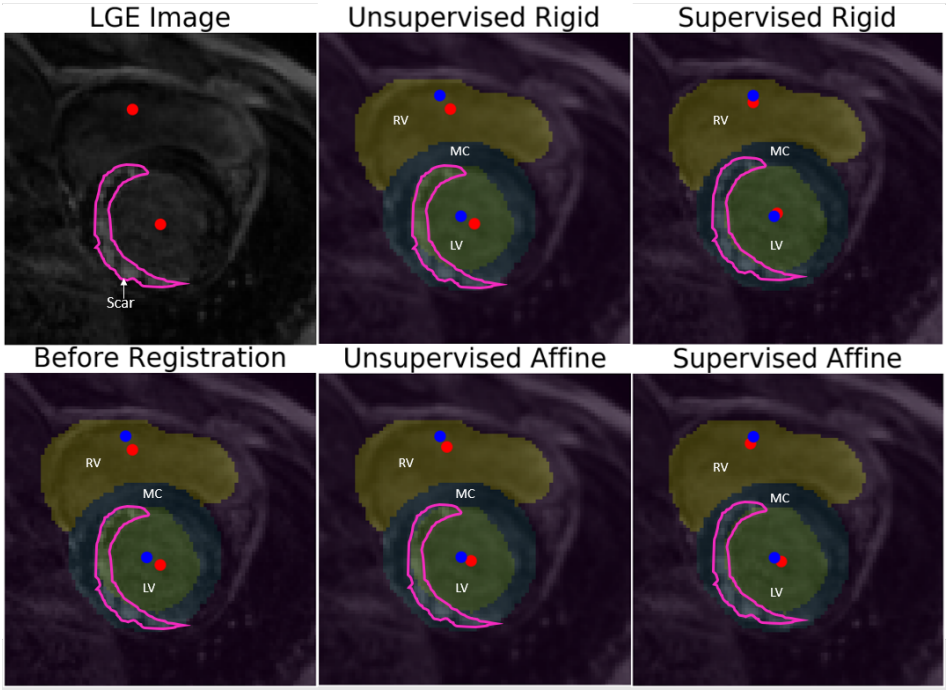
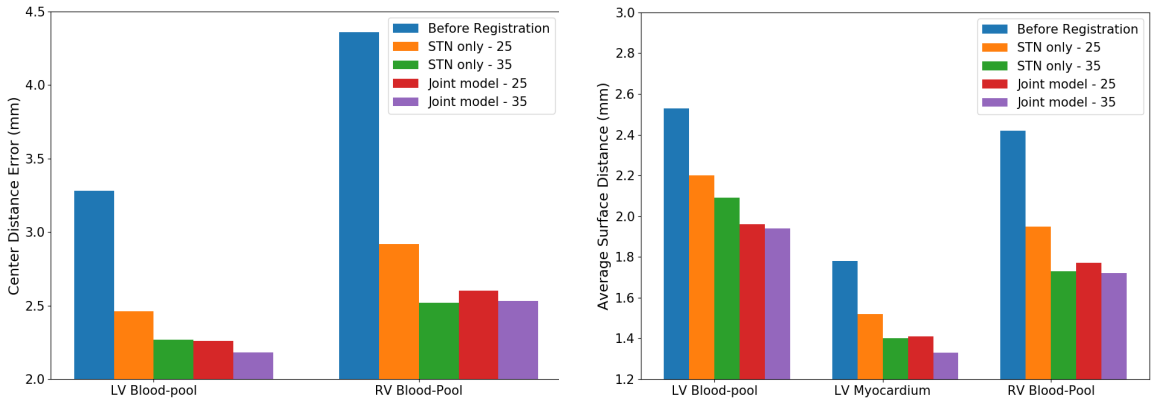


Figure 4.5: Panel 1-1: Unregistered LGE CMR image and associated hyper-enhanced regions marked by pink contour and LV and RV blood-pool centers marked by red dots; Panel 2-1: before registration (CD: 2.72 *mm*, ASD: 2.24 *mm*); overlaid unregistered LGE CMR image and features (from Panel 1-1) onto the cine CMR image showing the RV and LV blood-pools and their centers (marked by blue dots) and the LV myocardium (MC) marked on the cine CMR image (Note: The hyper-enhanced regions enclosed by pink contour erroneously appear over the LV blood-pool, not the LV myocardium, where they truly belong); Panel 1-2: overlaid LGE CMR image onto the cine CMR image following unsupervised rigid registration (CD: 2.56 *mm*, ASD: 2.20 *mm*); Panel 2-2: unsupervised affine registration (CD: 2.52 *mm*, ASD: 2.18 *mm*); Panel 1-3: supervised rigid registration (CD: 1.56 *mm*, ASD: 1.64 *mm*); and Panel 2-3: supervised affine registration(CD: 1.68 *mm*, ASD: 1.76 *mm*). (Note: The accurate overlay of the hyper-enhanced regions marked by the pink contour over the LV myocardium, as well as significantly improved LV and RV blood-pool center-to-center distance following supervised registration in Panel 1-3 and Panel 2-3).

the stand-alone STN based registration when trained using 25 image datasets (p-value < 0.1 for RV blood-pool CD, and LV blood-pool CD and ASD). We can also observe that when the joint deep learning model is trained using 35 image datasets, the LV blood-pool CD and LV myocardium ASD is significantly lower than the rest of the models (Fig. 4.6). In Fig. 4.7, we show an example of the manual annotations of the cardiac chambers of a bSSFP cine CMR image overlaid on its corresponding LGE CMR image before registration and after registration

Table 4.2: Summary of registration evaluation for joint deep learning framework. Mean (std-dev) center-to-center distance (CD) and average surface distance (ASD) for LV blood-pool (LV), LV myocardium (MC) and RV blood-pool (RV). The best evaluation metrics achieved are labeled in **bold**. Statistically significant differences between the registration metrics *before* and *after* registration were evaluated using the Student t-test and are reported using * for $p < 0.05$ and ** for $p < 0.005$.

	LV CD (mm)	LV ASD (mm)	MC ASD (mm)	RV CD (mm)	RV ASD (mm)
Before Registration	3.28 (1.83)	2.53 (1.23)	1.78 (0.78)	4.36 (3.79)	2.42 (1.20)
Stand-alone STN Model Training: 25 patients	2.46 (1.31)**	2.20 (1.23)*	1.52 (0.84)*	2.92 (2.18)**	1.95 (1.02)*
Stand-alone STN Model Training: 35 patients	2.27 (1.38)**	2.09 (1.14)**	1.40 (1.12)*	**2.52 (2.66)****	1.73 (1.02)**
Joint Model Training: 25 patients	2.26 (1.34)**	1.96 (0.93)**	1.41 (0.71)*	2.60 (2.02)**	1.77 (0.84)**
Joint Model Training: 35 patients	**2.18 (1.46)****	**1.94 (0.93)****	**1.33 (0.73)****	2.53 (2.14)**	**1.72 (0.97)****



(a) (b)

Figure 4.6: Comparison of (a) mean CD and (b) mean ASD values before registration, stand-alone STN based supervised registration (training data: 25 patients), stand-alone STN based supervised registration (training data: 35 patients), joint deep learning model (training data: 25 patients) and joint deep learning model (training data: 35 patients)

using both stand-alone STN-based RoI-guided CNN model and the joint deep learning model.

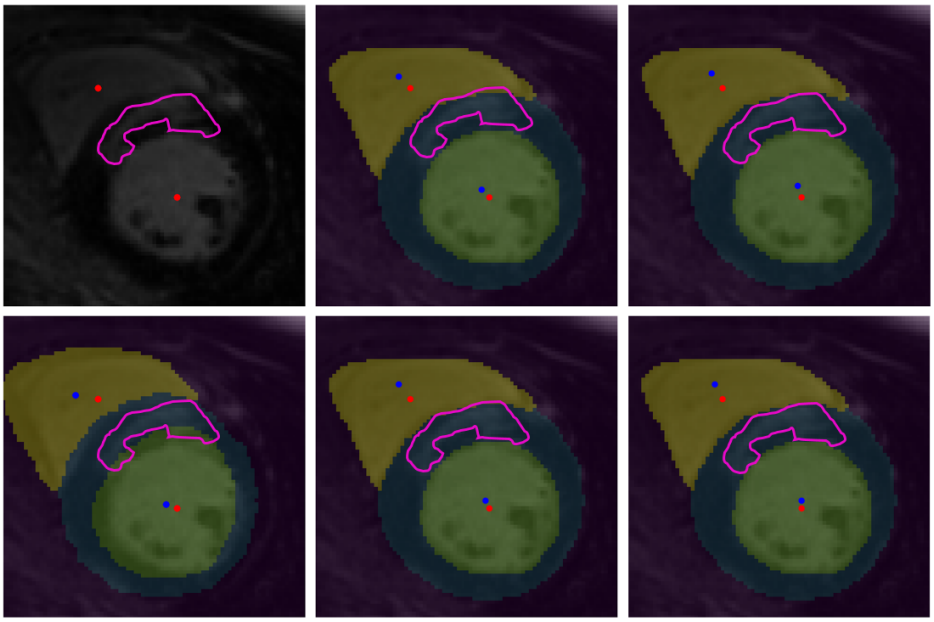


Figure 4.7: Panel 1-1: LGE CMR image and associated hyper-enhanced regions marked by pink contour and LV and RV blood-pool centers marked by red dots; Panel 2-1: before registration (CD: 3.46 *mm*, ASD: 1.87 *mm*); overlaid unregistered LGE CMR image and features (from Panel 1-1) onto the bSFFP image showing the RV blood-pool (yellow) and LV blood-pool (green) and their centers (marked by blue dots) and the LV myocardium (blue) marked on the cine CMR image; Panel 1-2: overlaid LGE CMR image onto the cine CMR image following stand-alone STN model registration using 25 patients for training (CD: 2.77 *mm*, ASD: 1.62 *mm*); Panel 2-2: stand-alone STN model registration using 35 patients for training (CD: 2.72 *mm*, ASD: 1.45 *mm*); Panel 1-3: joint deep learning model registration using 25 patients for training (CD: 2.77 *mm*, ASD: 1.51 *mm*); and Panel 2-3: joint deep learning model registration using 35 patients for training (CD: 2.65 *mm*, ASD: 1.47 *mm*).

4.4 Discussion

Primarily, we present a STN inspired CNN architecture to register the bSSFP cine CMR images to its corresponding LGE CMR images in a supervised manner using a dual-loss (weighted Dice loss and weighted cross-entropy loss) cost function. Our experiments show a statistically significant reduction of the CD between the bSSFP cine CMR images and the LGE CMR images in LV blood-pool from 3.28 *mm* before registration to 2.22 *mm* after supervised rigid registration and 2.27 *mm* after supervised affine registration, and in RV blood-pool from 4.36 *mm* before registration to 2.69 *mm* after supervised rigid registration and 2.52 *mm* after supervised affine registration. We also observed a statistically significant improvement in

the ASD between the bSSFP and LGE MRI images in LV blood-pool from 2.53 mm before registration to 2.14 mm after supervised rigid registration and 2.09 mm after supervised affine registration, in LV myocardium from 1.78 mm before registration to 1.42 mm after supervised rigid registration and 1.40 mm after supervised affine registration, and in RV blood-pool from 2.42 mm before registration to 1.72 mm after supervised rigid registration and 1.73 mm after supervised affine registration. These results are achieved with minimal pre-processing i.e. resampling all the images to a slice thickness of 5 mm (using spline interpolation), pixel spacing of 0.75 $mm \times 0.75$ mm and then resizing them to a common resolution of 224×224 pixels. Another major advantage of our proposed method is the time required for training (80 seconds to train each epoch).

Our proposed supervised method outperforms both the unsupervised rigid and unsupervised affine registration methods. The registration results of unsupervised methods are obtained by training our network using NMI as cost-function. We also experimented with structural similarity image measure (SSIM) loss as a cost function, which yielded similar results. The manual annotations of LV blood-pool, LV myocardium and RV blood-pool used during supervised training enables the network to focus on registering the images accurately around the regions of interest, improving the overall registration accuracy.

We also experimented using only the manual annotations of LV blood-pool and LV myocardium, however the rotational transformation of the registration fails due to the circular nature of the LV blood-pool and LV myocardium. Hence, one potential drawback is the need for annotations of cardiac structures for training. While the LV blood-pool, LV myocardium and RV blood-pool labels used here were available through the challenge, there exist numerous sufficiently accurate and robust cardiac image segmentation methods, such as [16, 17, 18, 19, 20], that can be leveraged to annotate the desired structures from the bSSFP CMR images to provide sufficient rotational asymmetry for optimal registration. Lastly, another minor drawback is the possibility of losing certain critical information around the cardiac structures during image resampling prior to registration, but such challenges have always been faced in image registration, an example being atlas construction from images featuring different

in-plane resolution and slice thickness.

We would like to address the insignificant difference between the supervised rigid and supervised affine registration approach. Note that both the bSSFP cine CMR and the LGE CMR images are acquired during the same imaging exam, using the same scanner, without changing the patient position, and while also employing ECG gating for end-diastole image capture, resulting in similar shapes and sizes of the cardiac structures.

We showed that a STN inspired RoI-guided CNN architecture can be reliably used to register the bSSFP cine CMR and LGE CMR images. The major drawback of the method is the need for annotations of cardiac structures for large number of training data. Hence, we investigate whether the joint deep learning framework is a viable option for registration of LGE and bSSFP cine CMR images. Our results reveal that the proposed joint deep learning model leverages the weights learnt from the segmentation task to improve the registration accuracy and produces reliable registration results using lesser number of training data and manual annotations.

The mean Dice scores achieved by the segmentation branches of the bSSFP cine CMR images and LGE CMR images are 84.73% and 71.49%, respectively. The poor results of the segmentation branches are due to the limited number of the training data, however, the weights learnt from these segmentation branches improve the registration accuracy in the joint deep learning model. The computational time required for each epoch for a stand-alone STN model is around 63 seconds and 67 seconds to train 25 and 35 of the 45 available image datasets, respectively, whereas the joint deep learning model requires around 110 seconds and 155 seconds to train 25 and 35 of the 45 available image datasets, respectively. It is worth to be noted that although the stand-alone STN model takes less training time for the registration, it nevertheless requires manual annotations, while the joint model requires relatively more time for training, the number of manual annotations needed are fewer.

This study was conducted using LGE and bSSFP cine MRI images from patients diagnosed with cardiomyopathy. Nevertheless, the proposed methods are useful for registering any gadolinium-enhanced and cine MRI images of any patient provided their cardiac conditions

is visible and appropriate for gadolinium-enhanced imaging during diagnosis. Such conditions include, but are not limited to, myocarditis or myocardial infarction, or other diseases that show hyper-enhancement of the compromised myocardial regions.

To further reduce the reliance on manual annotations to conduct the RoI-guided registration, in our future work we plan to investigate the use of previously validated machine learning-based segmentation techniques [16, 17, 18, 19, 20] to automatically extract the required ROI labels (as manual annotated labels are not typically available for large datasets), then proceed with the registration.

Another area of improvement for our future work is to correct for slice misalignment during MR image acquisition prior to stacking up the segmented and registered image slices in the effort to build 3D models that help quantify and visualize the compromised myocardial regions using 3D maps, leveraging the methods by Dangi *et al.* [27].

4.5 Conclusion

In this work, we first show that the proposed STN based RoI-guided CNN can be used to register bSSFP cine CMR sequence and LGE CMR sequence accurately and in a time-efficient manner. Our proposed method outperforms unsupervised deep learning algorithms trained using popular similarity metrics such as NMI.

Next, we present a joint deep learning model for registration of LGE and bSSFP cine CMR images, and the segmentation of cardiac chambers from the LGE CMR and the bSSFP CMR images. The coupling of the segmentation and the registration tasks enables a multi-task training and results in obtaining reliable registration results using a lower number of training datasets, reducing the need for a large number of manual annotations.

As part of our future work, we will be investigating other variants of U-Net architecture to improve the segmentation performance of the joint deep learning model and these obtained segmentation masks can be used to further fine-tune the registration in case of sparsely annotated datasets, resulting in a weakly-supervised method for registration. Ultimately, we intend to build 3D models that help quantify and visualize the compromised myocardial regions.

Bibliography

[1] Emelia J Benjamin, Michael J Blaha, Stephanie E Chiuve, Mary Cushman, Sandeep R Das, Rajat Deo, J Floyd, M Fornage, C Gillespie, CR Isasi, et al. Heart disease and stroke statistics-2017 update: a report from the american heart association. *Circulation*, 135(10):e146–e603, 2017. 3.1, 4.1

[2] Dong Wei, Ying Sun, Ping Chai, Adrian Low, and Sim Heng Ong. Myocardial segmentation of late gadolinium enhanced MR images by propagation of contours from cine MR images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 428–435. Springer, 2011. 1.3.2.2, 4.1, 4.1

[3] Laura Jimenez Juan, Andrew M Crean, and Bernd J Wintersperger. Late gadolinium enhancement imaging in assessment of myocardial viability: techniques and clinical applications. *Radiologic Clinics*, 53(2):397–411, 2015. 4.1

[4] Xiahai Zhuang. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):2933–2946, 2018. 1.3.2, 4.1, 4.2.1

[5] Xiahai Zhuang. Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 581–588. Springer, 2016. 1.3.2, 4.1, 4.2.1

[6] Steffen E Petersen, Miguel A González Ballester, and Karim Lekadir. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhance-

ment cardiac MRI. In *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers*, page 290. Springer Nature. 4.1

[7] Yashu Liu, Wei Wang, Kuanquan Wang, Chengqin Ye, and Gongning Luo. An automatic cardiac segmentation framework based on multi-sequence MR image. *arXiv preprint arXiv:1909.05488*, 2019. 4.1

[8] Xumin Tao, Hongrong Wei, Wufeng Xue, and Dong Ni. Segmentation of multimodal myocardial images using shape-transfer gan. *arXiv preprint arXiv:1908.05094*, 2019. 4.1

[9] Chen Chen, Cheng Ouyang, Giacomo Tarroni, Jo Schlemper, Huaqi Qiu, Wenjia Bai, and Daniel Rueckert. Unsupervised multi-modal style transfer for cardiac MR segmentation. *arXiv preprint arXiv:1908.07344*, 2019. 4.1

[10] Yasmina Chenoune, Constantin Constantinides, Racha El Berbari, Elodie Roullot, Frédérique Frouin, Alain Herment, and Elie Mousseaux. Rigid registration of delayed-enhancement and cine cardiac MR images using 3D normalized mutual information. In *2010 Computing in Cardiology*, pages 161–164. IEEE, 2010. 1.3.2.1, 4.1

[11] Fumin Guo, Mengyuan Li, Matthew Ng, Graham Wright, and Mihaela Pop. Cine and multicontrast late enhanced MRI registration for 3D heart model construction. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 49–57. Springer, 2018. 1.3.2.1, 1.3.2.3, 4.1

[12] Azira Khalil, Siew-Cheok Ng, Yih Miin Liew, and Khin Wee Lai. An overview on image registration techniques for cardiac diagnosis and treatment. *Cardiology research and practice*, 2018, 2018. 1.3.2.3, 4.1

[13] Matthew CH Lee, Ozan Oktay, Andreas Schuh, Michiel Schaap, and Ben Glocker. Image-and-spatial transformer networks for structure-guided image registration. In *International*

Conference on Medical Image Computing and Computer-Assisted Intervention, pages 337–345. Springer, 2019. 4.1

[14] Hessam Sokooti, Bob De Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, and Marius Staring. Nonrigid image registration using multi-scale 3D convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–239. Springer, 2017. 4.1

[15] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2758–2766, 2015. 4.1

[16] Shusil Dangi, Cristian A Linte, and Ziv Yaniv. A distance map regularized CNN for cardiac cine MR image segmentation. *Medical physics*, 46(12):5637–5651, 2019. 4.1, 4.4

[17] Roshan Reddy Upendra, Shusil Dangi, and Cristian A Linte. An adversarial network architecture using 2D U-net models for segmentation of left ventricle from cine cardiac MRI. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 415–424. Springer, 2019. 4.1, 4.4

[18] Roshan Reddy Upendra, Shusil Dangi, and Cristian A Linte. Automated segmentation of cardiac chambers from cine cardiac MRI using an adversarial network architecture. In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11315, page 113152Y. International Society for Optics and Photonics, 2020. 4.1, 4.4

[19] Shusil Dangi, Ziv Yaniv, and Cristian A Linte. Left ventricle segmentation and quantification from cardiac cine MR images via multi-task learning. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 21–31. Springer, 2018. 1.3.1.3, 4.1, 4.4

[20] SM Kamrul Hasan and Cristian A Linte. Condenseunet: a memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation. In *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 11315, page 113151J. International Society for Optics and Photonics, 2020. 3.1, 3.2.4.2, 3.3.0.2, 4.1, 4.4

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention. Lecture Notes in Computer Science*, volume 9351, pages 234–241. Springer, 2015. 2.2.3, 2.2.4.1, 4.1

[22] Chen Qin, Wenjia Bai, Jo Schlemper, Steffen E Petersen, Stefan K Piechnik, Stefan Neubauer, and Daniel Rueckert. Joint learning of motion estimation and segmentation for cardiac MR image sequences. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 472–480. Springer, 2018. 1.3.1.3, 1.3.3.2, 3.1, 4.1, 4.2.4

[23] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015. 1.3.2.3, 3.2.3.1, 4.2.2

[24] Yi-Hsuan Tsai, Ming-Hsuan Yang, and Michael J Black. Video segmentation via object flow. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3899–3908, 2016. 4.2.4

[25] Jingchun Cheng, Yi-Hsuan Tsai, Shengjin Wang, and Ming-Hsuan Yang. Segflow: Joint learning for video object segmentation and optical flow. In *Proceedings of the IEEE international conference on computer vision*, pages 686–695, 2017. 4.2.4

[26] Roshan Reddy Upendra, Richard Simon, and Cristian A Linte. A supervised image registration approach for late gadolinium enhanced MRI and cine cardiac MRI using

convolutional neural networks. In *Annual Conference on Medical Image Understanding and Analysis*, pages 208–220. Springer, 2020. 4.2.5.2

[27] Shusil Dangi, Cristian A Linte, and Ziv Yaniv. Cine cardiac MRI slice misalignment correction towards full 3D left ventricle segmentation. In *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10576, page 1057607. International Society for Optics and Photonics, 2018. 3.2.2, 4.4

Chapter 5

A Deep Learning Framework for Image Super-Resolution for Late Gadolinium Enhanced Cardiac MRI

Cardiac magnetic resonance imaging (MRI) provides 3D images with high-resolution in-plane information, however, they are known to have low through-plane resolution due to the trade-off between resolution, image acquisition time and signal-to-noise ratio. [1] *This results in anisotropic 3D images which could lead to difficulty in diagnosis, especially in late gadolinium enhanced (LGE) cardiac MRI, which is the reference imaging modality for locating the extent of myocardial fibrosis in various cardiovascular diseases like myocardial infarction and atrial fibrillation. To address this issue, we propose a self-supervised deep learning-based approach to enhance the through-plane resolution of the LGE MRI images. We train a convolutional neural network (CNN) model on randomly extracted patches of short-axis LGE MRI images and this trained CNN model is used to leverage the information learnt from the high-resolution in-plane data to improve the through-plane resolution. We conducted experiments on LGE MRI dataset*

[1] This chapter is adapted from:
Upendra R.R. et al., "A convolutional neural network-based deformable image registration method for cardiac motion estimation from cine cardiac MR images." In 2020 Computing in Cardiology (pp. 1-4). IEEE.
Upendra R.R. et al., "A 3D Convolutional Neural Network with Gradient Guidance for Image Super-Resolution of Late Gadolinium Enhanced Cardiac MRI." In 2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC) (pp. 1707-1710). IEEE.

made available through the 2018 *atrial segmentation challenge. While the proposed 2D patch-based method improves the through-plane resolution, they ignore the global context information that is crucial for accurate segmentation and also, can be inefficient as it requires the use of patches even during the inference stage leading to inconsistencies during the fusion process. Additionally, the 2D slice-by-slice training employed by many researchers to train the CNN does not take advantage of the 3D information provided by the cardiac MRI images.*

To that end, we propose a 3D convolutional neural network (CNN) framework with two branches: a super-resolution branch to learn the mapping between low-resolution and high-resolution LGE-MRI volumes, and a gradient branch that learns the mapping between the gradient map of low-resolution LGE-MRI volumes and the gradient map of high-resolution LGE-MRI volumes. The gradient branch provides structural guidance to the CNN-based super-resolution framework. To assess the performance of the proposed CNN-based framework, we train two CNN models with and without gradient guidance, namely, dense deep back-projection network (DBPN) and enhanced deep super-resolution (EDSR) network. We train and evaluate our method on the 2018 atrial segmentation challenge dataset. Additionally, we also evaluate these trained models on the left atrial and scar quantification and segmentation challenge (LAScarQS) 2022 dataset to assess their generalization ability. Finally, we investigate the effect of the proposed CNN-based super-resolution framework on the 3D segmentation of the left atrium (LA) from these cardiac LGE-MRI image volumes.

5.1 Introduction

Cardiac MRI is the current gold standard to assess cardiac function and diagnose various cardiovascular diseases. MRI images provide dynamic 3D images of the heart with high-resolution in-plane information. In clinical cardiac MRI, due to the limitations of the maximal breath-hold time achievable by the patient, high-resolution 2D stacks of images are typically acquired resulting in anisotropic 3D volumes of the heart. Therefore, these 3D volumes usually have low through-plane resolution (i.e., slice thickness). For example, in a typical LGE cardiac MRI, which is widely used to assess the myocardium viability in post-infarct patient and study

the extent of fibrosis in the atria of patients with atrial fibrillation [1], etc., has a high in-plane resolution of 1 to 1.5 mm, but a low through-plane resolution of 5 to 10 mm [2]. In order to localize and quantify LA fibrosis with high precision, high-resolution 3D LGE-MRI images are necessary. Therefore, the anisotropic 3D LGE-MRI images with poor through-plane resolution impose challenges on downstream tasks, such as the 3D segmentation of LA cavity and fibrosis quantification.

Conventional interpolation methods such as bilinear, spline and Lancoz resampling methods can be used to upsample the low-resolution volumes to high-resolution volumes, however, these methods often cause artifacts, like blurring, and cannot recover the missing high-frequency semantic and structural information. In order to address this limitation, several researchers proposed learning-based super-resolution methods that learn the structural information between slices using low- and high-resolution image pairs [3, 4, 5]. In recent years, increasing research efforts proposed deep learning-based super-resolution algorithms for medical images [6], especially to enhance the through-plane resolution of anisotropic brain MRI images [7, 8, 9, 10, 11].

Subsequently, these deep learning-based super-resolution techniques have been applied to alleviate the through-plane resolution degradation in 3D cardiac MRI volumes. In response to the challenge of acquiring high resolution isotropic images, Steeden et al. [12] demonstrated the potential of a convolutional neural network (CNN)-based approach for super-resolution reconstruction of balanced steady state free precession (bSSFP) cardiac MRI images using synthetic training data. Masutani et al. [13] explored the feasibility of both single frame and multi-frame CNN models to generate super-resolution bSSFP cine cardiac MRI images. Basty et al. [14] showed that recurrent neural networks (RNN) can be used to reconstruct super-resolution cardiac cine MRI long-axis slices from low-resolution acquisitions by using the temporal recurrence, thereby, using the temporal context to improve the resolution of cardiac cine MRI volumes. Sander et al. [15] proposed an unsupervised deep learning-based approach to enhance the through-plane resolution of cine cardiac MRI volumes by leveraging the latent space interpolation ability of the autoencoders; however, large variations in anatomy

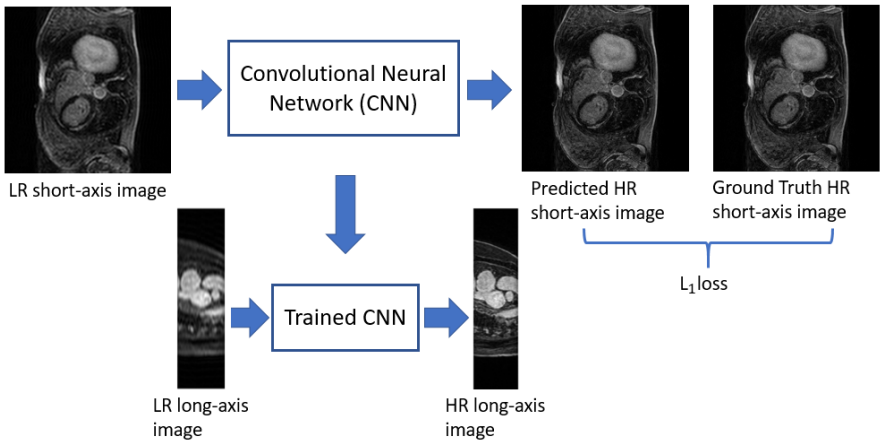


Figure 5.1: Self-supervised deep learning framework to improve through-plane resolution in LGE cardiac MRI

between adjacent slices affect the performance of the method. Zhao et al. [16] proposed a 2D CNN-based super-resolution method that takes advantage of the high-resolution information from the in-plane data to improve the through plane resolution. They applied this technique on cine cardiac, neural and tongue MRI images. These methods successfully improve the through-plane resolution of cine cardiac MRI images, however, limited efforts have been made to improve the resolution of LGE cardiac MRI images using CNN-based methods.

In order to improve the through-plane resolution of LGE cardiac MRI images to aide LA segmentation for AF patients, we first propose a 2D CNN-based method to enhance the through-plane resolution of LGE cardiac MRI images by leveraging the information learnt by training short-axis 2D patches to learn the mapping from simulated low-resolution in-plane data and their corresponding high-resolution in-plane data and using this learnt information to enhance the poor through-plane resolution 5.1.

In addition to the 2D self-supervised deep learning framework to compute super-resolution LGE MRI images, we developed a 3D CNN-based architecture with gradient guidance to generate super-resolution cardiac LGE-MRI images [17]. Inspired by the 2D structure-preserving super-resolution (SPSR) method [18], we extended a 3D gradient branch that guides our 3D CNN model to "pay more attention" to the 3D structure of the tissues in the LGE-MRI images, as illustrated in Fig. 5.2. Our main contributions in this work can be summarized as follows:

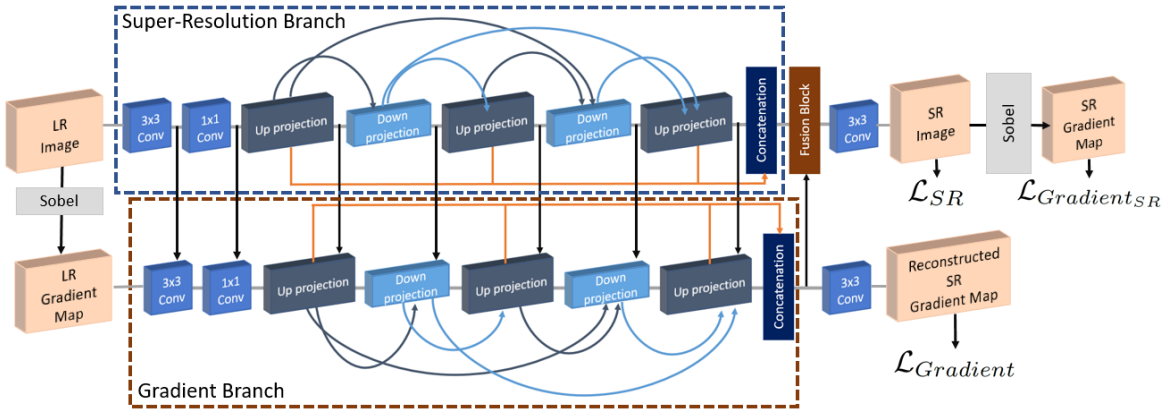


Figure 5.2: Framework of the proposed dense DBPN-based super-resolution (SR) method with gradient guidance.

1. Firstly, to enhance the resolution of 3D cardiac LGE-MRI images, we present a 3D deep learning-based framework with two branches: a super-resolution branch with 3D dense deep back-projection network (DBPN) [19] as the backbone of our CNN architecture and an auxiliary gradient branch. As illustrated in Fig. 5.2, while the super-resolution branch learns the mapping between the low-resolution input image and its corresponding high-resolution image, the gradient branch learns the mapping between the gradient map of the low-resolution input image and the gradient map of its corresponding high-resolution image. We evaluate the performance of the proposed super-resolution method by training and testing them on the 2018 atrial segmentation challenge dataset [20].

2. Secondly, we further assess the performance of the proposed gradient guidance method by replacing the dense DBPN model with the enhanced deep super-resolution (EDSR) [21] network as the backbone of our deep learning framework.

3. Thirdly, in addition to evaluating our methods by training and testing them on the 2018 atrial segmentation challenge dataset [20], we also evaluate the generalization ability of our trained models by testing them on an external test set, the left atrial and scar quantification and segmentation challenge (LAScarQS) 2022 dataset [22, 23, 24].

4. Finally, we investigate the effect of the proposed super-resolution framework on the

downstream segmentation task, i.e., the segmentation of the LA from these LGE-MRI volumes.

5.2 Materials and Methods

5.2.1 Data

A set of 154 3D LGR-MRI volumes obtained from 60 patients with AF was available through the 2018 atrial segmentation challenge [20]. These clinical images were acquired with either a Siemens 1.5 Tesla Avanto or a 3.0 Tesla Verio scanner. We split the available 154 LGE-MRI volumes to 80 for training, 20 for validation and 54 for testing.

The spatial dimensions of these LGE-MRI volumes are either 576x576x88 or 640x640x88, and feature an isotropic voxel spacing of 0.625x0.625x0.625 mm^3. In order to train our CNN models, we need to simulate low-resolution LGE-MRI volumes from the available high-resolution LGE-MRI volumes. Therefore, we first center-crop the high-resolution images to 224x224x88 and downsample them randomly on-the-fly during training using either Fourier or Gaussian downsampling, with uniform distribution. In Fourier downsampling, we block the high frequency information in the w-axis of the Fourier domain by truncating the k-space in order to simulate the low-resolution data acquisition process in the Fourier domain. In Gaussian downsampling, we simulate low-resolution images using Gaussian blur with a standard deviation in the range of $[0.5, 1.5]$ and downsample them using linear interpolation in the z-axis direction, i.e., slice-encoding direction. We downsample the LGE-MRI volumes using a scale factor of 2, 4, and 8. Subsequently, to train our 3D CNN models, we generate 3D LGE-MRI patches of size 96x96x44, 96x96x22 and 96x96x11, respectively, with an overlap of 33.33%.

Additionally, to evaluate the generalization ability of our trained models, we use a subset (only the high-resolution isotropic LGE-MRI data) of the training dataset from the LAScarQS 2022 segmentation challenge dataset [22, 23, 24] as our external test set. These clinical images feature a spatial resolution of 1.4x1.4x1.4 mm^3 and were acquired using a Philips Achieva 1.5 Tesla scanner from patients with AF. We test the models trained on the 2018 atrial

segmentation challenge [20] dataset on these 30 LGE-MRI volumes from the LAScarQS 2022 segmentation challenge dataset [22, 23, 24]

5.2.2 Self Supervised Super-Resolution Framework

In order to generate low-resolution data for training, we first blur the images in the x-axis to obtain low-resolution in-plane images. This is done by Fourier downsampling to simulate data acquisition process in MRI and to ensure no high frequency information on the u-axis in the Fourier domain. Now, we have the low-resolution in-plane data and their corresponding high-resolution in-plane data to train the CNN. This CNN model will be trained to learn the mapping between the low-resolution and the high-resolution data. We also repeat the Fourier downsampling process in z-axis to obtain low-resolution through-plane data. This low-resolution through-plane data is used as the test dataset for our experiments.

To train the CNN model, we first extract patches of dimensions 640x88 pixels from the low-resolution short-axis images in both horizontal and vertical directions. These low-resolution patches are input to a 2D CNN, an encoder-decoder network with skip connections (U-Net [25]). The output of the CNN and the corresponding high-resolution patch is used to compute a L_1 loss function to backpropagate the CNN, thereby, learning the mapping from low-resolution in-plane data to high-resolution in-plane data. This mapping is subsequently applied to long-axis images to improve the through-plane resolution (Fig. 5.1).

In our experiments, we split the total 154 LGE MRI dataset to 100 for training and 54 for testing in a 3-fold cross-validation strategy. The networks are trained using the Adam optimizer with a learning rate of 10^{-4} and a gamma decay of 0.99 every alternate epoch for fine-tuning, a batch size of 20 patches, for 50 epochs on a machine equipped with NVIDIA RTX 2080 Ti GPU with 11GB of memory.

5.2.3 3D CNN Super-Resolution Framework with Gradient Guidance

As illustrated in Fig. 5.2, our proposed CNN framework to generate super-resolution cardiac LGE-MRI volumes consists of two branches: a super-resolution branch and a gradient branch.

5.2.3.1 Super-Resolution Branch

The super-resolution branch takes in low-resolution images I^{LR} as input and aims to generate super-resolution images I^{SR} as output, given the high-resolution images I^{HR} as ground-truth. In our work, we use the dense DBPN [19] model to super-resolve the low-resolution LGE-MRI volumes. The dense DBPN model illustrated in Fig. 5.2 can be split into three parts: initial feature extraction, back-projection and reconstruction. In the initial feature extraction stage, we construct the initial low-resolution feature maps from I^{LR} using 32 filters, which is then further reduced to 16 filters before entering the back-projection stage. Following the initial feature extraction step, in the back-projection stage, we have a sequence of three up-projection and two down-projection blocks, wherein, each block has access to outputs of all the previous blocks (Fig. 5.2). This enables the generation of effective feature maps [19]. Here, the up- and down-projection blocks are alternating between the construction of low-resolution and high-resolution feature maps and the number of filters used in each projection block is set to 16. Finally, in the reconstruction stage, all the high-resolution feature maps from the up-projection blocks are concatenated, along with the output of the gradient branch, to generate I^{SR} as output.

5.2.3.2 Gradient Branch

The target of the gradient branch is to learn the mapping between the gradient map of the low-resolution images $G(I^{LR})$ and the gradient map of their corresponding high-resolution images $G(I^{HR})$, thereby, reconstructing a super-resolution gradient map $I^{SR}_{gradient}$. Here, $G(\cdot)$ stands for the operation that extracts the gradient map of the images, which in our case is a Sobel filter. Similar to the super-resolution branch, we use a dense DBPN model to learn the mapping in the gradient branch. As illustrated in Fig. 5.2, the gradient branch incorporates feature map representations from the super-resolution branch at every level, as opposed to incorporating features only at the up-projection level, as shown in our earlier work [17]. The advantage of this step is that it enables the reconstruction of the higher-resolution gradient map using the rich structural information from the super-resolution branch (strong

prior) and reduces the number of parameters needed for the gradient branch. Next, the high-resolution feature maps from the up-projection blocks of the gradient branch are concatenated and integrated with the super-resolution branch to guide reconstruction of the super-resolution 3D LGE-MRI images. The motivation behind the integration of high-resolution feature maps from the gradient branch to the super-resolution branch is that it can implicitly echo if the recovered region should be sharp or smooth, thus preserving the structure of the tissue as the CNN concentrates more on the spatial relationships of the outlines in the gradient branch. Meanwhile, the concatenated high-resolution feature maps from the up-projection blocks of the gradient branch are used to reconstruct super-resolution gradient map $I^{SR}_{gradient}$.

5.2.3.3 Objective Function

Our proposed CNN model is trained using the following objective function:

$$\mathcal{L} = \alpha.\mathcal{L}_{SR} + \beta.\mathcal{L}_{Gradient} + \gamma.\mathcal{L}_{Gradient_{SR}}\,, \tag{5.1}$$

where $\mathcal{L}_{SR}$ is the L_1 loss computed between I^{SR} and I^{HR}, $\mathcal{L}_{Gradient}$ is the L_1 loss computed between $I^{SR}_{gradient}$ and $G(I^{HR})$, and $\mathcal{L}_{Gradient_{SR}}$ denotes L_1 loss between $G(I^{SR})$ and $G(I^{HR})$. In Equation (5.1), the α, β and γ represent the scalar weights associated with the $\mathcal{L}_{SR}$, $\mathcal{L}_{Gradient}$ and $\mathcal{L}_{Gradient_{SR}}$ loss functions, respectively.

5.2.3.4 Experiments

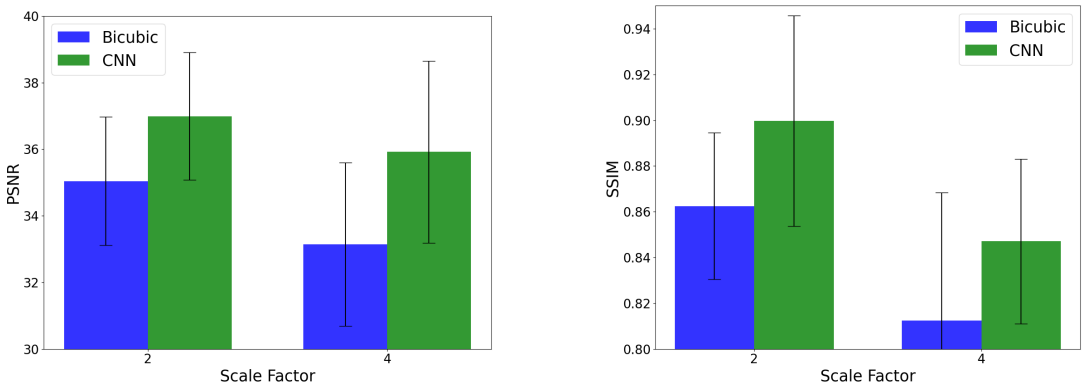
In order to evaluate the effectiveness of our proposed method, three experiments were designed.

In the first experiment, we compare the results of our proposed framework with bicubic interpolation and dense DBPN model without gradient guidance. Additionally, to assess the effectiveness of the gradient branch, we use the EDSR model [21] as the back-bone network in our proposed framework and compare it with the EDSR network without gradient guidance. Here, we split the 154 LGE-MRI dataset made available through the 2018 atrial segmentation challenge [20] into 80 datasets for training, 20 datasets for validation and 54 datasets for testing. We refer to this test set as the internal test set since the training and test sets belong

to the same dataset. As discussed in Section 5.2.1, we train the models on low-resolution LGE-MRI volumes obtained by downsampling high-resolution volumes using a scale factor of 2, 4 and 8, respectively. We train these models on a NVIDIA RTX 2080 Ti GPU with 11 GB memory using the Adam optimizer with a learning rate of 10^{-4} and a gamma decay of 0.5 every 15 epochs, for 50 epochs.

In the second experiment, we apply the models trained on the 2018 atrial segmentation challenge [20] to test them on a subset of the LAScarQS 2022 segmentation challenge dataset [22, 23, 24], which we refer to as the external test set. This is done to assess the generalization ability of the proposed framework. We test the trained models on 30 LGE-MRI volumes from the LAScarQS dataset, which is the total number of high-resolution isotropic LGE-MRI data available in the training set of the LAScarQS dataset.

In our third experiment, we use the super-resolved LGE-MRI volumes generated by each of the above-mention algorithms to train the U-Net [25] models to segment the LA chamber, and compare the segmentation results, in order to investigate the effect of the proposed super-resolution framework on the downstream segmentation tasks. We use the similar split, i.e., 80 for training, 20 for validation and 54 for testing from the 2018 atrial segmentation challenge [20] dataset to train the U-Net models. The ground-truth LA segmentation masks provided by the 2018 atrial segmentation challenge dataset [20] were manually annotated by experts in the field and were used to train the U-Net models. While training, we augment the LGE-MRI images randomly on-the-fly using translation, rotation and gamma correction operations. These U-Net models are trained on the NVIDIA RTX 2080 Ti GPU with 11 GB memory using the Adam optimizer with a learning rate of 10^{-5} and a gamma decay of 0.99 every alternate epoch, for 100 epochs.



(a) (b)

Figure 5.3: Comparison of (a) mean PSNR and (b) mean SSIM values achieved by bicubic interpolation and the proposed CNN framework

Table 5.1: Mean (std-dev) peak signal-to-noise-ratio (PSNR) and structural similarity index measure (SSIM) achieved using bicubic interpolation and our proposed self supervised CNN framework for downsampling scale factor of 2 and 4, respectively. The best evaluation metrics achieved are labeled in **bold**. Statistically significant differences were evaluated using the Student t-test and are reported using * $p < 0.005$.

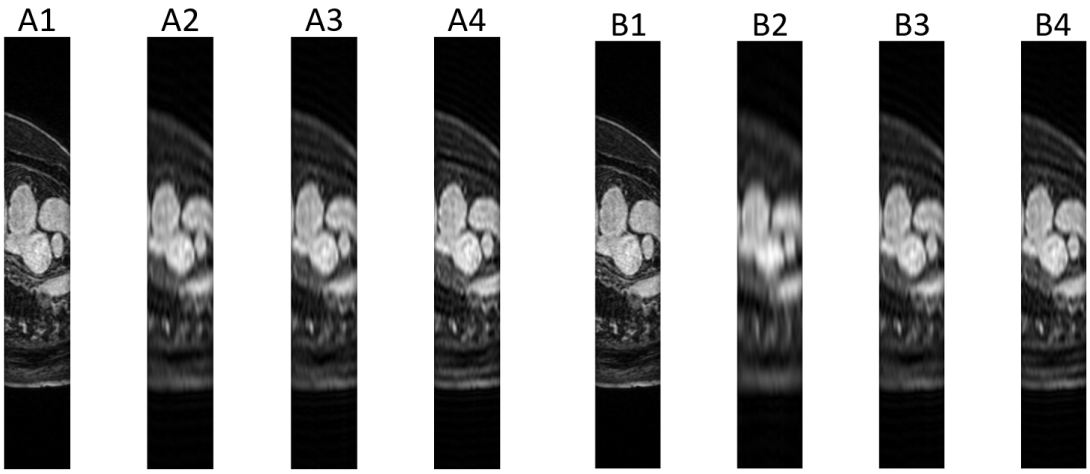
Methods	Scale Factor: 2		Scale Factor: 4	
	PSNR	SSIM	PSNR	SSIM
Bicubic Interpolation	35.04 (1.93)	0.86 (0.03)	33.14 (2.45)	0.81 (0.05)
CNN	**36.99 (1.91)***	**0.90 (0.04)***	**35.92 (2.73)***	**0.84 (0.03)***

5.3 Results

5.3.1 Evaluation of Self Supervised Super-Resolution Framework

To evaluate our results, we compute the mean PSNR and mean SSIM between the super-resolution long-axis images obtained from the proposed method and the ground truth high-resolution long-axis images. We then compare the computed PSNR and SSIM with the results obtained by bicubic interpolation.

Table 5.1 shows a comparison of the mean PSNR and mean SSIM between our proposed self supervised CNN method and the bicubic interpolation for low-resolution images simulated by



(a) (b)

Figure 5.4: (a) Long axis views of LGE cardiac MRI with (A1) ground-truth high-resolution, (A2) low-resolution with downsampling scale factor 2 (PSNR: 27.34, SSIM: 0.61), (A3) the LR image upsampled by bicubic interpolation (PSNR: 29.93, SSIM: 0.73) and (A4) the super-resolution image from CNN model (PSNR: 32.43, SSIM: 0.81); and (b) long axis views of LGE cardiac MRI with (B1) ground-truth high-resolution, (B2) low-resolution with downsampling scale factor 4 (PSNR: 23.42, SSIM: 0.48), (B3) the LR image upsampled by bicubic interpolation (PSNR: 28.38, SSIM: 0.69) and (B4) the super-resolution image from self supervised CNN model (PSNR: 31.04, SSIM: 0.79)

a downsampling scale factor of 2 and 4. We achieved a mean PSNR of 36.99 and 35.92 using our trained CNN model on images downsampled by a scale factor of 2 and 4, respectively, compared to 35.04 and 33.14, respectively, using bicubic interpolation alone. Similarly, we achieved a mean SSIM of 0.9 and 0.84 using our trained CNN model on images downsampled by a scale factor of 2 and 4, respectively, compared to 0.86 and 0.81, respectively, using bicubic interpolation alone (Fig. 5.3). We show an example of the improved through-plane resolution for low-resolution images simulated by a downsampling scale factor of 2 and 4 in Fig. 5.4a and Fig. 5.4b, respectively.

5.3.2 Evaluation of 3D CNN Super-Resolution Framework with Gradient Guidance

5.3.2.1 Validation on Internal Test Set

The performance of the proposed 3D CNN framework is evaluated by computing the mean peak signal-to-noise ratio (PSNR) and mean structural similarity index (SSIM) between the super-resolution 3D LGE-MRI volumes and ground-truth high-resolution 3D LGE-MRI volumes. In Table 5.2, we show the super-resolution results of different methods, namely, bicubic interpolation, EDSR model with and without gradient guidance, and dense DBPN model with and without gradient guidance, respectively, on the internal test set. The super-resolution models are trained and tested on simulated low-resolution images downsampled by a scale factor of 2, 4 and 8, respectively. To compare the performance of the experimented models, we conducted a statistical significance (Student's t-test) test. While the CNN models outperform the bicubic interpolation ($p < 0.01$) for all the three downsampling scale factors, our experiments show higher PSNR and SSIM for the EDSR and dense DBPN models with gradient guidance compared to their stand-alone counterparts. In order to show the improvement in through-plane resolution using the proposed CNN framework, in Fig. 5.5, we show an example of a Gaussian downsampled low-resolution LGE cardiac MRI slice along the z-axis with its corresponding high-resolution image slice, and the super-resolution images obtained using the above-mentioned methods for all three downsampling scale factors.

5.3.2.2 Generalization Testing: Validation on External Test Set

The generalization ability of our method is evaluated using the external test set (Table 5.3). The 3D CNN models trained on the 2018 atrial segmentation challenge [20] were tested on this external test set (LAScarQS 2022 dataset [22, 23, 24]). Similar to the internal test set, low-resolution images were simulated by a downsampling scale factor of 2, 4 and 8, respectively, in order to compute super-resolution LGE-MRI volumes. The super-resolution results on the external dataset are analogous to the internal dataset, i.e., the CNN models outperform the bicubic interpolation results and the PSNR and SSIM achieved by the CNN models with

Table 5.2: Internal Test Set Evaluation: Mean (std-dev) peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) achieved using bicubic interpolation, EDSR model with and without gradient guidance, and dense DBPN model with and without gradient guidance for a downsampling scale factor of 2, 4 and 8, respectively. The best evaluation metrics achieved are labeled in **bold**. Statistically significant differences were evaluated between the CNN models with and without gradient guidance using the Student's t-test and are reported using * $p < 0.05$ and ** for $p < 0.01$.

Methods	Scale Factor: 2		Scale Factor: 4		Scale Factor: 8	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic Interpolation	26.55 (2.04)	0.785 (0.085)	22.56 (1.54)	0.613 (0.083)	19.86 (1.68)	0.378 (0.070)
EDSR	29.80 (1.60)	0.881 (0.064)	24.41 (1.70)	0.695 (0.071)	20.69 (1.62)	0.395 (0.043)
EDSR with Gradient Guidance	29.97 (2.04)	0.880 (0.069)	25.13 (1.48)	0.740 (0.065)**	20.80 (1.51)	0.417 (0.047)*
Dense DBPN	29.59 (1.80)	0.879 (0.065)	25.36 (1.55)	0.730 (0.061)	20.74 (1.48)	0.395 (0.046)
Dense DBPN with Gradient Guidance	**30.93 (1.79)***	**0.916 (0.065)****	**26.63 (1.48)***	**0.763 (0.061)****	**20.81 (1.45)**	**0.421 (0.048)**

gradient guidance are significantly higher than the stand-alone CNN models. We also show an example of the improved through-plane resolution on the external dataset in Fig. 5.6.

5.3.2.3 Effect of Super-Resolution on Downstream Segmentation Task

In order to show the effect of super-resolution on downstream segmentation task, we train 3D U-Net models to segment the LA from LGE-MRI images. Since the dense DBPN models has resulted in higher PSNR and SSIM values compared to the EDSR models, we train the U-Net models on the 2018 atrial segmentation challenge [20] dataset by simulating low-resolution images and upsampling them by dense DBPN model with gradient guidance, and comparing them with the segmentation results obtained by training U-Net models on images generated using dense DBPN model without gradient guidance and bicubic interpolation. We summarize the segmentation performance on these upsampled images in Table 5.4 using Dice score. We can see that the segmentation results obtained by training the images upsampled using dense DBPN model with gradient guidance is significantly better than bicubic interpolation ($p < 0.01$) and dense DBPN model without gradient guidance ($p < 0.05$), for all the three

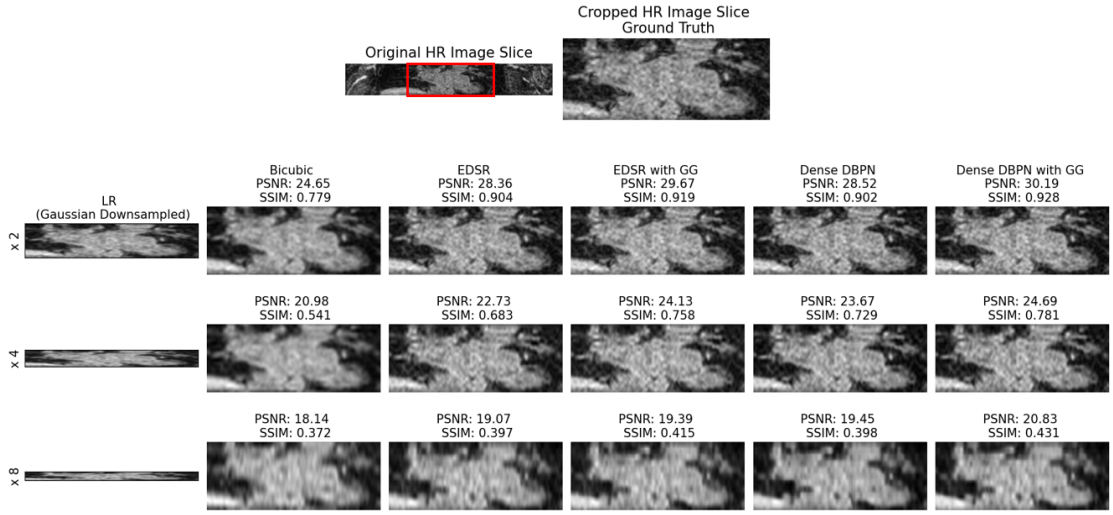


Figure 5.5: Visual assessment of reconstructed super-resolution images from **internal dataset**. Row 1: High-resolution (HR) LGE cardiac MRI slice along z-axis and its cropped version (ground truth). Low-resolution (LR) Gaussian downsampled image with downsampling scale factor 2 upsampled by bicubic interpolation, the super-resolution image from EDSR model, super-resolution image from EDSR model with gradient guidane (GG), the super-resolution image from dense DBPN model, and the super-resolution image from dense DBPN model with GG, respectively. Row 3: LR with downsampling scale factor 4 upsampled by the above-mentioned methods. Row 4: LR with downsampling scale factor 8 upsampled by the above-mentioned methods. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) evaluated between each super-resolved image and the original high-resolution grounds truth image is also labeled for interpretation vis-a-vis visual assessment.

downsampling scale factors.

Ablation Studies

In Table 5.5, we show the results of the ablation studies we performed to understand the contributions of the various components of the loss functions and the network. These studies were performed using the dataset from 2018 atrial segmentation challenge [20] and downsampling them by a factor of 4. We show the super-resolution results on the dense DBPN model with only $\mathcal{L}_{SR}$ as a loss function and without gradient guidance, followed by dense DBPN model with $\alpha.\mathcal{L}_{SR} + \gamma.\mathcal{L}_{Gradient_{SR}}$ as loss function and without gradient guidance, where $\mathcal{L}_{SR}$ is the L_1 loss computed between the generated super-resolution images, I^{SR} and ground-truth high-resolution images, I^{HR}, and $\mathcal{L}_{Gradient_{SR}}$ denotes L_1 loss between $G(I^{SR})$ and $G(I^{HR})$, where

Table 5.3: Generalization Evaluation on External Test Set: Mean (std-dev) peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) achieved using bicubic interpolation, EDSR model with and without gradient guidance, and dense DBPN model with and without gradient guidance for a downsampling scale factor of 2, 4 and 8, respectively. The best evaluation metrics achieved are labeled in **bold**. Statistically significant differences were evaluated between the CNN models with and without gradient guidance using the Student's t-test and are reported using * $p < 0.05$ and ** for $p < 0.01$.

Methods	Scale Factor: 2		Scale Factor: 4		Scale Factor: 8	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic Interpolation	27.60 (2.21)	0.798 (0.094)	22.86 (1.59)	0.629 (0.089)	20.21 (1.78)	0.377 (0.064)
EDSR	30.13 (1.51)	0.868 (0.039)	25.65 (1.50)	0.713 (0.047)	20.94 (1.49)	0.399 (0.044)
EDSR with Gradient Guidance	30.44 (1.49)	0.902 (0.038)*	25.94 (1.42)	0.760 (0.040)**	**21.57 (1.47)**	**0.424 (0.045)**
Dense DBPN	30.46 (1.27)	0.904 (0.033)	24.69 (1.68)	0.718 (0.059)	20.91 (1.64)	0.421 (0.056)
Dense DBPN with Gradient Guidance	**31.69 (1.73)***	**0.925 (0.038)***	**26.39 (1.39)***	**0.784 (0.034)****	21.03 (1.53)	0.423 (0.044)

$G(\cdot)$ stands for the gradient map of the images. Next, we show the super-resolution results of dense DBPN model with gradient guidance at only up-projection levels [17], followed by dense DBPN model with gradient guidance at all the levels (ours). As evident in Table 5.5, the dense DBPN model with gradient guidance at all the levels (ours) yields the highest PSNR and SSIM values across all tested models.

5.4 Discussion

In this paper, we first presented a self supervised CNN-based super-resolution framework to improve the through-plane resolution of LGE cardiac MRI images without the need for external training data to train the network. The CNN model is trained to learn the mapping of simulated short-axis low-resolution patches to their corresponding ground truth short-axis high-resolution patches. This information learnt from the in-plane data is used to improve the through-plane resolution. Our experiments show significantly improved PSNR and SSIM compared to the results obtained from bicubic interpolation. Lastly, the resulting super-

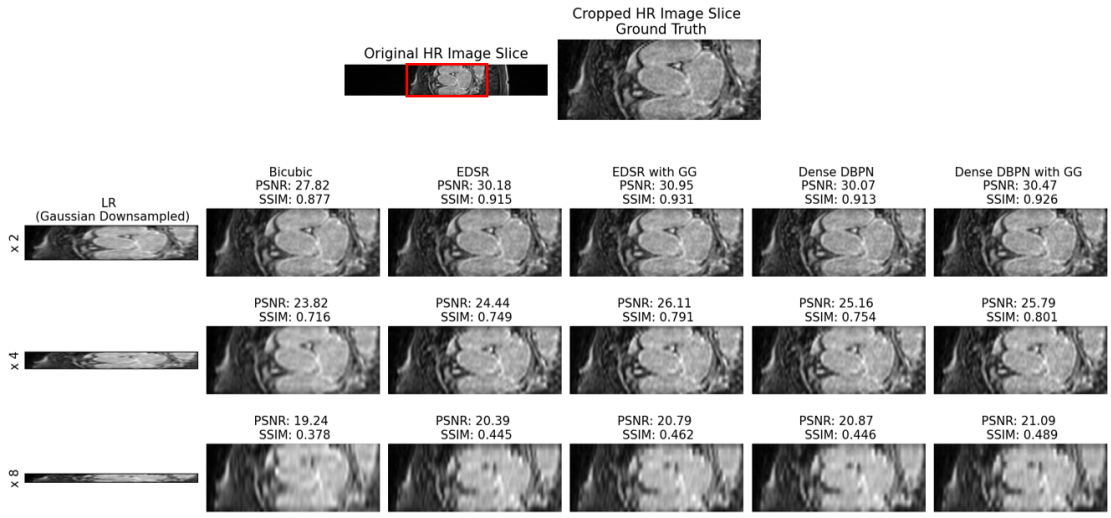


Figure 5.6: Visual Assessment of reconstructed super-resolution images from **external dataset**. Row 1: High-resolution (HR) LGE cardiac MRI slice along z-axis and its cropped version (ground truth). Low-resolution (LR) Gaussian downsampled image with downsampling scale factor 2 upsampled by bicubic interpolation, the super-resolution image from EDSR model, super-resolution image from EDSR model with gradient guidance (GG), the super-resolution image from dense DBPN model, and the super-resolution image from dense DBPN model with GG, respectively. Row 3: LR with downsampling scale factor 4 upsampled by the above-mentioned methods. Row 4: LR with downsampling scale factor 8 upsampled by the above-mentioned methods. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) evaluated between each super-resolved image and the original high-resolution grounds truth image is also labeled for interpretation vis-a-vis visual assessment.

resolution images featured less blurring and information loss than the bicubic interpolated images. While the proposed 2D patch-based method improves the through-plane resolution, it ignores the global context information, as well as the 3D information provided by the LGE-MRI images that is crucial for accurate segmentation; lastly, it also requires the use of 2D patches during the inference stage, which can lead to inconsistencies during the fusion process.

Therefore, we presented a novel 3D CNN-based framework with gradient guidance for super-resolution of cardiac LGE-MRI data. There are four major contributions in this work. First, a 3D deep learning-based architecture with two branches: a super-resolution branch and a gradient branch is presented, wherein, the dense DBPN model is the backbone of the CNN architecture. We exploited the structural information learnt by the gradient branch to provide structural guidance to the super-resolution branch of our CNN, thus preserving the

Table 5.4: Left atrium segmentation evaluation, mean (std-dev) Dice score (%) using the 2018 atrial segmentation challenge dataset. U-Net models are trained on LGE-MRI volumes downsampled by a scale factor of 2, 4 and 8, respectively, and upsampled using bicubic interpolation and dense DBPN model with and without gradient guidance. Statistically significant differences were evaluated between the dense DBPN model with and without gradient guidance using the Student's t-test and are reported using by * $p < 0.05$

Methods	Scale Factor: 2 Dice Score	Scale Factor: 4 Dice Score	Scale Factor: 8 Dice Score
Bicubic Interpolation	90.46 (1.98)	88.52 (1.47)	86.58 (2.49)
Dense DBPN	94.60 (3.33)	91.96 (1.47)	91.24 (1.83)
Dense DBPN with Gradient Guidance	**95.43 (2.01)***	**93.47 (2.02)***	**93.31 (2.07)***

Table 5.5: Ablation Study Results: Mean (std-dev) peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) achieved on the internal test set for a downsampling scale factor of 4. The best evaluation metrics achieved are labeled in **bold**.

Methods	Scale Factor: 4 PSNR	 SSIM
Dense DBPN L_{SR}	25.36 (1.55)	0.730 (0.061)
Dense DBPN $\alpha.\mathcal{L}_{SR} + \gamma.\mathcal{L}_{Gradient_{SR}}$	25.35 (1.36)	0.729 (0.058)
Dense DBPN Gradient Guidance at Up-projection	25.58 (1.59)	0.740 (0.065)
Dense DBPN with Gradient Guidance at all levels (Ours)	**26.63 (1.48)**	**0.763 (0.061)**

3D cardiac structures in the LGE-MRI images. To the best of our knowledge, this is the first work that presents a 3D CNN-based method to improve the through-plane resolution of cardiac LGE-MRI data. Second, we established that the presented gradient guidance method could improve the through-plane resolution of LGE-MRI data using other CNN models, such as EDSR model, as the backbone of the proposed deep learning framework. Third, we demonstrated the generalization ability of the proposed method by testing it on an external dataset that was not used while training our models. Fourth, we investigated and showed the effect of the proposed super-resolution framework on the downstream segmentation task by training a vanilla U-Net model to segment the LA from the super-resolved LGE-MRI volumes.

In our first experiment (Table 5.2), we compared the results of bicubic interpolation method

and the CNN models, with and without gradient guidance on the internal dataset. We can observe that the results of the CNN models with gradient guidance are significantly better than their stand-alone counterparts for all the three downsampling scale factors, wherein, the dense DBPN model with gradient guidance provides the best super-resolution results. This is more obvious in the SSIM comparison, as it is a combination of three comparison measures: luminance, contrast and structure [26], and as mentioned earlier, the gradient branch provides structural guidance to the proposed CNN framework.

In our second experiment (Table 5.3), we evaluated the above-mentioned models on the external dataset to assess their generalization ability. We can observe that the results are similar to the results achieved on the internal datasets, where CNN models with gradient guidance outperform their stand-alone counterparts, thus, generalizing well. It can be observed that the PSNR and SSIM values of the models when evaluated on the external dataset are higher than those achieved when the models were evaluated on the internal dataset. This could be attributed to the fact that the internal dataset features a high isotropic spatial resolution of 0.625x0.625x0.625 mm^3, whereas the external dataset features a relatively low isotropic spatial resolution of 1.4x1.4x1.4 mm^3. In Fig. 5.5 and 5.6, we show an example of an image slice along the z-axis from both internal and external dataset, respectively. The super-resolution images generated using the CNN models with gradient guidance show improved through-plane resolution; the reconstructed images feature less blurring and look sharper than the images upsampled using bicubic interpolation and CNN models without gradient guidance. In both figures, this observation is most obvious for the super-resolution images generated from the low-resolution images downsampled by a scale factor of 4.

In our final experiment (Table 5.4), we investigated the effect of the super-resolution models on the downstream segmentation task by training 3D U-Net models to segment LA from the super-resolved LGE-MRI volumes. Since the dense DBPN models performed better than the EDSR models in improving the through-plane resolution of LGE-MRI volumes, we used the super-resolved images obtained using dense DBPN models with and without gradient guidance to train the U-Net models and compared them. To serve as baseline comparison, we also train

the U-Net models using the images obtained using bicubic interpolation. We can observe that the Dice score of the LA segmentation computed by training super-resolved images obtained from the dense DBPN models with gradient guidance is significantly higher than the Dice score computed by training super-resolved images obtained from the stand-alone dense DBPN models and bicubic interpolation methods, for all the three downsampling scale factors, thus, corroborating our hypothesis that providing structural guidance using the gradient branch to improve the through-plane resolution of LGE-MRI volumes could improve the segmentation of cardiac structures.

One of the limitations of our work, as shown in Table 5.2 and 5.3, is that the PSNR and SSIM values for LGE-MRI volumes downsampled by a scale factor of 8 are poor due to the immense loss of information by downsampling, however, there is a considerable improvement in the Dice score for segmentation of LA from the super-resolved images. Another limitation is the use of simulated low through-plane resolution images due to the lack of real-world high-resolution and low-resolution image pair datasets. Therefore, we could not comprehensively assess our method on real-world low-resolution images. However, we employed two downsampling methods: Gaussian and Fourier downsampling methods, and demonstrated the effectiveness of our method to reliably super-resolve the simulated low-resolution images.

5.5 Conclusion

Here, we showed that the proposed self-supervised CNN-based super-resolution framework can be used to improve the through-plane resolution of LGE cardiac MRI images. Furthermore, we presented a 3D CNN architecture for image super-resolution and demonstrated that the proposed method can be used to improve the through-plane resolution of LGE-MRI volumes, which in turn, enables accurate segmentation of the cardiac chambers. We also demonstrated the generalization ability of our proposed approach. Our approach takes advantage of the information learnt from the gradient branch to provide structural guidance to the super-resolution branch, thereby, generating super-resolution 3D LGE-MRI images while preserving the cardiac structure information.

Bibliography

[1] Christopher McGann, Nazem Akoum, Amit Patel, Eugene Kholmovski, Patricia Revelo, Kavitha Damal, Brent Wilson, Josh Cates, Alexis Harrison, Ravi Ranjan, et al. Atrial fibrillation ablation outcome is predicted by left atrial remodeling on mri. *Circulation: Arrhythmia and Electrophysiology*, 7(1):23–30, 2014. 5.1

[2] Aurelien Bustin, Damien Voilliot, Anne Menini, Jacques Felblinger, Christian de Chillou, Darius Burschka, Laurent Bonnemains, and Freddy Odille. Isotropic reconstruction of mr images using 3d patch-based self-similarity learning. *IEEE transactions on medical imaging*, 37(8):1932–1942, 2018. 5.1

[3] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *Proceedings of the IEEE international conference on computer vision*, pages 1920–1927, 2013. 5.1

[4] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*, pages 111–126. Springer, 2014. 5.1

[5] Samuel Schulter, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3791–3799, 2015. 5.1

[6] Y Li, Bruno Sixou, and F Peyrin. A review of the deep learning methods for medical images super resolution problems. *Irbm*, 42(2):120–133, 2021. 1.3.4, 1.3.4.2, 5.1

[7] Irina Sánchez and Verónica Vilaplana. Brain MRI super-resolution using 3D generative adversarial networks. *arXiv preprint arXiv:1812.11440*, 2018. 5.1

[8] Yuhua Chen, Yibin Xie, Zhengwei Zhou, Feng Shi, Anthony G Christodoulou, and Debiao Li. Brain MRI super resolution using 3d deep densely connected neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 739–742. IEEE, 2018. 5.1

[9] Yuhua Chen, Feng Shi, Anthony G Christodoulou, Yibin Xie, Zhengwei Zhou, and Debiao Li. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 91–99. Springer, 2018. 5.1

[10] Can Zhao, Aaron Carass, Blake E Dewey, Jonghye Woo, Jiwon Oh, Peter A Calabresi, Daniel S Reich, Pascal Sati, Dzung L Pham, and Jerry L Prince. A deep learning based anti-aliasing self super-resolution algorithm for MRI. In *International conference on medical image computing and computer-assisted intervention*, pages 100–108. Springer, 2018. 5.1

[11] Anitha Priya Krishnan, Roshan Reddy Upendra, Aniket Pramanik, Zhuang Song, and Richard AD Carano. Multimodal super resolution with dual domain loss and gradient guidance. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 91–100. Springer, 2022. 5.1

[12] Jennifer A Steeden, Michael Quail, Alexander Gotschy, Kristian H Mortensen, Andreas Hauptmann, Simon Arridge, Rodney Jones, and Vivek Muthurangu. Rapid whole-heart CMR with single volume super-resolution. *Journal of Cardiovascular Magnetic Resonance*, 22(1):1–13, 2020. 5.1

[13] Evan M Masutani, Naeim Bahrami, and Albert Hsiao. Deep learning single-frame and multiframe super-resolution for cardiac MRI. *Radiology*, 295(3):552, 2020. 5.1

[14] Nicolas Basty and Vicente Grau. Super resolution of cardiac cine MRI sequences using deep learning. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 23–31. Springer, 2018. 5.1

[15] Jörg Sander, Bob D de Vos, and Ivana Išgum. Unsupervised super-resolution: creating high-resolution medical images from low-resolution anisotropic examples. In *Medical Imaging 2021: Image Processing*, volume 11596, pages 82–88. SPIE, 2021. 5.1

[16] Can Zhao, Muhan Shao, Aaron Carass, Hao Li, Blake E Dewey, Lotta M Ellingsen, Jonghye Woo, Michael A Guttman, Ari M Blitz, Maureen Stone, et al. Applications of a deep learning method for anti-aliasing and super-resolution in MRI. *Magnetic resonance imaging*, 64:132–141, 2019. 5.1

[17] Roshan Reddy Upendra and Cristian A Linte. A 3d convolutional neural network with gradient guidance for image super-resolution of late gadolinium enhanced cardiac mri. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1707–1710. IEEE, 2022. 5.1, 5.2.3.2, 5.3.2.3

[18] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7769–7778, 2020. 5.1

[19] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 1, 5.2.3.1

[20] Zhaohan Xiong, Qing Xia, Zhiqiang Hu, Ning Huang, Cheng Bian, Yefeng Zheng, Sulaiman Vesal, Nishant Ravikumar, Andreas Maier, Xin Yang, et al. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical Image Analysis*, 67:101832, 2021. 1, 3, 5.2.1, 5.2.3.4, 5.3.2.2, 5.3.2.3, 5.3.2.3

[21] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2, 5.2.3.4

[22] Lei Li, Veronika A Zimmer, Julia A Schnabel, and Xiahai Zhuang. AtrialGeneral: Domain generalization for left atrial segmentation of multi-center LGE MRIs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 557–566. Springer, 2021. 3, 5.2.1, 5.2.3.4, 5.3.2.2

[23] Lei Li, Veronika A Zimmer, Julia A Schnabel, and Xiahai Zhuang. AtrialJSQnet: A new framework for joint segmentation and quantification of left atrium and scars incorporating spatial and shape information. *Medical Image Analysis*, 76:102303, 2022. 3, 5.2.1, 5.2.3.4, 5.3.2.2

[24] Lei Li, Veronika A Zimmer, Julia A Schnabel, and Xiahai Zhuang. Medical image analysis on left atrial LGE MRI for atrial fibrillation studies: A review. *Medical Image Analysis*, page 102360, 2022. 3, 5.2.1, 5.2.3.4, 5.3.2.2

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1.6, 1.3.1.3, 2.1, 2.2.4.2, 3.2.2, 5.2.2, 5.2.3.4

[26] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 1.3.5.3, 5.4

Chapter 6

Discussion, Conclusion, and Future Research Directions

This chapter revisits the challenges to be tackled to generate patient-specific cardiac anatomical models from cardiac MRI data and our proposed methods to overcome these challenges. Also, this chapter provides some of the future directions along with the impact on the field.

6.1 Thesis Motivations, Summary and Contributions

In **Chapter 1**, we established the importance of image-based patient-specific anatomical models of the heart for clinical decision support. To this end, the overall aim of this dissertation was to develop deep learning-based tools and methodologies to generate patient-specific anatomical models of the heart using cardiac MRI data. Specifically, we focus our efforts on obtaining accurate and robust segmentation maps of the cardiac chambers from cine MRI images, accurate cardiac motion estimation from 4D cine MRI data, and integrating the scar anatomy from LGE MRI images with the myocardial anatomy from cine MRI images. In the process of developing these deep learning frameworks to build patient-specific anatomical models, we realized that the anisotropic nature of the cardiac MRI images with low through-plane resolution imposes major challenges in achieving accurate segmentation and registration results. Therefore, we focused our efforts on deep learning-based super-resolution algorithms towards

the end of this PhD dissertation.

In order to analyze ventricular structure and function from short-axis cine MRI data, it is essential to compute cardiac structural and functional indices, which requires delineation of cardiac chambers from the cine MRI images. Manual segmentation can be a arduous and time-consuming task subjected to intra- and inter-observer variability. Studies have showed that the intra- and inter-observer variability could lead to substantial difference between the clinical cardiac function indices computed from the manual segmentation and the true clinical values of these cardiac function indices. Therefore, a number of automated segmentation algorithms have been proposed and developed to aid cardiologists. In recent years, deep learning techniques have gained immense popularity in medical image segmentation, including cardiac MRI segmentation. A majority of the segmentation methods introduced in the past few years are based on the U-Net model. As presented in **Chapter 2**, in order to further improve the segmentation of the cardiac chambers from cine cardiac MRI, we integrated U-Net models in an adversarial framework called SegAN. We initially employed this method to segment the LV blood-pool on the 2017 ACDC dataset. Here, we used a multi-scale L_1 loss function to train the SegAN network, which enables the model to learn features at pixel-, low-, mid- and high-level. We evaluated the LV blood-pool segmentation results of three different U-Net models in the SegAN framework against the stand-alone U-Net models and obtained significantly better segmentation results than the stand-alone U-Net models. We then extended the SegAN framework for multi-class segmentation of LV blood-pool, LV myocardium and RV blood-pool. Additionally, we showed the effect of the proposed segmentation method on clinical cardiac parameters, such as ventricular stroke volumes, ejection fraction and myocardial mass. The adversarial nature of the architecture and the multi-resolution approach resulted in significantly improved segmentation of LV and RV blood-pool, which in turn, resulted in a more accurate computation of critical clinical parameters such as ventricular stroke volumes and ejection fraction.

In addition to global function parameters like stroke volumes, ejection fraction and myocardial mass, regional myocardial function parameters such as myocardial motion, strain,

torsion and wall thickness are also paramount to detect and predict the progression of the cardiovascular diseases. Therefore, accurate cardiac motion estimation from cine MRI images is important, as it helps assess the kinematic and contractile properties of the myocardium. In **Chapter 3**, we proposed and evaluated a VoxelMorph-based 4D deformable registration technique for consistent motion estimation from 4D cine MRI images. The presented workflow includes a slice misalignment correction step to alleviate the impact of the out-of-plane motion in the slice stack on the image registration. We also evaluated and compared the effect of the Laplacian-based operator versus the gradient-based operator for smoothing the registration field produced by the VoxelMorph-based registration network. Our experiments revealed that the Laplacian-based smoothing operator produces significantly better registration results as it considers global properties of the objective function more than the gradient-based smoothing operator. Additionally, we illustrated our intended application of the VoxelMorph-based 4D deformable registration technique, i.e., to build dynamic patient-specific LV myocardial models across subjects with different pathologies by propagating the LV myocardium isosurface and volume meshes (generated using the segmentation of end-diastole frame) at end-diastole cardiac frame to subsequent frames of the cardiac cycle. Lastly, we also demonstrated the potential of the VoxelMorph-based 4D deformable registration technique to build dynamic patient-specific RV models.

In order to assess myocardial viability, it is crucial to include scar anatomy in the personalized anatomical models of the heart. A number of unsupervised deep learning-based registration algorithms have been proposed to register cine MRI images to their corresponding LGE MRI images, however, their registration accuracy is limited by the cost function used, as the similarity measures used are the same as the traditional approaches. An alternative deep learning-based approach to quantify scar tissue in LGE MRI images is to generate synthetic LGE images with improved soft tissue contrast from cine MRI images using adversarial networks, and using the generated synthetic images to segment the scar tissues. However, this is a time-intensive approach as it includes training adversarial networks to generate synthetic LGE MRI images followed by training a segmentation algorithm on the synthetic LGE

MRI images. In order to address these limitations, we proposed and evaluated a RoI-guided CNN-based registration approach to register cine MRI images to its corresponding LGE MRI images, as presented in **Chapter 4**. Our proposed method fully exploits the RoIs of the cardiac anatomy (i.e., LV blood-pool, LV myocardium and RV blood-pool) and therefore, outperforms the unsupervised CNN-based algorithms. It requires minimal preprocessing and does not necessitate the need to train adversarial networks to generate synthetic LGE MRI images. Our experiments showed that the proposed RoI-guided CNN-based registration approach can be reliably used to register the cine MRI and LGE MRI images. The major limitation of the proposed method is the need for annotations of the RoI for large number of training data. Hence, we proposed and evaluated a joint deep learning framework that involves coupling of segmentation and registration tasks by sharing weights. The proposed joint deep learning model produces reliable registration results using lesser number of training data and manual annotations by leveraging the weights learnt from the segmentation task to improve the registration accuracy.

Lastly, in **Chapter 5**, to address the limitation of anisotropic 3D cardiac MRI images (in-plane resolution of 1 to 1.5 mm and through-plane resolution of 5 to 10 mm), which imposes challenges in the segmentation and registration models, and therefore in cardiac image visualization, analysis and diagnosis, we first proposed and evaluated a self-supervised 2D deep learning algorithms to compute super-resolution LGE cardiac MRI images. This proposed method leverage the information it learns between mapping the simulated low-resolution in-plane data and its corresponding high-resolution in-plane data to enhance the low through-plane resolution. However, this proposed self supervised patch-based method does not take advantage of the 3D information provided in the cardiac MRI images and ignores the global context information. Also, since the method requires the use of 2D patches in the inference stage, it can lead to inconsistencies during the fusion process. Therefore, we presented a 3D CNN-based super-resolution framework with gradient guidance to compute super-resolution cardiac LGE MRI images. The proposed 3D CNN framework produces reliable super-resolution results by taking into advantage of the 3D information in the LGE MRI images and also, the gradient

guidance ensures the 3D CNN model "pays more attention" to the 3D structure of the tissues in the LGE MRI images.

6.2 Future Work

Although significant research effort has been put forth in this dissertation to generate patient-specific anatomical models of the heart using cardiac MRI images, some key challenges still remain. One such challenge is the application of the developed deep learning-based pipeline on paediatric cardiac MRI. A comprehensive review and future research directions regrading the role of deep learning in paediatric cardiac MRI, and how the deep-learning models, such as the ones developed in this thesis work can be used for automated bi-ventricular segmentation, automated diagnosis, patient-specific models and precision medicine is explained in [1].

Furthermore, this section will briefly mention some of the future research directions that could lead to a more accurate personalized anatomical models of the heart.

Cardiac MRI Segmentation

Most of the current deep learning-based cardiac segmentation models are supervised learning, thereby, requiring ground-truth segmentation masks. One of the major challenges in cardiac MRI segmentation tasks is the limited availability of clinical image datasets accompanied by expert annotations. To this end, pre-training and data augmentation are two crucial data-driven methods that better uses the existing data [2]. Pre-training includes training the model on a larger dataset before training it on your dataset, to broaden the model's horizon and enhance its robustness. While we used spatial and intensity-based data augmentation in all our deep learning models, it will be interesting to explore adversarial augmentation techniques, in order to expose the model to higher variability and thereby, increase the model's robustness.

In this dissertation, we described the various U-Net models used to segment cardiac chambers from cardiac MRI images. Hence, in Chapter2, we proposed an integration of U-Net models in an adversarial framework to further improve the segmentation of cardiac chambers from cine cardiac MRI. With the advent of self-attention-based architectures in medical seg-

mentation, we propose a hybrid U-Net and self-attention-based framework for cardiac MRI segmentation [3].

Cardiac Motion Estimation

In the past decade, deep learning models have gained increased popularity in medical image registration [4]. In this dissertation, we presented a deep learning-based 4D deformable registration method for cardiac motion estimation from cine cardiac MRI dataset by leveraging the VoxelMorph framework [5]. Additionally, we demonstrated the application of the VoxelMorph-based cardiac motion estimation method to build dynamic patient-specific left ventricle (LV) myocardial models across subjects with different pathologies [6]. Although the CNN-based cardiac motion estimation presented in this work showed promising performance, the CNN-based approaches usually exhibit limitations in modeling explicit long-range spatial relations due to the limited receptive fields of convolution operations [7]. Therefore, the large variations in shape and size of the cardiac chambers can affect the registration performance of the CNN-based cardiac motion estimation methods.

In recent years, self-attention-based architectures (Transformer-based), due to their great success in sequence-to-sequence prediction in natural language processing have gained increasing interests in computer vision tasks [8], including medical image segmentation [7] and registration [9]. These current research studies show that fusing the self-attention mechanism with the CNN models overcome the limitation of the convolution operation in learning global semantic information, which is critical for the image registration task in cardiac motion estimation from the cine cardiac MRI images. Therefore, we propose a hybrid CNN-Vision Transformer architecture for consistent cardiac motion estimation from 4D cine cardiac MRI images.

Multimodal Registration: Cine MRI and LGE MRI

One of the core challenges in multimodal registration is the use of appropriate similarity metrics. Mutual information, cross-correlation ratios, and their variants have be successively

used, including our work in this dissertation. However, mutual information is dependent only on the intensity distributions of different images and cross-correlation ratio uses the functional relationship among modalities. In the course of our research on various similarity metrics for multimodal registration, we hypothesize that a combination of weighted self-similarity structure vector and texture weight maps that describes the local structure information can be used for multimodal registration [10].

Cardiac MRI Super-Resolution

The major limitation of cardiac MRI super-resolution is the need for high-resolution (HR) images while training. In order to overcome this limitation, it is crucial to develop an unsupervised deep learning framework to generate super-resolution images from only low-resolution (LR) images. To this end, we propose to extend the unsupervised deep learning framework described in [11] as a means to improve the resolution of both cine and LGE cardiac MRI images.

Bibliography

[1] Andrew M Taylor. The role of artificial intelligence in paediatric cardiovascular magnetic resonance imaging. *Pediatric Radiology*, 52(11):2131–2138, 2022. 6.2

[2] Shizhan Gong, Weitao Lu, Jize Xie, Xiaofan Zhang, Shaoting Zhang, and Qi Dou. Robust cardiac mri segmentation with data-centric models to improve performance via intensive pre-training and augmentation. In *Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers: 13th International Workshop, STACOM 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Revised Selected Papers*, pages 494–504. Springer, 2023. 6.2

[3] Yi Wang, Yuan-Zhe Li, Yin-Hui Huang, Ping Xiang, Wen-Xi Liu, Qing-Quan Lai, Yiyuan Gao, Maosheng Xu, and Yi-Fan Guo. Rsu-net: U-net based on residual and self-attention mechanism in the segmentation of cardiac magnetic resonance images. *Computer Methods and Programs in Biomedicine*, page 107437, 2023. 6.2

[4] Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1):1–18, 2020. 1.3.2.3, 6.2

[5] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019. 3.2.3.1, 6.2

[6] Roshan Reddy Upendra, Brian Jamison Wentz, Richard Simon, Suzanne M Shontz, and Cristian A Linte. CNN-based cardiac motion extraction to generate deformable geometric

left ventricle myocardial models from cine MRI. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 253–263. Springer, 2021. 6.2

[7] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. TransUnet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 6.2

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 6.2

[9] Junyu Chen, Yufan He, Eric C Frey, Ye Li, and Yong Du. ViT-V-Net: Vision transformer for unsupervised volumetric medical image registration. *arXiv preprint arXiv:2104.06468*, 2021. 6.2

[10] Yifan Wang, Tianyu Fu, Chan Wu, Jian Xiao, Jingfan Fan, Hong Song, Ping Liang, and Jian Yang. Multimodal registration of ultrasound and mr images using weighted self-similarity structure vector. *Computers in Biology and Medicine*, page 106661, 2023. 6.2

[11] Yikang Liu, Eric Z Chen, Xiao Chen, Terrence Chen, and Shanhui Sun. An unsupervised framework for joint mri super resolution and gibbs artifact removal. *arXiv preprint arXiv:2302.02849*, 2023. 6.2