Title
Automated Model Construction for Image-Based Cardiac Computational Simulations

Author
Kong, Fanwei

Publication Date
2022

Peer reviewed|Thesis/dissertation

Automated Model Construction for Image-Based Cardiac Computational Simulations

by

Fanwei Kong

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Shawn Shadden, Chair
Professor Somayeh Sojoudi
Professor John Canny

Spring 2022

Automated Model Construction for Image-Based Cardiac Computational Simulations

Abstract

Automated Model Construction for Image-Based Cardiac Computational Simulations

by

Fanwei Kong

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Shawn Shadden, Chair

This dissertation seeks to develop novel data-driven algorithms to automatically construct simulation-suitable meshes from volumetric medical image data to enable high-throughput, large-cohort analyses of patient cardiac function from medical image data. Image-based cardiac modeling derives geometries from patient medical image data, and can simulate a variety of aspects of cardiac function, including electrophysiology, hemodynamics, and tissue mechanics to explore improvements in cardiovascular diagnoses and treatments, and the biomechanical underpinnings of diseases. However, creating accurate models of the heart from patient image data requires significant time and human efforts and is the primary challenge of image-based modeling.

Deforming-domain computational fluid dynamics (CFD) simulations of the intracardiac hemodynamics, in particular, require both the geometry and the deformation of the heart from a sequence of image snapshots throughout the cardiac cycle. In the first part of the thesis, we present a two-stage approach to automatically generate CFD-ready left ventricle (LV) models from patient image data. This approach first uses deep-learning-based automatic segmentation and then uses geometry processing algorithms to robustly create CFD-suitable LV models from image data.

In the remainder of this dissertation, we developed end-to-end deep-learning algorithms to directly construct the surface meshes of the whole heart from volumetric medical image data without the need for a multistage approach. These algorithms leverage shape templates, shape deformation methods, and regularization functions during optimization to generate meshes that are suitable for computational simulations. Namely, we trained a graph convolutional neural network to deform a pre-defined mesh template to fit the whole heart in a target 3D image volume. Various mesh deformation methods were combined with deep learning to create whole heart surfaces, including direct displacements on meshes, free-form deformation by control lattices as well as control-handle-based deformation using bi-harmonic coordinates. We demonstrated the application of our method in constructing a dynamic

whole heart mesh from time-series CT image data to simulate the cardiac flow driven by the cardiac motion. We also demonstrated the capability to switch template meshes to accommodate different modeling requirements.

The algorithms developed in this thesis were able to construct a 4D dynamic simulation-suitable mesh of the heart within a minute on a standard desktop computer whereas prior methods require hours of human efforts and multiple programs. The code was implemented in Python, fully open-sourced, and can be conveniently executed from the command line.

To my parents, 齐帅 and 孔冬青,
for your unconditional love and support.
To my grandparents, 曲绵英 and 孔宪凯,
for instilling my life with care and warmth.
To my grandparents, 邓秀娟 and 齐兆谦,
for being in my sweetest memories and for being my inspiration.

Contents

List of Figures

List of Tables

Acknowledgments

During my Ph.D. studies at Berkeley, I have been fortunate to learn and grow with the help of numerous talented researchers and professionals. I am grateful to these mentors, colleagues, and friends who helped me to become who I am today.

First and foremost, I would like to thank my research advisor Shawn Shadden for all of his support, knowledge, patience, and kindness throughout my Ph.D. Shawn is an amazing teacher and mentor who always gives very insightful advice. He is always supportive during the exploration of my research interests and directions and always gives constructive feedback. Not only has he supported the development of this thesis, but he has also guided and inspired me to improve myself to become a better researcher. His positive influence will continue to benefit me in my future career.

I would also like to thank other mentors during my Ph.D. for their help and support. I first want to thank Nathan Wilson for his help and support while I was working on the software development of SimVascular. He was always patient while coaching me to become an effective programmer and provided critical feedback that pushed me to improve my research design and programming skills. I also want to thank Alison Marsden for being super supportive and for always providing thoughtful and helpful feedback for my research. I am also grateful to my peer mentors, Adam Updegrove and Jiacheng Wu. Adam was always there to help me out during my first year while we worked together on SimVascular and he spent so much time patiently guiding me to develop knowledge and skills that were important to the rest of my Ph.D. Jiacheng helped me develop my research interests in machine learning and provided valuable feedback during the early stage of my Ph.D. research. Besides mentors in my Ph.D. studies, I am grateful to my undergraduate research advisors Wei Sun and Ross Ethier. Their guidance during my undergraduate years has prepared me well for my Ph.D. studies and they continue to inspire me along my path to becoming a good researcher.

I would like to thank the professors on my Ph.D. dissertation and qualifying exam committees, Somayeh Sojoudi, John Canny, Lisa Pruitt, Grace O'Connell, and Omer Savas. Your helpful feedback and comments have been a valuable source of improvement to this thesis.

This dissertation would not have been possible without our collaborators. Thank you, Shone Almeida, Amirhossein Arzani, and Kashif Shaikh for providing the time-series CT image data, and Dr. Theodore Abraham for providing the time-series MR image data. I also sincerely thank Vijay Vedula and Angjoo Kanazawa for their helpful comments concerning my research and for pointing me to resources related to my research. I also want to thank other members from the SimVascular and SimCardio development teams, Gabriel Maher, Martin Pfaller, Dave Parker, and Oguz Tikenogullari for all their help and feedback regarding my research and for the good times we spent together preparing for workshop demos.

I would not have survived my Ph.D. without the support from all my friends and colleagues, the laughs, and the good moments we spent together. I especially want to thank Shadden lab members, Siavash Ameli, Kirk Hansen, Adam Updegrove, Jiacheng Wu, Jianxun Wang, Jeffrey Pyne, Sarah Frank, Debanjan Mukherjee, Miguel Rodriguez, Mehran Mir-

ramezani, Chi Zhu, Numi Sveinsson, Arjun Narayanan, and Theophile Sautory. Thank you everyone for being so nice, kind, helpful, and supportive.

Finally and most importantly, I would like to express my greatest gratitude to my family for their unconditional love and support. I also thank Wenke Zhang and closest friends Thu Tran and Yufei Wang for always being on my side, understanding me, and supporting me during the ups and downs.

Chapter 1

Introduction

1.1 Motivations

The heart is a complex physical system involving the coupling of electrophysiology, solid mechanics, and fluid mechanics. During a heartbeat, the sinoatrial node, which is a small bundle of specialized cells in the right atrium, initiates a voltage signal or action potential that is produced periodically. Due to this reason, the sinoatrial node is also called the natural "pacemaker" of the heart. Next, the action potential travels to the atrioventricular node through the Bundle of His, causing the atrium to contract and forcing part of the blood in the atrium into the ventricles. The electrical signals further travel down the branches of the Purkinje Fibers to the myocardium, causing the ventricles to contract, pumping blood out of the heart.

Heart disease is the leading cause of death worldwide and in the United States alone, 659,000 people die from heart disease each year, which is 1 in every 4 deaths [158]. Understanding the malfunctioning of the heart is difficult due to the complex interaction of biomechanics, electrophysiology, and hemodynamics within the heart. Treatments of cardiac diseases are further complicated by the interaction between the heart and implantable medical devices, the variations of cardiac physiology among patients, and the currently limited capability of extracting information about the (patho)physiological process for individual patients [100]. Current treatment strategies for cardiac diseases have been primarily based on randomized clinical trials which study therapies that are beneficial at the population level. However, there has been a growing realization that patient groups are less uniform and can show variability in response to the same medical therapies [23, 76].

Consequently, one area of research that currently receives considerable attention is computational modeling of the heart from patient-specific clinical data to provide tailored understanding and treatment that suits an individual's pathology. This approach derives personalized geometries of the heart from image data, integrates the patient's diagnostic clinical measurements, and numerically solves mathematical equations that describe various physiology on discretized computational domains. A few review articles have thoroughly summa-

rized computational modeling of electrophysiology [152, 151, 149, 102], hemodynamics [9, 94, 177, 147], tissue mechanics [160, 103] and multi-physics coupling [123, 100, 104]. Such "digital twin" modeling of a patient's heart can provide information that cannot be readily measured to facilitate diagnosis and treatment planning for individual patients [2, 171, 56, 122, 135, 146, 71]. It can also be used further quantify biomechanical underpinnings of cardiac diseases [37, 119, 92, 145]. This paradigm has motivated numerous research efforts on a wide range of clinical applications, such as simulations of the stress and strain of cardiac tissues when interacting with implantable cardiac devices[69], the cardiac flow pattern after surgical corrections [56, 71], and cardiac rhythm outcome after ablation surgery [171, 122].

Generating simulation-suitable models of the heart from image data has remained a time-consuming and labor-intensive process. While earlier computational cardiac modeling often used generic idealized shapes of the heart, personalized shape models provide necessary geometric details for personalized treatment planning and optimization. Segmentation of cardiac tissues is often obtained manually or semi-automatically from medical imaging techniques (computerized tomography, magnetic resonance, X-ray, and echocardiogram), followed by manual construction of computational meshes and complicated steps to label various surfaces or regions for the assignment of boundary conditions or parameters for computational simulations. Much of the challenge is related to the entwined nature of the heart, difficulty differentiating individual cardiac structures from each other and the surrounding tissue, and the large deformations of these structures over the cardiac cycle. Consequently, model construction is the major bottleneck limiting large-cohort validations and clinical translations of functional computational heart modeling[94, 144, 75, 107]. Indeed, prior studies have been limited to only a few subjects[94, 56, 97, 20, 14, 98, 61, 131, 121, 5].

Deep learning methods can train deep artificial neural networks from existing data to automatically process information. As their names suggest, deep neural networks are inspired by the connections of biological neurons in the human brain. Namely, they are often designed to include many layers of artificially connected "neurons" or computational nodes that have associated weights and operations to produce the corresponding outputs given the input information that they received from the previous layers. The computational nodes will send information to the next layer if their outputs activate some non-linear activation functions. The weights within a neural network will be optimized using a set of training data so that the outputs from the neural network will approximately minimize an objective function. During the past decade, a class of artificial neural networks, convolutional neural networks (CNNs) have demonstrated great success in visual pattern recognition, powered by the recent availability of big data. They were inspired by the organization of the animal visual cortex to use layers of localized convolutional kernels with trainable weights that are shared across the entire image to hierarchically process image features from low levels to high levels [35, 73, 172].The effectiveness of deep learning using CNNs was first demonstrated in image classification [70, 124] but then rapidly expanded to a wide range of computer vision applications, including segmentation [138, 17], object detection [125, 127], image restoration[89, 175], super-resolution[74, 28], and synthesis[40, 59, 178].

Deep learning methods using CNNs have since been widely adopted in many medical

image analysis applications [82]. Among the most popular methods are nnU-Net for segmentation [51] and VoxelMorph for registration [8] of anatomical and/or pathological structures from patient medical scans. Within the field of whole heart segmentation that this dissertation is related to, several deep-learning-based approaches using CNNs have achieved higher segmentation precision of the heart than prior model- or atlas-based segmentation approaches and are much faster in practice [6, 182, 10, 99]. The prior model- or atlas-based segmentation approaches work by fitting surface meshes of the heart to target images using local optimization to match with tissue boundaries on input images [30, 31, 117]. Therefore, they are often sensitive to initialization and require complicated steps and manual efforts to construct a mean template of the heart.

Most deep learning methods for medical image segmentation have, however, focused on segmentation (pixel classification) rather than the construction of a computer model of the heart, usually represented by tessellated meshes. They are often optimized for pixel-wise accuracy and thus lack awareness of the overall shape and anatomy of the target organs. A few prior studies have combined deep learning and CNNs with atlas propagation or geometry-aware shape refinements to incorporate anatomical shape priors in deep-learning-based segmentations [29, 47, 162]. Others have leveraged carefully designed objective functions to encourage more anatomically plausible segmentations [79, 87, 93]. However, how to use deep neural networks to directly produce heart geometries that are suitable for various types of functional computational modeling of the heart remains to be addressed.

Geometric deep learning has extended the concepts of the convolutional neural network on irregular graphs to work with 3D shapes [25, 11, 44]. There are two major types of graph convolutional neural networks (GCNs) and different GCNs along with their applications have been summarized in [174]. The first type defines graph convolutions in the spectral domain based on the graph Fourier transform [25], and the second type defines convolution in the spatial domain based on aggregations of node features within the node neighborhoods [44]. Computational heart models require surface or volumetric meshes, which can be represented as irregular graphs. Following the success of GCNs, a few learning-based approaches have been proposed to automatically reconstruct 3D objects in the form of surface meshes from 2D camera images [159, 54, 36, 111, 81]. However, these methods cannot be directly applied to generate meshes from medical images, which are inherently different from camera images. The major challenge is the high dimensionality of medical image data, which is often available as 3D volumes where various internal tissues or organs appear with different signal intensity values. Each 3D medical image volume often contains a few million voxels, which poses difficulties in efficiently generating high-resolution 3D meshes of complex cardiac structures within the limits of computational resources.

1.2 Contributions

The goal of this dissertation is to develop novel data-driven algorithms to automatically construct simulation-suitable meshes from volumetric medical image data to enable high-

throughput, large-cohort analyses of patient-specific cardiac function. Among various functional simulations, the dissertation targets the deforming-domain CFD simulations of the intracardiac hemodynamics, which is particularly labor-intensive since it requires reconstructing temporally-consistent deformations of the heart from a sequence of images snapshots [94, 97, 20, 14, 98, 61, 131].

Chapter 2 first presents a multi-stage framework to automatically generate CFD-ready left ventricle (LV) models from patient image data. The framework first uses a deep convolutional neural network to automatically and accurately segment cardiac structures and then uses geometry processing algorithms to robustly create CFD-suitable LV models from segmentations, as well as registration algorithms to obtain the motion of the heart over time.

The intermediate segmentation step often produces extraneous regions containing topological anomalies that are unphysical and unintelligible for simulation-based analyses. A major contribution of this dissertation is thus to address the *end-to-end* learning between volumetric medical images and simulation-suitable surface meshes of the whole heart, namely, heart chambers and great vessels. We propose to *directly* reconstruct the *surface meshes* of the whole heart from volumetric medical image data without the need for a multistage approach that involves an intermediate segmentation step. Despite the entwined nature of the heart, we observe that all of the individual cardiac structures are homeomorphic to spheres. We thus propose to reconstruct whole heart surface meshes by learning to deform meshes of spheres and thereby naturally produce topologically correct cardiac structures. Chapter 3 presents our approach that directly reconstructs the surface meshes of multiple cardiac structures in a 3D image volume by using graph convolutional neural networks to deform pre-defined mesh templates of spheres.

Meshes for computational simulations need to be smooth (i.e. polygons are smoothly connected except along the feature edges) and free of topological artifacts. These requirements can only be weakly enforced with mesh regularizers (Laplacian, edge length, normal consistency, etc) in the objective functions, as done in chapter 3. Furthermore, image-based computational simulations have various requirements for what cardiac structures to include in the computational domains, mesh resolutions, and ways to define boundary geometries. Chapters 4 and 5 present our approaches to directly construct meshes that are suitable for cardiac computational simulations from patient image data. Namely, our approaches combine space deformation techniques with neural networks to ensure the smoothness of the predicted meshes intrinsically and learn to deform the space enclosing a whole heart template to construct meshes that match with patient image data and satisfy various modeling requirements. In contrast to learning per-vertex mesh displacements directly, as described in chapter 3, the neural networks learn the translations of a small set of control points that parameterize the deformation of the space and thereby smoothly deform the template mesh. The space parameterization method can be either free-form deformation by control lattices described in chapter 4 or, in chapter 5, control-handle-based deformation using biharmonic coordinates to support more detailed local deformation with fewer control points.

The approaches presented in this dissertation are the first to demonstrate capability in the direct generation of simulation-suitable meshes of the heart from medical images. We

demonstrated the application of our methods in constructing dynamic meshes of the heart from time-series CT image data to simulate the cardiac flow driven by the cardiac motion. We also demonstrated the capability to readily switch template meshes to accommodate different modeling requirements after training.

1.3 Prior Publication and Software Availability

Much of the content presented in this dissertation has been previously published or is currently under review. These publications are the results of joint efforts with my advisors Shawn Shadden and Nathan Wilson. My contributions include conceptualization, methodology development, conducting experiments, and preparing the drafts of the papers. All co-authors have provided written consent for the inclusion of their work in this dissertation. A list of papers used to develop Chapters 2, 3, 4, and 5 are listed below.

Chapter 2 Fanwei Kong and Shawn C. Shadden. "Automating Model Generation for Image-Based Cardiac Flow Simulation". In: *Journal of Biomechanical Engineering* 142.11 (Sept. 2020). 111011. ISSN: 0148-0731. DOI: 10.1115/1.4048032

Chapter 3 Fanwei Kong, Nathan Wilson, and Shawn Shadden. "A Deep-Learning Approach For Direct Whole-Heart Mesh Reconstruction". In: *Medical Image Analysis* (2021), p. 102222. ISSN: 1361-8415. DOI: 10.1016/j.media.2021.102222

Chapter 4 Fanwei Kong and Shawn C. Shadden. "Automatic Whole Heart Meshes Generation For Image-Based Computational Simulations By Learning Free-From Deformations". In: *International Conference on Medical Image Computing and Computer Assisted Intervention* (2021)

Chapter 5 Fanwei Kong and Shawn C. Shadden. "Learning Whole Heart Mesh Generation From Patient Images For Computational Simulations". In: *ArXiv* abs/2203.10517, Under review in IEEE Transactions on Medical Imaging (2022)

The code for all the chapters is open-source and available on GitHub [1]. The code for Chapter 2 has been integrated into SimVascular [155], which is an open-source software suite for cardiovascular simulation. The code for the rest chapters will also be integrated into SimVascular in the future.

[1] https://github.com/fkong7

Chapter 2

Automating Model Generation For Image-Based Left Ventricle Flow Simulation

2.1 Introduction

Image-based modeling of blood flow is an important research area in biomedical engineering. It is based on applying computational fluid dynamics (CFD) to image-based computer models of the heart, arteries, or veins to compute patient-specific blood flow information that is not measurable *in vivo*. This paradigm has also recently gained broad clinical use for coronary artery disease diagnosis [19].

The vast majority of applications of image-based hemodynamics modeling have been in vascular domains. Cardiac applications, while existing, are far less common. This is despite the fact that intracardiac hemodynamics are known to be important in the initiation and progression of heart diseases, e.g., [115, 34, 90, 38, 114, 32, 33]. There are two main approaches to modeling intracardiac hemodynamics. The first approach tracks the deformation of the heart from time-resolved imaging and imposes this motion to the fluidic domains inside the heart, which leads to a deforming-domain CFD problem[20, 14, 98, 61, 131, 9]. The second approach couples electrophysiology, structural mechanics and fluid dynamics in the heart so that the heart motion is solved for rather than measured [130, 56, 4]. This second approach is formidable and is generally unnecessary if the purpose of the model is to derive intracardiac hemodynamics. Therefore, a deforming-domain CFD approach is considered the *de facto* method to derive patient-specific modeling of LV hemodynamics; it is, however, not without its challenges. Namely, in comparison to most vascular applications, cardiac structures undergo large deformations, and individual cardiac structures can be difficult to differentiate from each other and the surrounding tissue. This makes generating CFD-ready cardiac models from medical image data a substantial challenge and this is regarded as the largest bottleneck in image-based CFD modeling of cardiac hemodynamics [94].

Figure 2.1 shows a typical model construction workflow of CFD simulations of LV flow. It starts with the delineation of endocardial surfaces by manual or semi-automated image-segmentation. Most studies target left ventricle (LV) hemodynamics, and thus segment the LV, and often portions of the left atrium (LA) and aorta (Ao) inflow and outflow tracks. This is done for a sequence of image snapshots of the heart throughout the cardiac cycle, resulting in a sequence of segmentations over time. The segmented regions from a chosen time instant are used to generate a reference volumetric mesh of the fluidic domain using appropriate mesh generation software. Then a registration process is performed to deform this reference volumetric mesh so that its boundary is consistent with the image segmentation sequence. These steps are generally performed using separate software tools, which further complicates the workflow and data management. And the manual nature of the process is prone to operator-dependent errors that are unpredictable, complicating reliability and reproducibility.

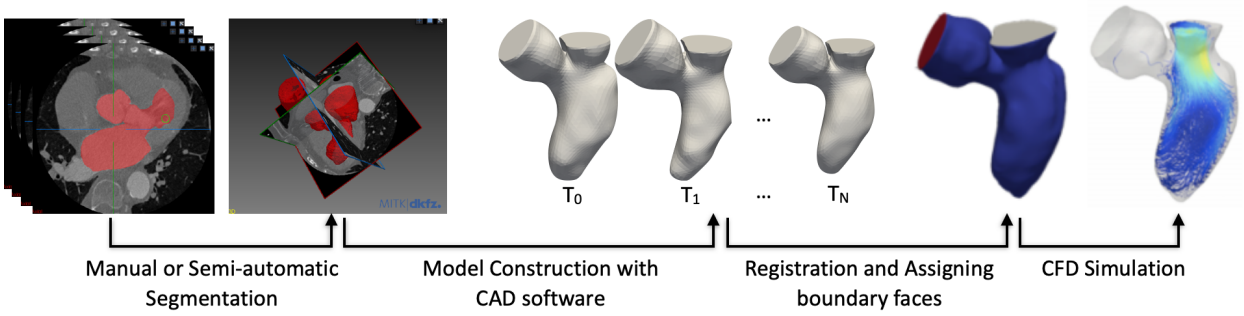


Figure 2.1: A typical model construction workflow of CFD simulations of LV flow.

Some recent works have sought to accelerate part of the model construction process. Schenkel, et al. [132] accelerated LV segmentation by fitting LV contours that depended on manual seed placements and manual segmentation of the valve ring. Nguyen, et al. [101] presented a semi-automatic, minimal operator involvement approach for LV meshing, smoothing and reconstruction but used simplified LV geometries generated from a closed-source software. Khalafvand, et al. [60] developed a semi-automatic pipeline using automated multi-atlas segmentation and statistic shape modeling of the LV, but only studied the effect of shape changes on LV flow and did not apply the method to patient-specific image data. Vellguth, et al.[157] developed an efficient pipeline using semi-automatic segmentation and geometric modeling packages but applied the pipelines to only one set of patient data. These recent approaches, while accelerating some part of the model construction process, still require various operator-dependent steps, employ closed-source software packages, or have been tested with very few examples. A need remains to develop an automated method to reliably generate patient-specific LV CFD models directly from image data.

Following recent developments in deep learning, automated segmentation of cardiac structures using convolutional neural networks (CNNs) has gained momentum. However, since most deep learning approaches focus on voxel-based classification, e.g., without consideration

of the overall topology of segmented structures, the utility of automatic CNN-based segmentation for computational modeling remains unclear. Building on prior work in the area of learning-based image classification, we present here an automated framework to generate CFD-ready models from cardiac CT or MR scans. Our framework proposes a CNN-based method to perform automatic segmentation of cardiac tissues from clinical imaging. We demonstrate that our method outperforms several recently-published grand-challenge segmentation algorithms [182]. We further develop automated surface processing and image registration to generate deforming volumetric computational models suitable for deforming mesh CFD simulations. The proposed framework can be executed from the command line (i.e., requires no visual interventions from the user) as an automated process and has only open-source software dependencies. We validate our model construction using both CT and MR benchmark image data sets, and we demonstrate the viability of using the models to perform CFD simulations of LV hemodynamics.

2.2 Methods

The proposed automated framework consists of three major steps to generate CFD-compatible models for LV flow simulations: segmentation, mesh generation and registration. For reference, the proposed process is over-viewed in Figure 2.2.

Patient Image Data

The MMWHS dataset was recently established as part of a grand challenge to evaluate different algorithms of whole heart segmentation[182]. The CT images were obtained from routine cardiac CT angiography and the MR images were acquired by using 3D b-SSFP sequences. The mean axial in-plane resolution for CT was 0.78 x 0.78 mm and the average slice thickness was 1.6 mm. The image resolution for the MR data was re-sampled to around 2 mm along each direction. The imaging window generally spanned from the upper abdomen to the aortic arch. The patients imaged by MR had a variety of cardiovascular diseases, including myocardium infarction, atrial fibrillation, tricuspid regurgitation, aortic valve stenosis, Alagille syndrome, Williams syndrome, dilated cardiomyopathy, aortic coarctation, and Tetralogy of Fallot [182]. The dataset includes 60 CT and 60 MR scans. 20 CT and 20 MR scans were released as training data, which contained manual segmentation of seven cardiac structures: LV, LA, RA, RV, myocardium, aorta and pulmonary artery. The remaining 40 CT and 40 MR scans were considered test data, with no provided manual segmentation as ground truth. Some datasets included segmentation of the coronary arteries and/or LA appendage, but these were not considered for the analyses herein.

Image-based CFD simulation of LV hemodynamics requires cardiac motion over one or more cardiac cycles. The MMWHS dataset only contains CT or MR scans of a single time frame. This is sufficient for testing the accuracy of our automated segmentation process, which is expected to be the most critical step in developing accurate models. However, to

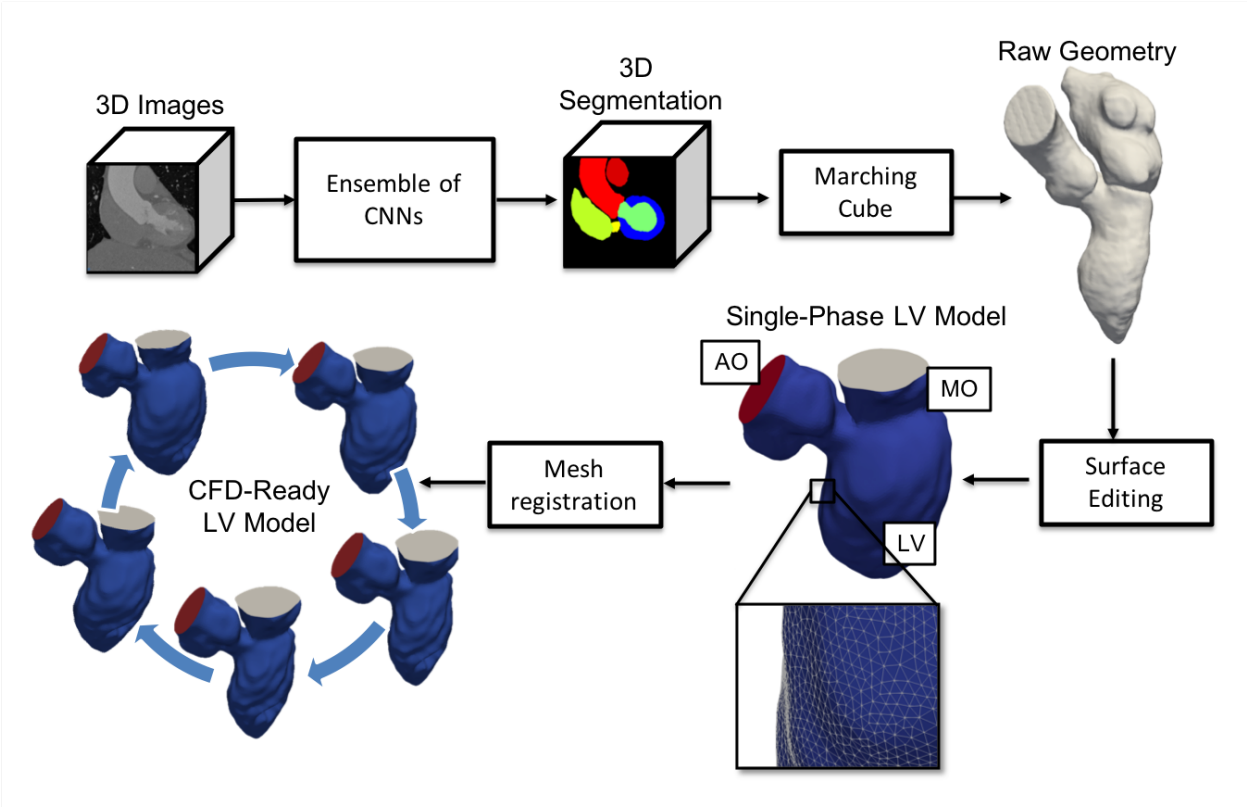


Figure 2.2: Diagram of the automated model generation framework for LV CFD simulations

test the registration process (and ultimately run CFD simulations) time-resolved image data is required. For this purpose, time-resolved CT data sets were used to verify the complete framework. The time-resolved CT data came from a 74-year-old male patient and a 73-year-old male patient, respectively. Both patients had left ventricular diastolic dysfunction. The data was de-identified and previously collected for other purposes. The mean axial in-plane resolution was 0.44 x 0.44 mm and the slice thickness was 1.25 mm. The temporal resolution was around 100 ms and 10 time frames were constructed for one cardiac cycle.

Intensity normalization and re-sampling were applied to all 3D image volumes to obtain consistent image dimensions, pixel spacing and pixel intensity range. The pre-processed image volumes can then serve as inputs to an automatic segmentation framework to generate segmentations. We first normalized pixel intensity values of each image volume such that they ranged from -1 to 1. Namely, CT intensity values, nominally ranging from around -1000 to 3000, were clipped to intensity values from -750 to 750. The cardiac tissues are well within this range while the intensity variations from bones or background noise could be mostly removed. The intensity values were then divided by 750 such that they ranged from -1 to 1. For MR, the intensity values depend not only on tissue properties but also on the MR signal intensity in individual patient scans. Therefore, for each patient's MR scan,

the pixels intensity values were clipped between 0 and the 99th percentile to reduce bright artifacts. The pixel values were then normalized by the maximum intensity after clipping and then shifted such that they ranged from -1 to 1. The 3D image volumes were resampled to have isotropic spacing and resized to 256×256×256, which maintained image resolution with a manageable computational cost.

Automated Segmentation

Automatic Segmentation Using an Ensemble of CNNs

Our framework employed an ensemble of CNN models as described here. Broadly, using the images as inputs, a CNN-based model outputs the probability of each image pixel belonging to a particular anatomical domain (LV, LA, RA, RV, myocardium, aorta and pulmonary artery). Limited by their high memory consumption and computational cost, CNN-based 3D segmentation algorithms usually require down-sampling the input data or adopting a sliding-window strategy to avoid running out of memory. Such compromises may lead to either low spatial resolution of the segmentation results or high time complexity, respectively. Since 2D CNN-based algorithms can be directly end-to-end trained, it is possible to slice 3D image data into a number of 2D slices and then use a 2D-based algorithm on each slice. However, 2D CNN-based algorithms ignore the spatial connection between adjacent slices and thus are not able to fully explore inter-slice information as compared to 3D CNNs. Therefore, to overcome the memory constraint of performing a 3D CNN, and information loss of performing 2D CNNs, we utilized an ensemble of 2D CNNs to generate a 3D segmentation (Figure 2.3a). Since deep neural network models generally have high prediction variance, ensemble learning with deep neural networks can reduce the variance and thus better generalize to unseen data [53, 176]. We sliced the 3D image volumes along the axial, sagittal or coronal axis to obtain corresponding 2D image datasets. A CNN model was trained for each 2D dataset to predict the probability of each pixel belonging to each cardiac structure. To automatically segment a new 3D image volume, we sliced the image volume into 2D images along the axial, coronal and sagittal axes, respectively, and utilized the corresponding trained 2D CNN model to predict the 2D probability maps of the sliced images in each viewing axis. The 2D predictions for slices along the same viewing axis were stacked together to form 3D predictions. These three 3D predictions, each derived from a different viewing axis, were then averaged to obtain the final probability prediction. The determination of each 3D anatomical domain was then achieved by finding the regions with the largest probability for each pixel. This automatic segmentation process is summarized in Algorithm 1.

Network Architecture

The 2D CNN models were implemented based on the U-Net architecture specialized for medical image segmentation[128] (Figure 2.3b). The network architecture included a down-sampling path (left side) to extract features from input images and an up-sampling path

Algorithm 1: Automatic Segmentation Using an Ensemble of CNNs

Input 3D CT/MR image I; 2D CNNs trained separately for three views
Output 3D segmentation S
Initialize a 3D probability volume P with zeros
for *each view i* **do**
  for *each 2D slice s in view i* **do**
    Compute the 2D probability map with the 2D CNN trained for view i
  end
  Assemble 2D probability maps into a 3D probability volume P_i
  $P \leftarrow P + P_i$
end
$P \leftarrow P/3$; `// Compute average probability map`
for *each voxel k in S* **do**
  $S(k) \leftarrow$ segmentation domain with the highest probability value in $P(k)$
end

(right side) to reconstruct segmentation from extracted features. The down-sampling path included five convolution blocks. Each convolution block consisted of repeated convolutions with multiple 3×3 convolution kernels, followed by activation functions and 2×2 max pooling operations. The activation function was the rectified linear unit (ReLU), $f(x) = \max(0, x)$. The max-pooling operation selected the maximum value within a 2×2 window applied across the activation output and thus halved the spatial resolution of the output. The up-sampling path included four convolution blocks and each block consisted of a transpose convolution and repeated convolutions with multiple 3×3 convolution kernels, followed by ReLU activation functions. The transpose convolutions utilized 2×2 kernels with trainable weights to recover the spatial dimension of the intermediate output. An additional convolution layer was applied at the end to generate an 8-channel probability map, with each channel corresponding to each cardiac domain. Skip connections concatenated the intermediate output from each convolution block of the down-sampling path to the corresponding convolution block input of the up-sampling path.

CNN Optimization

The training of our CNN models was supervised by the manual ground truth segmentation in the MMWHS dataset. We considered a hybrid loss function that contained both multi-class cross-entropy and dice-score loss. Namely, let $L(I, G)$ denote the loss between the CNN prediction P for image I and the corresponding one-hot coded ground truth segmentation G. The hybrid loss function was

$$L(I,G) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{x\in I} G_i(x)\log(P_i(x)) + N - \sum_{i=1}^{N}\frac{2\sum_{x\in I} G_i(x)P_i(x)}{\sum_{x\in I} G_i(x) + \sum_{x\in I} P_i(x)} \tag{2.1}$$

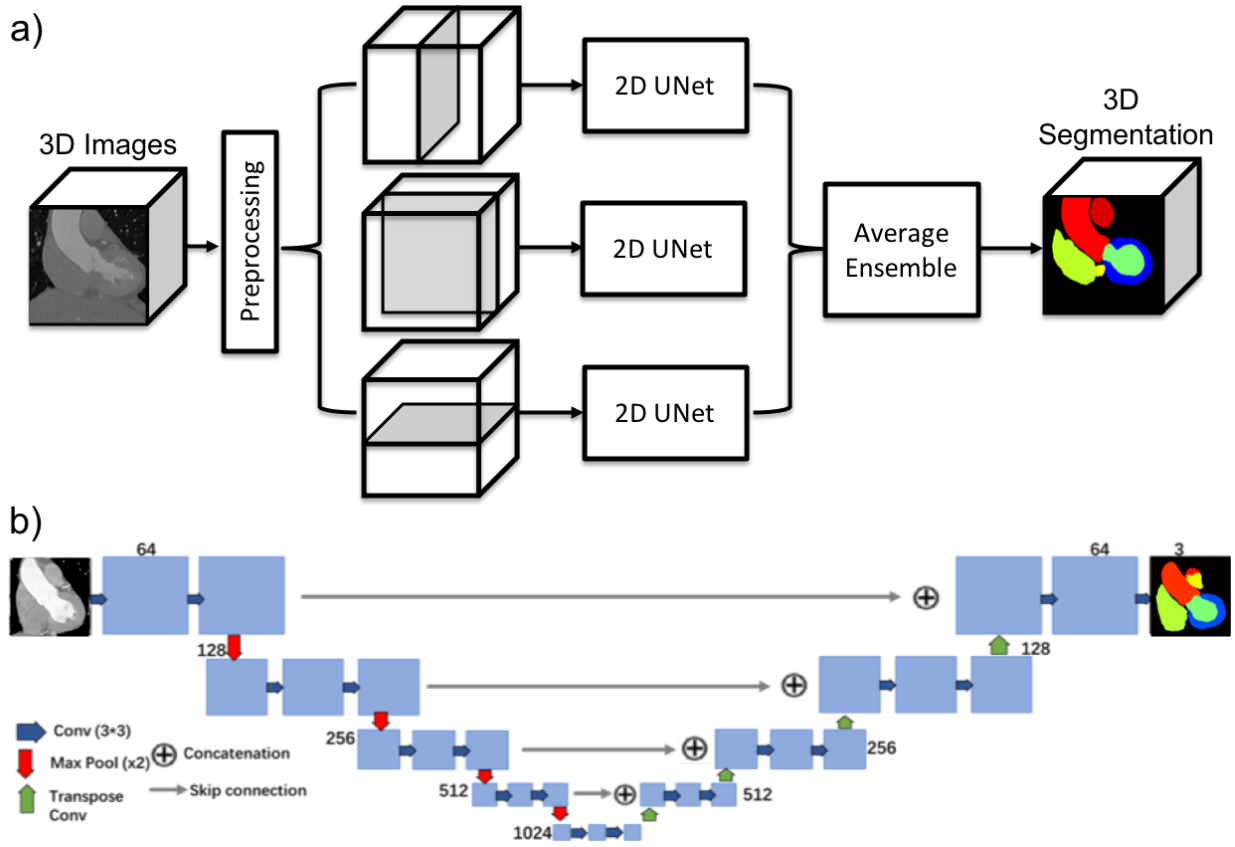


Figure 2.3: a) Diagram of the proposed automatic segmentation approach using an ensemble of CNNs. b) Network architecture of the 2D U-Net CNN model. Numbers illustrate the number of convolution kernels used.

where N denotes the total number of the segmentation (anatomical) domains, while x denotes the pixel in the input image I. $P_i(x)$ represents the predicted probability of pixel x belonging to the segmentation domain i. Weights of the convolution kernels we computed by minimizing the above loss function using the Adam stochastic gradient descent algorithm[62]. The initial learning rate was set to be 0.0001, while β_1 and β_2 for the Adam algorithm were set to 0.9 and 0.999, respectively. Among the 20 CT and 20 MR patient scans, 16 CT and 16 MR scans were randomly chosen as training data. The other CT and MR scans were considered as validation data to select the best-performing model. Dice score was evaluated on the validation data after each training epoch and the CNN model was saved after one epoch only if the validation dice score had improved. Therefore, only the CNN model with the best validation dice score was chosen for future evaluation on the held-out test dataset (which contained another set of 40 CT and 40 MR scans). We adopted a learning rate schedule where the learning rate was reduced by 20% if the validation dice score had not improved for 5 epochs. The minimum learning rate was 5×10^{-6}. The CNN models were trained on a GeForce GTX 1080 Ti GPU

until the validation dice score converged. Data augmentation techniques of random flipping, random shifting, random scaling and random intensity changes were also applied during training to improve robustness. This automated segmentation algorithm was implemented using the functionality of TensorFlow (version 1.12)[1].

Evaluation Metrics And Statistical Methods

Segmentation accuracy was evaluated with an executable provided by MMWHS organizers [182], which computed the surface-to-surface distance errors as well as dice and jaccard scores between our segmentation results and the (hidden) ground truth as determined by the MMWHS organizers. Dice and jaccard scores are similarity indices that range from 0 to 1 as given by

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{2.2}$$

$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.3}$$

, where $|A|$ and $|B|$ are the cardinalities of the two sets of pixels belonging to a certain tissue class in the groud truth segmentation and in the CNN-produced segmentation. Differences in segmentation accuracy among segmentation domains were quantified using paired t-tests.

Geometry Reconstruction and Mesh Generation

We automated the geometry reconstruction and mesh generation process as described here. The entire process was implemented in Python with open-source, Python scriptable dependencies, VTK and SimVascular. For each segmentation region (LV, LA, Ao), the largest connected region of each segmentation domain was extracted to remove any disconnected islands. These segmentations were then smoothed by a closing filter that filled any sharp corners and holes with a diameter smaller than 5 mm to correct non-physical segmentation artifacts. Conversely, an opening filter was applied to remove any extrusions with diameters smaller than 5 mm. Boundaries of the Ao and LA were identified. Since the ground truth segmentation results did not consider the tissue thickness of the LA or Ao, LA and Ao segmentations were sometimes connected, leading to an incorrect fusion between the constructed LA and Ao surfaces. Therefore, LA and Ao segmentations at the location of any shared boundary were eroded until they were distinctly separated. The segmentation volume containing LA, Ao and LV segmentations was then converted to a binary segmentation volume and re-sampled with a resolution of $1.2 \times 1.2 \times 1.2$ mm. Illustrations of the above segmentation processing is shown in Figure 2.4.

A marching cube algorithm [86] was applied to the processed segmentations to generate a watertight surface mesh of the LV, LA and aorta. In order to define appropriate inlet and outlet surfaces, clipping of the Ao and LA is generally necessary. To achieve this, the boundary between the LV and LA (LV and Ao) was identified as the set of points

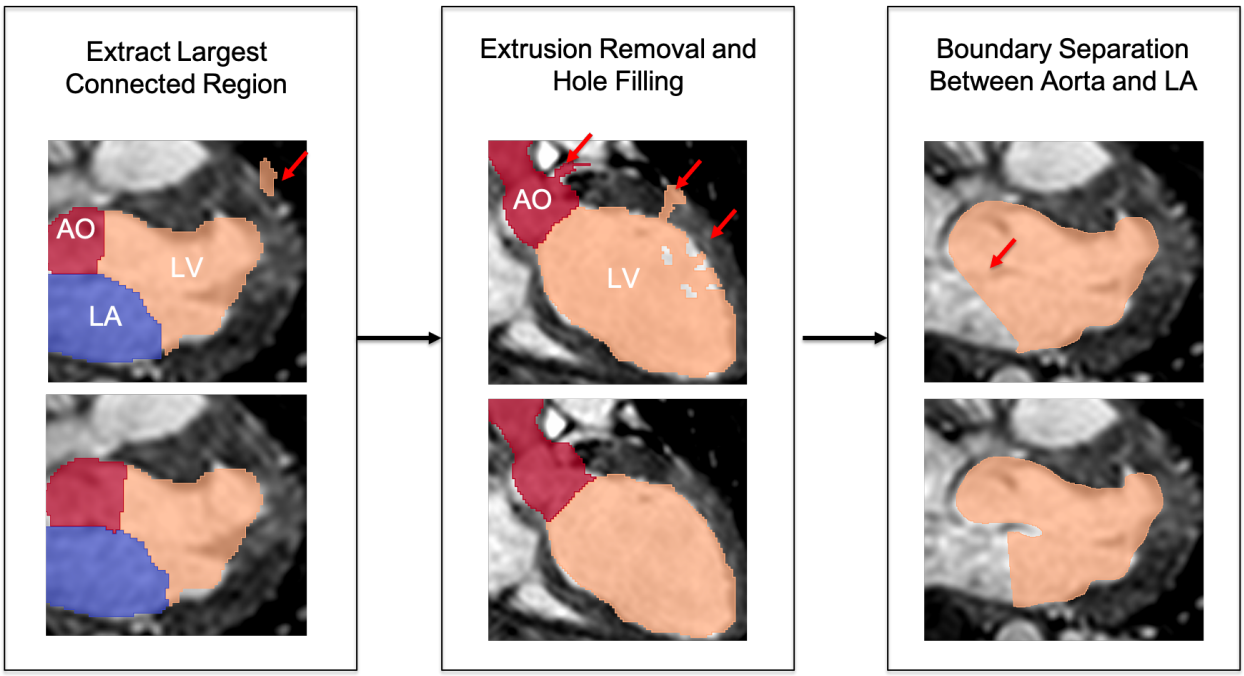


Figure 2.4: Illustrations of segmentation errors corrected by segmentation processing. The top and bottom images display segmentation results before and after described segmentation processing. Red arrows indicate locations of artifacts that segmentation processing corrects. Orange, red and blue shadings on the left and middle panels represent respectively, LV, AO and LA segmentations. The orange shading on the right panel shows the processed binary segmentation of LV with parts of AO and LA combined.

shared between the respective segmentations. The mitral plane (aortic plane) was fitted through those points. The mitral plane origin (aortic plane origin) was defined by the centroid. The LA was clipped by defining a clipping plane parallel to, and 22 mm from, the mitral plane. The aorta was clipped using a plane parallel to, and 45 mm from, the aortic plane origin. Based on our observations, these distances were large enough to avoid substantial boundary effects but small enough to avoid the computational model from being unnecessarily large. We note that naive trimming would generally truncate other parts of the model than intended. We, therefore, constructed trimmers that isolated, respectively, the LA or Ao regions, trimmed these isolated regions, and mapped the results back to the unified model. The resulting mitral opening (MO) and aortic opening (AO) were smoothed by projecting the mesh vertices to their fitted plane and applying Laplacian smoothing on nearby mesh elements. The obtained MO and AO were then triangulated using a constrained 2D Delaunay algorithm.

SimVascular meshing functionality [155], utilizing a combination of custom code, MMG, VMTK and TetGen, was used to generate high-quality surface and volume meshes. First,

the trimmed model is remeshed with a maximum mesh edge size of 1.0 mm. Second, a volume mesh is generated with a maximum mesh edge size of 1.5 mm and a boundary layer meshing near walls. Third, non-rigid image registration of the segmentations was performed automatically (by python scripting of SimpleElastix [64]), to appropriately deform the volume mesh over time. During this process, points located on the MO or AO were projected to their least-square fit plane to ensure that they remained co-planar. The mesh generation process is shown schematically in Figure 2.5.

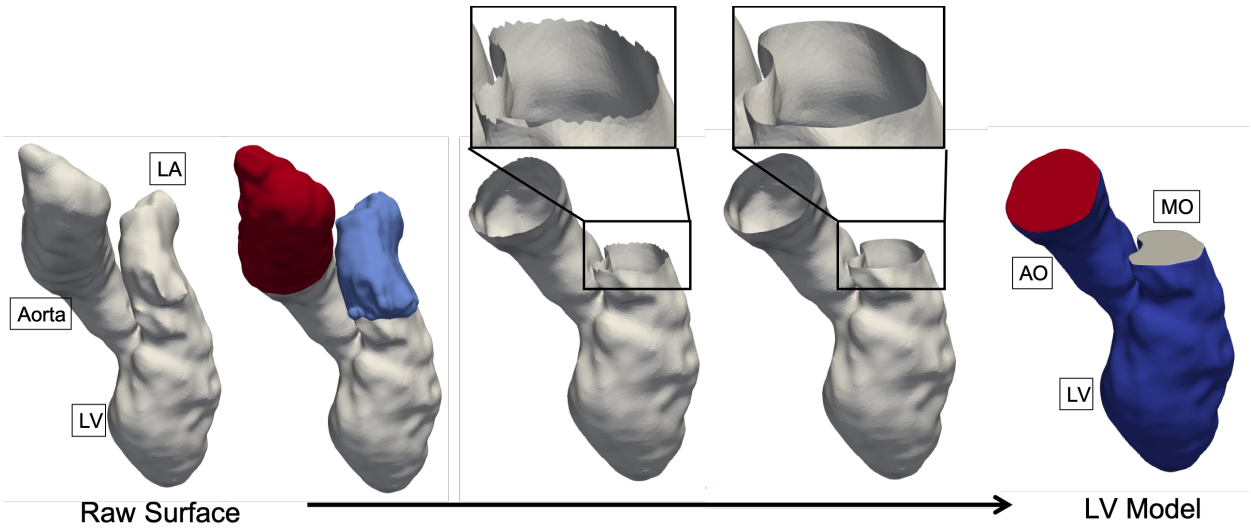


Figure 2.5: Surface processing strategies for converting raw surfaces processed by the marching cube algorithm to CFD-suitable LV meshes. From left to right showing a raw surface generated by the marching cube algorithm, the raw surface with its aorta (red) and LA (light blue) trimmers, the trimmed LV model with parts of aorta and LA, the LV model with smoothed aortic and mitral openings, and the completed LV model with identified boundary faces of LV wall, mitral opening (MO) and aortic opening (AO) in dark blue, gray and red.

Image-Based LV CFD Simulations

We applied the Arbitrary Lagrangian-Eulerian (ALE) formulation of the incompressible Navier-Stokes equations to simulate the intraventricular flow and account for deforming volumetric mesh. The weak formulation of the Navier-Stokes equations defined in ALE coordinates for the 3D moving domain $\Omega(t) \in \mathbb{R}^3$ is given as follows:

Find fluid velocity $\mathbf{v} \in \mathcal{S}_v$ and pressure $p \in \mathcal{S}_p$, such that for all test functions $\mathbf{w} \in \mathcal{V}_v$

and $q \in \mathcal{V}_p$,

$$B(\{\mathbf{w}, q\}, \{\mathbf{v}, p\}, \hat{\mathbf{v}}) = F(\mathbf{w}) \quad \text{where,} \tag{2.4a}$$

$$\begin{aligned} B(\{\mathbf{w}, q\}, \{\mathbf{v}, p\}, \hat{\mathbf{v}}) = &\langle \mathbf{w}, \rho\left(\dot{\mathbf{v}} + (\mathbf{v} - \hat{\mathbf{v}}) \cdot \nabla \mathbf{v}\right)\rangle_{\Omega(t)} \\ &- \langle \nabla \cdot \mathbf{w}, p\rangle + \langle \nabla^s \mathbf{w}, 2\mu \nabla^s \mathbf{v}\rangle_{\Omega(t)} + \langle q, \nabla \cdot \mathbf{v}\rangle_{\Omega(t)} \end{aligned} \tag{2.4b}$$

$$F(\mathbf{w}) = \langle \mathbf{w}, \mathbf{h}\rangle_{\Gamma(t)} \tag{2.4c}$$

where $\langle ., . \rangle$ represents the integral inner product over domain $\Omega(t)$, ∇^s is the symmetrization of gradient operator ∇, $\hat{\mathbf{v}}$ represents the mesh velocity defined as $\hat{\mathbf{v}} = \frac{\mathbf{x_{t+1}} - \mathbf{x_t}}{\Delta t}$. $\mathbf{h}$ represents boundary traction. Blood was assumed to have a viscosity μ of $4.0 \times 10^{-3} Pa \cdot s$ and a density ρ of $1.06 g/cm^3$. The equations were solved with the open-source svFSI solver from the SimVascular project [155, 141].

We note that LV surface and volume meshes were created at the end of diastole and propagated to different time frames. Since the time resolution of the image data is too coarse to be used directly in time-stepping of the Navier-Stokes equations, cubic spline interpolation of the mesh motion was applied to generate 2000 interpolated meshes. The mesh motions computed from these interpolated meshes were imposed as Dirichlet boundary conditions on walls, and to the MO during systole, or to the AO during diastole. Neumann (prescribed pressure) boundary conditions were applied to the mitral inlet during diastole or to the aortic inlet during systole. Diastole and systole phases were determined based on the increase and decrease of LV volume.

2.3 Results

We tested segmentation accuracy on the 40 patient CT scans and 40 patient MR scans from the MMWHS test data set. These data were not used in any way to train the model. Table 2.1 displays the dice and Jaccard scores and average surface distance errors of the LV, LA and Ao produced by our automated segmentation framework. The MMWHS grand challenge [182] reports these measures for whole-heart (WH) segmentation results (which includes all seven segmented cardiac tissue domains) from challenge participants. In Table 2.1 we compare our WH segmentation accuracy with the *top performing* algorithm from the grand challenge. For CT data, our WH segmentations outperformed the top-performing algorithm in all metrics–mean dice score, Jaccard score and mean average surface distance error. For MR data, our WH segmentation results achieved better mean average surface distance errors than the top-performing algorithm but had slightly lower mean dice and Jaccard scores.

To provide further details on segmentation accuracy, the box plots in Fig. 2.6 give the distributions of the segmentation accuracy measures for LV, Ao, LA and WH segmentation. For CT data, both LV and Ao segmentations were more accurate than WH segmentation in terms of dice score ($p < 0.01$) and Jaccard score ($p < 0.001$). For MR data, LV segmentation was more accurate than WH segmentation across all metrics ($p < 0.001$). Altogether, our

		LV	LA	Ao	WH	WH (top) [182]
CT	Dice	0.938±0.042	0.936±0.027	0.95±0.02	0.92±0.022	0.908±0.086
	Jaccard	0.886±0.07	0.88±0.047	0.905±0.036	0.852±0.036	0.823±0.037
	SD (mm)	0.84±0.647	0.941±0.318	0.498±0.177	0.978±0.283	1.117±0.25
MR	Dice	0.915±0.051	0.871±0.064	0.869±0.083	0.871±0.05	0.874±0.039
	Jaccard	0.847±0.077	0.776±0.089	0.776±0.113	0.775±0.072	0.778±0.06
	SD (mm)	1.155±0.667	1.393±0.524	2.384±1.758	1.612±0.577	1.631±0.58

Table 2.1: Dice and jaccard scores and surface distance (SD) accuracy of our LV, LA, Ao and WH segmentations. Our WH segmentation accuracy is compared with the top-performing algorithm from the MMWHS grand challenge [182]. All accuracy measures are represented by mean ± standard deviation, which are computed over different patients.

segmentation algorithm performance was comparable to or better than the most accurate grand challenge algorithms in terms of WH segmentation, and moreover, our LV segmentation, in particular, was generally more accurate than WH segmentation.

Testing the accuracy of the segmentation process is important, but only assesses the accuracy of pixel classification. This does not directly assess if segmentations will lead to valid model geometries, or domains that can be effectively meshed for CFD purposes and therefore post-processing of the segmentation results was necessary. Thus we evaluated the robustness of our geometry reconstruction and mesh generation process on the 40 CT and 40 MR scans in the MMWHS test set. We evaluated the accuracy of the post-processed LV segmentation, whether there were any errors in geometry construction or volumetric meshing, and finally visually inspected the models for obvious artifacts. Segmentation post-processing slightly improved the LV dice score and average surface distance errors for MR data. For CT data where our segmentation framework already generated more accurate LV segmentation than MR data, post-processing slightly reduced the LV segmentation accuracy. Figure 2.7 displays the segmentation and the constructed LV models from the CT data within the 10th, 50th, and 90th percentiles and the "worst case" scenario encountered in terms of LV surface distance errors. For all but one segmentation result obtained from the 40 CT patient scans, our framework was able to generate the reconstructed LV geometry with LA and Ao extensions and produce a valid volumetric mesh. The remaining case had segmentation that contained holes too large to be removed by our baseline, automatic post-processing. Although model geometry and mesh were constructed, visual inspection detected large artifactual indentations. Not surprisingly this model was the one that exhibited the largest average LV surface distance error of 4.44 mm among the CT test set.

Figure 2.8 displays the segmentation and the corresponded LV models for MR patient data with the 10th, 50th, 90th percentiles and the "worst-case" that had the largest LV

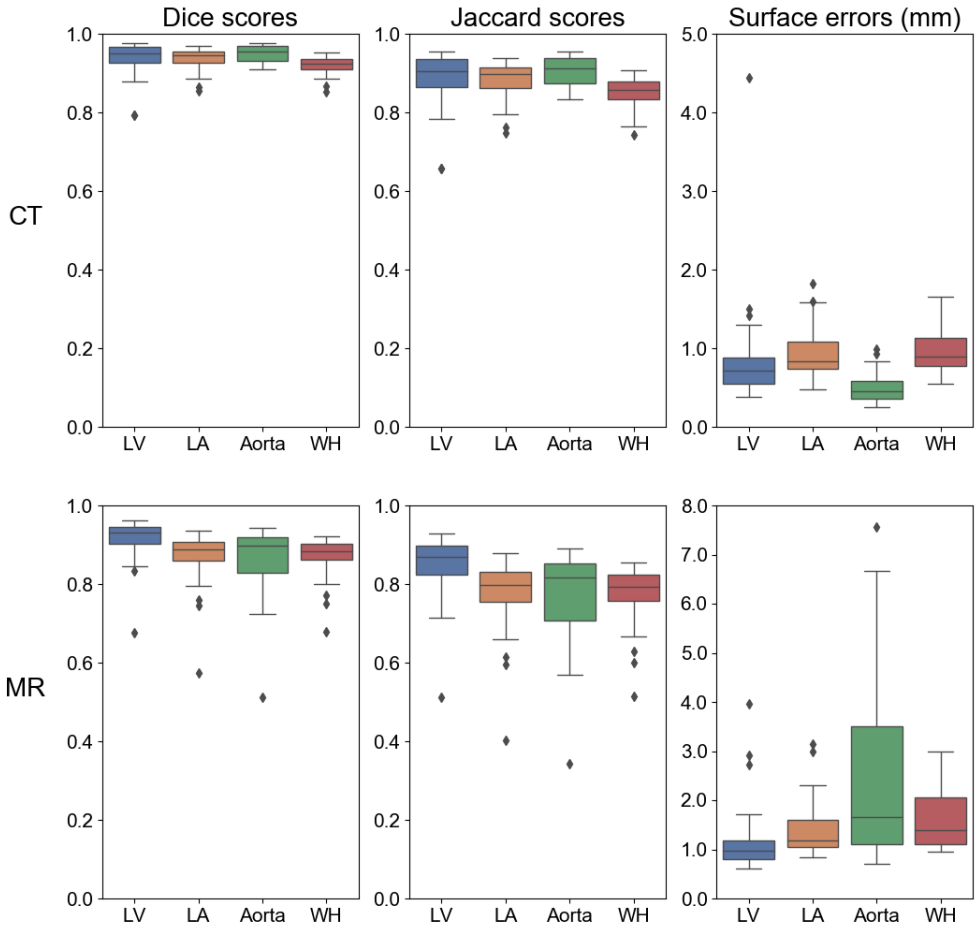


Figure 2.6: Box plots of dice scores, jaccard scores and surface distance errors for LV, LA, Ao and WH segmentation results from the MMWHS test data sets.

segmentation surface distance errors. Our framework was able to successfully generate CFD-suitable LV meshes automatically for 36 out of the 40 MR patient data. For 3 out of the 4 cases, we observed missing segmentation in the middle of the aorta due to poor image quality, causing the framework to be unable to identify the aortic outlet. However, this particular problem was readily corrected in practice by moving the cutting plane of the aorta towards LV to reduce the length of the aorta required to generate aortic outflow. Indeed, by decreasing the aortic cutting plane locations, our framework was able to succeed in these 3 cases. However, our framework was not able to generate CFD-suitable LV meshes for one

remaining case without manual correction of the segmentation results. The failed case had erroneous segmentations that missed part of the cardiac structures and corresponded to the lowest WH dice scores of 0.679 and the largest mean LV surface distance errors of 3.96mm.

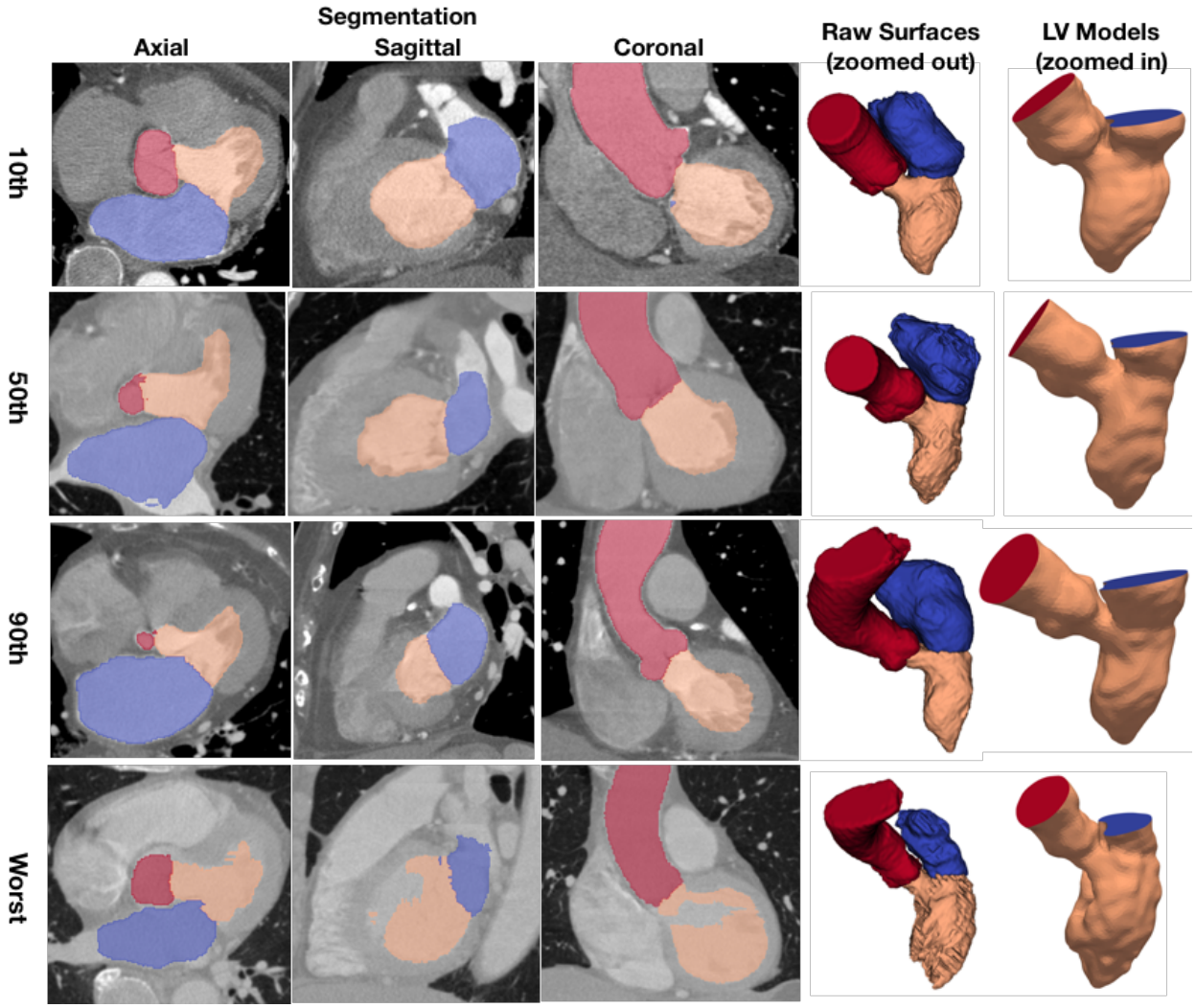


Figure 2.7: Segmentation results, raw surfaces and constructed models for CT test cases with the 10th, 50th, 90th percentiles and the largest average LV segmentation surface distance errors. Segmentation and raw surfaces of LA, AO and LV are shown in blue, red and orange, respectively. The identified boundary faces of the LV wall, mitral opening and aortic opening on the constructed models are shown in orange, blue and red, respectively.

Figure 2.9 displays the distribution of model construction time, which is the time required to go from image data to a volumetrically meshed 3D model for single-phase patient CT or MR scans in the MMWHS test dataset. A 3.5 GHz Intel Core i7 CPU processor was

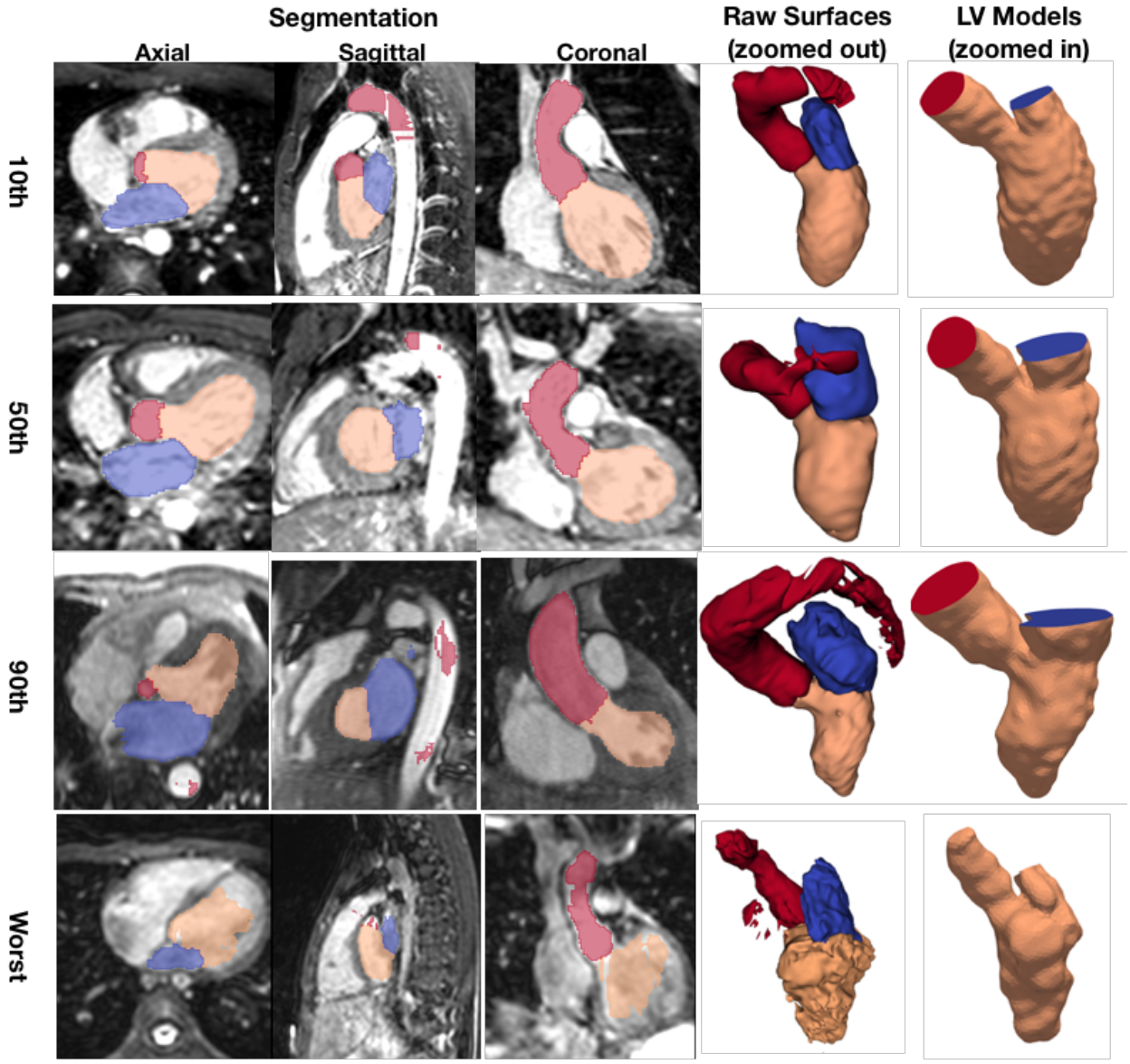


Figure 2.8: Segmentation results, raw surfaces and constructed models for MR test cases with the 10th, 50th, 90th percentiles and the largest average LV segmentation surface distance errors. Segmentation and raw surfaces of LA, AO and LV are shown in blue, red and orange, respectively. The identified boundary faces of the LV wall, mitral opening and aortic opening on the constructed models are shown in orange, blue and red, respectively.

used to evaluate geometry and mesh construction time, and an Nvidia Tesla K80 GPU was used to evaluate CNN segmentation time. The maximum, median and minimum total model

generation times were 172, 126 and 102 seconds respectively for CT data and were 188, 138 and 71 seconds respectively for MR data. Model generation time for MR data was significantly longer than for CT data ($p < 0.01$) due to a longer segmentation post-processing time. On average, the percentages of time spent in segmentation, segmentation post-processing, geometry reconstruction and meshing were 42%, 3%, 5% and 50%, respectively for CT data and were 38%, 8 %, 5 % and 49%, respectively for MR data.

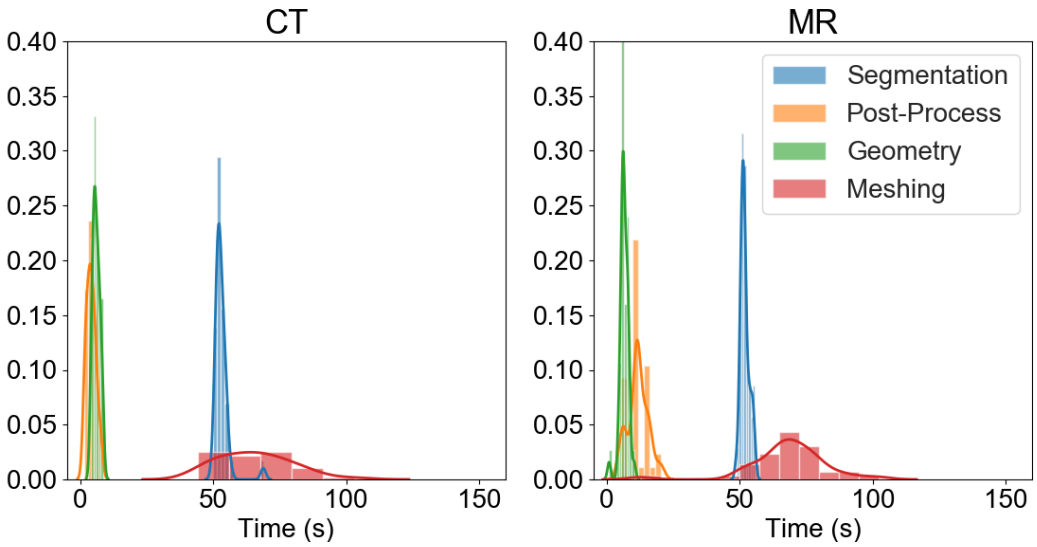


Figure 2.9: Histogram distribution of the time spent in segmentation, post-processing, geometry reconstruction and meshing for CT and MR data

CFD-ready LV models were automatically generated from the time-series CT data of two patients with diastolic dysfunction, as shown in figure 2.10. Non-rigid image registration took an average of 160 seconds to propagate the constructed LV model to the next time frame on a 3.5 GHz Intel Core i7 CPU processor. Interpolating the registered meshes and writing the CFD-ready model to SimVascular input files took another 158 seconds on average. To evaluate the accuracy of the reconstructed LV geometries on these time-resolved data, we used manual and semi-automatic segmentation tools provided by the open-source software SimVascular to generate a ground-truth segmentation of each time frame. We also constructed ground truth LV models with LA and Ao extensions from the ground truth segmentation. Compared with the ground truth models, the *maximum* value of the average surface distance errors among all time frames was 1.40 mm for patient A, and was 1.87 mm for patient B. Fig. 2.10b shows that volume curves computed from the interpolated meshes generated with our framework were similar to those computed from the ground truth for both patients and the maximum percentage differences among all time frames were 3.6% and 6.3% for patient A and B, respectively. Fig. 2.11 shows simulated velocity streamlines of LV flow at different time frames of the cardiac cycle. During the ejection phase, the

velocity streamlines of LV demonstrated a converging flow pattern for both patients. The maximum outflow velocity during systole for patient A and patient B was 1.36 and 1.25 m/s, respectively. During diastole, we observed early filling, diastasis and atrial filling phases for both patients as shown at times C, E and F in Fig. 2.11. The mitral jet entered LV during early filling, changed direction due to impact with the LV wall and formed circulatory flow within the LV. During diastasis, patient A had a dominant LV vortex. With a smaller LV diameter to length ratio, patient B had a more complicated flow pattern with two major circulations, in the upper and lower parts of the LV, respectively. The maximum inflow velocity during early filling and atrial fillings were 0.69 and 0.37 m/s for patient A and 0.56 and 0.28 m/s for patient B, respectively.

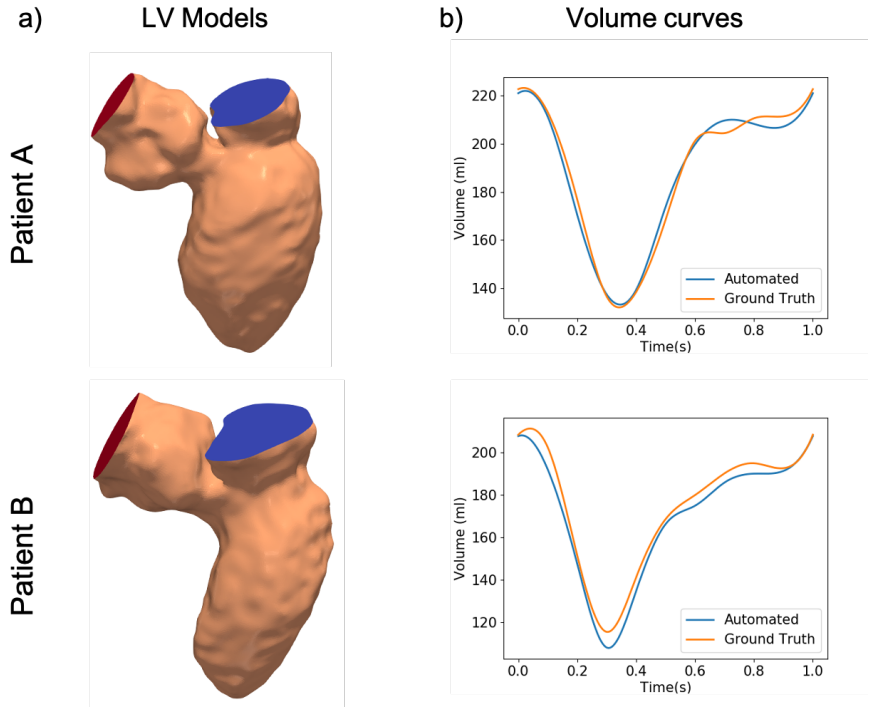


Figure 2.10: Patient-specific CFD-ready LV models: a) constructed model geometries at middle diastole and b) comparisons of volume curves computed from interpolations of LV models generated with our framework (blue) and LV models generated manually (yellow) during one cardiac cycle.

2.4 Discussion

Image-based CFD simulations of LV flow, although powerful in understanding patient cardiac hemodynamics, usually require significant user interactions in the model generation process. Prior studies have thus involved only a single or very few patient-based models. In the present

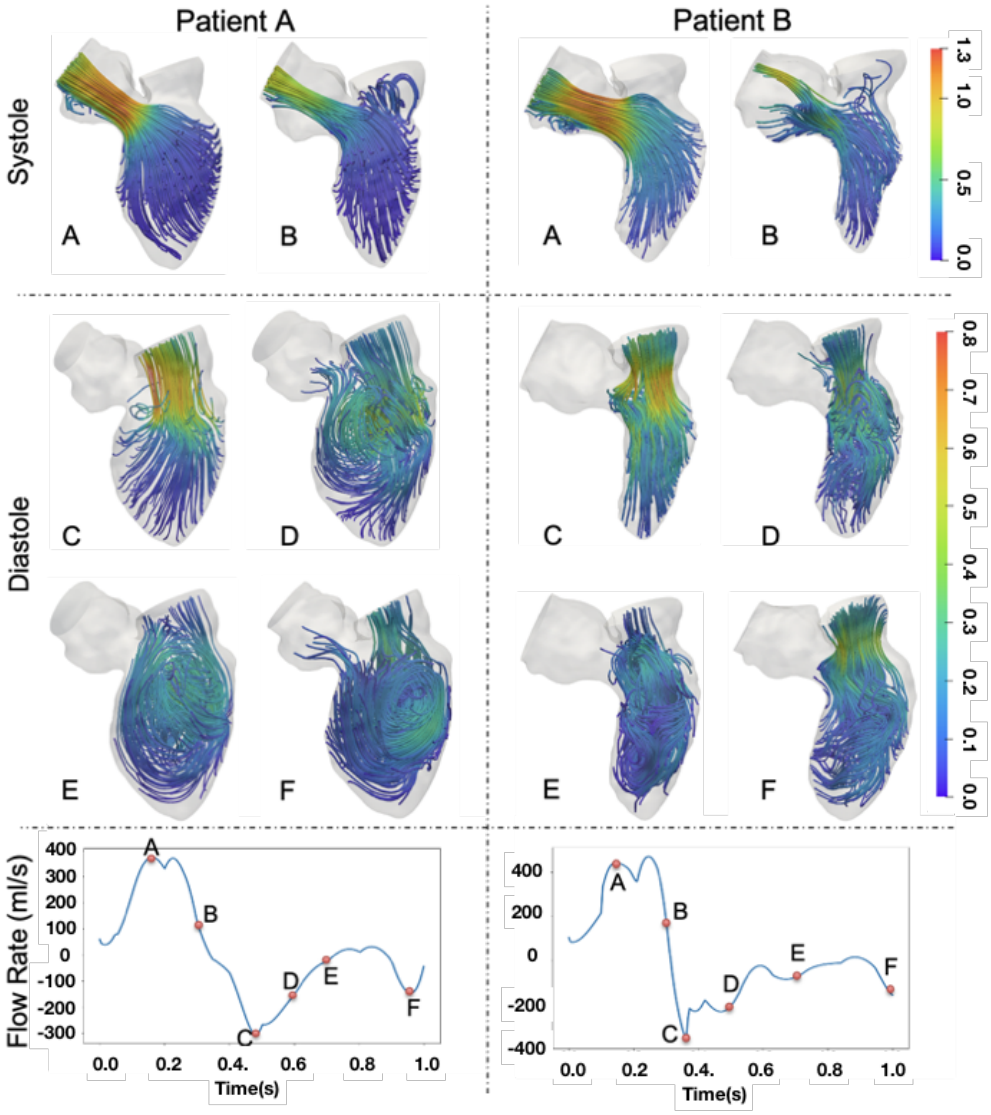


Figure 2.11: Patient-specific CFD simulation results from time-resolved patient CT data. Top and middle panels show velocity streamlines at five different time frames, peak systole (A), late systole (B), early diastolic filling (C), diastasis (D) and atrial filling (E), as labeled on the flow rate curves on the bottom panel. The color map represents velocity magnitude (m/s).

study, we demonstrated an automated framework to efficiently generate CFD-suitable LV models from patient data from two common imaging modalities (CT and MR) using a novel combination of deep-learning-based segmentation, geometric processing and image registration techniques.

The automated segmentation framework based on an ensemble of CNNs demonstrated

promising accuracy for both CT and MR scans. Testing on the same benchmark CT data set, our framework outperformed the previous best algorithm by Payer et al. who used a two-stage 3D CNN pipeline[182, 113]. Besides increased accuracy, another advantage of our segmentation pipeline was the increased resolution of segmentation results (33% along each dimension). By using three 2D CNNs rather than 3D CNNs, our segmentation pipeline reduced the computation and memory requirements during training and thus was able to handle a larger input image size of $256 \times 256 \times 256$. A higher segmentation resolution is beneficial to the downstream model generation process for CFD simulations since it helps to avoid the staircase artifacts due to poor image resolutions, which can affect computed hemodynamics. Moreover, most prior deep-learning-based automatic segmentation algorithms for cardiac structures have been trained on a single imaging modality, except for Tong et al. who trained on both MR and CT data but did not achieve very good performance [150, 182]. We demonstrated that it is feasible to train a single system of CNN models on MR and CT data simultaneously while still achieving good performance for both modalities. This could be explained by the high capacity of CNNs due to the large number of parameters they possess [95]. With this advantage, our framework did not require manual specification of which imaging modality to operate on and may store only one set of CNN model parameters for both MR and CT data. Compared with CT data, MR data presented larger intensity variation, acquisition field of view, image quality and uncertainties in ground truth segmentation [182]. Consistent with prior segmentation algorithms, our framework performed better in segmentation for CT data than for MR data [182]. More training data of MR scans may be required for the deep neural networks to better capture the inherently more diverse distribution of MR data.

Although deep-learning-based algorithms have been extensively applied to LV segmentation, to our knowledge, our framework was the first to explore the downstream geometry reconstruction procedures required to generate CFD-suitable models. Maher et al. constructed vascular models for CFD simulations from segmentation generated by deep-learning algorithms[88]; however, LV models have different geometric considerations than those of vascular models. In contrast to atlas-based segmentation approaches that attempt to map an existing atlas to new images, deep-learning-based segmentation algorithms are usually trained to optimize the voxel-wise segmentation accuracy between predicted segmentation and the ground truth, thus have no constraints on the shapes of the segmented structures. The lack of shape constraints encourages better generalization to the new and diverse data and avoids the tremendous computational cost related to atlas registration [48]; however, it poses challenges to the down-stream LV geometry reconstruction process required for generating a CFD-suitable geometry since the segmentation results are not guaranteed to have valid global topology. Indeed, it is possible to have good segmentation accuracy based on conventional "closeness metrics" but not have a segmentation suitable for CFD model construction. In this study, we demonstrated that with a reliable deep-learning-based segmentation framework, along with simple and automatic post-processing techniques, we were able to successfully construct LV CFD models for the vast majority of cases considered. Specifically, from the segmentation results of LV, LA and aorta, small and isolated regions

needed to be removed by extracting the largest connected region; boundaries between LA and aorta needed to be clearly separated to generate anatomically correct LA and aorta geometries with the marching cube algorithms; noisy extrusions or holes within and on the boundary of segmentation needed to be removed or filled. These operations were conveniently achieved by a combination of image foreground dilation or erosion functions and were successful on 39 out of 40 tested CT image data and 39 out of 40 tested MR image data. Not surprisingly, reliable segmentation results were essential to obtain accurate LV model geometries. The failed cases were due to erroneous segmentation results with large surface distance errors, missing structures or extrusions, or holes too large to be corrected by the post-processing algorithms. For the small number of failed cases, manual corrections may be required to generate acceptable LV model geometries. Since our deep-learning-based segmentation algorithm was developed based on a limited number of samples, it may not generalize to all kinds of image abnormalities, such as low image quality, artifacts or extreme tissue intensity. Indeed, it is impossible to guarantee an automated approach will always produce a valid result when image quality is not controlled.

Although this study focused on the image-to-volume-mesh process, and not the analysis of intraventricular hemodynamics, we did demonstrate the capability to perform ALE-based CFD simulations of LV hemodynamics from the generated models. The fact that the time-resolved CT image data were from clinical scans and a completely different source than the MMWHS dataset demonstrates the potential robustness of the approach. Compared with ground truth LV models created through manual efforts, LV models created by our framework had a relatively small percentage of volume difference over the cardiac cycle. We note that the ground truth models were created by only one observer with no repeats. Zhuang et al. reported that while LV was the least challenging cardiac structure to segment manually among others, the inter-and intra- observer variabilities were 6.3% and 5.8% for LV segmentation in MR data [182]. Therefore, the volume differences of the LV models generated automatically are comparable to inter-and intra- observer variabilities. Our CFD simulations provided detailed LV flow patterns throughout the cardiac cycle and the converging flow pattern during systole and the circulatory flow patterns were consistent with prior studies [61, 97]. Although both patients had diastolic LV dysfunction, their LV shapes were different especially in terms of LV sphericity. We observed a dominating flow circulation in the patient LV with higher sphericity but not in the patient LV with lower sphericity. Such difference was in agreement with Martinez-Legazpi et. al., who demonstrated that reduced LV chamber sphericity could reduce vortex contribution to the diastolic filling of LV [91].

Our framework was able to generate meshed LV model for a single phase in around two minutes on average, using a modern desktop computer with the help of a GPU. When including the time spent in segmentation registration to propagate LV model to other time frames, interpolating meshes and creating compatible input files for SimVascular svFSI solver, our framework took around 30 minutes in total to construct CFD-ready LV simulation input files for one set of time-resolved patient image data with 10 time frames. Segmentation registration was the most time-consuming step in our model construction framework. However, this step could be parallelized in the future to reduce the total model generation time[136].

In contrast, prior approaches of generating one CFD-ready LV model from images could take anywhere from 20–50 hours of work and significant human efforts[94, 4]. Using the semi-automatic segmentation techniques and geometry processing algorithms in SimVascular, we spent around 10 hours constructing each LV model on the same patient data, which we consider to be typical for an "experienced" user. Therefore, compared with prior model construction approaches using manual or semi-automatic techniques, our framework could save on the order of hours of time and human efforts.

Limitations of the current automated LV CFD model generation framework include the lack of explicit mitral valve and aortic valve structures. However, valve leaflets are generally not resolvable from clinical scans. Therefore, patient-specific geometry reconstruction of valves is very challenging due to limited image resolution and large deformation of the valve structures. Recent advances in machine-learning-based approaches to obtain heart valve geometry based on statistical information could be applied in the future to improve our framework[120, 80]. Similar to the valve leaflet, the papillary muscles and trabeculae structures of the LV were not modeled in our framework since these are generally not resolvable from clinical imaging. Although smoothed LV geometry is a common simplification adopted by many prior studies, recent studies have demonstrated that these structures could lead to improved apical washout, enhanced viscous dissipation rate, increased intra-ventricular pressure drop and reduced the wall shear stress and thus should be incorporated for better simulation accuracy[129, 156].

With improved insights into the importance of RV dysfunction in the pathogenesis and outcomes of cardiovascular diseases over recent years, there has been a growing interest in understanding the intraventricular flow pattern in RV [112, 137, 106, 24]. Image-based CFD simulations of RV flow may provide patient-specific, spatially and temporally well-resolved analysis of RV hemodynamics. We note that although this study focused on LV, we expect the proposed framework could be readily adapted for the automated construction of patient-specific CFD-ready RV models. Our framework was able to automatically produce segmentations of RV, RA and pulmonary arteries. From those segmentations, similar segmentation post-processing and surface reconstruction procedures could be applied to reconstruct the RV geometry with appropriate inlet and outlet structures.

2.5 Conclusions

We have developed a streamlined framework to automatically generate CFD-ready LV models from patient image data. The framework leveraged a novel combination of deep-learning-based automatic segmentation algorithms and geometry processing algorithms to robustly create CFD-suitable LV models from both CT and MR image data. We utilized an ensemble of 2D CNNs to achieve high 3D segmentation resolution and outperformed previous automatic segmentation approaches evaluated on the same dataset. To support CFD simulation of LV hemodynamics using an ALE formulation, the framework can automatically identify boundary faces of mitral and aortic opening, and LV as well as compute displace-

ment information of the mesh vertices throughout the cardiac cycle using image registration techniques. Compared with prior manual or semi-automatic methods, our framework offers orders of magnitude savings in time and human efforts in developing image-based CFD simulations of LV flow. The entire framework was implemented in Python and can be conveniently executed from the command line as a program with all dependencies (TensorFlow, VTK, SimVascular and SimpleElastix) being open-source and Python-scriptable. The above advantages may enable our framework to aid in future higher throughput, large-cohort analyses of patient-specific LV hemodynamics in relation to LV dysfunction.

Chapter 3

A Deep-Learning Approach For Direct Whole-Heart Mesh Reconstruction

3.1 Introduction

In the previous chapter, we presented a multistage approach whereby 3D segmentations of cardiac structures are first obtained from image volumes, meshes of the segmented regions are then generated using marching cube algorithms, and finally manual surface post-processing or editing is performed. However, the quality of reconstructed surfaces highly depends on the quality of segmentation and the complexity of the anatomical structures. Automatic heart segmentation has been a popular research topic and previously published algorithms have been summarized in detail [180, 182, 116, 43]. Generally, there are two common approaches to whole heart segmentation: multi-atlas segmentation (MAS) [7, 179, 181] and deep-learning-based segmentation [128, 21]. Compared with MAS, deep-learning-based approaches have become more popular as they have demonstrated higher segmentation precision [182, 113] and are much faster in practice. While a couple of recent studies have reduced the processing time of MAS approaches down to a couple of minutes [12, 13], deep-learning-based approach can generally process a whole heart segmentation within a couple of seconds. However, while deep-learning-based methods may produce segmentations that achieve high average voxel-wise accuracy, they can contain extraneous regions and other nonphysical artifacts. As we described in the previous chapter, correcting such artifacts would require a number of carefully designed post-processing steps and sometimes manual efforts [66]. Indeed, since the CNN-based segmentation methods are based on the classification of each image voxel to a particular tissue class, the neural networks are often trained to reduce the voxel-wise discrepancy between the predicted segmentation and the ground truth and therefore lack awareness of the overall anatomy and topology of the target organs.

Moreover, CNN-based 3D segmentation methods are memory intensive and therefore require downsampling of the data to fit within memory and thus can only generate segmentation with limited resolution. However, high-resolution geometries are often required

for downstream applications such as computational simulations, and direct or low-resolution segmentation will often produce surfaces with staircase artifacts that require additional post-processing [166, 165, 154]. Compared with representing whole heart geometries as segmentations on a dense voxel grid, representing the geometries as meshes is a more compact representation, as only point coordinates on the organ boundaries need to be stored. This advantage may enable efficient reconstruction of high-resolution surface meshes on a limited memory budget and avoid the stair-case artifacts of surfaces constructed from low-resolution 3D segmentation. Moreover, for low-resolution input images, voxel-wise segmentation would be a coarse representation of the underlying cardiac structures, but a surface mesh representation can still function as a smoother and more realistic representation of the shapes as the mesh vertices are defined in a continuous coordinate space and do not have to align with the input voxel grid.

Some studies have adopted a model-based approach to directly fit surfaces meshes of the heart to target images [30, 31, 117]. Such approaches deform a template mesh using local optimization to match with tissue boundaries on input images. However, they are often sensitive to initialization and require complicated steps and manual efforts to construct a mean template of the heart. A recent study by [173] proposed deep learning to learn the initialization of the active contour method–a model-based approach–to help solve for the contours of the target tissues. Alternatively, others have turned to pure deep learning methods that do not require test-time optimization.[170] proposed a deep learning approach to jointly predict the segmentation and the geometry of the left ventricle in the form of a point cloud from the image data.

Recent progress in geometric deep learning has extended the concepts of the convolutional neural network on irregular graphs [25, 11]. Recent deep-learning-based approaches have shown promise in reconstructing shapes as surface meshes from image data using graph convolutional neural networks [159, 167, 118]. However, these approaches have focused on reconstructing a single shape from a 2D camera image and thus cannot be directly applied to reconstructing multiple anatomical structures from volumetric medical image data. A recent study from [168] extended the work of [159] to 3D volumetric medical image data and demonstrated improved segmentation results. However, their method demonstrated success only on simple geometries such as the liver, hippocampus, and synaptic junction but not on the whole heart that involves multiple cardiac structures with widely varying shapes.

To overcome these shortcomings, we explore the problem of using a deep-learning-based approach to directly predict surface meshes of multiple cardiac structures from volumetric image data. Our approach leverages a graph convolutional neural network to predict deformation on mesh vertices from a pre-defined mesh template to fit multiple anatomical structures in a 3D image volume. The mesh deformation is conditioned on image features extracted by a CNN-based image encoder. Since cardiac structures such as heart chambers are homeomorphic to a sphere, we use spheres as our initial mesh templates, which can be considered as a topological prior of the cardiac structures. Compared with classification-based approaches, our approach can reduce extraneous regions that are anatomically inconsistent. Using a generic initial mesh also enables our approach to be easily adapted to other anatom-

ical structures.

The key contributions of our work are as follows:

1. We propose the first end-to-end deep-learning-based approach to predicte multiple anatomical structures in the form of surfaces meshes from 3D image data. We show that our method was able to better produce whole-heart geometries from both CT and MR images compared to classification-based approaches.

2. We investigate and compare the impact of dataset size and variability on whole-heart reconstruction performance to different methods. When having trained on both small and larger training datasets, our method demonstrated better Dice scores for most of the cardiac structures reconstructed than prior approaches.

3. As cardiac MR image data often have large variations across different data sources, we compare different methods and demonstrate the advantage of our approach on MR images with varying through-plane resolution as well as on low-resolution MR images that differ significantly from our training datasets.

4. Since our approach predicts deformation from a template mesh, we show that our reconstructions generally have point correspondence across different time frames and different patients by consistently mapping mesh vertices on the templates to similar structural regions of the heart. We demonstrate the potential application of our method in efficiently constructing 4D whole heart dynamics that captures the motion of a beating heart from a time series of images.

3.2 Methods

Dataset Information

Since cardiac medical image data is sensitive to a number of factors, including differences in vendors, modalities and acquisition protocols across clinical centers, deep-learning-based methods can be easily biased to these factors. Therefore, we aimed to develop our models using whole heart image data collected from different sources, vendors and imaging modalities. We included data from four existing public datasets that contain contrast-enhanced CT images or MR images that cover the whole heart. These four datasets are from the multi-modality whole heart segmentation challenge (MMWHS) [182], orCalScore challenge [169], left atrial wall thickness challenge (SLAWT) [57] and left atrial segmentation challenge (LASC) [148]. The use of such diverse data enables us to not only better evaluate the reconstruction accuracy of our trained model but also evaluate the impact of dataset size and variability on model performance.

Additional time-series CT and MR images were collected to evaluate the performance of our trained neural network models on time-series image data acquired from different data sources from the training data. The time-series CT data were from 10 patients with left

	CT data				MR data		
	MMWHS [182]	OrCaScore [169]	SLAWT [57]	time-series CT	MMWHS [182]	LASC [148]	cine MR
Vendor	Philips	GE, Philips, Siemens and Toshiba	Philips Achieva 256 iCT	GE	1.5T Philips and 1.5T Siemens Magnetom Avanto	1.5 T Philips Achieva	1.5 T Philips
# of clinical sites involved	2	4	1	1	2	1	1
# of 3D image volumes	60	72	10	100	60	27	200
# of patients involved	60	72	4	10	60	27	10
In-place resolution (mm)	0.78 by 0.78	0.4-0.5 by 0.4-0.5	0.4 by 0.4	0.44 by 0.44	1.6-2.0 by 1.6-2.0	1.25 by 1.25	0.65-1.75 by 0.65-1.75
Slice thickness (mm)	1.6	0.5-0.625	0.8-1.0	0.625	2.0-3.2	2.7	8-10
Temporal resolution (ms)	N/A	N/A	N/A	100	N/A	N/A	50
Public or private	public	public	public	private	public	public	private

Table 3.1: Summary of data characteristics for whole heart CT and MR data included.

ventricular diastolic dysfunction. The 9 sets of cine cardic MR data were from 5 healthy subjects and 4 patients with cardiac diseases. All data was de-identified and previously collected for other purposes. The details of the datasets used and collected are described in the following sub-sections and summarized in Table 3.1. We followed the same method of [182] to manually delineate seven cardiac structures: LV, LA, RA, RV, myocardium, aorta and pulmonary artery for the collected image data that did not have ground truth annotations of the whole heart.

Geometry Reconstruction From Volumetric Images

Our framework consists of three components to predict the whole-heart meshes from a volumetric input image: (1) an image encoding module that extracts and encodes image features, (2) a mesh deformation module that combines features from images and meshes to predict deformation of mesh vertices, and (3) a segmentation module that predicts a binary segmentation map to allow additional supervision using ground truth annotations. Figure 3.1 shows the overall architecture.

Image Encoding Module

For an input image data, the image encoding module uses a series of 3D convolutional layers to extract volumetric image feature maps at multiple resolutions. These feature maps are required by the following mesh deformation module to predict whole-heart geometries.

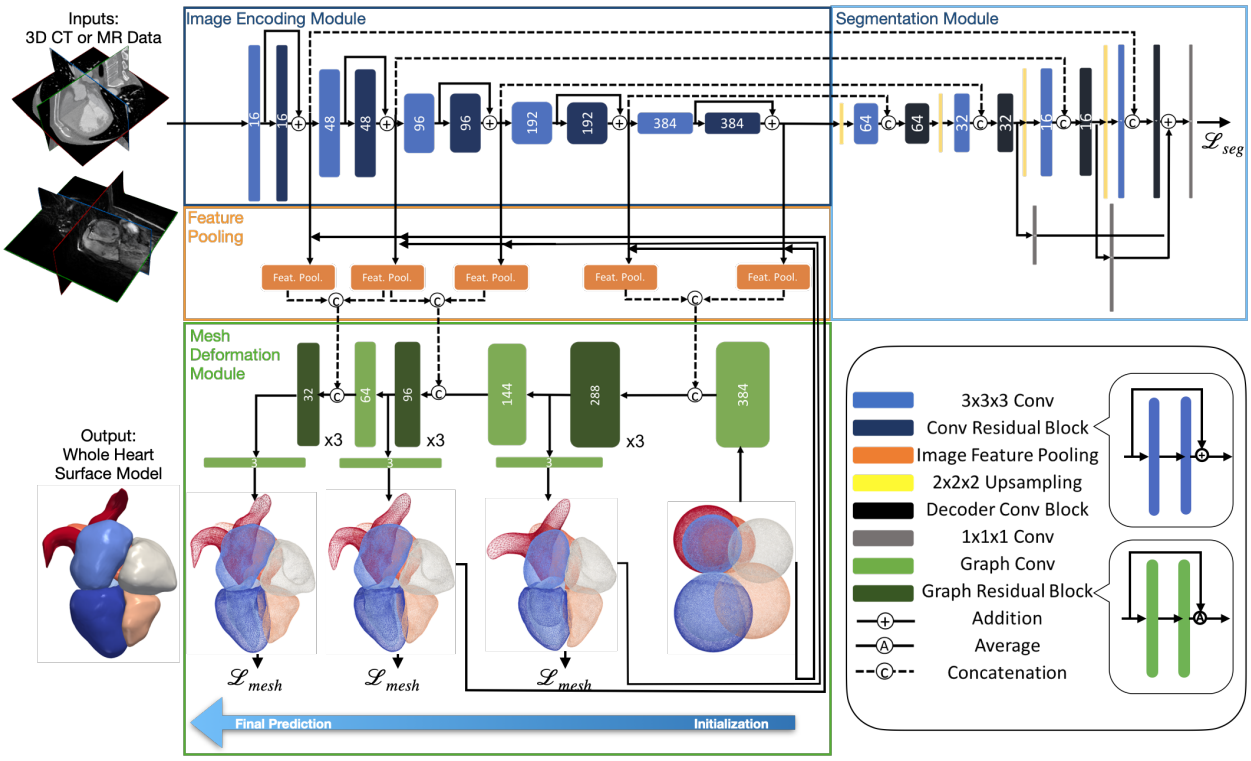


Figure 3.1: Diagram of the proposed automatic whole heart reconstruction approach. The framework uses 3D convolutional layers (shown in blue) to encode image features and predict a binary segmentation map from an input image volume. The corresponding image features are sampled by pooling layers (shown in orange) based on the vertex coordinates of the template mesh. From the combined image and mesh features, graph convolutional layers (shown in green) are then used to predict the deformation of mesh vertices to generate the final mesh predictions

Therefore, the image encoder should both be effective for better geometric reconstruction and be memory-efficient to process a $128 \times 128 \times 128$ volumetric input image in a single pass. Our image feature encoder is based on an improved 3D UNet architecture that was designed to work effectively for large volumetric image data [50]. Briefly, the feature encoder architecture consists of multiple levels of residual blocks that encode increasingly abstract representations of the input. Residual connections are known to facilitate gradient propagation during training and improve generalization [45]. Each residual block contains two $3 \times 3 \times 3$ convolutional layers and a dropout layer before the last convolutional layer. The input to the first convolutional layer is then added to the output of the last one. After each residual block, we use a $3 \times 3 \times 3$ convolutional layer with input stride 2 to reduce the resolution of the feature maps.

Segmentation Module

While our purpose is to reconstruct surface meshes directly from image data, the ground truth segmentation can function as an additional supervision to the network to further facilitate training. From our experiments, including the segmentation module helped avoid non-manifold geometries due to local minimums and thus improve reconstruction accuracy. Since the ground truth mesh is a sparse representation of the cardiac structures compared with a volumetric segmentation, including the segmentation as a dense supervision with skip connections to the image feature encoder can improve gradient propagation to the image encoding module to better interpret the full volumetric input data. However, since we are only interested in reconstructing meshes, rather than predicting segmentations for all cardiac structure, our segmentation module is trained to predict only a binary segmentation representing the occupancy of the heart in the input image. The adopted network architecture is simplified from the decoder architecture of [50] with only a small number of filters in the convolutional layers. Briefly, the segmentation module contains multiple levels of decoder convolutional blocks that correspond to the residual blocks from the image encoding module to reconstruct segmentation from extracted features. Following a $3 \times 3 \times 3$ convolution of the up-sampled intermediate output, a decoder convolutional block concatenates the current output with the corresponding output from the residual blocks of the image encoding module and then uses a $1 \times 1 \times 1$ convolutional layer to process the concatenated features. Binary segmentation predictions were generated from three different levels of the segmentation module and added together to form the final prediction.

Graph Convolution on Mesh

Our neural network uses graph convolutions on a template mesh to predict deformation vectors on its vertices. Unlike for structured data such as images, convolution in the spatial domain is not well defined for manifold structures such as meshes. Therefore, we apply graph convolution in the frequency domain following recent process in graph convolutional neural networks [11, 25]. Briefly, our template mesh is represented by a graph $\mathcal{M} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_i\}_{i=1}^{N}$ is the set of N vertices and $\mathcal{E} = \{e_i\}_{i=1}^{E}$ is the set of E edges that define the connections among mesh vertices. The graph adjacency matrix $A \in \{0,1\}^{N \times N}$ is a sparse matrix that defines the connection between each pair of vertices, with $A_{ij} = 0$ if vertices v_i and v_j are not connected and $A_{ij} = 1$ if the two vertices are connected. The degree matrix D is a diagonal matrix that represents the degree of each vertex, with $D_{ii} = \sum_j A_{ji}$. Therefore, the graph Laplacian matrix is a real and symmetric matrix defined as $L = D - A$, which can then be normalized as $L_{norm} = I - D^{-1/2} A D^{-1/2}$. The normalized Laplacian matrix can be diagonalized by the Fourier basis on graph $U \in \mathbb{R}^{N \times N}$ as $L_{norm} = U \Lambda U^T$. The columns of U are the orthogonal eigenvectors of L and Λ is a diagonal matrix containing the corresponding eigenvalues. The Fourier transform of a function defined on mesh vertices, $f \in L^2(\mathcal{V})$, is thus described by $\hat{f} = U^T f$ and the inverse Fourier transform is $f = U\hat{f}$. Therefore, convolution between f and $g \in L^2(\mathcal{V})$ is described as $f * g = U((U^T f) \odot (U^T g))$.

If we parameterize g with learnable weights, a graph convolution layer can then be defined as $f_{out} = \sigma(U g_\theta(\Lambda) U^T f_{in})$, where f_{in} and f_{out} are the input and output and σ is the ReLU activation function.

The above expression is computationally expensive for meshes with a large number of vertices, since U is not sparse and the number of parameters required can be as many as the number of vertices. Therefore, we followed [25] to approximate $g_\theta(\Lambda)$ using Chebyshev polynomials so that $U g_\theta(\Lambda) U^T = \sum_{k=0}^{K} \theta_k T_k(\tilde{L})$, where $\tilde{L}$ is the scaled sparse Laplacian matrix $\tilde{L} = 2L_{norm}/\lambda_{max} - I$, where λ_{max} is the maximum eigenvalue of L_{norm}. θ_k is the parameter for the kth order Chebyshev polynomial and T_k is the kth order polynomial that can be computed recursively as $T_0 = I$, $T_1 = \tilde{L}$ and $T_k(\tilde{L}) = 2\tilde{L}T_{k-1}(\tilde{L}) - T_{k-2}(\tilde{L})$. We chose $K = 1$ since a lower order polynomial can effectively avoid fitting the noise on our ground truth surfaces and reduce the amount of parameters to learn. Therefore, the graph convolution on the mesh using a first-order Chebyshev polynomial approximation is described as $f_{out} = \sigma(\theta_0 f_{in} + \theta_1 f_{in}\tilde{L})$, where $\theta_0, \theta_1 \in \mathbb{R}^{d_{out} \times d_{in}}$ are trainable weights. $f_{out} \in \mathbb{R}^{d_{out} \times N}$ and $f_{in} \in \mathbb{R}^{d_{in} \times N}$ are, respectively, the input and output feature matrices, where d_{in} and d_{out} are, respectively, the input and output dimensions of the mesh features.

Mesh Initialization

Our method uses a single network to simultaneously deform multiple sphere templates to corresponding cardiac structures on the input image. Since the relative locations and scales of different cardiac structures of the heart are generally consistent across a population, we leverage this prior knowledge into our neural network by scaling and positioning the corresponding initial sphere mesh template based on the relative sizes and locations of the cardiac structures. We then used a graph convolution layer to augment the coordinates of the initial meshes such that they have comparable contribution as the image features, in terms of the length of feature vectors, to the following deformation block. Namely, after pre-processing the volumetric training data and obtaining the corresponding ground truth meshes as described in detail in section 3.2, we computed the corresponding image coordinates of the vertices of the surface meshes in the volumetric training image data. For each cardiac structure, we then computed the average centroid location and the average length between surface and centroid, across all the ground truth meshes in the training data. For each input image, we then used this approximated center and radius to initialize each sphere. By having a closer initialization compared with using centered unit spheres as in prior approaches [168, 159], our network can have reduced distance between predictions and ground truths and thus avoid large deformation during the early phase of training. From our experiments, this is an important and effective technique to avoid getting stuck in local minimums and achieve faster convergence.

Mesh Deformation Module

Our proposed mesh deformation module consists of three deformation blocks with graph convolutional layers that progressively deform our initial template meshes based on both existing mesh vertex features and image features extracted from the image encoding module. Meshes of all different cardiac structures are deformed simultaneously by these shared mesh deformation blocks. The volumetric feature maps have increasing level of abstraction but decreasing spatial resolution as we progress deeper in the image encoding module. Therefore, as shown in Figure 3.1, we used more abstracted, high-level image feature maps for the initial mesh deformation blocks to learn the general shapes of cardiac structures while using low-level, high-resolution feature maps for the later mesh deformation blocks to produce more accurate predictions with detailed features. For each mesh deformation block, we project image features from the image encoding module to the mesh vertices and then concatenate the extracted image feature vector with the existing vertex feature vector. As we deform the mesh through multiple deformation blocks, we decrease the size of the graph convolutional filters to reduce the dimension of mesh feature vectors to match with the reduced number of filters used in upper levels of the image encoding module. Within each mesh deformation block, the concatenated feature vectors are processed by three graph residual blocks, which contains two graph convolutional layers with residual connections. We then use an additional graph convolutional layer to predict deformation as 3D feature vectors on mesh vertices and add those with the vertex coordinates of the initial mesh or the mesh from the previous deformation block to obtain the current predicted vertex coordinates. To project corresponding image features onto mesh vertices, from the vertex locations of the initial or previously deformed mesh, we compute the corresponding image coordinates in the volumetric image feature maps. We then tri-linearly interpolate the feature vectors that correspond to the 8 neighboring voxels of the computed image coordinates in the volumetric feature maps.

Loss functions

The training of our networks was supervised by 3D ground truth meshes of the whole heart as well as a binary segmentation indicating occupancy of the heart on the voxel grid that corresponds to the input image volume. The whole heart meshes were extracted from segmentation of cardiac structures using the marching cube algorithm and the binary segmentation was also obtained from segmentation by setting all non-background voxels to 1 and the rest to 0. We used two categories of loss functions, geometry consistency losses and regularization losses in the training process. The geometry consistency losses include point and normal consistency losses while the regularization losses include edge length and Laplacian losses.

Segmentation loss

We used a hybrid loss function that contained both cross-entropy and dice-score losses. This loss has been used in training UNets and has demonstrated promising results on various medical image segmentation tasks [51]. Namely, let $L_{occupancy}(I_p, I_g)$ denote the loss of between the predicted occupancy probability map I_P and the ground truth binary segmentation of the whole heart I_G. The hybrid loss function was

$$\begin{aligned}\mathcal{L}_{occupancy}(I_P, I_G) = &- \sum_{x \in I_G} \left(I_G(x) \log(I_P(x)) + (1 - I_G(x)) \log(1 - I_P(x))\right) \\ &- \frac{2 \sum_{x \in I} I_G(x) I_P(x)}{\sum_{x \in I} I_G(x) + \sum_{x \in I} I_P(x)}\end{aligned} \tag{3.1}$$

where x denotes the pixel in the input image I.

Point loss

We used Chamfer loss to regulate the accuracy of the vertex locations on predicted meshes. For a point from the predicted mesh or the ground truth mesh, Chamfer loss finds the nearest vertex in the other point set and adds up all pair-wise distances. The point loss is defined by,

$$\mathcal{L}_{point}(\mathbf{P}_i, \mathbf{G}_i) = \sum_{\mathbf{p} \in \mathbf{P}_i} \min_{\mathbf{g} \in \mathbf{G}_i} ||\mathbf{p} - \mathbf{g}||_2^2 + \sum_{\mathbf{g} \in \mathbf{G}_i} \min_{\mathbf{p} \in \mathbf{P}_i} ||\mathbf{p} - \mathbf{g}||_2^2 \tag{3.2}$$

where $\mathbf{p}$ and $\mathbf{g}$ are, respectively, points from the vertex sets of the predicted mesh $\mathbf{P}_i$ and the ground truth mesh $\mathbf{G}_i$ of cardiac structure i.

Normal loss

We used a normal consistency loss to regulate the accuracy of the surface normal on the predicted meshes. For each point, the surface normal is estimated by the cross product between two edges of a face connected to the point. The predicted surface normal is then compared with the ground truth surface normal at the nearest vertex. Namely,

$$\mathcal{L}_{normal}(\mathbf{P}_i, \mathbf{G}_i) = \sum_{\mathbf{p} \in \mathbf{P}_i; \mathbf{g} = \underset{\mathbf{g} \in \mathbf{G}_i}{\arg\min} ||\mathbf{p} - \mathbf{g}||_2^2} \|(\mathbf{p}_1 - \mathbf{p}) \times (\mathbf{p}_2 - \mathbf{p}) - \mathbf{n}_\mathbf{g}\|_2^2 \tag{3.3}$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the two vertices sharing the same face with vertex $\mathbf{p}$.

Edge length loss

We used an edge length loss to encourage a more uniform mesh density on the predictions. That is, we regularize the difference between each edge length and an estimated average edge length μ_i of the corresponding cardiac structure $\mathbf{G}_i$. Namely, we compute the average

surface area of our ground truth mesh for each cardiac structure and estimate the average edge length based on the surface area ratio between the template and ground truth meshes, leading to

$$\mathcal{L}_{edge}(\mathbf{P}_i) = \sum_{\mathbf{p}\in\mathbf{P}_i} \sum_{\mathbf{k}_p\in\mathcal{N}(\mathbf{p})} \left| \|\mathbf{p}-\mathbf{k}_\mathbf{p}\|_2^2 - \mu_i^2 \right| , \tag{3.4}$$

where $\mathcal{N}(\mathbf{p})$ represents the neighborhood of vertex $\mathbf{p}$.

Laplacian loss

To encourage a smoother mesh prediction, we used a Laplacian loss to regularize the difference between a vertex location $\mathbf{p}$ and the mean location of its neighboring vertices $\mathbf{k}_\mathbf{p}$ as

$$\mathcal{L}_{lap}(\mathbf{P}_i) = \sum_{\mathbf{p}\in\mathbf{P}_i} \left\| \mathbf{p} - \sum_{\mathbf{k}_\mathbf{p}\in\mathcal{N}(\mathbf{p})} \frac{1}{||\mathcal{N}(\mathbf{p})||} \mathbf{k}_\mathbf{p} \right\|_2^2 . \tag{3.5}$$

Total loss

Prior approaches of mesh reconstruction from images commonly formulated the total loss function as a weighted sum of multiple loss functions [159, 168]. However, for multi-loss regression problems, different loss functions are different in scales. Manually tuning the weight assigned to each loss function is difficult and expensive since losses can differ by orders of magnitude. Therefore, we express the total loss on predicted meshes as a weighted geometric mean of the individual losses so that the gradient for an individual loss function can be invariant to its scale relative to other loss functions [18]. Thus, for predicted meshes $\mathbf{G}$ and ground truth meshes $\mathbf{P}$ with N cardiac structures, the total mesh loss is expressed as,

$$\mathcal{L}_{mesh}(\mathbf{P},\mathbf{G}) = \sum_{i}^{N} \mathcal{L}_{point}(\mathbf{P}_i,\mathbf{G}_i)^{\lambda_1} \mathcal{L}_{normal}(\mathbf{P}_i,\mathbf{G}_i)^{\lambda_2} \mathcal{L}_{edge}(\mathbf{P}_i)^{\lambda_3} \mathcal{L}_{lap}(\mathbf{P}_i)^{\lambda_4} , \tag{3.6}$$

where each λ is a hyperparameter to weight each individual loss based on its importance without being affected by its scale. We can thus choose hyperparameters from a consistent range for all the losses. We generated 8 sets of random numbers ranging from 0 to 1 and chose the best out of the 8 sets of hyperparameters that produced the smallest point loss on the validation data. The chosen hyperparameters are $\lambda_1 = 0.3$, $\lambda_2 = 0.46$, $\lambda_3 = 0.16$ and $\lambda_4 = 0.05$. For total loss, we added up losses from all three deformation blocks as well as the binary segmentation loss:

$$\mathcal{L}_{total} = \mathcal{L}_{mesh}(\mathbf{P}^{B1},\mathbf{G}) + \mathcal{L}_{mesh}(\mathbf{P}^{B2},\mathbf{G}) + \mathcal{L}_{mesh}(\mathbf{P}^{B3},\mathbf{G}) + \mathcal{L}_{occupancy}(I_p, I_g) . \tag{3.7}$$

The network parameters were computed by minimizing the total loss function using the Adam stochastic gradient descent algorithm [62].

Implementation Details

Image Pre-Processing

Intensity normalization and resizing were applied to all 3D image volumes to obtain consistent image dimensions and pixel intensity range. We followed the procedures in [66] to normalize pixel intensity values of each CT or MR image volume such that they ranged from -1 to 1. The 3D image volumes were then resized using linear interpolation to a dimension of $128 \times 128 \times 128$, which maintained image resolution with a manageable computational cost. The ground truth meshes were generated by applying the Marching Cube algorithm [86] on the segmentations, followed by 50 iterations of Laplacian smoothing.

Image Augmentation

Data augmentation techniques were applied during training to improve the robustness of the neural network models to the variations of input images. Specifically, we applied random scaling (-5% to 5%), random rotation (-5° to 5°), random shearing (-10° to 10°) as well as elastic deformations [140] on the input images. For elastic deformations, 16 control points were placed along each dimension of the 3D image volume and were randomly perturbed. The input images are then warped according to the displacements of the control points using the B-spline interpolation.

Training

The model parameters were computed by minimizing the total loss function using the Adam stochastic gradient descent algorithm [62]. The initial learning rate was set to be 0.001, while β_1 and β_2 for the Adam algorithm were set to 0.9 and 0.999, respectively. Point losses were evaluated on the validation data after each training epoch and the model was saved after one epoch only if the validation point loss had improved. We adopted a learning rate schedule where the learning rate was reduced by 20% if the validation point losses had not improved for 10 epochs. The minimum learning rate was 5×10^{-6}. The network was implemented in TensorFlow and the training was conducted on a Nvidia GeForce GTX 1080 Ti graphics processing unit (GPU) until the validation loss converged.

Evaluation Metrics

We used Dice, Jaccard scores as well as average symmetric surface distance (ASSD) and Hausdorff distance (HD) to evaluate the accuracy of our reconstructions. Dice and Jaccard scores are similarity indices that range from 0 to 1 as given by

$$\text{Dice}(I_P, I_G) = \frac{2|I_P \cap I_G|}{|I_P| + |I_G|} \tag{3.8}$$

$$\text{Jaccard}(I_P, I_G) = \frac{|I_P \cap I_G|}{|I_P \cup I_G|} \tag{3.9}$$

The ASSD and HD measure the average and the largest inconsistency in terms of Euclidean distance between the reconstruction result and the ground truth, respectively. For reconstructed meshes $\mathbf{P}$ and the ground truth meshes $\mathbf{G}$, the ASSD and HD are given by

$$\mathrm{ASSD}(\mathbf{P},\mathbf{G}) = \sum_{\mathbf{p}\in\mathbf{P}} \min_{\mathbf{g}\in\mathbf{G}} \frac{||\mathbf{p}-\mathbf{g}||_2}{|\mathbf{P}|} + \sum_{\mathbf{g}\in\mathbf{G}} \min_{\mathbf{p}\in\mathbf{P}} \frac{||\mathbf{p}-\mathbf{g}||_2}{|\mathbf{G}|} \tag{3.10}$$

$$\mathrm{HD}(\mathbf{P},\mathbf{G}) = \max\left\{ \max_{\mathbf{p}\in\mathbf{P}} \min_{\mathbf{g}\in\mathbf{G}} ||\mathbf{p}-\mathbf{g}||_2, \max_{\mathbf{g}\in\mathbf{G}} \min_{\mathbf{p}\in\mathbf{P}} ||\mathbf{p}-\mathbf{g}||_2 \right\} \tag{3.11}$$

Normal discrepancy between the reconstruction result and the ground truth was evaluated by an average normal error (ANE). Namely, for $\mathbf{n}_x$, $\mathbf{n}_y$ being the vertex normals at points $\mathbf{x}$ and $\mathbf{y}$, respectively,

$$\mathrm{ANE}(\mathbf{P},\mathbf{G}) = \sum_{\mathbf{p}\in\mathbf{P};\mathbf{g}=\underset{\mathbf{g}\in\mathbf{G}}{\arg\min}\,||\mathbf{p}-\mathbf{g}||_2} \frac{1-\langle \mathbf{n}_p, \mathbf{n}_g\rangle}{|\mathbf{P}|} \tag{3.12}$$

Surface smoothness was evaluated by the average normalized Laplacian distance (ANLD). ANLD measures the Euclidean distances between the coordinates of mesh vertices and the mean coordinates of their neighbours, normalized by the average edge length between the mesh vertices and their neighbours. Namely,

$$\mathrm{ANLD}(\mathbf{P}) = \sum_{\mathbf{p}\in\mathbf{P}} \frac{\left\| \mathbf{p} - \sum_{\mathbf{k_p}\in\mathcal{N}(\mathbf{p})} \frac{1}{|\mathcal{N}(\mathbf{p})|}\mathbf{k_p} \right\|_2}{\frac{|P|}{|\mathcal{N}(\mathbf{p})|} \sum_{\mathbf{k_p}\in\mathcal{N}(\mathbf{p})} \|\mathbf{p}-\mathbf{k_p}\|_2} \,. \tag{3.13}$$

The percentage mesh self-intersection was calculated as the percentage of intersected mesh facets among all mesh facets. The intersected mesh facets were detected by TetGen [139].

3.3 Experiments and Results

Baselines

We considered the following three baselines to compare our method against: 2D UNet [128], a residual 3D UNet [50] and Voxel2Mesh [168]. The UNets are arguably the most successful architecture for medical image segmentation and thus can function as strong baselines. In particular, the 2D UNet is a part of the whole-heart segmentation framework implemented in [66] that recently demonstrated state-of-the-art performance on the MMWHS challenge dataset. The residual 3D UNet has demonstrated improved performance than a regular 3D UNet and won the KiTS2019 Challenge [49, 46]. To ensure a fair comparison, the same network architecture and convolutional filter numbers were used for the image encoding module between our method and the residual 3D UNet and the same image pre-processing

and augmentation methods were applied during the training of all methods. For Voxel2Mesh, we reduced the resolution of the template mesh such that the total memory consumption during training can fit within the memory available on our Nvidia GeForce GTX 1080 Ti GPU (11 GB). The final mesh resolution is thus halved compared to the original implementation [168] and contains 3663 vertices for each cardiac structures. In contrast, our method can process a template mesh with 11494 mesh vertices for each cardiac structures within the available GPU memory.

Whole Heart Reconstruction for CT and MR images

We first compare the performance of whole-heart reconstruction from our method against our baselines. In this experiment, we trained and validated our method using both CT and MR images collected from existing public datasets except for the held-out test dataset of the MMWHS challenge, which we used for test-time evaluation. Our training set thus contained 87 CT images and 41 MR images and the validation set contained 15 CT images and 6 MR images. The MMWHS held-out test dataset contained 40 CT images and 40 MR images. We analyzed the performance of our method against baselines in terms of both the accuracy and the quality of the surface reconstructions. We converted the surface predictions of our method and those of Voxel2Mesh to segmentations at the spatial resolution of the input image data, which is the same as the resolution of the segmentations produced by 2D UNet and 3D UNet. This allowed us to evaluate the accuracy of different methods at the same resolution against the ground truth segmentation using the executable provided by the MMWHS challenge organizers. We also manually labeled the testing images and compared this with the ground truth segmentation of the MMWHS challenge to provide a comparison between the evaluated reconstruction accuracy of our deep-learning-based method and the inter-observer variability in manual delineations. The surface quality was evaluated in terms of surface smoothness, normal consistency and topological correctness.

Table 3.2 shows the average Dice and Jaccard scores, average symmetric surface distance (ASSD) and Hausdorff distance (HD) of the reconstruction results of both the whole heart and individual cardiac structures for the MMWHS test dataset. For both CT and MR data, our method consistently outperformed our baselines in terms of Dice and Jaccard scores for both whole heart and all individual cardiac structures. In terms of surface ASSD and HD measures for the whole heart or individual cardiac structures, our method was the best or the second among the four deep-learning-based methods compared. To provide further details on segmentation accuracy, Figure 3.2 gives the distribution of different segmentation accuracy metrics for whole heart and individual cardiac structures. Overall, our method demonstrated advantages of whole heart reconstruction for both CT and MR images, and 2D UNet was the closest to ours compared with 3D UNet or Voxel2Mesh. All methods produced better reconstruction for CT images than for MR images. Furthermore, there are no significant differences between the evaluated Dice scores of our methods and those of our manual labeling, except for left ventricle epicardium ($p<0.05$). That is, the discrepancy between our predicted whole-heart reconstruction and the ground truths provided by the

			Epi	LA	LV	RA	RV	Ao	PA	WH
CT	Dice (↑)	Ours	**0.899**	**0.932**	**0.940**	**0.892**	**0.910**	**0.950**	**0.852**	**0.918**
		2DUNet	**0.899**	0.931	0.931	0.877	0.905	0.934	0.832	0.911
		3DUNet	0.863	0.902	0.923	0.868	0.876	0.923	0.813	0.888
		Voxel2Mesh	0.775	0.888	0.910	0.857	0.885	0.874	0.758	0.865
		Manual	0.919	0.938	0.941	0.894	0.917	0.955	0.854	0.925
	Jaccard (↑)	Ours	**0.819**	**0.875**	**0.888**	**0.809**	**0.837**	**0.905**	**0.755**	**0.849**
		2DUNet	0.817	0.872	0.873	0.787	0.828	0.879	0.726	0.837
		3DUNet	0.762	0.825	0.861	0.769	0.783	0.860	0.695	0.799
		Voxel2Mesh	0.638	0.801	0.839	0.754	0.795	0.778	0.619	0.763
		Manual	0.852	0.884	0.890	0.814	0.848	0.914	0.759	0.860
	ASSD (mm) (↓)	Ours	1.335	**1.042**	**0.842**	**1.583**	1.176	**0.531**	1.904	1.213
		2DUNet	**0.808**	1.049	0.905	1.719	**1.064**	0.645	**1.551**	**1.088**
		3DUNet	1.443	1.528	1.024	1.943	1.663	0.814	2.194	1.552
		Voxel2Mesh	1.714	1.696	1.266	2.020	1.492	1.341	3.398	1.848
		Manual	1.437	0.936	0.815	1.541	0.983	0.480	1.455	1.106
	HD (mm) (↓)	Ours	14.393	10.407	10.325	13.639	13.360	**9.407**	**26.616**	28.035
		2DUNet	**9.980**	8.773	**6.098**	13.624	10.016	10.013	27.834	28.727
		3DUNet	13.635	10.814	9.580	16.031	15.635	13.326	26.941	31.088
		Voxel2Mesh	13.564	**8.743**	6.248	**12.116**	**9.601**	12.080	26.252	**27.459**
		Manual	14.446	12.677	12.619	15.313	13.496	11.189	25.449	27.181
MR	Dice (↑)	Ours	**0.797**	**0.881**	**0.922**	**0.888**	**0.892**	**0.890**	**0.816**	**0.882**
		2DUNet	0.795	0.864	0.896	0.852	0.865	0.869	0.772	0.859
		3DUNet	0.761	0.852	0.879	0.866	0.828	0.742	0.764	0.840
		Voxel2Mesh	0.602	0.734	0.852	0.774	0.830	0.700	0.506	0.766
		Manual	0.830	0.885	0.925	0.887	0.894	0.885	0.807	0.887
	Jaccard (↑)	Ours	**0.671**	**0.791**	**0.858**	**0.801**	**0.812**	**0.805**	**0.697**	**0.790**
		2DUNet	0.668	0.765	0.817	0.752	0.771	0.774	0.641	0.757
		3DUNet	0.626	0.756	0.802	0.766	0.728	0.650	0.639	0.732
		Voxel2Mesh	0.443	0.584	0.752	0.635	0.721	0.552	0.352	0.626
		Manual	0.713	0.797	0.862	0.799	0.812	0.798	0.681	0.798
	ASSD (mm) (↓)	Ours	2.198	**1.401**	**1.183**	**1.611**	**1.333**	2.648	2.689	1.775
		2DUNet	**1.830**	1.488	1.455	1.715	1.483	**2.447**	**1.820**	**1.690**
		3DUNet	2.175	2.503	1.836	1.890	2.871	4.092	1.952	2.037
		Voxel2Mesh	2.505	3.365	2.506	3.475	2.233	4.614	6.078	3.359
		Manual	1.837	1.301	1.070	1.463	1.218	2.159	1.581	1.485
	HD (mm) (↓)	Ours	**16.923**	11.723	10.891	14.810	13.463	**22.219**	19.345	**27.701**
		2DUNet	19.139	**10.781**	**9.958**	**14.530**	**13.082**	22.567	**16.721**	28.350
		3DUNet	28.159	23.640	21.494	18.949	21.095	37.937	17.055	43.022
		Voxel2Mesh	20.156	13.416	10.301	15.796	11.672	27.806	26.464	33.020
		Manual	15.854	12.444	12.125	14.376	13.145	21.783	13.754	25.336

Table 3.2: A comparison of prediction accuracy on MMWHS MR and CT test datasets from different methods.

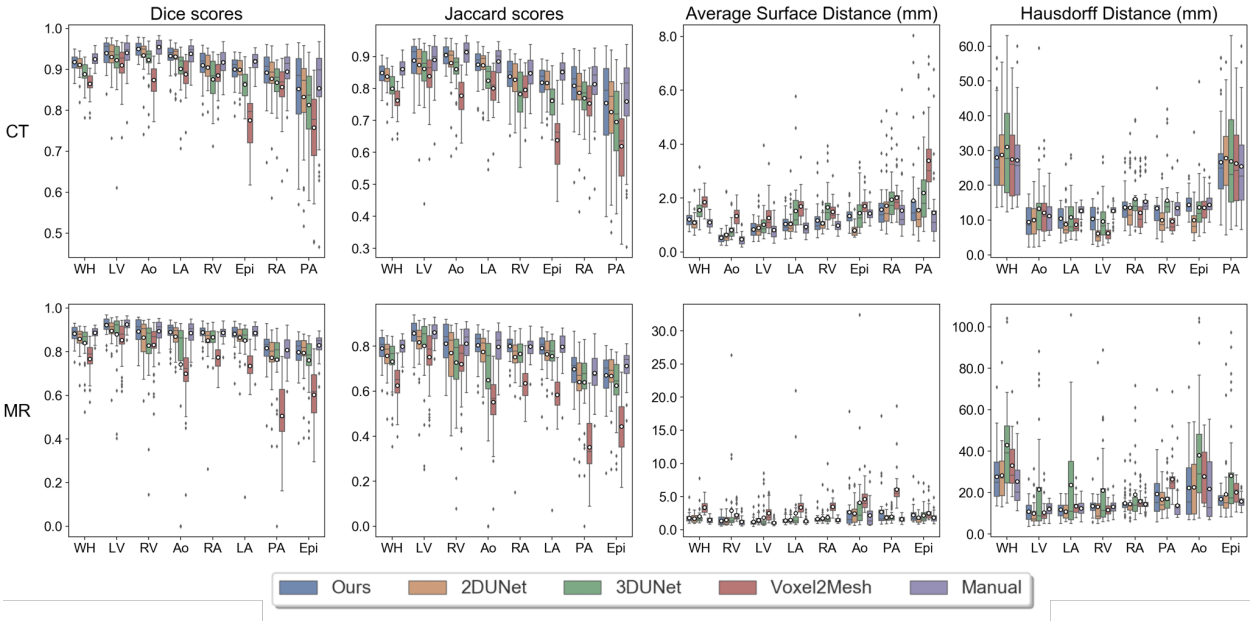


Figure 3.2: Comparison of segmentation accuracy for whole heart and individual cardiac structures from different methods. White circles on the boxes indicate mean values across patients. Cardiac structures are sorted based on the accuracy of our method.

MMWHS challenge is comparable to the inter-observer variability of manual whole-heart segmentation.

Figure 3.3 displays two examples of the reconstruction results for CT and MR from the MMWHS test dataset, including the surface meshes of individual cardiac structures. Despite starting from a generic template, our method is able to accurately map a template sphere to various cardiac structures with disparate shapes such as the left ventricle epicardium and the pulmonary artery. Moreover, we are able to generate smooth surface reconstruction with consistent normal while capturing the details of individual cardiac structures such as mitral annulus on the left ventricle epicardium, aortic outlet of the left ventricle and the aortic sinus.

Figure 3.4 and 3.5 visualize the median and worst results from the different methods for CT and MR images, respectively, from the MMWHS test dataset. The surface meshes of 2D UNet and 3D UNet were extracted from the segmentation results using the marching cube algorithm. As shown, our method is able to construct smooth geometries while segmentation based methods, such as 2D UNet or 3D UNet, produced surfaces with staircase artifacts. Such artifacts require surface post-processing techniques such as Laplacian smoothing that often also degrade true features. Generally, all four methods are able to produce reasonable median cases from CT data. For MR data, our method produced reasonable reconstructions, while the 2D UNet and 3D UNet produced reconstructions with

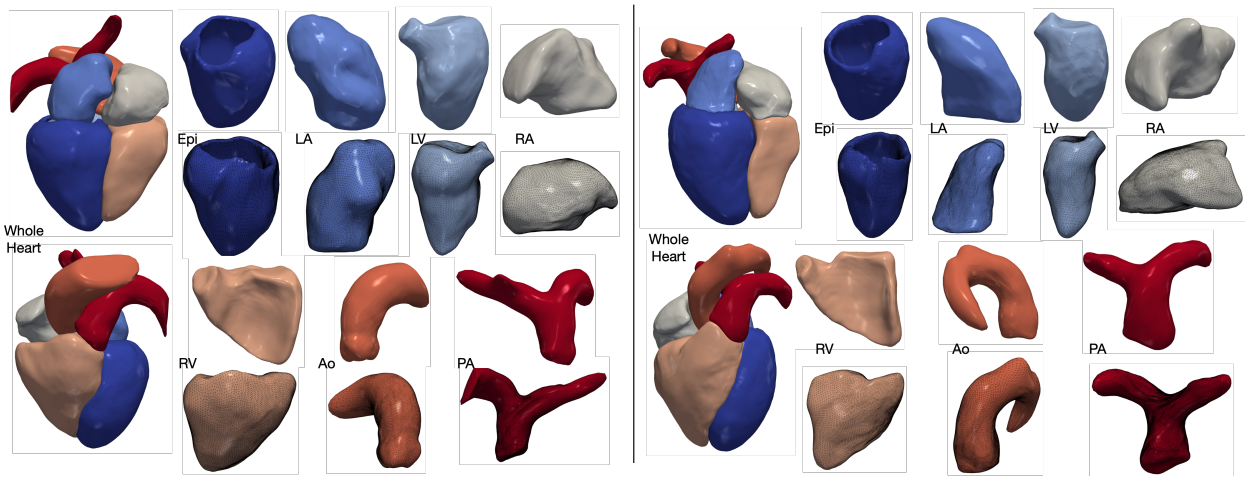


Figure 3.3: Example reconstructions from our method for CT (left) and MR (right) data selected from MMWHS test dataset. Our method reconstructs the whole heart consisting of seven cardiac structures, including the four heart chambers, left ventricle epicardium, aorta and pulmonary arteries. Geometry of each reconstructed cardiac structure is demonstrated in two different views, with the bottom view also displaying the meshes.

disconnected regions that would require post-processing to remove or connect. Voxel2Mesh was unable to capture detailed shapes of some structures such as the bifurcation of the pulmonary artery branches. In the worst cases for both CT and MR, our method nonetheless produced realistic shapes. However, 2D UNet and 3D UNet predicted geometries with missing parts, noisy surfaces, incorrect classifications and/or disconnected regions that would require significant post-processing. Voxel2Mesh predicted worst-case geometries that deviated largely from ground truths and had major surface artifacts. To provide quantitative comparison on the surface quality produced by different methods, Table 3.3 displays average normal error (ANE), average normalized Laplacian distance (ANLD), and percentage mesh self-intersection of the reconstruction results. Overall, our method demonstrated the best surface smoothness and normal consistency for all cardiac structures for CT data and for most cardiac structures for MR data. For topology correctness, our method produced meshes with a small number of self-intersections. In contrast, the segmentation-based approaches apply the Marching Cube algorithm to generate uniform and watertight surface meshes without self-intersection.

Figures 3.6 and 3.7 provide further qualitative comparisons of the results from the different methods. As shown in Fig. 3.7, our method was able to generate smoother reconstruction than the ground truth segmentation on MR images that have relatively large voxel spacing. In contrast, 2D UNet that produces segmentation on a slice-by-slice manner along the sagittal view, may suffer from inconsistency between adjacent slices, leading to coarse segmentation when looking from the axial view that the 2D UNet was not trained on. 3D UNet, limited

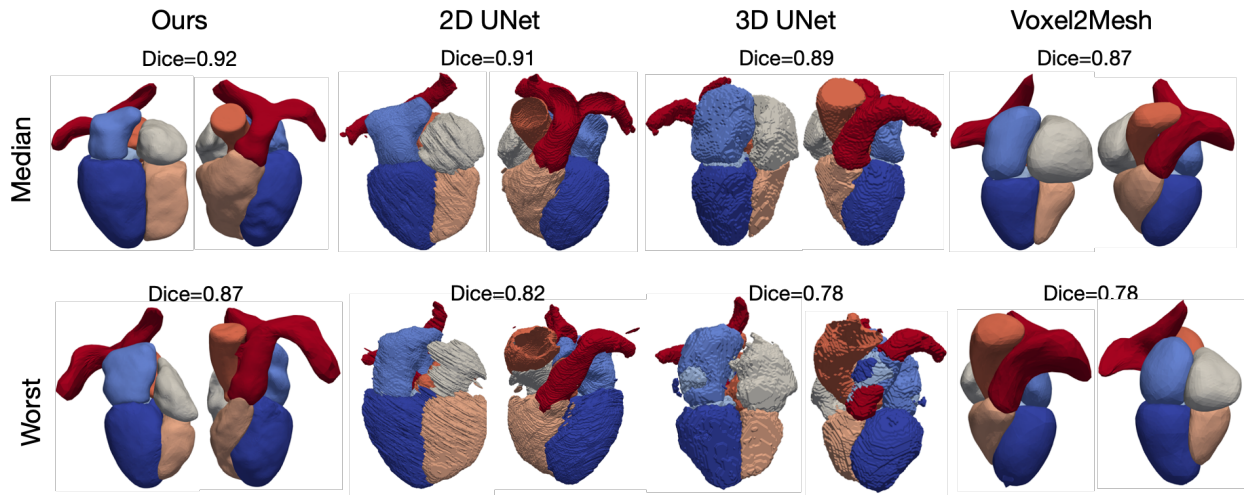


Figure 3.4: Visualizations of the median and worst reconstruction results among the MMWHS CT test dataset in terms of whole-heart Dice scores for all compared methods.

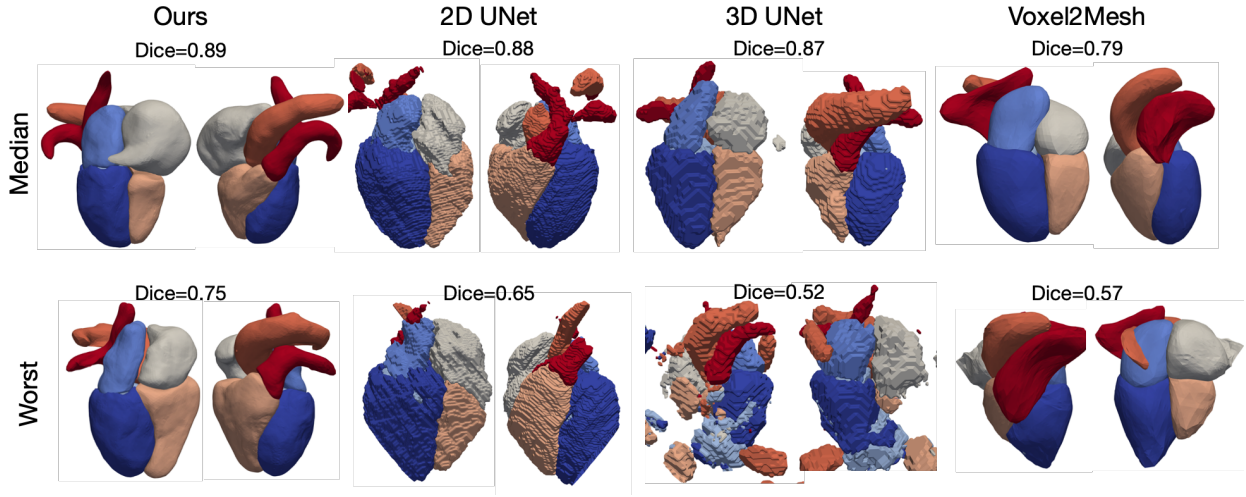


Figure 3.5: Visualizations of the median and worst reconstruction results among the MMWHS MR test dataset in terms of whole-heart Dice scores for all compared methods.

by the memory constrain of GPU, can only produce coarse segmentation on a down-sampled voxel grid of $128 \times 128 \times 128$ for high-resolution CT image data. Although Voxel2Mesh can also produce smooth surface meshes, it tends to predict surfaces that lack shape details and do not match well with the true boundary of many cardiac structures.

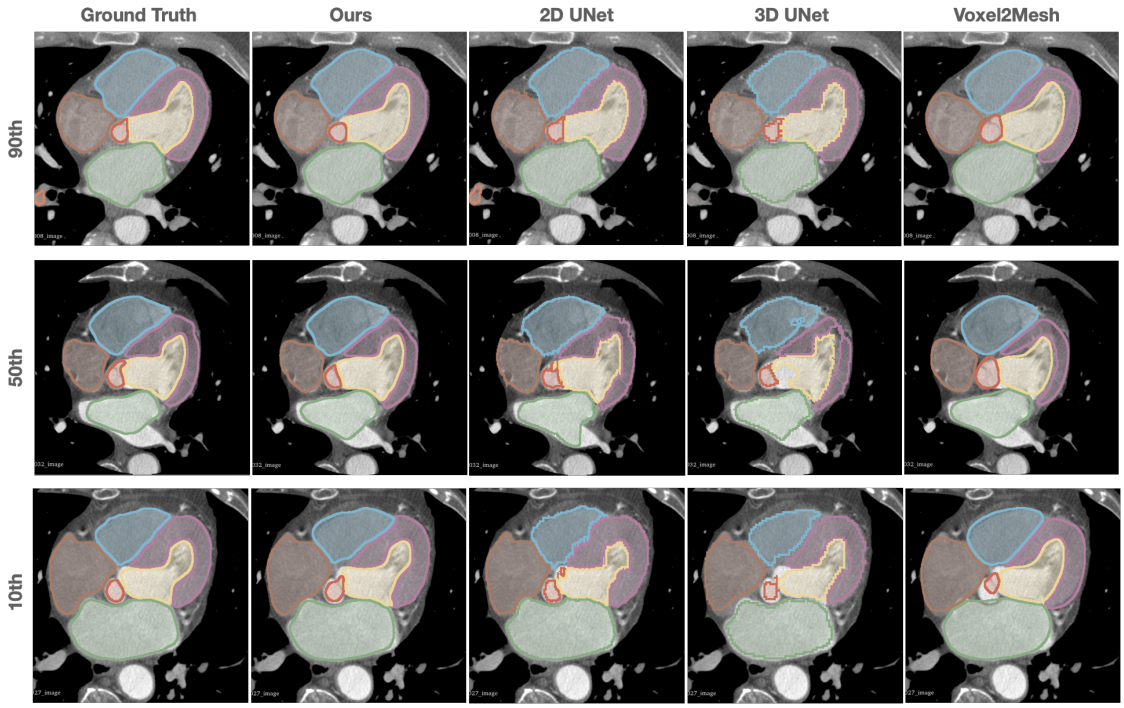
Figure 3.8 shows reconstruction results for the 10 most challenging CT and MR images

			Epi	LA	LV	RA	RV	Ao	PA
CT	ANE (↓)	Ours	**0.004**	**0.003**	**0.003**	**0.006**	**0.006**	**0.002**	**0.008**
		2DUNet	0.036	0.012	0.014	0.022	0.030	0.010	0.023
		3DUNet	0.033	0.015	0.014	0.017	0.021	0.010	0.016
		Voxel2Mesh	0.136	0.171	0.129	0.150	0.136	0.143	0.105
	ANLD (↓)	Ours	**0.091**	**0.078**	**0.085**	**0.085**	**0.080**	**0.076**	**0.090**
		2DUNet	0.287	0.280	0.287	0.282	0.286	0.265	0.278
		3DUNet	0.292	0.284	0.295	0.295	0.290	0.273	0.291
		Voxel2Mesh	0.113	0.119	0.129	0.134	0.126	0.140	0.160
	Intersection (%) (↓)	Ours	0.014	0.006	0.017	0.007	0.024	0.005	0.049
		2DUNet	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
		3DUNet	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
		Voxel2Mesh	0.269	**0.000**	**0.000**	0.003	**0.000**	**0.000**	0.020
MR	ANE (↓)	Ours	**0.015**	**0.012**	**0.007**	**0.010**	**0.013**	**0.015**	0.017
		2DUNet	0.057	0.016	0.022	0.024	0.033	0.026	0.018
		3DUNet	0.056	0.017	0.035	0.020	0.034	0.037	**0.014**
		Voxel2Mesh	0.104	0.189	0.130	0.136	0.150	0.123	0.160
	ANLD (↓)	Ours	**0.103**	**0.093**	**0.088**	**0.101**	**0.092**	**0.088**	**0.103**
		2DUNet	0.287	0.274	0.285	0.282	0.276	0.275	0.289
		3DUNet	0.296	0.283	0.299	0.297	0.288	0.296	0.304
		Voxel2Mesh	0.130	0.132	0.129	0.144	0.139	0.155	0.159
	Intersection (%) (↓)	Ours	0.069	0.018	0.023	0.069	0.069	0.108	0.134
		2DUNet	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
		3DUNet	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
		Voxel2Mesh	0.189	**0.000**	**0.000**	0.070	**0.000**	0.059	0.020

Table 3.3: A comparison of the quality of the whole heart surfaces from different methods on MMWHS MR and CT test datasets.

for which 2D UNet (the method that demonstrated closest performance to our method) predicted less accurate segmentations in terms of Dice scores compared with the rest images in the test datasets. For all the 10 MR images and 8 out of the 10 CT images, our method produced whole-heart reconstructions with improved Dice scores. For all these CT cases, we were able to generate accurate reconstruction with Dice scores above 0.87 and smooth surfaces without obvious artifacts. However, for the 10 MR cases, although we demonstrated improvement against 2D UNet predictions, we observed buckling and bumpiness on mesh surfaces of one or more cardiac structures for 5 out of 10 cases.

Interestingly, as indicated by the point-correspondence color maps in Figure 3.8, although we did not explicitly train our method to generate feature-corresponding meshes across different input images, our method was generally able to consistently deform template meshes to map mesh vertices to similar structural features of the heart for different images. This behavior allowed convenient generation of the mean whole heart shapes from the test dataset by computing the average coordinates of each vertex. Figure 3.9 demonstrates the mean



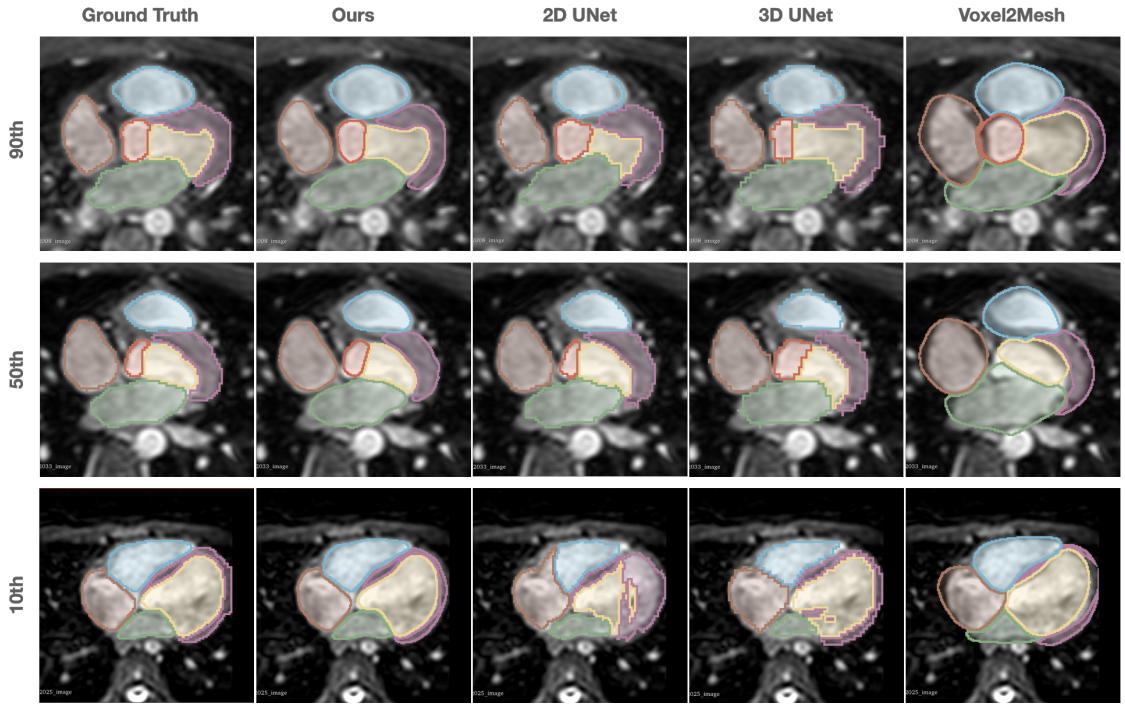
images and predictions from different test cases with the 10th, 50th, 90th percentiles of Dice scores based on our method.

Figure 3.6: Comparison of the predicted whole heart surfaces from different methods for CT test cases. Different rows demonstrated the zoomed-in axial view of the

whole heart shapes for MR and CT images from the MMWHS test dataset, respectively, and the distribution of the average surface distance errors on the whole heart compared with manual ground truths. For both CT and MR data, locations that suffer from higher surface errors include the ends of the aorta and pulmonary arteries, boundaries between the right ventricle and the pulmonary artery, boundaries between the right atrium and the ventricle, and the inferior vena cava region on the right atrium. We note that several of the locations of largest error are artificial boundaries, or arbitrary truncations of vessels extending away form the heart.

Generalization to Low-Resolution MR Images

Cardiac MR image data are often acquired in a slice-by-slice manner and thus often vary in through-plane resolution due to the use of different acquisition protocols and vendors. For MR images with low through-plane resolution, accurately constructing smooth surface geometries is challenging since a method would need to complete the cardiac structures that are not captured between the slices. Therefore, having trained our method on MR images with high through-plane resolution to produce detailed whole heart geometries, we



images and predictions from different test cases with the 10th, 50th, 90th percentiles of Dice scores based on our method.

Figure 3.7: Comparison of the predicted whole heart surfaces from different methods for MR test cases. Different rows demonstrated the zoomed-in axial view of the

evaluate the performance of our method on MR images with lower through-plane resolution and compare it with our baselines. To disentangle the effect of through-plane resolution from the effect of other variations of MR images, we first generate low-resolution MR images from our validation data by down-sampling the images to various slice thicknesses. We then evaluate the robustness of different methods to challenging real low-resolution MR images that significantly differ from our training datasets. Namely, we used data from our cine MR images, which were acquired with large slice thicknesses (8-10 mm), different acquisition planes, and from a different clinical center.

Synthetic Low-resolution MR data

Figure 3.10 displays an example of down-sampling an input image dataset along the longitudinal direction of the left ventricle to various slice thickness of 1 mm, 6 mm, and 10 mm, as well as the corresponding predictions from our method, 2D UNet and 3D UNet, respectively. For low through-plane resolution images, the same linear resampling method was applied as before to interpolate the 3D image volume to the sizes required by the neural network models. As the slice thickness was increased to up to 10 mm, while 2DUNet can generally

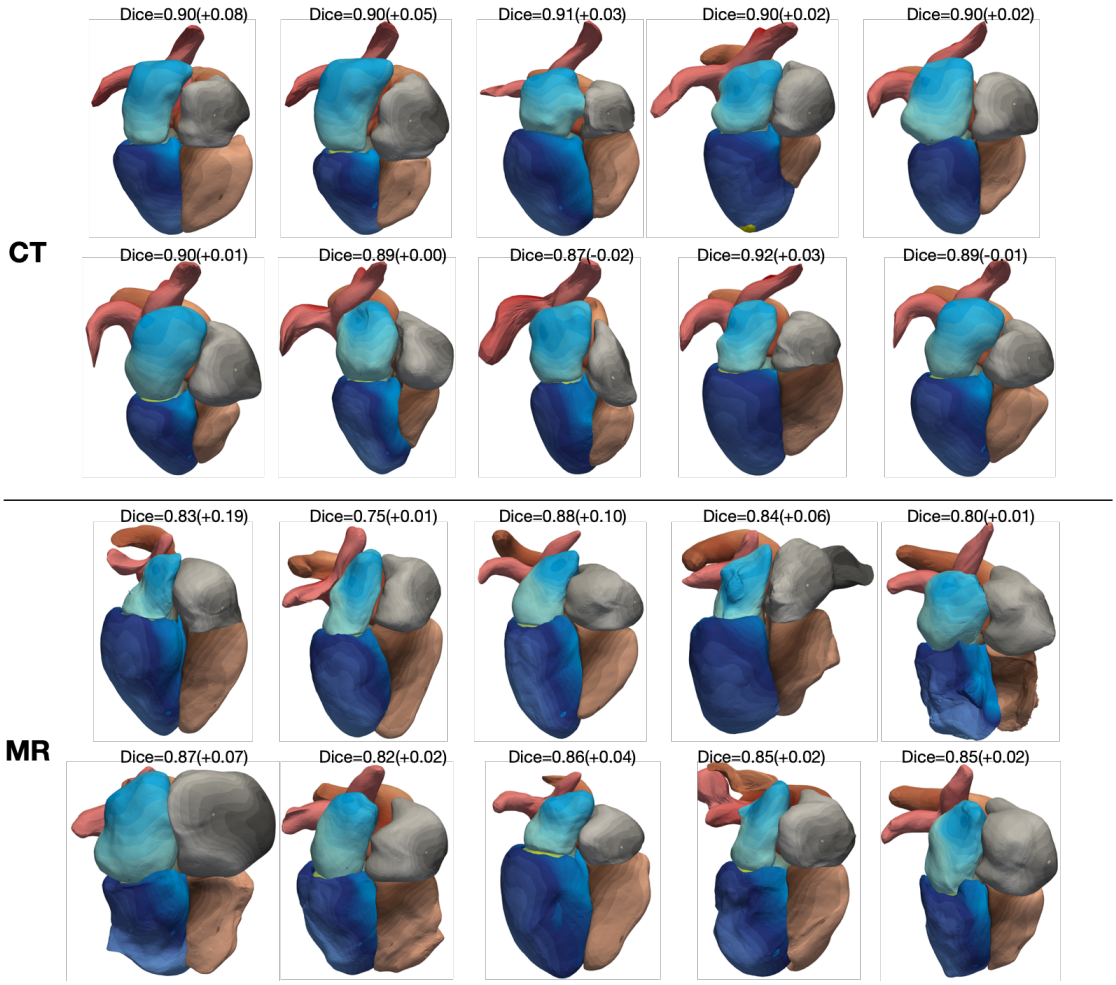


Figure 3.8: Whole heart reconstruction results from the 10 most challenging CT and MR images for which 2D UNet predicted less accurate segmentations in terms of Dice scores compared with the rest images in the MMWHS test datasets. On top of each case is the whole-heart Dice score of our result and the difference in whole-heart Dice score compared with 2D UNet reconstruction. The color map denotes the indices of mesh vertices and demonstrates the correspondence of mesh vertices across reconstructed meshes from different images.

produce consistent segmentation on 2D slices, it produces uneven 3D geometries due to poor inter-slice consistency. In contrast, the 3D UNet is able to produce smoother surfaces by accounting for inter-slice information. However, as slice thickness increases, the 3D UNet produces less accurate segmentation, such as incorrectly classifying a part of the RV into the RA and a part of the PA into the aorta, as shown by the arrows in Figure 3.10. Our method, however, for all different slice thicknesses, produces consistent reconstructions that

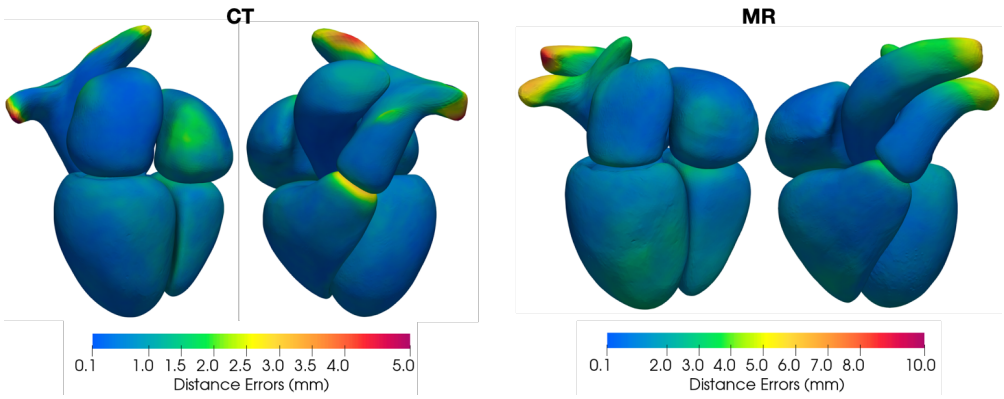


Figure 3.9: Distribution of the average surface distance errors on mean whole heart shapes from the CT and MR data in MMWHS test dataset.

closely resembles the ground truth surfaces and are free of any major artifacts. Figure 3.11 displays quantitative evaluations of the reconstruction performance on various image resolutions. Regardless of slice thickness values considered, our method out-performed 2D UNet and 3D UNet both in terms of Dice and ASSD. Moreover, as slice thickness increases from 1 mm to 10 mm, in general, we observed increasing improvement of our method compared with 2D UNet or 3D UNet. Furthermore, by taking a 3D image volume as the input, our method and 3D UNet are more robust to additional in-plane resolution changes than the 2D UNet. Both our method and the 3D UNet demonstrated a smaller reduction in accuracy with 4 times reduction of in-plane resolution.

Real Low-resolution MR data

We evaluated the robustness of our method on the challenging cine MR dataset, which significantly differs from our training datasets in terms of the through-plane resolution, imaging plane orientation and coverage of the heart. To generate ground truth segmentation and meshes from low-resolution MR data, we re-sampled such 3D image volume and linearly interpolated between the slices to have an isotropic spacing of 1 mm along all three axes. The ground truth segmentations were obtained by manually segmenting the interpolated image data and manually correcting artifacts due to low through-plane resolution based on prior human expert knowledge of the heart to obtain smooth and physiologically plausible geometries that match with the low-resolution image data as much as possible. Table 3.4 compares the reconstruction accuracy between our method and the baselines. The reconstruction accuracy was evaluated at two time frames, end diastole and end systole, for each patient. Overall, our method demonstrated high reconstruction accuracy and outperformed the other methods for most cardiac structures in terms of average Dice score and ASSD.

Figure 3.12 compares the whole-heart geometries reconstructed by our method with others for one example of cine cardiac MR images. Our method was able to produce clean surface

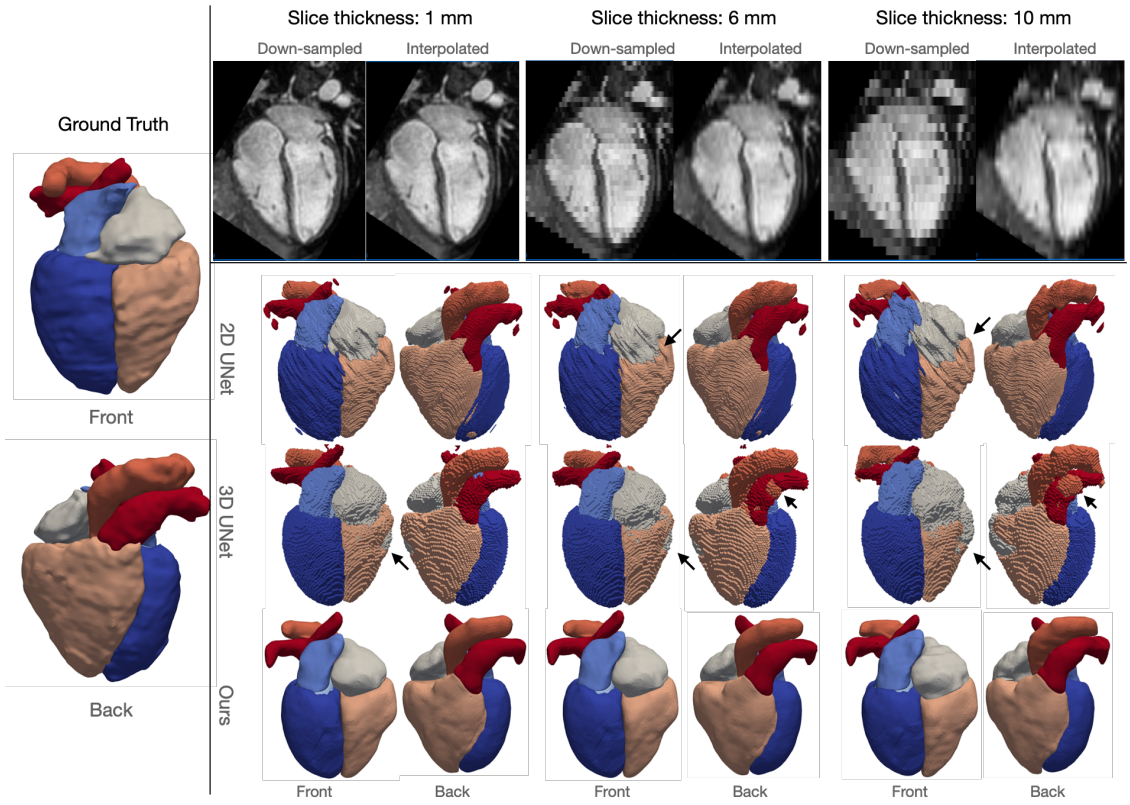


Figure 3.10: Robustness of different methods to through-plane resolution changes of MR images. Left panel shows the front and back views of the ground truth surfaces; top panel shows example slices along the down-sampling axis of images down-sampled to varying slice thicknesses, and bottom panel shows front and back views of predicted whole-heart surfaces from different methods corresponding to different slice thickness values.

meshes while at the same capture most of the cardiac structures with reasonable accuracy. In contrast, since these images were acquired on imaging planes that were different from those used in acquiring the training data, 2D UNet produced inaccurate reconstruction and disconnected surfaces. 3D UNet produces more complete reconstruction of the cardiac structures but often produced many disconnected false positive regions. Voxel2Mesh is able to produce clean surface meshes with generally correct topology but the predictions are not accurate. Furthermore, as changes in input images over different time frames are small, our method produced consistent reconstruction over different time phases. However, segentation-based methods, 2D UNet or 3D UNet, often produce inconsistent reconstruction with significant shape or topology changes, despite small changes in input images over different time frames.

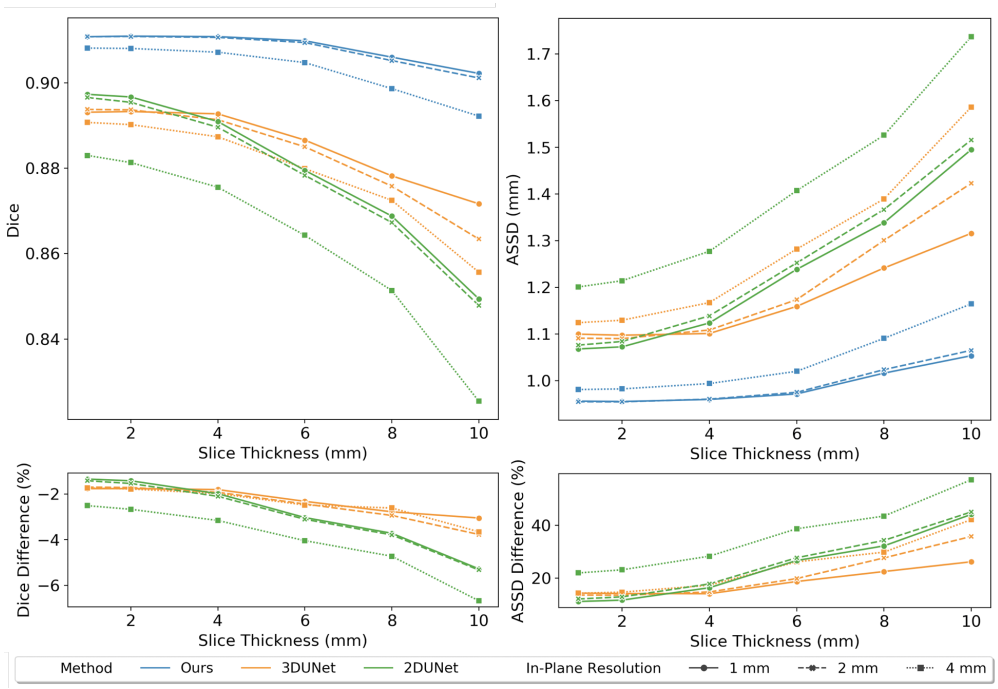


Figure 3.11: Relation of Dice and ASSD values of whole-heart surfaces to through-plane resolution of MR images. Comparison between different methods and different in-plane resolutions are indicated by lines with different color and different styles, respectively. The bottom panel shows the average percentage differences of Dice or ASSD values between our method and 2D UNet or 3D UNet across all validation images.

		Epi	LA	LV	RA	RV	Ao	PA	WH
Dice (↑)	Ours	**0.656±0.169**	**0.708±0.187**	**0.822±0.104**	**0.672±0.114**	0.643±0.228	**0.543±0.255**	0.445±0.225	**0.693±0.112**
	2D UNet	0.543±0.263	0.517±0.283	0.734±0.218	0.274±0.218	**0.644±0.184**	0.393±0.215	**0.487±0.286**	0.598±0.166
	3D UNet	0.546±0.244	0.702±0.22	0.782±0.134	0.598±0.169	0.631±0.144	0.495±0.175	0.285±0.249	0.627±0.131
	Voxel2Mesh	0.438±0.178	0.529±0.275	0.669±0.135	0.54±0.206	0.598±0.273	0.395±0.246	0.223±0.195	0.527±0.167
ASSD (mm) (↓)	Ours	4.009±1.118	**4.775±2.522**	4.534±2.195	**5.299±1.883**	5.468±1.856	**6.713±3.233**	**7.463±3.14**	**5.466±1.613**
	2D UNet	4.585±3.501	6.665±5.147	5.204±3.3	10.638±6.918	**4.12±2.493**	8.36±7.738	7.914±9.257	6.784±3.951
	3D UNet	**3.498±2.47**	4.841±5.061	**3.228±2.945**	8.537±5.393	5.234±2.466	10.022±6.599	11.643±8.608	6.715±3.091
	Voxel2Mesh	5.104±1.767	7.105±3.082	6.763±2.528	6.945±3.163	7.775±4.613	9.181±4.593	12.079±7.703	7.85±2.881

Table 3.4: A comparison of prediction accuracy on cine MR dataset from different methods. All accuracy measures are represented by mean ± standard deviation, which are computed over different patients and time frames.

Construction of Whole-Heart 4D Models from Motion Image Data

We further tested our method on time-series CT datasets. Table 3.5 compares the reconstruction accuracy between our method and the other baseline methods. Similar to above, the reconstruction accuracy was evaluated at two time frames, end diastole and end systole, for each patient. Overall, our method demonstrated high reconstruction accuracy and out-

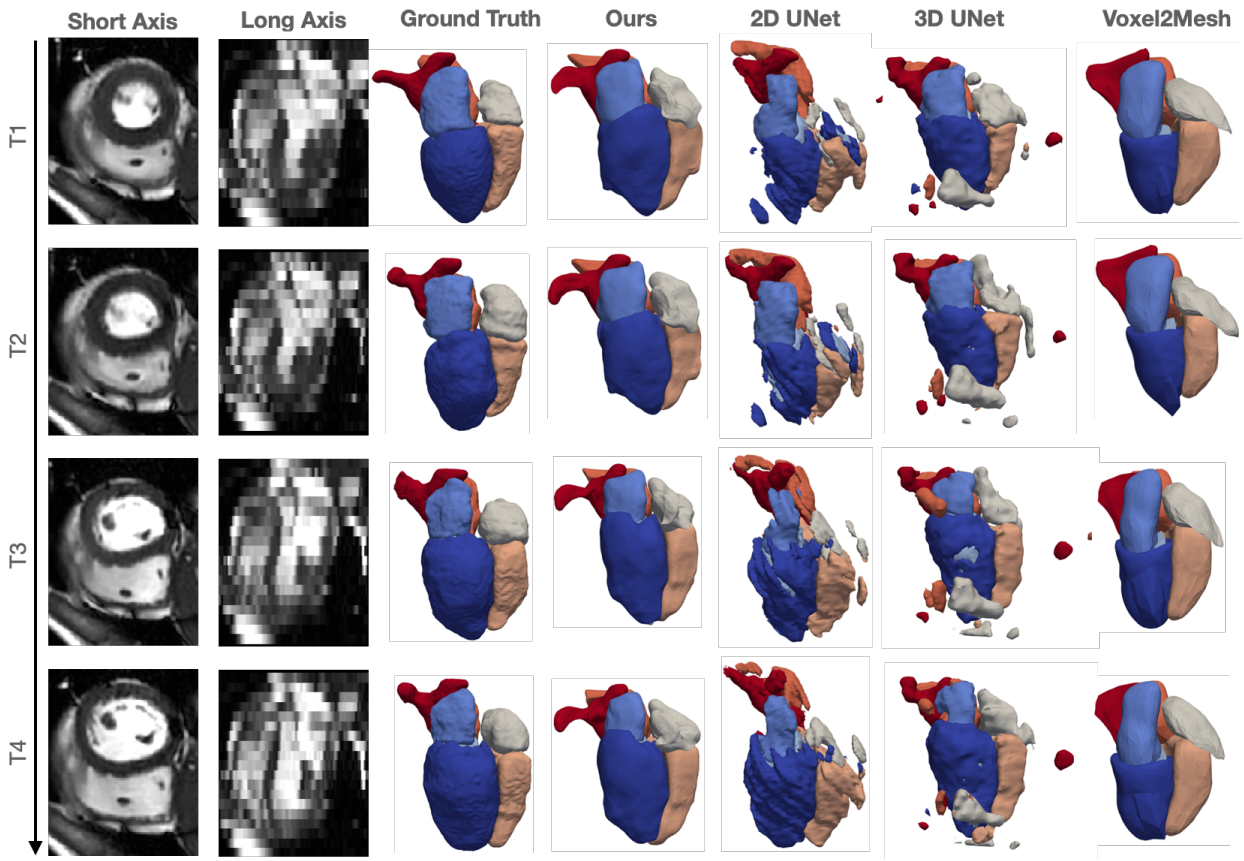


Figure 3.12: Short axis and long axis slices at different time frames for an example cine cardic MR data and the corresponding reconstructed whole heart surfaces from different methods.

performed the other methods for most cardiac structures in terms of average Dice score and ASSD.

Furthermore, we explore the potential capability of our method to reconstruct dynamic 4D whole-heart models to capture the motion of the heart from time-series image data. Figure 3.13 displays example whole-heart reconstruction results of our methods on time-series CT data that consisted of images from 10 time frames over the cardiac cycle for each patient. Although our model predicts mesh reconstructions independently from each time frame, it is able to consistently deform the template meshes such that the same mesh vertices on the template meshes are generally mapped to the same region of the reconstructed geometries across different time frames, as shown by the color maps of vertex IDs in Figure 3.13. Moreover, as demonstrated by the segmentation in Figure 3.13, our method is able to capture the minor changes between time frames. Therefore, our method can potentially be applied to efficiently construct 4D dynamic whole-heart models to capture the motion of a beating heart.

		Epi	LA	LV	RA	RV	Ao	PA	WH
Dice (↑)	Ours	0.902±0.035	**0.96±0.018**	0.956±0.033	**0.946±0.014**	**0.944±0.017**	**0.974±0.006**	0.798±0.129	**0.94±0.012**
	2D UNet	**0.913±0.028**	0.958±0.014	**0.957±0.023**	0.927±0.041	0.925±0.041	0.971±0.009	**0.867±0.114**	0.937±0.022
	3D UNet	0.884±0.03	0.935±0.012	0.946±0.03	0.928±0.019	0.92±0.02	0.955±0.01	0.831±0.059	0.922±0.014
	Voxel2Mesh	0.786±0.072	0.933±0.019	0.928±0.037	0.92±0.021	0.928±0.019	0.924±0.011	0.651±0.123	0.894±0.014
ASSD (mm) (↓)	Ours	0.697±0.308	**0.54±0.205**	0.574±0.399	**0.781±0.21**	**0.756±0.219**	**0.28±0.073**	2.714±3.079	0.906±0.5
	2D UNet	**0.634±0.281**	0.569±0.181	**0.538±0.25**	1.097±0.668	1.099±0.737	0.281±0.103	**1.155±1.019**	**0.767±0.291**
	3D UNet	0.811±0.34	0.871±0.277	0.711±0.381	0.993±0.325	1.017±0.267	0.504±0.19	1.598±1.183	0.929±0.259
	Voxel2Mesh	1.297±0.451	0.916±0.208	0.993±0.423	1.194±0.327	1.034±0.275	0.844±0.124	3.788±2.008	1.438±0.325

Table 3.5: A comparison of prediction accuracy on time-series CT dataset from different methods. All accuracy measures are represented by mean ± standard deviation, which are computed over different patients and time frames.

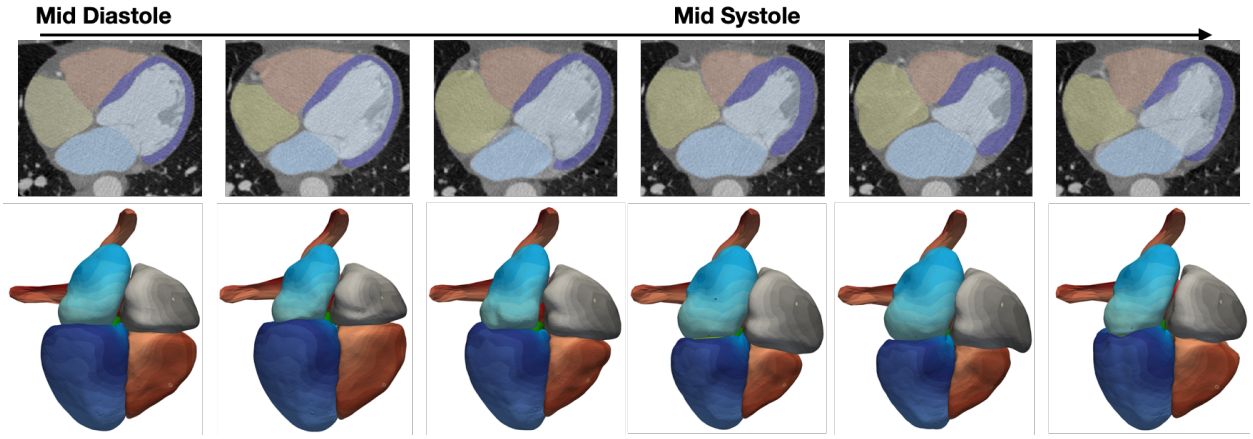


Figure 3.13: Whole-heart reconstruction results for time-series CT data. From left to right, each column displays results at one time frame from middle diastole to early diastole. The top row shows predicted segmentation overlaid with CT images and the bottom row shows the correspondence maps of mesh vertices across reconstructed meshes from different time frames, with same color denoting the same mesh vertices on reconstructed meshes.

Impact of Post-Processing on Reconstruction Performance

Post processing techniques have been commonly applied to correct prediction artifacts from segmentation-based deep-learning methods. Therefore, we investigated how the performance of our method compare with that of the 2D UNet and 3D UNet after post-processing. Namely, for each cardiac structure, we applied a median filter with a kernel size of $5 \times 5 \times 5$ voxels to fill any small gaps within the segmentation and smooth segmentation boundaries. We then removed any disconnected regions from the segmentation by computing the largest connected component for each cardiac structure. To correct for gaps between the predicted cardiac structures we leveraged the ability of our method to consistently map the same vertices to the similar regions of the heart. Thus, we can readily identify the vertices on the

			Epi	LA	LV	RA	RV	Ao	PA	WH
CT	Dice (↑)	Ours-Post	**0.902** (0.003)	**0.933** (0.001)	**0.940**	**0.892**	**0.910**	**0.950**	**0.856** (0.003)	**0.919** (0.001)
		2DUNet-Post	0.895 (-0.004)	0.924 (-0.006)	0.928 (-0.002)	0.878 (0.001)	0.904 (-0.001)	0.926 (-0.008)	0.831 (-0.001)	0.908 (-0.002)
		3DUNet-Post	0.864 (0.001)	0.903 (0.001)	0.930 (0.007)	0.871 (0.003)	0.877 (0.001)	0.920 (-0.003)	0.793 (-0.019)	0.889 (0.001)
	Jaccard (↑)	Ours-Post	**0.823** (0.004)	**0.876** (0.001)	**0.888**	**0.809**	**0.837** (-0.001)	**0.905**	**0.760** (0.005)	**0.850** (0.001)
		2DUNet-Post	0.812 (-0.006)	0.861 (-0.011)	0.869 (-0.004)	0.787	0.827 (-0.001)	0.864 (-0.015)	0.724 (-0.002)	0.833 (-0.004)
		3DUNet-Post	0.763 (0.001)	0.825	0.870 (0.009)	0.774 (0.005)	0.785 (0.002)	0.854 (-0.006)	0.678 (-0.017)	0.801 (0.001)
	ASSD (mm) (↓)	Ours-Post	0.874 (-0.461)	**1.020** (-0.022)	**0.823** (-0.020)	**1.549** (-0.034)	1.139 (-0.037)	**0.528** (-0.003)	1.896 (-0.009)	**1.112** (-0.100)
		2DUNet-Post	**0.863** (0.054)	1.125 (0.0750	0.960 (0.056)	1.681 (-0.038)	**1.129** (0.065)	0.819 (0.174)	**1.701** (0.149)	1.171 (0.083)
		3DUNet-Post	1.295 (-0.148)	1.455 (-0.073)	0.958 (-0.066)	1.906 (-0.036)	1.680 (0.017)	0.905 (0.090)	3.135 (0.941)	1.649 (0.097)
	HD (mm) (↓)	Ours-Post	13.978 (0.415)	**7.960** (-2.447)	**6.252** (-4.074)	**11.735** (-1.904)	10.958 (-2.401)	**9.044** (-0.363)	26.616	28.041 (0.006)
		2DUNet-Post	**9.194** (-0.786)	8.368 (-0.406)	6.287 (0.189)	12.243 (-1.381)	**9.750** (-0.266)	10.161 (0.148)	**26.100** (-1.734)	**26.900** (-1.826)
		3DUNet-Post	10.250 (-3.386)	9.828 (-0.986)	6.618 (-2.961)	13.251 (-2.779)	12.614 (-3.020)	12.500 (-0.826)	28.700 (1.759)	30.582 (-0.506)
MR	Dice (↑)	Ours-Post	**0.800** (0.002)	**0.879** (-0.002)	**0.921** (-0.001)	**0.888**	**0.892**	**0.889** (-0.001)	**0.817**	**0.881**
		2DUNet-Post	0.790 (-0.005)	0.850 (-0.014)	0.892 (-0.004)	0.842 (-0.010)	0.862 (-0.003)	0.862 (-0.008)	0.764 (-0.008)	0.854 (-0.005)
		3DUNet-Post	0.770 (0.009)	0.848 (-0.004)	0.881 (0.002)	0.868 (0.001)	0.830 (0.003)	0.817 (0.076)	0.761 (-0.003)	0.844 (0.004)
	Jaccard (↑)	Ours-Post	**0.674** (0.003)	**0.788** (-0.003)	**0.856** (-0.002)	**0.800** (-0.001)	**0.812**	**0.804** (-0.001)	**0.697**	**0.790**
		2DUNet-Post	0.661 (-0.007)	0.746 (-0.019)	0.811 (-0.006)	0.741 (-0.011)	0.766 (-0.005)	0.762 (-0.012)	0.632 (-0.009)	0.749 (-0.008)
		3DUNet-Post	0.635 (0.010)	0.752 (-0.004)	0.811 (0.009)	0.768 (0.002)	0.733 (0.006)	0.715 (0.065)	0.633 (-0.007)	0.737 (0.005)
	ASSD (mm) (↓)	Ours-Post	1.967 (-0.231)	**1.373** (-0.028)	**1.155** (-0.028)	**1.581** (-0.029)	**1.310** (-0.023)	2.650 (0.001)	2.692 (0.002)	**1.713** (-0.061)
		2DUNet-Post	**1.805** (-0.013)	1.699 (0.211)	1.520 (0.065)	2.008 (0.288)	1.523 (0.058)	2.747 (0.300)	**2.151** (0.331)	1.952 (0.286)
		3DUNet-Post	2.167 (-0.206)	2.151 (-0.318)	1.600 (-0.618)	1.658 (-0.338)	2.454 (-0.312)	**2.512**(-1.277)	2.209 (0.265)	2.042 (-0.073)
	HD (mm) (↓)	Ours-Post	16.516 (-0.406)	**9.658** (-2.065)	**8.070** (-2.820)	13.558 (-1.252)	**11.025** (-2.438)	**22.219**	19.319 (-0.026)	27.569 (-0.133)
		2DUNet-Post	**13.759** (-5.398)	11.185 (0.404)	9.972 (0.014)	13.825 (-1.005)	11.544 (-1.556)	24.912 (2.346)	17.056 (0.335)	28.024 (-0.273)
		3DUNet-Post	17.024 (-11.432)	11.564 (-12.263)	11.531 (-11.178)	**12.474** (-7.048)	12.699 (-8.295)	23.113 (-11.226)	**17.021** (0.140)	**27.065** (-15.400)

Table 3.6: A comparison post-processed prediction accuracy on MMWHS MR and CT test datasets from different methods. Numbers in parentheses display the accuracy differences (if any) before and after post processing.

adjacent surfaces between the cardiac structures from our training data. For test cases, we can then project each of these vertices to the closest vertex on the adjacent surface.

Table 3.6 compares the reconstruction accuracy for our method, 2D UNet, and 3D UNet after the above post-processing steps as well as the accuracy differences before and after post-processing for each method. For both CT and MR data, our method consistently outperformed the baselines for all cardiac structures in terms of Dice and Jaccard scores, and for most cardiac structures in terms of ASSD and HD measures, respectively. In general, post-processing techniques did not bring major improvements in Dice, Jaccard or ASSD measures for all the methods. Indeed, these post-processing techniques are designed to correct artifacts small in size and thus do not significantly contribute to the improvements in global accuracy measures. In contrast, for local accuracy measure HD, post-processing techniques brought a major improvement in HD measure for 3D UNet for MR data due to the removal of disconnected regions from the predictions. Figure 3.14 displays the segmentation and reconstruction results for a challenging MR case before and after post-processing. Segmentation-based approaches, 2D and 3D UNets, predicted topological incorrect LV myocardium geometries with large holes, whereas our template-based method predicted topological-correct geometries. Post-processing techniques were able to reduce, but not fully close these holes. For this MR case, our method produced a small gap between the LV and myocardium as these two structures are represented by individual surfaces. However, our post-processing method on the mesh was able to automatically seal this gap.

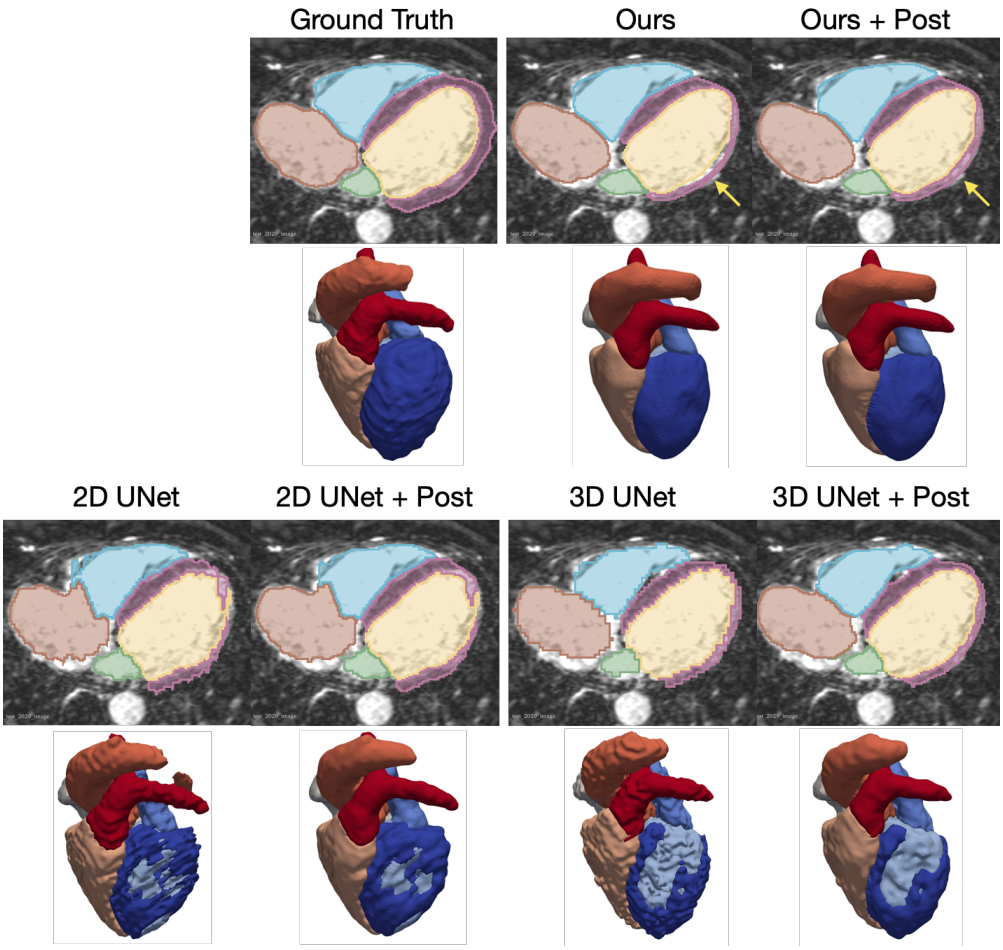


Figure 3.14: Example of whole heart segmentation and surface reconstruction results before and after post-processing.

Impact of Limited Training Data on Reconstruction Performance

We investigate how well our method can reconstruct whole-heart geometries using only a small number of training data. In this experiment, our neural network model was trained using only the training set of MMWHS challenge, which consists of 20 CT images and 20 MR images. 16 out of 20 image volumes from each modality were used for training and the rest were used for validation. We compared our method against the baseline methods for the same MMWHS test set described above. The baseline methods were trained using the same training and validation splits.

Table 3.7 compares the Dice and Jaccard scores, ASSD and HD of the reconstruction results for the methods trained with the reduced training set, as well as the accuracy differences compared with training models using more data, as described above. For CT data, our method consistently outperformed others in terms of Dice and Jaccard scores for the

whole heart and individual cardiac structures except for pulmonary arteries. In terms of ASSD and HD, our method outperformed 3D UNet and Voxel2Mesh and was comparable to 2D UNet. For MR data, our method demonstrated better performance than others in terms of whole heart Dice and Jaccard scores, as well as surface HD of whole heart. 2D UNet demonstrated the best whole heart ASSD performance. For individual cardiac structures, our method showed better Dice and Jaccard scores for Epi, LV, RA and RV, smaller ASSD values for Epi, LV, RA and smaller surface HD values for most of the cardiac structures except for LA and Ao. Figure 3.15 shows the distribution of different segmentation accuracy metrics for whole heart and individual cardiac structures among the MMWHS test dataset.

As shown in Table 3.7, when trained with a smaller training dataset, the methods generally showed reduced Dice or Jaccard scores and increased ASSD and HD values for both whole heart and individual cardiac structures compared with when trained with a larger dataset, as summarized in Table 3.2. Exceptions include the smaller HD values of Epi, LA, LV, RV and PA from our method for CT data and the better LV and aorta segmentation from 3D UNet for MR data in terms of all four metrics. Compared with CT data, all methods generally demonstrated more significant reduction of segmentation accuracy for MR data, in terms of average values of reduction for all four metrics. While performance drops due to reduced size of training data is consistent, the actually amount of performance drop is minor for our method, 2D UNet and 3D UNet. For example, although the number of CT training data was reduced from 87 to 16, we only observed a small average reduction (0.01-0.02) of whole heart Dice scores for 2D UNet, 3D UNet and our method. However, the performance drop for Voxel2Mesh in relation to the number of training data was much more significant, with a 0.27-0.28 reduction of whole-heart Dice scores for CT and MR data. Among all the cardiac structures, our method had the most significant performance reduction of PA reconstruction for both CT and MR data while segmentation based approaches, 2D UNet and 3D UNet, demonstrated a more uniform performance drop across all cardiac structures. Indeed, the shapes of the PA differ significantly from our initial sphere template mesh and therefore accurately capturing the shapes of PA might require more training data for our method. Figure 3.15 gives the distribution of different segmentation accuracy metrics for whole heart and individual cardiac structures.

Sensitivity Studies

We compare the effect of design choice changes on the whole heart reconstruction performance of our method. Namely, we trained another three models while, respectively, using reduced number of convolutional filters in the image encoding module, using reduced resolution of template meshes or excluding the elastic deformation from our image augmentation techniques. Specifically, the number of convolutional filters in the last four residual blocks were reduced from 48, 96, 192 and 384 to 32, 64, 128 and 256, respectively. The number of mesh vertices of the template mesh was reduced from 11494 to 3260. As shown in Table 3.8, reducing the number of filters or template mesh vertices mildly reduced the reconstruction accuracy of the whole heart or most cardiac structures compared with the our final model.

			Epi	LA	LV	RA	RV	Ao	PA	WH
CT	Dice (↑)	Ours	**0.880**	**0.926**	**0.931**	**0.868**	**0.885***	**0.945**	0.786*	**0.900***
		2DUNet	0.877*	0.916	0.926	0.855	0.876*	0.916	**0.805**	0.892*
		3DUNet	0.816*	0.916	0.914	0.848	0.878	0.923	0.793	0.877
		Voxel2Mesh	0.501*	0.748*	0.669*	0.717*	0.698*	0.555*	0.491*	0.656*
	Jaccard (↑)	Ours	**0.790**	**0.863**	**0.874**	**0.773**	**0.798***	**0.897**	0.666*	**0.819***
		2DUNet	0.784*	0.847	0.864	0.753	0.787*	0.850	**0.692**	0.807*
		3DUNet	0.696*	0.848	0.844	0.741	0.787	0.860	0.670	0.782
		Voxel2Mesh	0.337*	0.600*	0.510*	0.570*	0.543*	0.397*	0.337*	0.491*
	ASSD (mm) (↓)	Ours	1.357	**1.137**	0.966	1.750	**1.320**	**0.729***	2.020	1.333*
		2DUNet	**1.014***	1.141	**0.911**	**1.702**	1.433*	0.808	**1.754**	**1.240***
		3DUNet	1.809*	1.389	1.134	2.176	1.585	0.832	2.276	1.668
		Voxel2Mesh	3.412*	3.147*	4.973*	3.638*	4.300*	4.326*	5.857*	4.287*
	HD (mm) (↓)	Ours	13.789	10.362	9.628	**14.467**	**12.766**	12.740*	**25.362**	**27.567**
		2DUNet	**13.582**	**10.221**	**6.700**	14.788	16.608*	11.410	28.128	32.514
		3DUNet	15.044	40.157*	9.730	15.037	13.777	**10.821**	27.467	48.731
		Voxel2Mesh	15.526*	13.683*	22.146*	16.834*	18.390*	19.419*	35.322*	37.065*
MR	Dice (↑)	Ours	**0.773**	0.826*	**0.913**	**0.838***	**0.861**	0.824*	0.663*	**0.846***
		2DUNet	0.751	**0.831**	0.880	0.815	0.852	**0.838***	**0.747**	0.834
		3DUNet	0.733	0.811	0.885	0.827*	0.829	0.825	0.741	0.823
		Voxel2Mesh	0.282*	0.498*	0.515*	0.599*	0.539*	0.241*	0.300*	0.483*
	Jaccard (↑)	Ours	**0.639**	0.712*	**0.842**	**0.727***	**0.768**	0.715*	0.517*	**0.737***
		2DUNet	0.611*	**0.720**	0.793	0.702	0.753	**0.726***	0.608	0.719*
		3DUNet	0.588	0.695*	0.803	0.718*	0.727	0.718	**0.615**	0.709
		Voxel2Mesh	0.170*	0.339*	0.367*	0.442*	0.388*	0.144*	0.187*	0.327*
	ASSD (mm) (↓)	Ours	**2.385**	2.166*	**1.300**	**2.358***	1.812	3.243	3.138	2.235*
		2DUNet	2.692*	**1.688**	1.603	3.151*	**1.736**	**2.920**	2.281*	**1.897**
		3DUNet	2.713	3.866	1.551	2.475	1.931	4.049	**2.259**	2.120
		Voxel2Mesh	6.886*	5.987*	8.679*	6.173*	8.192*	7.877*	9.200*	7.419*
	HD (mm) (↓)	Ours	**16.804**	15.559*	**12.197**	**17.286***	**14.480**	26.012	**19.927**	**29.983**
		2DUNet	23.798	**14.887***	14.651	22.028*	22.810*	**24.237**	22.883*	39.724*
		3DUNet	20.136	32.978	13.643	23.735	22.351	31.900	21.363	43.475
		Voxel2Mesh	27.272*	22.748*	31.327*	24.456*	28.987*	29.381	33.637*	40.072*

Table 3.7: A comparison of prediction accuracy on MMWHS MR and CT test datasets from different methods trained with images from MMWHS training set. An asterisk $*$ indicates statistically significant accuracy differences, compared with Table 3.2, resulted from training on a smaller datset based on t-tests ($p<0.05$).

However, without elastic deformation augmentation, we observed a significant drop in reconstruction performance. Indeed, elastic deformation augmentation may improve model robustness to minor local perturbations of the anatomical structures, thereby facilitating accurate predictions of detailed cardiac structures.

Ablation Studies on Loss components

We performed an ablation study on the mesh loss functions to evaluate the contribution of individual loss components. Namely, we trained additional models with the normal loss,

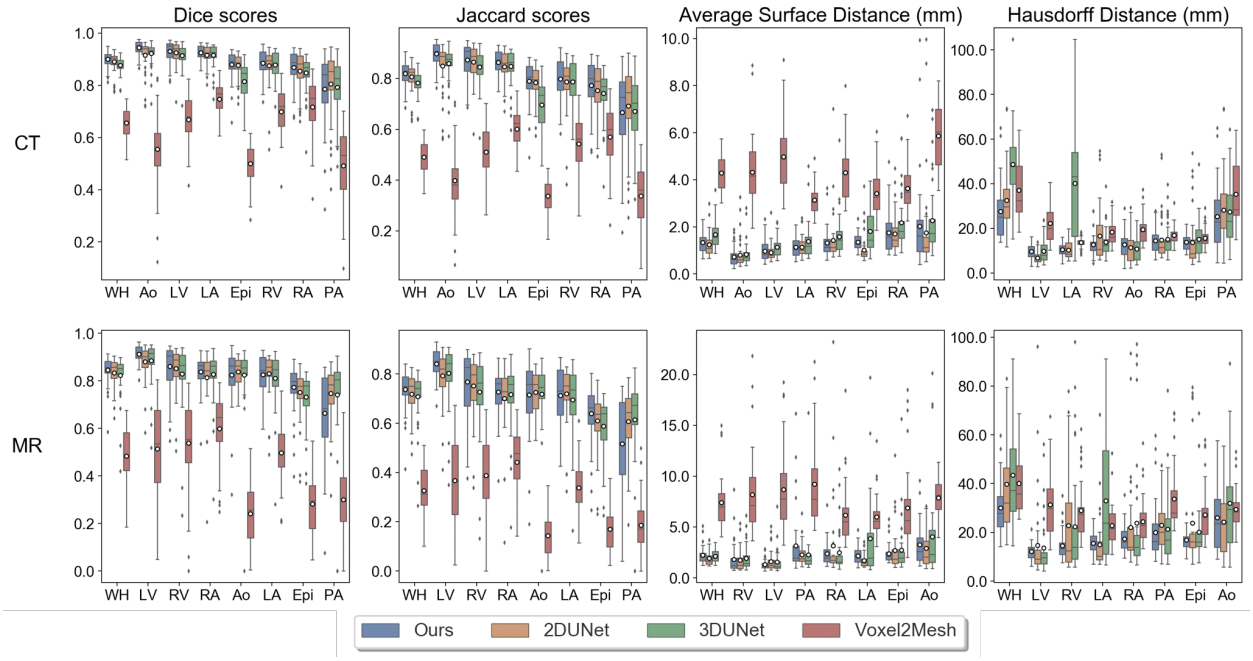


Figure 3.15: Comparison of segmentation accuracy for whole heart and individual cardiac structures from different methods trained using the small MMWHS training dataset. White circles on the boxes indicate mean values across patients. Cardiac structures are sorted based on the accuracy of our method.

			Epi	LA	LV	RA	RV	Ao	PA	WH
CT	Dice	Final Model	0.899	0.932	0.940	0.892	0.910	0.950	0.852	0.918
		Reduced Number of Filters	0.893	0.932	0.936	0.888	0.906	0.949	0.847	0.915
		Reduced Number of Vertices	0.842	0.929	0.931	0.888	0.892	0.943	0.837	0.899
		No Elastic Deformation Augmentation	0.546	0.882	0.878	0.801	0.757	0.882	0.552	0.773
	ASSD (mm)	Final Model	1.335	1.042	0.842	1.583	1.176	0.531	1.904	1.213
		Reduced Number of Filters	1.404	1.063	0.892	1.523	1.122	0.536	1.744	1.195
		Reduced Number of Vertices	1.768	1.126	1.029	1.664	1.425	0.583	1.768	1.378
		No Elastic Deformation Augmentation	3.055	1.742	1.731	2.479	3.188	1.117	5.823	2.920
MR	Dice	Final Model	0.797	0.881	0.922	0.888	0.892	0.890	0.816	0.882
		Reduced Number of Filters	0.813	0.873	0.919	0.888	0.881	0.876	0.789	0.879
		Reduced Number of Vertices	0.774	0.870	0.903	0.887	0.861	0.860	0.792	0.863
		No Elastic Deformation Augmentation	0.487	0.810	0.867	0.795	0.744	0.724	0.413	0.735
	ASSD (mm)	Final Model	2.198	1.401	1.183	1.611	1.333	2.648	2.689	1.775
		Reduced Number of Filters	2.053	1.556	1.238	1.488	1.429	2.143	2.205	1.645
		Reduced Number of Vertices	2.405	1.615	1.516	1.561	1.651	2.390	2.222	1.845
		No Elastic Deformation Augmentation	3.794	2.151	2.003	2.976	3.575	3.700	5.166	3.348

Table 3.8: A comparison of prediction accuracy on MMWHS MR and CT test datasets from different variants of our methods.

the edge loss or the Laplacian loss removed by changing λ_2, λ_3, or λ_4 to 0, respectively, in Equation 3.6. We also evaluated the effect of using $L2$ norm on the edge length loss rather than the $L1$ norm originally used in Equation 3.4. The rest of the hyperparameters were kept the same. As shown in Table 3.9, removing the Laplcaian loss caused the most accuracy drop on the MMWHS MR and CT test datasets, followed by the edge length loss, the normal loss and using the L2 edge length loss.

3.4 Discussion

Image-based reconstruction of cardiac anatomy and the concomitant geometric representation using unstructured meshes is important to a number of applications, including visualization of patient-specific heart morphology and computational simulations of cardiac function. Prior deep-learning-based approaches have shown great promise in automatic whole heart segmentation [182], however converting the segmentation results to topologically valid mesh structures requires additional, and often manual, post-processing, and is highly-dependent on the resolution of the image data. In this work, we present a novel deep-learning-based approach that uses graph convolutional neural networks to directly generate meshes of multiple cardiac structures of the whole heart from volumetric medial image data. Our approach generally demonstrated improved whole heart reconstruction performance compared with the baseline methods in terms of accuracy measures, Dice and Jaccard scores, ASSD and HD. Furthermore, our method demonstrated advantages in generating high-resolution, anatomically and temporally consistent geometries, which are not reflected by the accuracy measures.

Our method reconstructs cardiac structures by predicting the deformation of mesh vertices from sphere mesh templates. We have demonstrated the advantages of this approach over segmentation-based approaches in terms of both precision and surface quality. Namely, the use of a template mesh can introduce topological constraints so that predicted cardiac structure are homeomorphic to the template. Thus, our template based approach enables one to eliminate disconnected regions and greatly reduce erroneous topological artifacts often encountered with existing deep-learning-based segmentation methods. While the cardiac structures of interest were homeomorphic to spheres, the presented method has the potential to be generalized to organs with different topology, by using a different template mesh with the required surface topology.

When trained on a relatively large dataset with 87 CT and 41 MR images, our method was able to achieve comparable accuracy to manual delineations, which is considered the gold standard. Furthermore, since we explicitly regularized the surface smoothness and normal consistency, our method produced smooth and quality meshes while capturing the detailed features of the cardiac structures. Namely, these factors along with the use of a template enable our method to generate realistic cardiac structures even when image quality was poor and segmentation methods struggled to provide realistic topology. From our observations, the locations on the heart that our neural network models produced high surface errors are

			Epi	LA	LV	RA	RV	Ao	PA	WH
CT	Dice	Final	0.899	0.932	0.940	0.892	0.910	0.950	0.852	0.918
		L2 edge	0.871	0.926	0.932	0.886	0.904	0.938	0.840	0.907
		No normal	0.806	0.925	0.939	0.885	0.904	0.942	0.854	0.896
		No edge	0.631	0.866	0.880	0.809	0.793	0.881	0.511	0.788
		No Laplacian	0.439	0.870	0.803	0.799	0.760	0.870	0.546	0.746
	Jaccard	Final	0.819	0.875	0.888	0.809	0.837	0.905	0.755	0.849
		L2 edge	0.776	0.864	0.875	0.799	0.828	0.885	0.735	0.831
		No normal	0.676	0.861	0.887	0.797	0.826	0.891	0.754	0.812
		No edge	0.464	0.766	0.792	0.684	0.663	0.790	0.349	0.652
		No Laplacian	0.284	0.773	0.680	0.671	0.619	0.773	0.380	0.597
	ASSD (mm)	Final	1.335	1.042	0.842	1.583	1.176	0.531	1.904	1.213
		L2 edge	1.609	1.127	0.933	1.657	1.243	0.641	1.826	1.314
		No normal	2.039	1.154	0.838	1.701	1.200	0.726	2.147	1.469
		No edge	2.599	2.014	1.723	2.350	2.741	1.597	7.567	3.343
		No Laplacian	3.488	1.950	3.086	2.469	3.040	1.853	6.008	3.296
	HD (mm)	Final	14.393	10.407	10.325	13.639	13.360	9.407	26.616	28.035
		L2 edge	15.500	10.098	8.977	13.055	12.406	10.178	26.034	27.030
		No normal	14.261	11.269	10.027	13.502	11.768	12.500	27.737	29.066
		No edge	15.000	10.304	9.668	13.104	14.236	12.677	34.336	34.852
		No Laplacian	17.412	11.317	15.194	13.407	15.992	18.786	34.145	36.281
MR	Dice	Final	0.797	0.881	0.922	0.888	0.892	0.890	0.816	0.882
		L2 edge	0.794	0.871	0.915	0.876	0.873	0.867	0.776	0.868
		No normal	0.753	0.878	0.922	0.878	0.884	0.857	0.760	0.866
		No edge	0.505	0.745	0.853	0.818	0.789	0.783	0.498	0.743
		No Laplacian	0.450	0.765	0.846	0.786	0.772	0.747	0.471	0.733
	Jaccard	Final	0.671	0.791	0.858	0.801	0.812	0.805	0.697	0.790
		L2 edge	0.665	0.775	0.845	0.783	0.783	0.769	0.645	0.770
		No normal	0.609	0.787	0.857	0.786	0.798	0.755	0.629	0.765
		No edge	0.347	0.603	0.750	0.699	0.667	0.653	0.338	0.596
		No Laplacian	0.296	0.626	0.742	0.654	0.641	0.608	0.317	0.582
	ASSD (mm)	Final	2.198	1.401	1.183	1.611	1.333	2.648	2.689	1.775
		L2 edge	2.224	1.563	1.288	1.738	1.599	3.017	2.345	1.909
		No normal	2.330	1.458	1.149	1.774	1.396	2.691	2.978	1.923
		No edge	4.013	3.291	2.397	2.556	3.094	3.448	6.763	3.730
		No Laplacian	4.117	2.658	2.362	2.827	3.007	6.512	6.047	3.850
	HD (mm)	Final	16.923	11.723	10.891	14.810	13.463	22.219	19.345	27.701
		L2 edge	18.361	10.705	8.969	14.247	13.455	22.754	17.124	29.339
		No normal	15.460	12.190	10.354	16.143	13.493	23.968	21.291	29.490
		No edge	20.087	16.863	14.517	14.365	15.953	21.623	25.522	30.149
		No Laplacian	21.755	13.721	12.063	15.423	16.613	35.356	25.437	38.581

Table 3.9: Impact of using L2 edge length loss or removing edge length loss, Laplacian loss or normal loss on the prediction accuracy on MMWHS MR and CT test datasets

	Ours	Ours (Reduced Number of Filters)	Ours (Reduced Number of Vertices)	2D UNet	3D UNet	Voxel2Mesh
# of Parameters	16,765,112	8,474,257	16,765,112	31,110,152	18,556,552	9,124,521
Time (s)	0.425	0.378	0.240	1.555	0.367	3.492

Table 3.10: A comparison of neural network sizes and the average prediction time among our methods, 2D UNet, 3D UNet and Voxel2Mesh. The prediction time was measured on a Nvidia GeForce GTX 1080Ti GPU.

consistent with the locations that could suffer from high inter- or intra-observer variations, such as the arbitrary length of aorta and pulmonary arteries, boundaries between atria and ventricles and between the right atrium and the inferior vena cava. Indeed, these boundaries are not distinguishable by voxel intensity differences and are often subject to uncertainties even for human observers.

Compared with segmentation-based approaches, our method predicts whole heart surfaces directly in the physical space rather than on a voxel grid of the input image. The whole heart geometries are represented using surface meshes rather than a dense voxel grid. Hence, our method is able to generate high-resolution reconstructions (11K mesh vertices for each cardiac structure) efficiently on a limited memory budget and within a shorter or comparable run-time (table 3.10). Prior 3D segmentation-based approaches have sought to increase the segmentation resolution by training separate neural networks to first locate the region of interest or generate low resolution segmentations, and then generate refined segmentations within the localized region [113, 51]. Our method does not require training multiple neural networks and can make predictions directly from the entire down-sampled cardiac image volume. As we used a cascade of three mesh deformation blocks, we observed that the first deformation block can already effectively position and deform the meshes to the correct locations and the subsequent deformation blocks can further refine the predicted mesh vertex locations. High resolution segmentation may also be obtained by recent methods that represent geometries using implicit surfaces [63]. Namely, for each point in the physical space, this approach predicts the probability of this point belonging to a certain tissue class. Therefore, by sampling a large number of points in the physical space, these methods can also achieve high-resolution reconstruction that are not constrained by the voxel resolution of the input image or GPU memory. However, the inference process for such methods is computationally expensive [42] as it requires prediction on a large number of points. In contrast, our method represents the mesh as a graph (i.e., a sparse matrix) and takes less than a second to predict a high resolution whole heart mesh.

Compared with prior deep-learning-based mesh reconstruction methods from image data [159, 168], our method used a shared graph neural network to simultaneously predict surface meshes of multiple cardiac structures. This is made possible by initializing the template meshes at various scales and locations corresponding to individual cardiac structures. We observed that proper template initialization is essential to avoid local minimums due to large mesh deformation at the beginning stage of training. In contrast, prior approaches are designed for predicting a single geometry from image data and require training a separate

graph neural network for each anatomical structure and thus do not easily scale to reconstruct multiple cardiac structures at a high-resolution from a single image volume. Furthermore, while prior approaches proposed various up-sampling scheme to construct a dense mesh from a coarse template [168, 159], we directly deformed a high resolution template. Since the majority of weights is in the image feature encoder to process a dense volumetric input image, more mesh vertices can provide more effective gradient propagation to the image feature encoder. Indeed, using a coarse mesh with 3K mesh vertices for each cardiac structures, we observed a 2% reduction of whole heart dice score as shown in our supplemental materials (table 3.8). However, our method was still able to outperform Voxel2Mesh by 3% and 10% for CT and MR data using a coarse mesh template with a similar amount of mesh vertices. These design choices allowed our method to demonstrate promising generalization capabilities to unseen MR images and maintain good performance when trained with a smaller number of samples. In contrast, Voxel2mesh suffered from a large performance drop when trained on a smaller dataset.

When applied to time-resolved images, our method consistently deformed the template mesh such that mesh vertices were mapped to the similar regions of the heart across different time frames. Learning such semantic correspondence is purely a consequence of our model architecture and did not require any explicit training. This behavior of producing semantic corresponding predictions was also observed in DeepOrganNet, which reconstructed lung shapes from single-view X-ray images by deforming lung templates [161]. Point-corresponded meshes across different input images are required for numerous applications, such as building statistical shape models, constructing 4D dynamic whole-heart models for motion analysis and deriving boundary conditions for deforming-domain CFD simulations. Current approaches that construct feature corresponding meshes for the heart mostly use surface or image registration methods to deform a reference mesh so that its boundary is consistent with the target surfaces or image segmentation [108, 60, 66]. However, registration algorithms are often very computationally expensive to align high-resolution meshes and they often suffer from inaccuracies for complex whole heart geometries due to local minimums during the optimization process. In the case of time-series image data, our method naturally produces point corresponding meshes with high resolution (10K mesh vertices per cardiac structure) across time frames within a couple of seconds, while prior methods could require hours to generate a 4D dynamic whole-heart model at a similar resolution. Although not considered here, it is possible to include another loss function that minimizes the point distances between the vertex locations on the predicted meshes and ground truth landmarks when available to further enhance feature correspondence.

Limitations of the proposed method include a lack of diffeomorphic constraints to establish a differentiable mapping from the initial spheres and the predicted surfaces. While we used the Laplacian loss to regularize the smoothness of the meshes, a diffeomorphic constraints may help to further prevent face intersections. Recently, [42] proposed to learn neural ordinary differential equations to predict a diffeomorphic flow that maps a sphere mesh template to the target shapes, thus implicitly preserving the manifoldness of the template mesh without explicit regularizations. This approach could be combined with our

image-based whole-heart mesh prediction framework in the future to deform the whole heart geometry while preserving the manifoldness of the meshes so that they could be directly used in applications such as numerical simulations and 3D printing. Furthermore, while our method can simultaneously predict multiple structures from image data, those structures are not coupled to each other. Small intersections or gaps could appear between adjacent cardiac structures. While we have demonstrated that simple projection can generally correct such artifacts, future work could include more explicitly constraining the coupling of cardiac structures within the learning framework.

3.5 Conclusions

We have developed a deep-learning-based method to directly predict surface mesh reconstructions of the whole heart from volumetric image data. The approach leverages a graph convolutional neural network to predict deformation on mesh vertices from a predefined mesh template to fit multiple anatomical structures in a 3D image volume. The mesh deformation is conditioned on image features extracted by a CNN-based image encoder. The method demonstrated promising performance of generating accurate high-resolution and high-quality whole heart reconstructions and outperformed prior deep-learning-based methods on both CT and MR data. It also demonstrated robust performance when evaluated on MR or CT images from new data sources that differ from our the training datasets. Furthermore, the method produced temporally consistent predictions and feature-corresponding predictions by consistently mapping mesh vertices on the templates to similar structural regions of the heart. Therefore, this method can potentially be applied for efficiently constructing 4D dynamics whole heart model that captures the motion of a beating heart from time-series images data.

Chapter 4

Learning Free-Form Deformation For Whole Heart Mesh Generation

4.1 Introduction

Direct mesh reconstruction using geometric deep learning [25, 11] provides a recent avenue to address the end-to-end learning between volumetric medical images and simulation-ready surface meshes of the heart [168, 161, 68, 3, 30]. Particularly, our previous chapter leverages a graph convolutional neural network to predict deformation on mesh vertices from a pre-defined mesh template to fit multiple anatomical structures in a 3D image volume [68]. However, different structures were represented by decoupled mesh templates and thus still require post-processing to merge different structures for computational simulations involving multiple cardiac structures. Similarly, [3] uses deep neural networks and patient metadata to predict cardiac shape parameters of a pre-built statistical shape model of the heart. However, these approaches often assume the connectivity of the meshes. That is, the shape and topology of the predicted meshes from these approaches are pre-determined by the mesh template and cannot be easily changed to accommodate various mesh requirements for different cardiac simulations.

In this chapter, we are motivated to automatically and directly generate meshes that are suitable for computational simulations of cardiac function. We present herein, in contrast, an approach that deforms the space enclosing the mesh template. Once trained on the whole heart template, our network can deform alternative template meshes that represent a subset of the geometries in the template to accommodate different modeling requirements. Prior work by Wang et. al. [161] proposed to learn the *space* deformation by predicting the displacements of a control point grid to deform template meshes of the lung. However, it leveraged memory-intensive, fully-connected neural network layers to predict the displacements with a small number of control points from a 2D X-Ray image and thus cannot be directly applied to model complex whole-heart geometries with large geometric variations from 3D image volumes. A few studies have focused on learning space deformation fields.

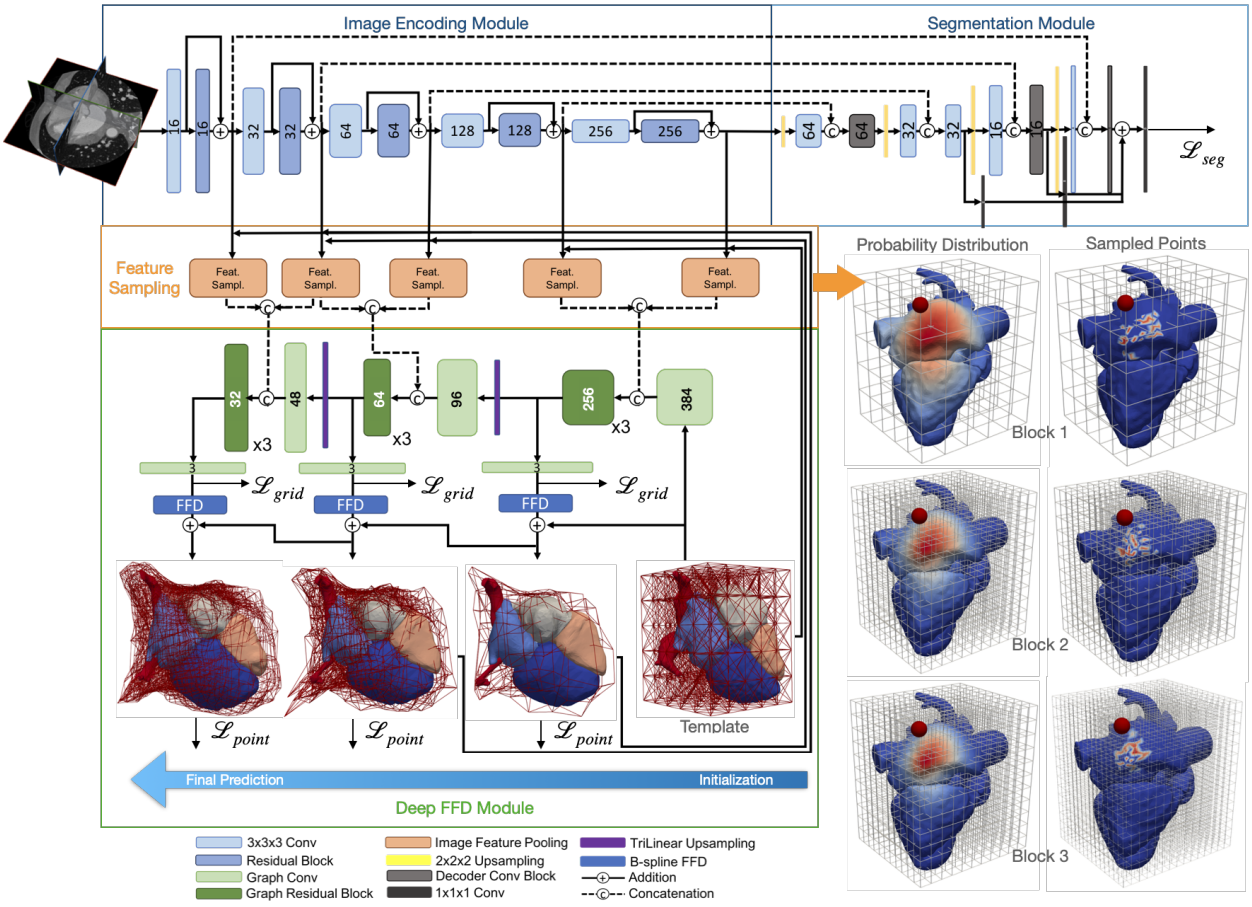


Figure 4.1: Diagram of the proposed automatic whole heart reconstruction approach.

[110] used a 3D UNet to predict the deformation field to deform only heart valve templates from CT images. To overcome these shortcomings, our method learns the multi-resolution B-spline free-form deformation of the space to produce detailed whole heart meshes from volumetric CT images by a novel graph convolutional module and effective feature sampling method. When applied to time-series image data, our approach is able to generate temporally consistent meshes that capture the motion of a beating heart and are suitable for cardiac hemodynamics simulations.

4.2 Methods

Fig 4.1 shows our proposed whole-heart mesh generation pipeline. Our framework consists of three components to predict the whole-heart meshes from a volumetric input image: (1) an image encoding module, (2) a image feature sampling module, (3) a deep free-form

deformation module that predicts control point displacements to deform the template mesh and (4) a segmentation module that predicts a binary segmentation map to allow additional supervision using ground truth annotations.

B-Spline Based FFD

We used a 3D tensor product of the 1D cubic B-spline formulation to deform the space. Namely, for a control point gird of $(l+1) \times (m+1) \times (n+1)$, the relation between a deformed mesh vertex $\mathbf{v}$ and the control points $\mathbf{p}$ is described by

$$\mathbf{v}(s,t,u) = \sum_{i=0}^{l} \sum_{j=0}^{m} \sum_{k=0}^{n} B_{i,3}(s) B_{j,3}(t) B_{k,3}(u) \mathbf{p}_{i,j,k} \tag{4.1}$$

, where $B_{i,3}$ is the cubic B-spline basis. Such relation can be expressed in the matrix form

$$\mathbf{V} = \mathbf{BP}, \quad \mathbf{V} \in \mathbb{R}^{N \times 3}, \quad \mathbf{B} \in \mathbb{R}^{N \times \psi}, \quad \mathbf{P} \in \mathbb{R}^{\psi \times 3} \tag{4.2}$$

, where N and ψ are the number of mesh vertices and control points, respectively. $\mathbf{B}$ is the trivariate B-spline tensor and can be pre-computed from the template mesh. $\mathbf{P}$ is the control point coordinates that the network will learn to predict. Compared with the Bernstein deformation tensor implemented in [161], the B-spline-based deformation matrix is sparse since the B-spline basis are defined locally and thus can greatly reduce the computational cost for high-resolution control point grids.

Deep FFD Module

Since the heart involves complicated geometries and significant shape variations during the cardiac cycle and across patients, a dense control point grid is necessary to produce accurate reconstruction of the cardiac structures. Therefore, flattening the image feature vectors of a 3D image and using fully connected layers to predicted control point displacements as proposed in [161] is no longer computationally feasible. We therefore propose to use a graph convolutional network (GCN) to predict the control point displacements based on sampled image feature vectors.

Graph Convolution On Control Grid

We represent the control point grid as a graph $\mathcal{M} = (\mathcal{V}, \mathcal{E})$. Each control point is connected with all its 26 neighbors (7, 11, and 17 neighbors at the corner point, edge point, and surface point, respectively). The graph convolution on a mesh follows [25] and [68]. Briefly, we used a first-order Chebyshev polynomial approximation described as $f_{out} = \sigma(\theta_0 f_{in} + \theta_1 f_{in} \tilde{L})$, where $\theta_0, \theta_1 \in \mathbb{R}^{d_{out} \times d_{in}}$ are trainable weights, $f_{in} \in \mathcal{R}^{d_{in} \times N}$, $f_{out} \in \mathcal{R}^{d_{out} \times N}$ are input and output feature matrices of a graph convolution layer applied on the control point grid, respectively, and $\tilde{L} = 2L_{norm}/\lambda_{max} - I$, $\tilde{L} \in \mathcal{R}^{N \times N}$is the scaled and normalized Laplacian matrix [25].

N is the number of control points. d_{in} and d_{out} are the input and output graph feature dimensions, respectively. The feature lengths of the intermediate layers match with the numbers in Fig 4.1, with 3 for displacements. Compared with conventional convolution with trainable filters, graph convolution requires far fewer parameters as the connection among vertices is encoded in the graph Laplacian matrix.

Deep Multi-Resolution FFD

Our proposed graph decoding module consists of three deformation blocks to progressively deform the template mesh. For the initial mesh deformation blocks, we used lower resolution control point grids conditioned on the more abstracted, high-level image feature maps while using high-resolution control point grids with low-level, high-resolution feature maps for the later mesh deformation blocks. Within each deformation block, we concatenate the sampled image feature vector with the vertex feature vectors on the control points and then use residual graph convolutional blocks to predict the displacements on the control points to deform the template meshes. Between two deformation blocks, we used used trilinear interpolation to upsample the features on lower-resolution control point grid to the same grid resolution as in the next deformation block. The numbers of control points along each dimension were 6, 12, and 16, respectively for the 3 deformation blocks.

Probability Sampling of Image Features

Effective sampling of the image features is essential for training a dense volumetric feature encoder. We randomly sample 16 points on the whole heart per control point based on a normal distribution centered at each control points with the covariance determined by the grid resolution (Fig 4.1). In each FFD block, we update the coordinates of sampled points based on FFD, sample image features at these coordinates and then compute the expectation of image features over the sampled points for each grid point. These image features are then concatenated with the grid features for displacement prediction using GCN. Control points in the low-resolution grid thus have a larger field of selection than those in the high-resolution grid. Fig 4.1 visualizes the probability distribution and sampled points correspond to one control point from control grid at different resolutions.

Loss functions

The training of our networks was supervised by 3D ground truth meshes of the whole heart as well as a binary segmentation indicating occupancy of the heart on the voxel grid that corresponds to the input image volume. The total loss is a weighted combination of a point loss, a grid elasticity loss and a segmentation loss over all deformation blocks B_i. Namely,

$$\mathcal{L}_{total} = \sum_{b}^{3} \mathcal{L}_{point}(\mathbf{P}^{B_i}, \mathbf{G}^{B_i}) + \alpha_1 \sum_{b}^{3} \mathcal{L}_{grid}(\Delta \mathbf{C}^{B_i}) + \alpha_2 \mathcal{L}_{seg}(I_p, I_g). \tag{4.3}$$

We used $\alpha_1 = 100$ selected from 10, 100, 1000 based on the validation accuracy. α_2 was initially set to 200 (selected from 10, 100 and 200) and decreased by 5% every 5 epochs during training. We use the Chamfer loss as the point loss

$$\mathcal{L}_{point}(\mathbf{P}_i, \mathbf{G}_i) = \sum_{\mathbf{p}\in\mathbf{P}_i} \min_{\mathbf{g}\in\mathbf{G}_i} ||\mathbf{p}-\mathbf{g}||_2^2 + \sum_{\mathbf{g}\in\mathbf{G}_i} \min_{\mathbf{p}\in\mathbf{P}_i} ||\mathbf{p}-\mathbf{g}||_2^2 \tag{4.4}$$

, where $\mathbf{p}$ and $\mathbf{g}$ are, respectively, points from vertex sets of the predicted mesh $\mathbf{P}_i$ and the ground truth mesh $\mathbf{G}_i$ of cardiac structure i. Since excessive deformation of the control points especially during early phase of training may introduce undesirable mesh artifacts, we use the grid elasticity loss to regularize the network to predict small displacements of the control points,

$$\mathcal{L}_{grid}(\Delta\mathbf{C}) = \sum_{\mathbf{c}\in\mathbf{C}} ||\Delta\mathbf{c} - \frac{1}{N}\sum_{\mathbf{c}\in\mathbf{C}} \Delta\mathbf{c}||_2^2. \tag{4.5}$$

We used a hybrid loss function $\mathcal{L}_{seg}(I_p, I_g)$ that sums the cross-entropy and the dice losses between the predicted occupancy probability map I_p and the ground truth binary segmentation of the whole heart I_g. The validation loss converged in 36 hrs on a GTX1080Ti GPU.

Cardiac flow simulation

We applied the Arbitrary Lagrangian-Eulerian (ALE) formulation of the incompressible Navier-Stokes equations to simulate the intraventricular flow and account for deforming volumetric mesh using the finite element method. The volumetric mesh was created automatically from our FFD predicted surface mesh using TetGen [139]. Blood was assumed to have a viscosity μ of $4.0 \times 10^{-3} Pa \cdot s$ and a density ρ of $1.06 g/cm^3$. The equations were solved with the open-source svFSI solver from the SimVascular project [155].

4.3 Experiments and Results

Dataset and Preprocessing

We applied our method to public datasets of contrast-enhanced CT images from both normal and abnormal hearts and mostly cover the whole hearts, MMWHS [182], orCalScore [169] and SLAWT [57]. Intensity normalization and resizing as well as data augmentation techniques, random scaling, rotation, shearing and elastic deformation were applied following the procedures in [68]. The training and validation datasets contained 87 and 15 CT images, respectively. The 40 CT images from MMWHS test dataset and 10 sets of times-series CT data [68] were left out for evaluations. The ground truth labels include the 4 heart chambers, aorta, pulmonary artery, parts of the pulmonary veins and venae cavae for the training and validation data.

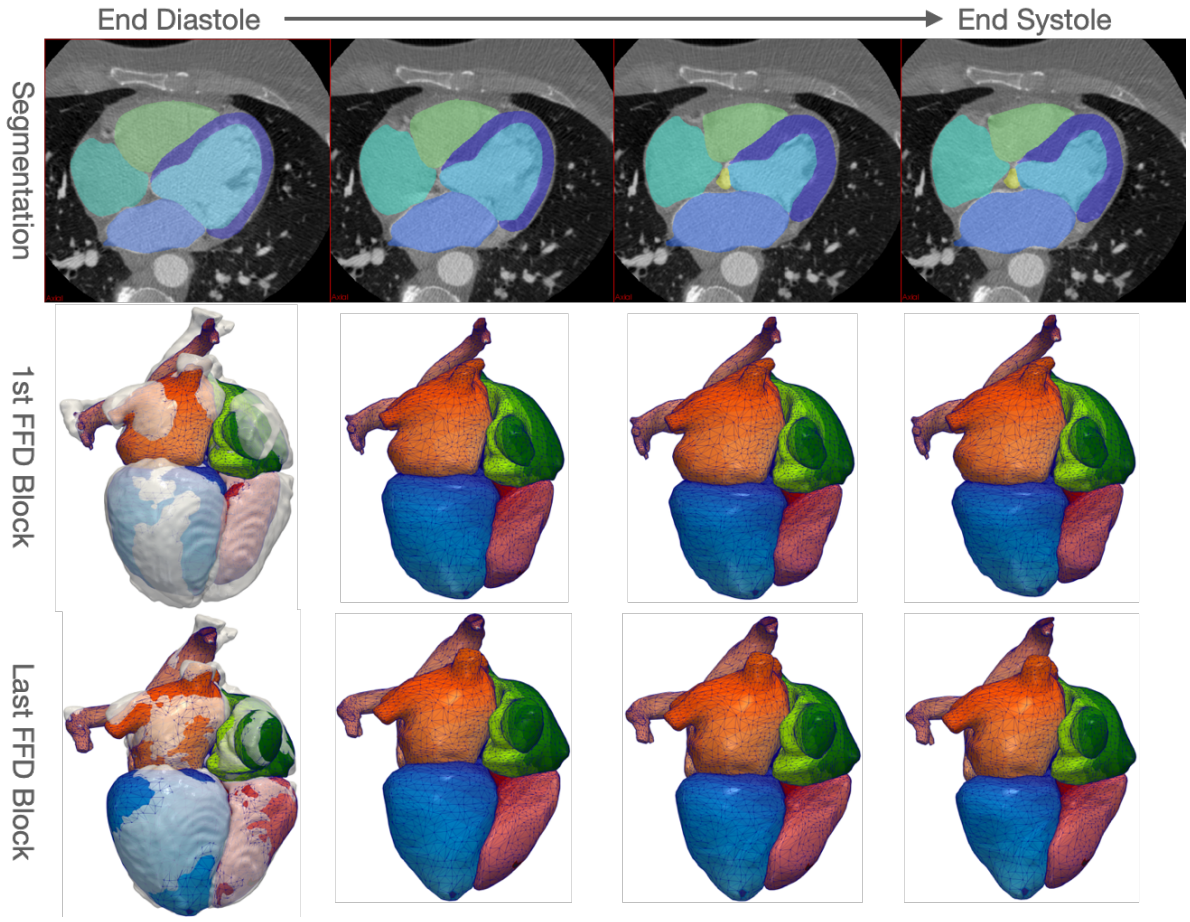


Figure 4.2: Whole-heart reconstruction results for time-series CT data. The first row shows predicted segmentation overlaid with CT images. The second and third rows shows mesh predictions from the first and the last deep FFD blocks. The predictions at the first time frame are overlaid with ground truths. Color maps denotes the mesh vertex IDs.

Generation of 4D Meshes for CFD Simulations

We applied our method on time-series CT image data that consisted of images from 10 time frames over the cardiac cycle for each patient. Fig 4.2 compares the predictions from the first and last FFD blocks. A low-resolution control point grid can capture the general shape and location of the heart in the image stack whereas the high-resolution control point grid can capture further detailed features. From the segmentation results in Fig 4.2, our method is able to capture the minor changes between time frames. Furthermore, as denoted by the color maps of vertex IDs, our method consistently deforms the template meshes such that predictions across different time frames have feature correspondence.

We also evaluated the potential of our method to generate simulation-ready meshes for deforming-domain CFD simulations of cardiac flow. We used a simulation ready template

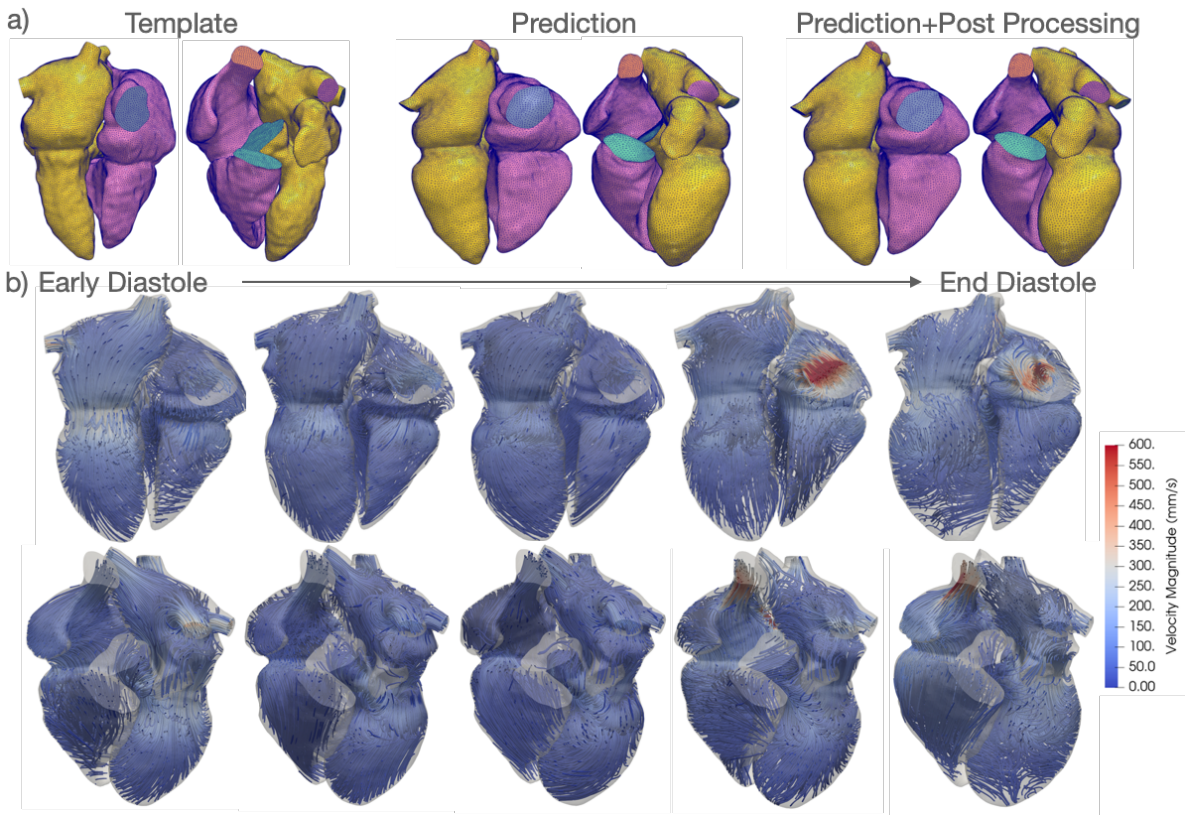


Figure 4.3: a) Simulation-ready templates and example predictions. b) CFD simulation results using the predicted 4D meshes.

with trimmed inlet/outlet geometries and tagged face IDs for prescribing boundary conditions. Fig 4.3a shows the template mesh and our prediction for a representative patient. As our method does not constrain the faces of inlets or outlets to be co-planer, we post-processed our prediction using automatic scripts to project the points on the tagged inlet and outlet faces to fitted planes. We used the resulting meshes to simulate the filling phase of heart after interpolating the 4D meshes to increase the temporal resolution to 0.001s. For the fluid domain, Dirichlet (displacements) boundary conditions were applied on the chamber walls as well as on aorta and pulmonary outlets, while Neumann (pressure) boundaries conditions were applied on pulmonary vein and vena cava inlets. Fig 4.3b displays the simulation results of the velocity streamlines at multiple time steps during diastole. Videos of the predicted meshes and simulation results of more cases are in our supplementary materials.

Comparison of different methods

We compared the whole heart reconstruction performance of different FFD strategies, namely, 1) using the same resolution of control point grid for all deformation blocks as in [161] and

		Epi	LA	LV	RA	RV	Ao	PA	WH
Dice	Singe-Res+US	0.72±0.09	0.82±0.08	0.81±0.07	0.79±0.07	0.81±0.06	0.78±0.09	0.69±0.14	0.79±0.05
	Multi-Res+US	0.81±0.06	0.89±0.05	0.88±0.07	0.84±0.07	0.86±0.05	0.87±0.07	0.76±0.13	0.86±0.04
	Multi-Res+WHS	**0.84±0.05**	**0.91±0.04**	**0.89±0.07**	**0.86±0.06**	**0.88±0.04**	**0.91±0.04**	**0.8±0.1**	**0.88±0.03**
ASSD (mm)	Singe-Res+US	2.28±0.76	2.61±0.97	2.72±1.02	2.87±0.93	2.42±0.56	2.41±0.96	3.55±1.62	2.69±0.69
	Multi-Res+US	1.61±0.43	1.5±0.56	1.62±0.66	2.09±0.92	1.59±0.4	1.29±0.52	2.5±1.38	1.75±0.41
	Multi-Res+WHS	**1.41±0.38**	**1.4±0.47**	**1.46±0.68**	**1.87±0.84**	**1.49±0.42**	**1.08±0.35**	**2.19±1.17**	**1.54±0.34**
HD (mm)	Singe-Res+US	15.44±2.42	11.54±3.47	10.45±3.13	15.22±5.35	12.02±2.5	14.9±7.32	26.41±11.49	27.87±10.59
	Multi-Res+US	14.45±2.49	9.25±2.92	8.04±2.25	13.55±5.84	10.86±2.59	12.54±5.53	**23.97±12.62**	**25.67±11.51**
	Multi-Res+WHS	**13.51±2.59**	**8.58±2.87**	**7.66±2.61**	**12.75±5.46**	**10.09±2.48**	**12.24±6.86**	24.79±12.52	26.76±11.17

Table 4.1: A comparison of prediction accuracy on MMWHS CT test datasets from different deep FFD methods.

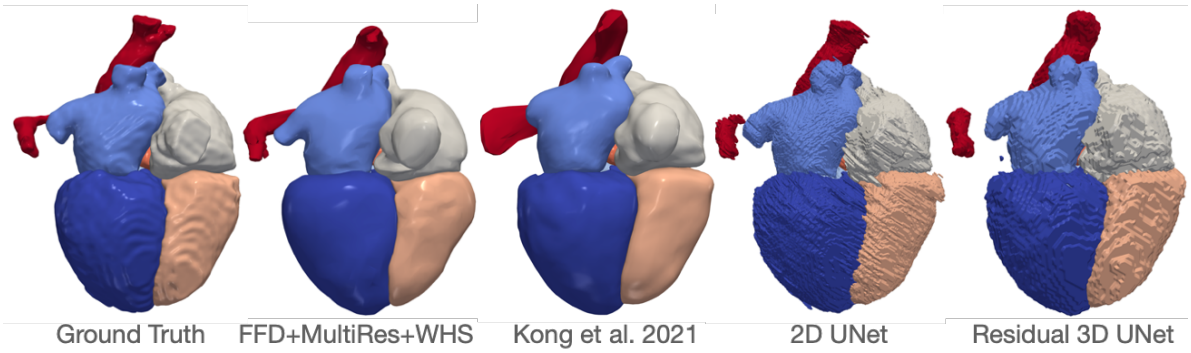


Figure 4.4: Qualitative comparisons among different methods

uniformly sample the image feature space based the the coordinates of the control points (Single-Res + US), 2) using multi-resolution control point grids and uniformly sample the image feature space based on the coordinates of the high resolution grid (Multi-Res + US) and 3) our final model that uses multi-resolution control point grids and probability sampling of the image feature space based on the coordinates of the whole heart template (Multi-Res + WHS). Furthermore, we compared these FFD-based methods with prior whole-heart reconstruction or segmentation methods, Kong et al [68], 2DUNet [128, 66], residual 3D UNet [49] and Voxel2Mesh[168]. We followed procedures described in [68] to implement those methods.

Table 4.1 presents the accuracy scores of different FFD-based methods evaluated on the MMWHS test dataset. Since the ground truths of MMWHS test dataset do not contain the pulmonary veins or venae cavae, we used a template without those structures to generate our predictions. For FFD-based methods, Multi-Res consistently produced more accurate geometries for all cardiac structures (Table 4.1). Multi-Res + WHS produced more accurate geometries with less surface artifacts than Multi-Res + US, indicating the contribution of our proposed sampling method. Fig 4.4 displays qualitative whole-heart reconstruction results from different methods. Mesh-deformation based methods, ours and [68], are able to generate smoother and more anatomically consistent geometries compared with segmentation-

		Epi	LA	LV	RA	RV	Ao	PA	WH
Dice	FFDNet	0.839	0.905	0.894	0.863	0.877	0.906	0.802	0.878
	Kong *et al.* [68]	0.899	0.932	0.940	0.892	0.910	0.950	0.852	0.918
	2DUNet [128]	0.899	0.931	0.931	0.877	0.905	0.934	0.832	0.911
	3DUNet [49]	0.863	0.902	0.923	0.868	0.876	0.923	0.813	0.888
	Voxel2Mesh [168]	0.775	0.888	0.910	0.857	0.885	0.874	0.758	0.865
Jaccard	FFDNet	0.725	0.829	0.814	0.764	0.783	0.831	0.680	0.784
	Kong *et al.* [68]	0.819	0.875	0.888	0.809	0.837	0.905	0.755	0.849
	2DUNet [128]	0.817	0.872	0.873	0.787	0.828	0.879	0.726	0.837
	3DUNet [49]	0.762	0.825	0.861	0.769	0.783	0.860	0.695	0.799
	Voxel2Mesh [168]	0.638	0.801	0.839	0.754	0.795	0.778	0.619	0.763
ASSD (mm)	FFDNet	1.406	1.404	1.455	1.870	1.488	1.080	2.193	1.544
	Kong *et al.* [68]	1.335	1.042	0.842	1.583	1.176	0.531	1.904	1.213
	2DUNet [128]	0.808	1.049	0.905	1.719	1.064	0.645	1.551	1.088
	3DUNet [49]	1.443	1.528	1.024	1.943	1.663	0.814	2.194	1.552
	Voxel2Mesh [168]	1.714	1.696	1.266	2.020	1.492	1.341	3.398	1.848
HD (mm)	FFDNet	13.508	8.584	7.665	12.753	10.089	12.243	24.794	26.765
	Kong *et al.* [68]	14.393	10.407	10.325	13.639	13.360	9.407	26.616	28.035
	2DUNet [128]	9.980	8.773	6.098	13.624	10.016	10.013	27.834	28.727
	3DUNet [49]	13.635	10.814	9.580	16.031	15.635	13.326	26.941	31.088
	Voxel2Mesh [168]	13.564	8.743	6.248	12.116	9.601	12.080	26.252	27.459

Table 4.2: Quantitative comparison of the geometric accuracy of the reconstruction results between the proposed method and prior methods, [68], [128], [49] and [168] on MMWHS CT test dataset. Detailed implementation of the prior methods has been described in [68] and we used the same image pre-processing and training data for all methods. Our method is able to provide similar or slightly reduced level of accuracy compared with prior methods while having the additional advantage of directly support various cardiac simulations.

based approaches, which produced surfaces with staircase artifacts and disconnected regions (Fig 4.4). However, Kong, et al., 2021 produced overly smoothed pulmonary veins and vena cava geometries, likely because these are elongated structures and that method deforms spheres rather than a more fitting template of each structure as used in our method here. Quantitatively, our method is generally able to produce similar geometric accuracy compared with prior state-of-the-art methods, while having the additional advantage of directly support various cardiac simulations 4.2. Nonetheless, we did observe slightly reduced level of accuracy (see supplementary materials), likely because it is challenging to use a single whole-heart template to fully capture the geometric variations across patients.

We conducted an ablation study of individual loss components and the use of GCN decoder. As table 4.3 shows, removing L_{grid} or L_{seg} as well as using a classical $3 \times 3 \times$ convolution resulted in drop of geometric accuracy of the whole heart reconstruction on MMWHS CT test dataset. Table 4.4 shows that our method is robust to different initialization and data splits.

		Epi	LA	LV	RA	RV	Ao	PA	WH
Dice	Final FFDNet	0.839	0.905	0.894	0.863	0.877	0.906	0.802	0.878
	CNN Decoder	0.806	0.872	0.874	0.826	0.846	0.843	0.764	0.845
	No Grid Loss	0.772	0.889	0.855	0.833	0.850	0.879	0.772	0.843
	No Segmentation Loss	0.781	0.884	0.859	0.837	0.851	0.874	0.780	0.845
Jaccard	Final FFDNet	0.725	0.829	0.814	0.764	0.783	0.831	0.680	0.784
	CNN Decoder	0.679	0.777	0.781	0.710	0.737	0.736	0.628	0.734
	No Grid Loss	0.636	0.804	0.755	0.720	0.743	0.790	0.642	0.731
	No Segmentation Loss	0.647	0.796	0.761	0.726	0.745	0.783	0.651	0.734
ASSD (mm)	Final FFDNet	1.406	1.404	1.455	1.870	1.488	1.080	2.193	1.544
	CNN Decoder	1.709	1.782	1.668	2.281	1.846	1.650	2.470	1.898
	No Grid Loss	1.899	1.574	2.009	2.165	1.770	1.257	2.599	1.904
	No Segmentation Loss	1.816	1.638	1.865	2.084	1.711	1.250	2.473	1.832
HD (mm)	Final FFDNet	13.508	8.584	7.665	12.753	10.089	12.243	24.794	26.765
	CNN Decoder	14.076	9.713	7.694	14.037	10.905	12.886	24.533	26.413
	No Grid Loss	14.330	9.521	10.086	14.163	12.013	12.607	25.361	27.688
	No Segmentation Loss	14.268	9.122	9.154	13.567	10.792	13.408	23.862	26.036

Table 4.3: Ablation study: a comparison of prediction accuracy on MMWHS CT test dataset after removing grid loss L_{grid}, removing segmentation loss L_{seg}, and using conventional convolution with $3 \times 3\times$ filters rather than graph convolution in the decoder.

	Epi	LA	LV	RA	RV	Ao	PA	WH
Dice	0.835±0.004	0.905±0.003	0.895±0.001	0.86±0.004	0.878±0.005	0.904±0.004	0.813±0.007	0.878±0.002
Jaccard	0.72±0.005	0.829±0.004	0.815±0.002	0.76±0.005	0.786±0.007	0.827±0.008	0.697±0.01	0.784±0.004
ASSD (mm)	1.459±0.036	1.404±0.035	1.439±0.015	1.932±0.043	1.468±0.043	1.005±0.054	2.077±0.102	1.536±0.031
HD (mm)	13.954±0.264	8.61±0.132	7.647±0.014	12.971±0.226	10.447±0.228	11.753±0.758	23.439±0.815	25.592±0.722

Table 4.4: Mean geometric accuracy measures and their standard deviations on the MMWHS CT test dataset for our proposed FFDNet trained using 5 different random initialization and training/validation data splits.

4.4 Conclusion

We proposed a novel deep-learning approach to directly construct whole heart meshes from image data. We learn to deform a template mesh to match the input data by predicting displacements of multi-resolution control point grids. To our knowledge, this is the first approach that is able to directly generate whole heart meshes for computational simulations and allows switching template meshes to accommodate different modeling requirements. We demonstrated application of our method on constructing a dynamic whole heart mesh from time-series CT image data to simulate the cardiac flow driven by the cardiac motion. Our method was able to construct such meshes within a minute on a standard desktop computer (3 GHz Intel Core i5 CPU) whereas prior methods can take hours of time and human efforts to generate simulation-ready 4D meshes.

Chapter 5

Learning Whole Heart Mesh Generation From Patient Images For Computational Simulations

5.1 Introduction

The previous chapter combined free-form deformation (FFD) with deep learning to predict the displacement of a control point grid to deform the space enclosing a simulation-ready whole heart template[65]. However, predicting the deformation fields requires many degrees of freedom to produce accurate results. Namely, the control handles are distributed in the physical space, rather them directly on the template meshes. The deformation is thus biased by the initial locations of mesh vertices in the Euclidean space rather than on the manifold. For example, cardiac structures that are close together in the template are not freely separable during the deformation. Consequently, our prior method required a dense grid of thousands of control points to achieve acceptable whole heart reconstruction accuracy. Herein we demonstrate that using control-handle-based deformation with biharmonic coordinates achieves higher reconstruction accuracy while using far fewer control points than the FFD-based approach.

Shape deformation using low-dimensional control of deformation fields has been extensively studied for decades in computer graphics and has been ubiquitously used in animated characters. These methods usually interpolate the transformation of a sparse set of control points to all points on the shape. Besides the free-form-deformation that uses a regular control point lattice to deform the shape enclosed within the lattice [133], more recent approaches include cage-deformation that uses a convex control cage that encloses the shape [105], as well as control-handle-based approaches that directly place control points on the surface of the shape [142, 52, 164]. These approaches addressed some of the abovementioned drawbacks of FFD and enabled more flexible shape deformations.

Recent studies have shown success in integrating these shape deformation methods in

deep-learning frameworks for automated mesh reconstruction from single-view camera images[72], generative shape modeling[83] as well as deformation transfer [163]. However, these approaches were designed to take 2D camera images or 3D meshes as input and used memory-intensive CNNs or fully connected neural networks to predict the transformation of control points. They thus cannot be directly applied to deform complicated whole heart structures from high-resolution 3D medical image data. We propose to use graph convolutional networks (GCN) and sparsely sample the volumetric image feature map to predict control point translations and thus can efficiently produce meshes from 3D medical images.

To overcome these shortcomings, in this chapter, we propose to learn to deform the space enclosing a whole heart template mesh to automatically and directly generate meshes that are suitable for computational simulations of cardiac function. Here we propose to leverage a control-handle-based shape deformation method to parameterize the smooth deformation of the template with the displacements of a small set of control handles and their biharmonic coordinates. Our approach learns to predict the control handle displacements to fit the whole heart template to target image data. We also introduce a few effective learning biases as objective functions to produce meshes that better satisfy the modeling requirements for computational simulation of cardiac flow. The contributions of this work are summarized as follows:

1. We propose a novel end-to-end learning method combining deformation handles to predict the deformation of whole heart mesh templates from volumetric patient image data. We show that our approach achieves comparable geometric accuracy for whole heart segmentation as prior state-of-the-art segmentation methods.

2. We introduced novel mesh regularization losses on vessel inlet and outlet structures to better satisfy the meshing requirements for CFD simulations. Namely, our method predicts meshes with coplanar vessel caps that are orthogonal to vessel walls for CFD simulations.

3. We validated our method for creating 4D dynamics whole heart and left ventricle meshes for CFD simulation of cardiac flow. Our method can efficiently generate simulation-ready meshes with minimal post-processing to facilitate large-cohort computational simulations of cardiac functions.

5.2 Methods

Shape Deformation Using Biharmonic Coordinates

Our approach constructs whole heart meshes by deforming a pre-defined whole heart mesh template. We parameterize deformations of whole heart meshes with the translations of a small set of deformation handles sampled from the mesh template. Given a set of mesh

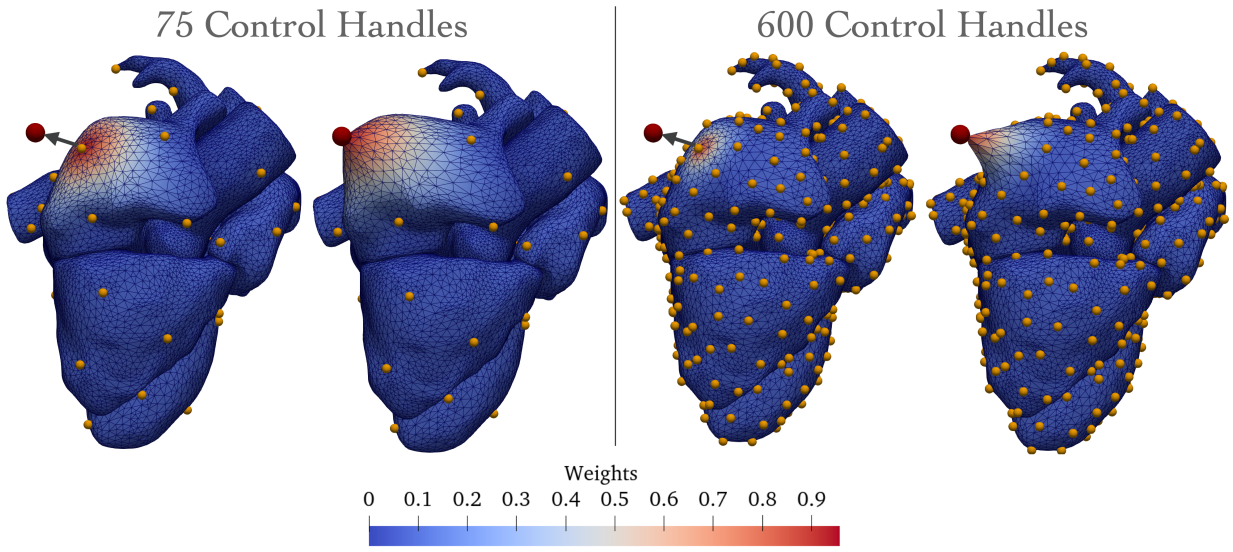


Figure 5.1: Example mesh deformation using control handles and biharmonic coordinates. Color map indicates the weights on mesh vertices corresponding to the translated control handle colored in red.

vertices $V \in \mathbb{R}^{n\times 3}$ and a set of control points $P \in \mathbb{R}^{c\times 3}$, we compute the biharmonic coordinates $W \in \mathbb{R}^{n\times c}$, which is a linear map, $V = WP$. W is defined based on biharmonic functions and can be pre-computed by minimizing a quadratic deformation energy function while satisfying the handle constraints with linear precision [164]. Namely, let $Q \in \mathbb{R}^{c\times n}$ be the binary selector matrix that selects rows of X corresponding to the control handles, and let $T \in \mathbb{R}^{(n-c)\times n}$ be the complementary selector matrix of Q corresponding to the free vertices. W is computed by

$$V = \arg\min_{X\in\mathbb{R}^{n\times 3}} \frac{1}{2}\text{trace}(X^T A X), \quad \text{subject to } QX = P \tag{5.1}$$

$$V = \underbrace{(Q^T - T^T(TAT^T)^{-1}TAQ^T)}_{W} P \tag{5.2}$$

where A is positive semi-definite quadratic form based on the squared Laplacian energy to encourage smoothness [164]. Under this framework, displacements of the control handles can smoothly deform the underlying mesh template. Figure 5.1 visualizes the deformation of a whole heart mesh template by moving a single control handle using biharmonic coordinates. The density of control handles affects the range of mesh vertices that each handle can control. That is, a higher control handle density (Figure 5.1, right) allows more localized deformation of the template, while fewer control handles (Figure 5.1, left) spreads the deformation over a larger area.

Network Architecture

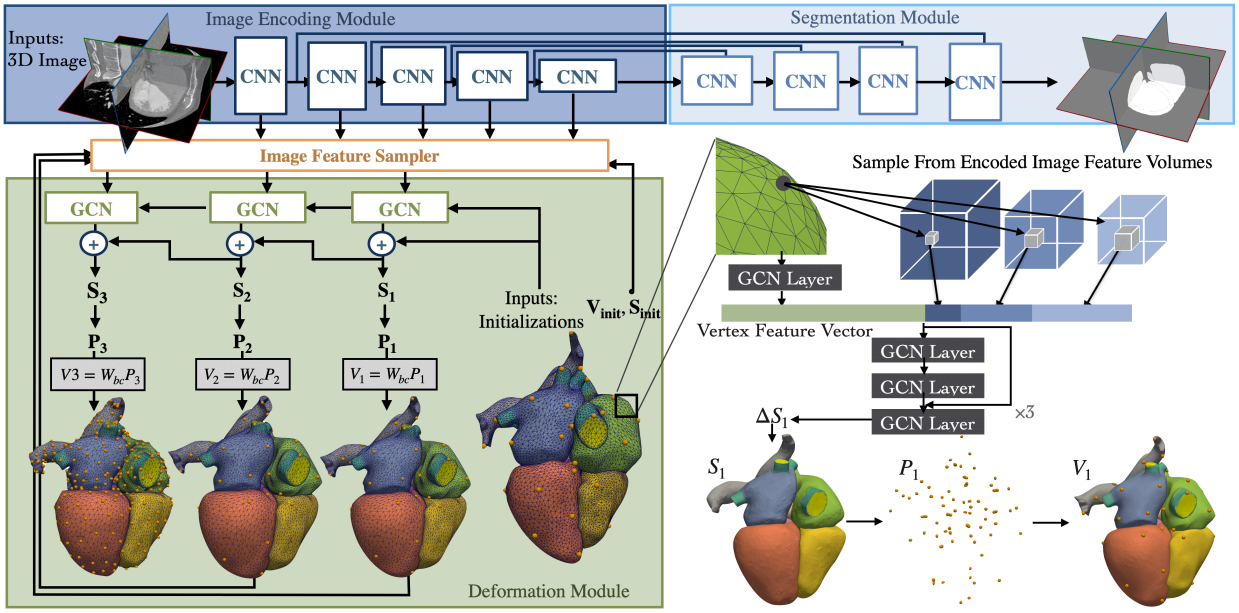


Figure 5.2: Diagram of the proposed automatic whole heart reconstruction approach. A total of three deformation blocks were used to progressively deform the mesh templates, using increasing number of control handles (75, 75 and 600, respectively).

Figure 5.2 shows the overall architecture of our mesh reconstruction network. The central architecture is the novel control-handle-based mesh deformation module, which learns to predict the displacements of control handles based on image features, so that the underlying mesh templates can be smoothly deformed to match the input 3D image data.

Image Encoding and Segmentation Modules

We applied a residual 3D CNN backbone to extract and encode image features at multiple resolutions [49]. The CNN backbone involves 4 down-sampling operations so that image feature volumes at 5 different resolution are obtained. These image features volumes are used as inputs to GCN layers to predict the displacements of control handles. Similar to [68], we also use a segmentation module that predicts a binary segmentation map to enable additional supervision using ground truth annotations. This module is only used during training and can be removed during testing.

Mesh Deformation Module

Biharmonic coordinates constrain the displaced control handles to be located on the deformed mesh template. Therefore, regardless of which set of control handles are sampled, these handles will be located at the corresponding positions on the template mesh. We use a neural network to update the coordinates of all points ($S \in \mathbb{R}^{n\times 3}$) on the mesh template and obtain the coordinates of the selected control handles ($P \in \mathbb{R}^{c\times 3}$) from the updated mesh vertex locations to deform the template using the pre-computed biharmonic coordinates. This design allows picking arbitrary sets of control handles to deform the template at various resolutions after training. Furthermore, training to predict the coordinates of every mesh vertex provides additional supervision that can speed up training.

Since the mesh template can be represented as a graph, a GCN was used to predict the mesh vertex displacements. We chose to approximate the graph convolutional kernel with a first order Chebyshev polynomial of the normalized graph Laplacian matrix [25]. At each mesh vertex, we extract the image feature vectors at the corresponding image coordinates from multiple image feature volumes with various resolution. These image feature vectors are then concatenated with the mesh feature vectors following a GCN layer. The combined vertex feature vectors are then processed by three graph residual blocks. We then use an additional GCN layer to predict displacements as 3D feature vectors on mesh vertices.

Network Training

The training of the networks was supervised by 3D ground truth meshes of the whole heart as well as a binary segmentation indicating occupancy of the heart on the input image grid. We used the following loss functions (illustrated in figure 5.3) in training to produce accurate whole heart geometries while ensuring the resulting mesh is suitable to support simulations.

Geometric Consistency Losses

The geometric consistency loss $\mathcal{L}_{geo}$ is the geometric mean between the point and normal consistency losses to supervise the geometric consistency between the prediction and the ground truth[68]. We note that edge length and Laplacian regularization losses, such as used in [68], are not necessary since the smoothness of the mesh prediction is naturally constrained by the biharmonic coordinates used to deform the template. Since only the selected control points were used to deform the mesh template while the displacements of all mesh points were predicted, we need to regularize the $L2$ distances between the mapped mesh points ($S \in \mathbb{R}^{n\times 3}$) and the corresponding mesh vertices on the deformed mesh template ($V \in \mathbb{R}^{n\times 3}$). This consistency loss between the points and the mesh ensures that coordinates of other unselected control points also result in reasonable deformations. In other words, the deformation results should not be sensitive to the choice of pre-selected control points.

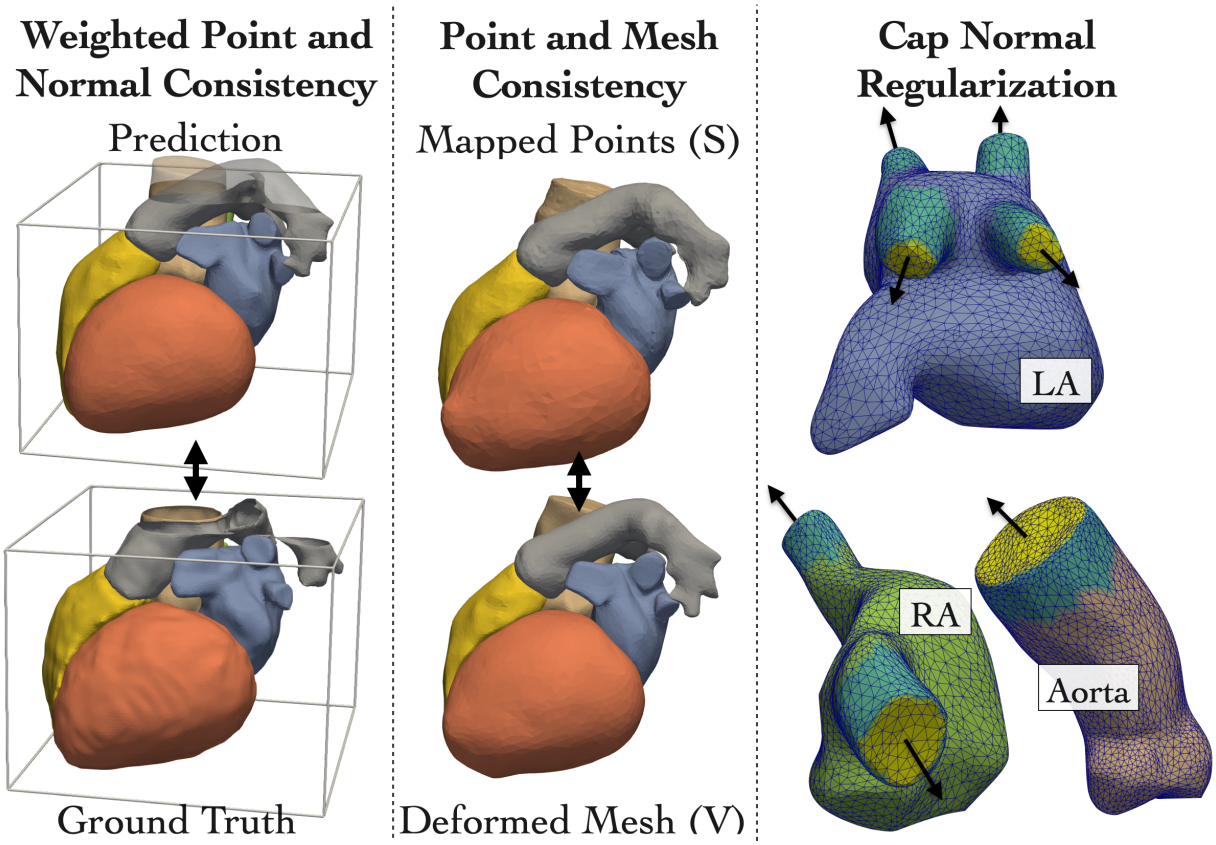


Figure 5.3: Graphical illustration of different loss functions. Yellow and teal on the right panel shows the caps and walls to apply the mesh regularization losses, respectively, and arrows shows cap normal vectors.

Mesh Regularization for CFD Simulations

Cardiac models generally included portions of the great vessel connected to the heart (e.g., pulmonary veins and arteries, venae cavae, and aorta). For CFD simulation of cardiac flow, locations where these vessels are truncated (so-called inflow and outflow boundaries, or "caps") should be planar and nominally orthogonal to the vessel. On our training template, we labeled these caps, as well as the associated vessel walls. The caps were identified semi-automatically on the template by manually selecting seeds on the centers of the caps and then automatically selecting connected mesh faces with normal vectors close to that of the seed point. Figure 5.3 shows the identified cap and wall faces on left atrium (LA), right atrium (RA) and aorta. We applied a co-planar loss on each cap that penalizes the $L2$

differences of the surface normals on the cap. Namely,

$$\mathcal{L}_{coplanar} = \sum_{k} \sum_{j \in \mathcal{C}_k} ||\mathbf{n}_j - \frac{1}{|C_k|} \sum_{j \in \mathcal{C}_k} \mathbf{n}_j||_2^2 \tag{5.3}$$

where C_k is the set of mesh faces for the kth cap and $\mathbf{n}_j$ is the normal vector for the jth face on C_k. For mesh vertices that are on the vessel walls near the caps, we minimize the absolute value of the dot products between their surface normal vectors and the surface normal vector of the caps to encourage orthogonality. Namely,

$$\mathcal{L}_{orthogonal} = \sum_{k} \sum_{j \in \mathcal{W}_k} |\langle \mathbf{n}_k, \frac{1}{|C_k|} \sum_{p \in \mathcal{C}_k} \mathbf{n}_p \rangle| \ , \tag{5.4}$$

where $\mathcal{W}_k$ is the set of mesh faces on the vessel wall that corresponds to the kth cap.

Weighted Mesh Losses

Patient images may not always contain the targeted cardiac structures. As shown in Figure 5.3 (left), cardiac structures such as, pulmonary veins, pulmonary arteries and the aorta are often not captured in full, although these structures can be necessary for simulations. We thus aim to predict a complete whole heart shape from incomplete image data. Namely, we compute the bounding box of the ground truth meshes and assign zero weights within the geometric consistency loss for predicted mesh vertices that located outside the bounding box. Furthermore, as the geometry of inlet vessels are important to the accuracy of CFD results, we applied a higher weight for the geometric consistency loss on mesh vertices that are located on vessel walls near the inlets.

Total Losses

The total loss on a predicted mesh M is

$$\mathcal{L}_{mesh}(M, G, W) = \sum_{i} \mathcal{L}_{geo}(M_i, G_i, \mathbf{w}_i) + \alpha \sum_{k} (\mathcal{L}_{coplanar}(M) + \beta \mathcal{L}_{orthogonal}(M)) \tag{5.5}$$

where G_i represents the ground truth mesh for individual cardiac structure, and $\mathbf{w}_i$ represents the weighting vector for each mesh point. M can be both the deformed whole heart mesh template V and the mesh obtained from mapping all mesh points S. The total loss for training is a weighted sum of the mesh losses and the segmentation loss, which is the sum of the binary cross-entropy and the dice losses between the predicted occupancy probability map I_p and the ground truth binary segmentation of the whole heart I_g.

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{mesh}(S, G, W) + \lambda_2 \mathcal{L}_{mesh}(V, G, W) + \lambda_3 ||S - V||_F^2 + \mathcal{L}_{seg}(I_g, I_p) \tag{5.6}$$

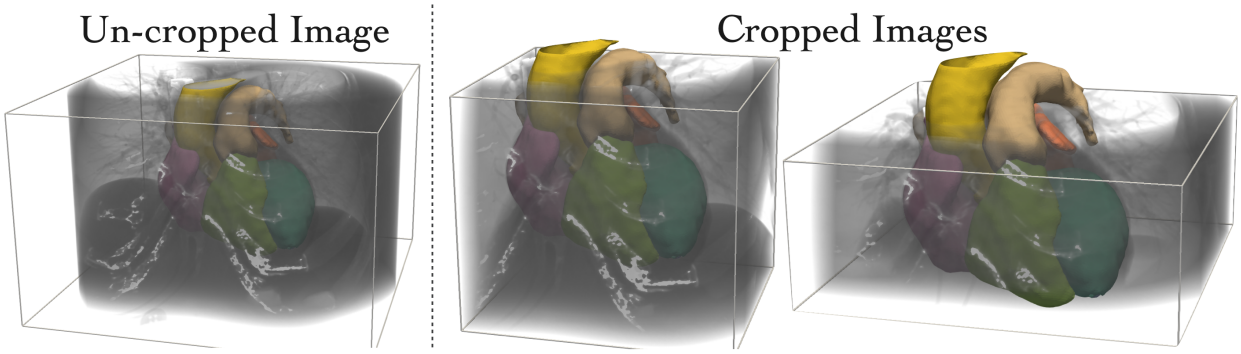


Figure 5.4: Visualization of augmented input image crops and the corresponding ground truth meshes.

Image Augmentation for Shape Completion

Leveraging the mesh template, we train our method to predict the geometries of the whole heart represented by the template mesh when the images do not cover the complete cardiac structures. Since CT images often do not cover the whole heart, we selected CT images that did cover the whole heart (n=10) from our training set and then generated 10 random crops for each image while keeping the ground truth meshes to be the same. Figure 5.4 visualizes example image crops and the corresponding ground truth meshes. We also applied random shearing, rotations and elastic deformations, following the same augmentation strategies described in our prior work [68].

Deforming-Domain CFD simulations of Cardiac Flow

We applied the Arbitrary Lagrangian-Eulerian (ALE) formulation of the incompressible Navier-Stokes equations to simulate the intraventricular flow and account for deforming volumetric mesh using the finite element method. For the fluid domain, displacements boundary conditions were applied on the chamber walls as well as on vessel outlets, while pressure boundaries conditions were applied on vessel inlets. Blood was assumed to have a viscosity μ of $4.0 \times 10^{-3} Pa \cdot s$ and a density ρ of $1.06 g/cm^3$. T he volumetric mesh was created automatically from our predicted surface mesh using TetGen [139]. The equations were solved with the open-source svFSI solver from the SimVascular project [155].

5.3 Experiments and Results

Datasets and Experiments

Task-1: Whole Heart Segmentation for 3D Images

We applied our method to public datasets of contrast-enhanced 3D CT images and 3D MR images from both normal and abnormal hearts. Namely, 87 CT images and 41 MR images from the multi-modality whole heart segmentation challenge (MMWHS)[182], orCalScore challenge[169], left atrial wall thickness challenge [57] and left atrial segmentation challenge [148] were used for training. 15 CT images and 6 MR images were used for validation and hyper-parameter tuning. The geometric accuracy of the whole heart reconstruction was first evaluated on the 40 CT and 40 MR images from the MMWHS test dataset. We followed the MMWHS dataset to create ground truth segmentation of seven cardiac structure including the blood cavities of left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), LV myocardium (Epi), aorta (Ao), and pulmonary artery (PA) for all images in our training and validation datasets.

We compared the geometric accuracy of the reconstructed whole heart surfaces against prior deep-learning methods that demonstrated promising performance of segmenting whole heart geometries from 3D medical images. Namely, we considered HeartFFDNet[65], our prior work that generates simulation-ready whole heart surface meshes from images by learning free-form deformation from a template mesh, MeshDeformNet [68] that predicts displacements on sphere mesh templates, as well as 2D UNet [128] and a residual 3D UNet [49] that are arguably the most successful neural network architecture for image segmentation. All methods were trained on the same dataset and used the same pre-processing and augmentation procedures to ensure a fair comparison.

Task-2: Whole Heart Mesh Construction for 4D Images

We applied our method on time-series CT images to evaluate its performance on creating whole heart meshes for CFD simulations. Since the MMWHS dataset does not include pulmonary veins, LA appendage or venae cavae, we prepared another set of ground truth segmentations to include these structures. The geometric accuracy and the mesh quality of the reconstructed meshes for CFD simulations were then evaluated on 10 sets of time-series CT images against the learning-based mesh reconstruction baselines, HeartFFDNet and MeshDeformNet.

Task-3: CFD Simulations

We conducted CFD simulations of cardiac flow using the predicted whole heart meshes from time-series CT images. Since our predicted structure do not contain heart valves, only diastolic flow was simulated. We also conducted CFD simulations for the LV and simulated the LV flow for the entire cardiac cycle. Figure 5.5 visualizes the simulation-ready templates

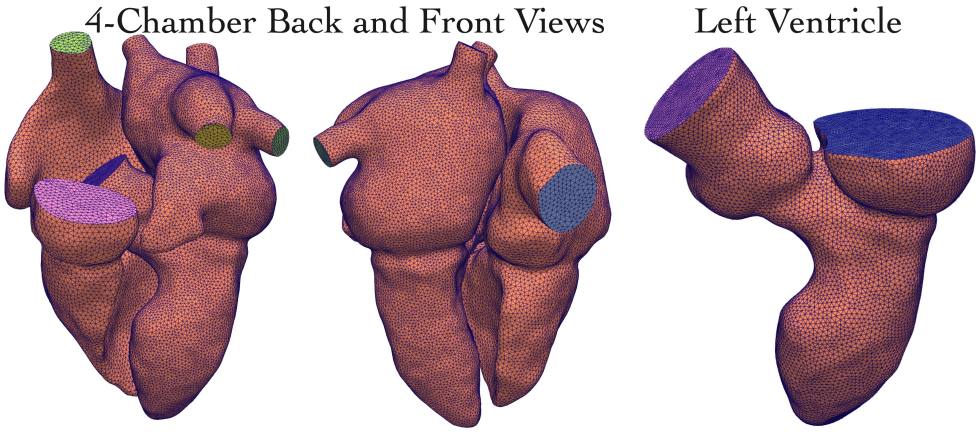


Figure 5.5: Visualization of simulation-ready templates with trimmed inlet/outlet geometries and tagged face IDs for prescribing boundary conditions.

of the 4 heart cambers and the LV with trimmed inlet/outlet geometries and tagged face IDs for prescribing boundary conditions. We linearly interpolated the pre-computed biharmonic coordinates onto the new templates so that our trained models can readily deform these new templates. The simulation results were compared against time-series ground truth meshes created manually in a surface processing software, SimVascular. We also compared simulation results among our method, HeartFFDNet, and a conventional semi-automatic model construction pipeline based on image registration, where a manually created ground truth mesh was morphed based on transformations obtained from registering images across different time points.

Evaluation Metrics

We used Dice similarity coefficient (DSC) and Hausdorff Distance (HD) to measure segmentation accuracy. For mesh-based methods, we converted the predicted surface meshes to segmentation prior to evaluation. Mesh quality was compared in terms of the percentage self-intersection, which measures the local topological correctness of the meshes, orthogonality of the vessel caps with respect to the vessel walls, as well as the coplanarity of the vessel caps. The percentage mesh self-intersection was calculated as the percentage of intersected mesh facets detected by TetGen [139] among all mesh facets. The orthogonality between vessel caps and walls (Cap-Wall-Orthogonality) was measured by the normal consistency between the mean cap normal vector and the vector connecting the centroids of the mesh points on the cap and on the wall, respectively. Namely,

$$CWO = \sum_{k} 1 - \langle \frac{1}{|\mathcal{W}_k|} \sum_{i \in \mathcal{W}_k|} \mathbf{n}_i, \frac{1}{|\mathcal{C}_k|} \sum_{j \in \mathcal{C}_k|} \mathbf{n}_j \rangle \tag{5.7}$$

where $\mathcal{W}_k$ and $\mathcal{C}_k$ represent the sets containing the mesh vertices on a vessel wall and the corresponding vessel cap. Vessel caps coplanarity was measured by the projected distance between the mesh vertices on the cap and the best fit plane over those mesh vertices. For CFD simulations, we compared the integrative measures during a cardiac cycle, namely, LV volume and volume averaged kinetic energy density $KE' = \frac{1}{2V_{LV}} \int\int\int \rho u^2 dV$, where V_{LV} is the volume of the LV and u is the flow velocity. We also compared the mean velocity near the mitral valve opening (MO) and aortic valve opening (AO) during a cardiac cycle. Paired t-test was used for statistical significance.

Comparative Studies on MMWHS Dataset

		CT								MR							
		Epi	LA	LV	RA	RV	Ao	PA	WH	Epi	LA	LV	RA	RV	Ao	PA	WH
DCS	Ours (S)	**90.07**	93.18	93.47	**89.48**	**91.48**	93.33	**85.6**	91.76	**80.45**	86.98	91.61	88.08	88.09	85.76	78.14	87.41
	Ours (V)	88.38*	92.53*	91.99*	88.76*	90.592*	91.25*	84.73*	90.538*	78.62*	86.27*	89.38*	87.79*	87.2*	83.3*	77.55*	86.04*
	HeartFFDNet	83.85*	90.55*	89.38*	86.33*	87.65*	90.65*	80.2*	87.82*	70.67*	83.27*	86.92*	84.47*	82.77*	79.71*	69.68*	81.33*
	MeshDeformNet	89.94*	**93.23**	**93.98**	89.18	91.0	**94.98***	85.22	**91.8**	79.71	**88.13**	**92.23***	**88.82**	**89.24**	**88.98***	**81.6***	**88.17***
	2DUNet	89.87	93.08	93.06*	87.71	90.49*	93.43	83.23*	91.09*	79.47*	86.41	89.61*	85.21	86.48	86.94	77.24	85.94*
	3DUNet	86.34*	90.17*	92.28*	86.77*	87.58*	92.34*	81.29*	88.78*	76.11*	85.2*	87.9*	86.63*	82.77*	74.18*	76.38	84.04*
HD	Ours (S)	14.41	10.72	10.41	13.75	11.63	9.49	26.68	27.92	16.39	12.12	11.23	13.52	13.5	**19.77**	**16.15**	**23.86**
	Ours (V)	14.4*	**8.18**	6.87*	**12.41***	**9.55***	9.5	27.02*	28.23	**15.96***	**10.15***	**8.27***	**12.27***	**11.13***	20.44*	16.4	23.96
	HeartFFDNet	13.51*	8.58*	7.66*	12.75*	10.09*	12.24*	**24.79***	**26.76**	16.59	11.06	10.8	13.55	12.41	22.2*	18.74*	25.27
	MeshDeformNet	14.39	10.41	10.33	13.64	13.36	**9.41**	26.62	28.03	16.92	11.72	10.89	14.81	13.46	22.22	19.34*	27.7*
	2DUNet	**9.98***	8.77*	**6.1***	13.62*	10.02	10.01	27.83	28.73	19.14	10.78	9.96	14.53*	13.08	22.57	16.72	28.35
	3DUNet	13.64	10.81	9.58*	16.03	15.63*	13.33*	26.94	31.09	28.16*	23.64*	21.49*	18.95*	21.09*	37.94*	17.05	43.02*

Table 5.1: Comparison of Whole-Heart Segmentation Performance, DSC (↑)and HD (mm) (↓), from different methods on the MMWHS CT and MR test datasets.* Denotes Significant Difference Of "Ours (S)" From the Others (p-Values¡0.05)

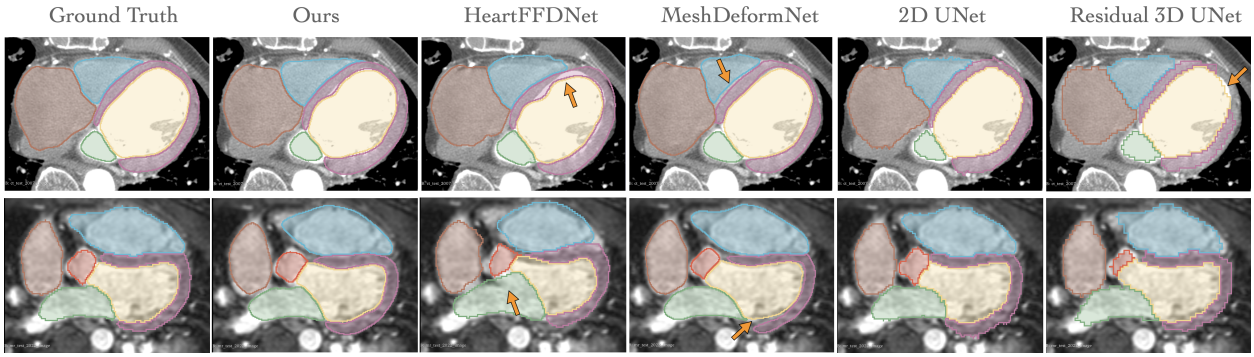


Figure 5.6: Example segmentation results for CT (top row) and MR (bottom row) images from different methods. The CT or MR images that had median Dice scores among the test data were selected, thus illustrating typical segmentation results.

Table 5.1 compares the average Dice scores and Hausdorff distances of the reconstruction results of both the whole heart and the individual cardiac structures for the MMWHS test dataset. For our approach, we show the accuracy of deforming the template by mapping all mesh points (S) and by interpolating the mesh deformation using only 600 uniformly-sampled control handles (V). Mapping all points consistently achieved higher dice scores than using 600 selected control handles, but the HDs are worse for some cardiac structures. For both CT and MR data, in terms of Dice scores, our method consistently outperformed HeartFFDNet and 3D UNet for all cardiac structures and achieved comparable performance with MeshDeformNet and 2D UNet for most cardiac structures. Our method achieved the best HDs for LA, RA and RV for CT data and for all cardiac structures for MR data.

Figure 5.6 presents typical segmentation results of CT and MRI images and provides qualitative comparisons of the results from the different methods. As shown, mesh-based approaches, ours, HeartFFDNet and MeshDeformNet produced smooth and anatomically consistent cardiac geometries while segmentation-based approaches, 2D UNet and residual 3D UNet produced segmentations with stair-case artifacts and missing parts. However, MeshDeformNet produced gaps between adjacent cardiac structures by deforming un-coupled spheres. Our method and the HeartFFDNet were able to avoid this limitation by deforming the space enclosing a whole heart template, preserving the connections among cardiac structures.

Construction of Cardiac Meshes for CFD Simulations

Ablation Study of Individual Loss Components on Vessel Inlet/Outlet Structures

		CoP+ Ortho+HW	CoP+Ortho	CoP	None
Cap-Wall Orthogonality (↓)	LA	0.128±0.121	**0.032**±0.012	0.365±0.265	0.273±0.266
	RA	0.023±0.008	**0.012**±0.008	0.105±0.038	0.066±0.026
	Ao	0.019±0.023	**0.005**±0.006	0.467±0.117	0.127±0.024
Cap Coplanarity (mm) (↓)	LA	0.228±0.041	0.256±0.029	**0.12**±0.024	0.312±0.058
	RA	0.34±0.073	0.339±0.055	**0.185**±0.043	0.466±0.114
	Ao	0.447±0.115	0.429±0.063	**0.263**±0.068	0.852±0.16
Wall Chamfer Distance (mm) (↓)	LA	2.093±0.803	2.715±1.105	2.487±0.898	**2.042**±0.857
	RA	2.021±1.176	2.66±1.301	2.231±0.983	**1.899**±0.952

Table 5.2: Ablation study of mesh regularization losses on vessel inlet and outlet structures on CT test dataset (N=20).

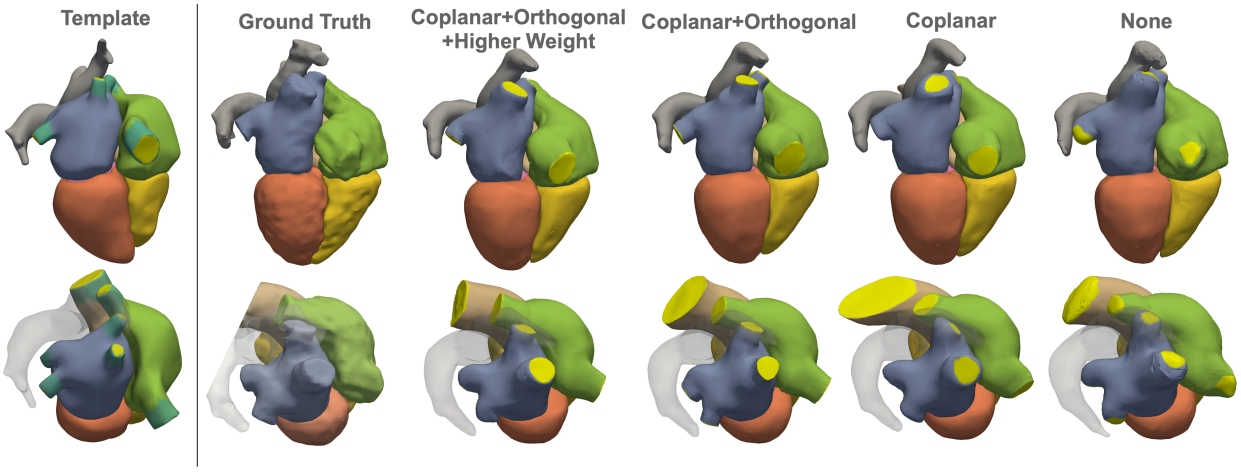


Figure 5.7: Visualization of example whole heart surface predictions following addition of regularization losses on vessel inlet/outlet structures.

CFD simulations of cardiac flow requires well-defined inlet and outlet vessel structures to prescribe boundary conditions for the inflow and outflow. Figure 5.7 and table 5.2 demonstrate the effect of applying individual regularization loss components on the predicted inlet and outlet geometries (pulmonary veins, vena cava, and aorta). The coplanar loss and the orthogonal loss succeeded in producing more planar cap geometries that are more orthogonal to vessel walls. We observed a trade-off between the geometric accuracy of these vessel structures and their quality for simulations. Namely, adding regularization losses improved the structural quality of inlet geometries but slightly reduced their geometric accuracy in terms of Chamfer distances. Applying a higher weight on the inlet mesh vertices in the geometric consistency loss was able to improve the accuracy of inlet geometries.

Comparison with Other Methods on Time-Series CT Data

Table 5.3 compares the reconstruction accuracy between our method and the other baseline methods on end-diastolic and end-systolic phases of a cardiac cycle. Overall, our method demonstrated high accuracy comparable to prior state-of-the-art approach MeshDeformNet, both in terms of Dice scores and Hausdorff distances. Figure 5.8 shows a qualitative comparison of the reconstructed whole heart surfaces at end-systolic and end-diastolic phases and the estimated surface motion by computing the displacements of mesh vertices over time. MeshDeformNet produced gaps between cardiac structures as well as overly smoothed pulmonary veins and vena cava geometries, since that method is biased by the use of sphere templates rather than a more fitting template of the whole heart. In contrast, our method produced high-quality geometries of the vessel inlets and outlets as well as whole heart geometries that best match with the ground truth. Furthermore, our method demonstrated a

		Epi	LA	LV	RA	RV	Ao	PA	WH
DCS	Ours (S)	89.53	93.30	94.48	92.91	94.32	96.20	**85.31**	93.14
	Ours (V)	88.27*	91.60*	93.21*	92.18*	93.21*	95.62*	83.48*	91.97*
	HeartFFDNet	84.37*	88.38*	91.41*	90.26*	90.19*	93.03*	70.44*	88.94*
	MeshDeformNet	**90.58***	**95.18***	**95.85***	**93.50***	**94.63**	**97.50***	80.21*	**93.94***
HD	Ours (S)	6.04	10.21	4.95	**10.04**	6.61	3.80	19.71	16.02
	Ours (V)	**5.91**	10.64*	5.36	10.32	7.03	4.16*	**19.14**	**15.69**
	HeartFFDNet	6.78	12.01*	6.37*	10.85	8.77*	5.13*	23.20*	16.46
	MeshDeformNet	5.98	**9.29**	**4.39**	10.42	**6.35**	**3.42***	23.25	15.77

Table 5.3: Comparison of DCS (↑) and HDs (mm) (↓) of predictions from different methods on 4D CT test images (n=20). * Denotes Significant Difference Of "Ours (S)" From the Others (p¡0.05)

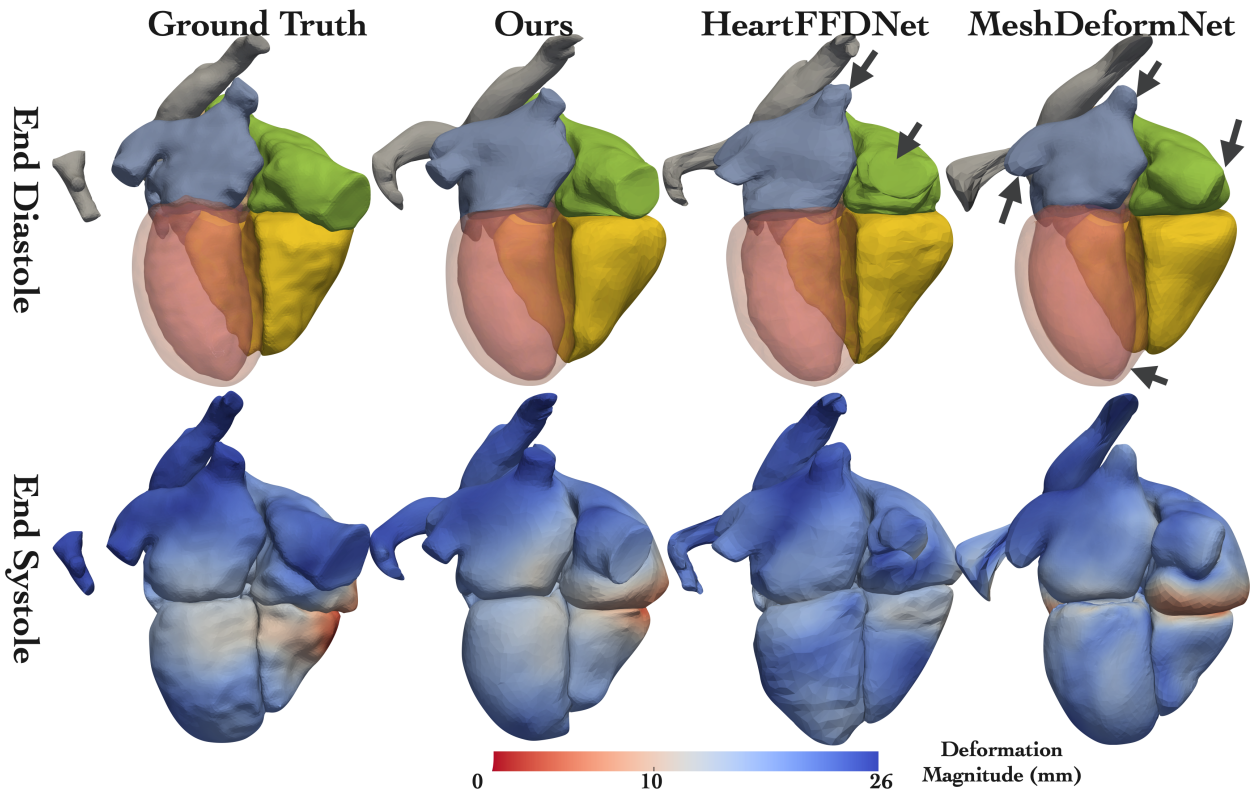


Figure 5.8: Qualitative comparison of whole heart surfaces from different methods at the end-diastolic phase and the end-systolic phase of a set of time-series image data. The colormap for end-systolic surfaces shows vertex displacement magnitude from end-systole to end-diastole.

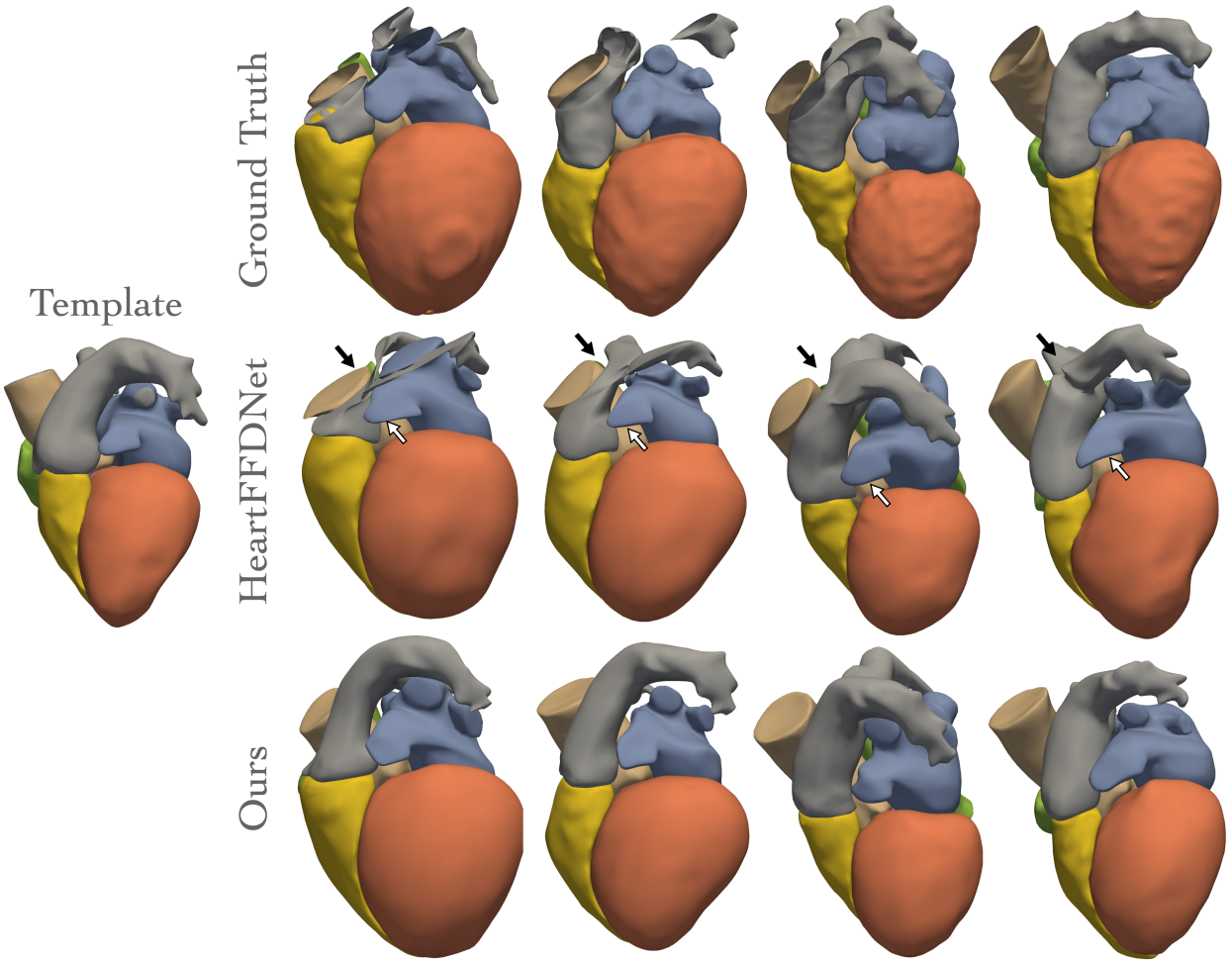


Figure 5.9: Comparison of whole heart surface predictions between using control handles as in our approach and using FFD as in HeartFFDNet.

more accurate estimation of surface deformation over time, which is required for prescribing boundary conditions for CFD simulations.

Figure 5.9 provides further qualitative comparisons between using FFD and using biharmonic coordinates to deform the template. Using biharmonic coordinates allows more flexible deformation and can thus more closely capture detailed geometries such as the left atrial appendage. In contrast, geometries of left atrial appendage predicted from HeartFFDNet were strongly biased by the geometries of the template, although it used far more control points (4096) than our method (600). Furthermore, our method was able to predict the shape of the heart that was not covered in the image data. Namely, thanks to our augmentation pipeline, our method generated reasonable geometries of the pulmonary arteries and pulmonary veins. In contrast, manual segmentation can only produce surface meshes of the cardiac structures

	Cap-Wall Orthogonality (↓)			Cap Coplanarity (mm) (↓)			% Face Intersection (↓)
	LA	RA	Ao	LA	RA	Ao	WH
Ours (V)	**0.038**±0.046	**0.013**±0.007	**0.013**±0.012	**0.22**±0.024	**0.284**±0.044	**0.292**±0.088	**0.018**±0.022
HeartFFDNet	0.137±0.08	0.228±0.182	0.494±0.386	0.398±0.068	0.45±0.125	0.949±0.557	0.262±0.191
MeshDeformNet	0.106±0.104	0.044±0.038	0.209±0.117	1.145±0.165	0.917±0.379	0.36±0.216	0.034±0.068
Manual	0.04±0.04	0.034±0.054	0.025±0.023	0.037±0.009	0.035±0.007	0.02±0.003	0.0±0.0

Table 5.4: A comparison of the quality of the inlet/outlet geometries and whole heart surface quality from different methods.

captured in the images and HeartFFDNet predicted flat and unphysiological geometries despite starting from a realistic whole heart template.

Table 5.4 compares the quality of the predicted inlet and outlet geometries as well as the percentage face intersection of the whole heart meshes. Besides comparing with our baselines, HeartFFDNet and MeshDeformNet, we also compared our method with the surface meshes generated from applying the Marching Cube algorithm on manual ground truth segmentations, where the vessel inlet and outlet geometries were manually trimmed by human experts. Our method produced significantly better vessel inlet and outlet geometries than HeartFFDNet and MeshDeformNet. In particular, our method outperformed the manual segmentation in terms of Cap-Wall Orthogonality. When deforming the mesh template using control handles, our method achieved the lowest percentage of face intersection than other deep-learning methods, and the small amount of face intersections that occurred could be easily corrected by a few iterations of Laplacian smoothing.

CFD Simulations of Cardiac Flow

We were able to successfully conduct CFD simulations using the automatically constructed LV meshes for all 10 patients, as well as for 9 of 10 patients with the 4-chamber meshes. The 1 failed case had structure penetrations between two pulmonary veins, causing the simulation to diverge. Figure 5.10 displays the simulation results of the velocity streamlines at multiple time steps during diastole for 2 different patients. The simulation results demonstrate the formation of typical vortex flow during ventricle filling.

Figure 5.11 provides quantitative comparisons of the accuracy of CFD simulation results of LV flow. Both our approach and HeartFFDNet significantly outperformed the image-registration-based approach in terms of all metrics. Namely, the image-registration-based method significantly underestimated the LV volume during diastole since the reconstructed meshes did not capture the large deformation of LV from systole to diastole. Our proposed approach demonstrated comparable or slightly better accuracy than HeartFFDNet in general, with smaller volume errors throughout a cardiac cycle and smaller errors in volume averaged kinetic energy density and mean aortic flow velocity during systole.

Figure 5.12 qualitatively compares the simulated LV flow pattern during both systole and diastole using meshes automatically constructed by our proposed approach and HeartFFD-

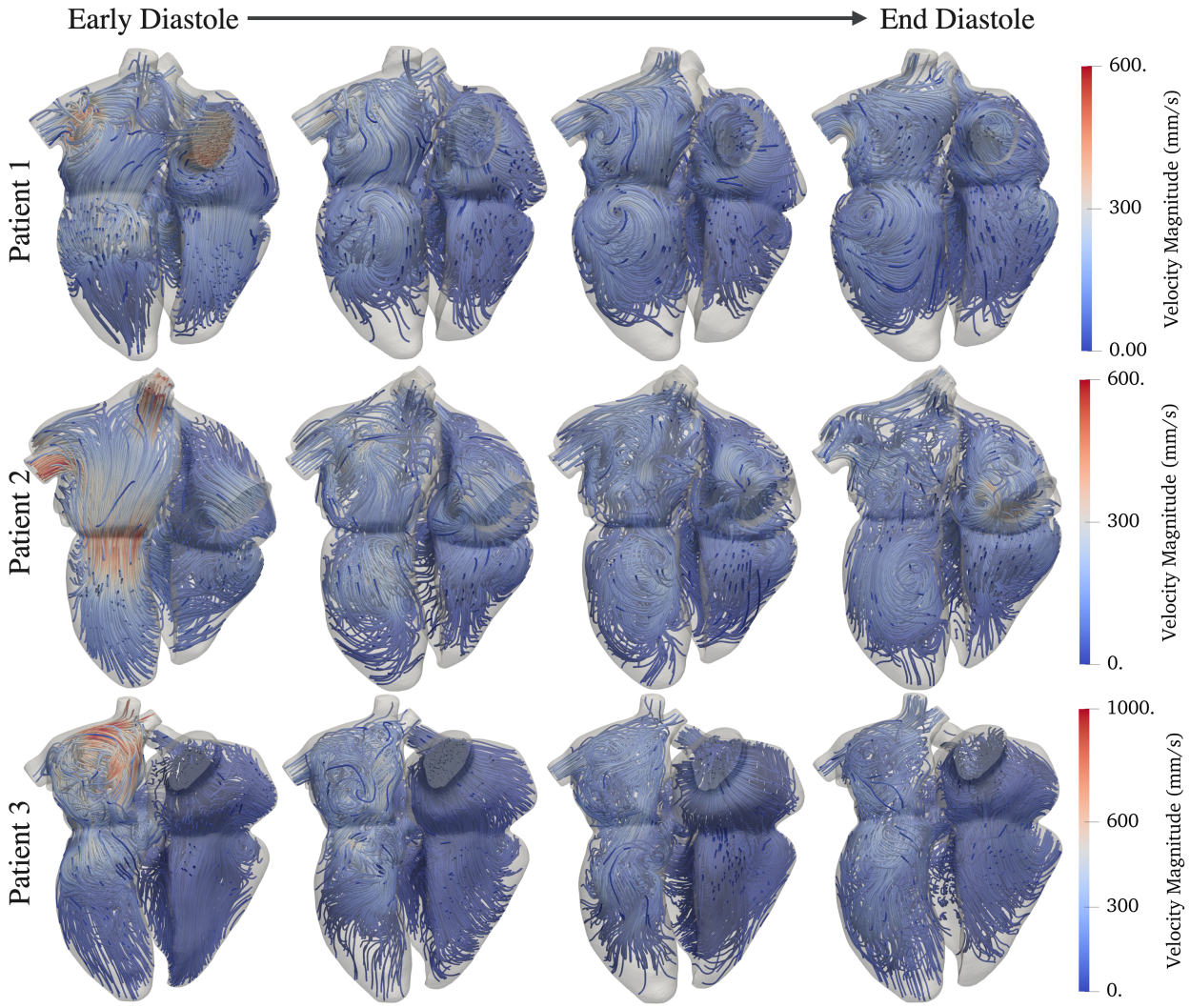


Figure 5.10: Velocity streamlines from CFD simulations of 2 different patients using the predicted 4D meshes.

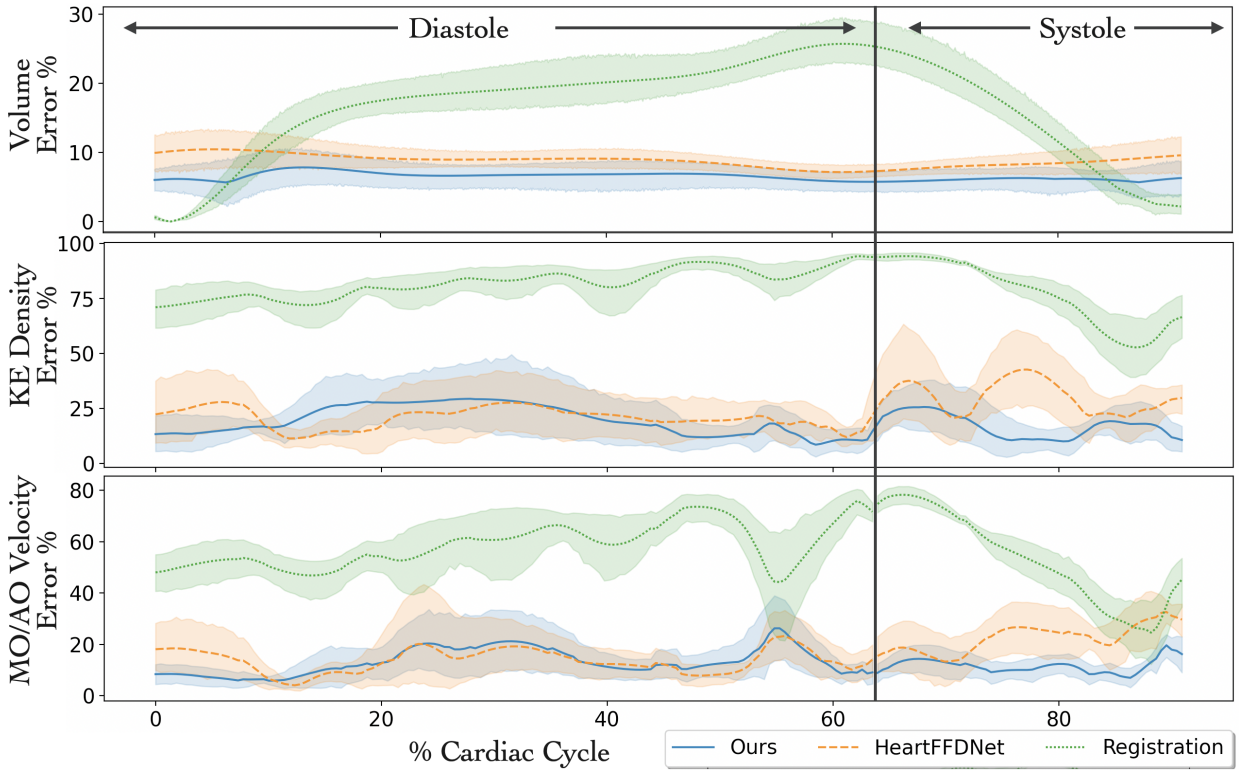


Figure 5.11: Quantitative comparisons of the % errors in LV volume, volume averaged KE density, mean velocity near the MO during diastole and mean velocity near the AO during systole among different methods. Lines show the mean values and shades show the 95% confidence intervals.

Net, semi-automatically constructed by conventional image registration and manually constructed by human observers. Image registration underestimated the LV expansion from end systole to diastole, leading to underestimated flow velocity and disparate flow pattern compared with the ground truth. Both of our approaches generally produced similar vortex structures during diastole and converging flow during systole, with moderate differences in flow velocity and vortex locations compared with the ground truth.

5.4 Discussion

Automated image-based reconstruction of cardiac meshes is important for computational simulation of cardiac physiology. While deep-learning-based methods have demonstrated success in tasks such as image segmentation and registration, few studies have addressed the end-to-end learning between images and meshes for modeling applications. Furthermore, prior learning-based mesh reconstruction approaches suffer from a number of limitations

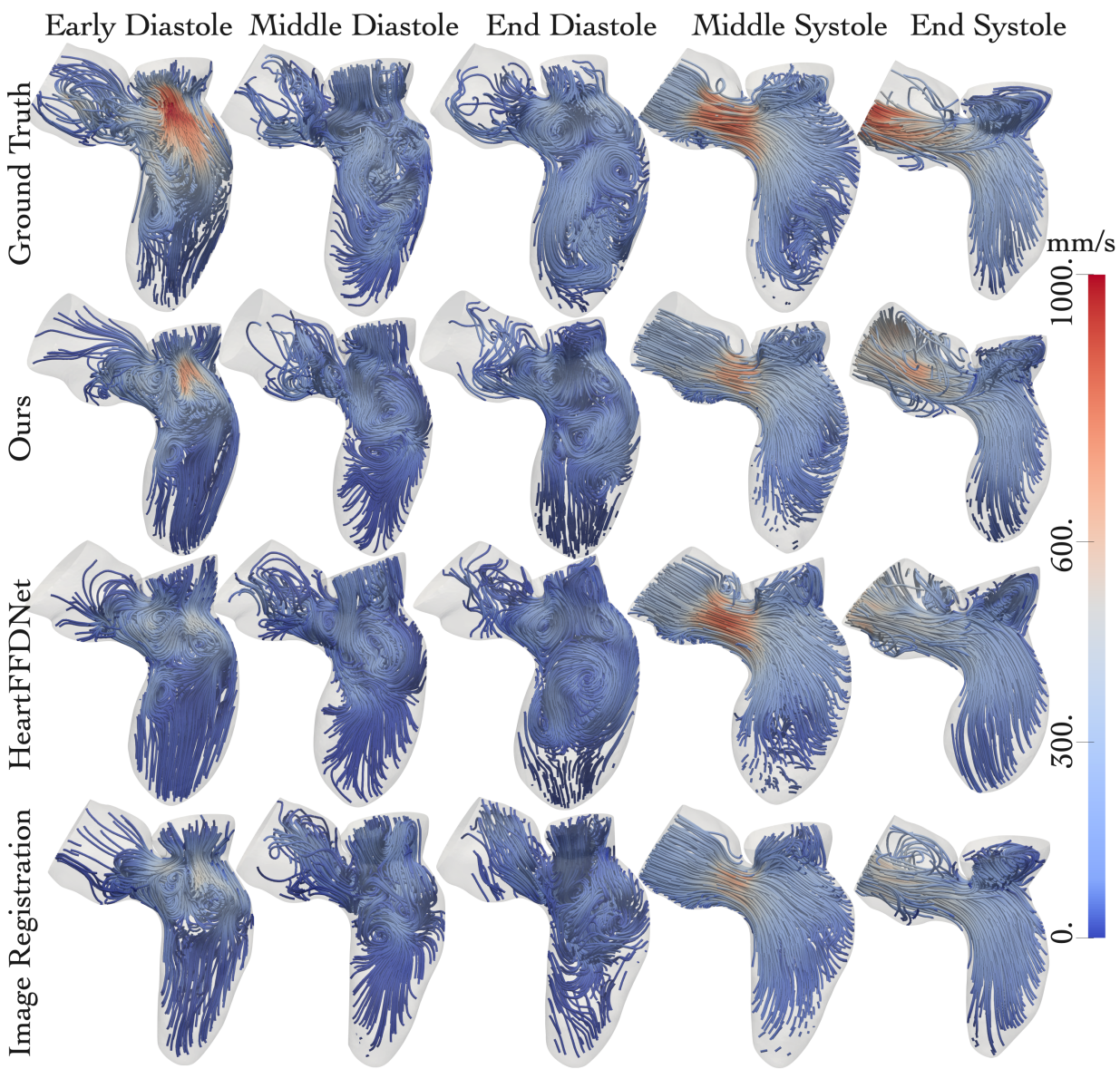


Figure 5.12: Qualitative comparisons of the simulated velocity streamlines from different methods at different time phases during a cardiac cycle. Color map shows the velocity magnitude.

	Ours	HeartFFDNet	MeshDeformNet	2D UNet	3D UNet
# of Parameters	8.7M	8.5M	16.8M	31.1M	18.6M
Training Time	18 hrs	26 hrs	32 hrs	7 hrs	37 hrs
Test Time	0.230s	0.177s	0.425s	1.555s	0.367s

Table 5.5: Comparison of model size, training and testing time.

such as using decoupled meshes of individual cardiac structures and assumed mesh topology, thus unable to directly support different cardiac simulations without additional efforts [168, 68]. We addressed this challenge herein using a novel approach that learns the translation of a small set of control handles to deform the space enclosing a whole heart template from volumetric image data. Our method demonstrated promising whole-heart reconstruction accuracy and was able to generate simulation-ready meshes from time-series image data for CFD simulations of cardiac flow.

Our approach achieved comparable geometric accuracy to the prior state-of-the-art whole heart mesh reconstruction method MeshDeformNet [68] while having the additional advantage of directly support various cardiac simulations. We note that our approach used fewer parameters in the CNN encoder compared to MeshDeformNet (Table 5.5) and the use of biharmonic coordinates naturally ensures the smoothness of deformation without using explicit mesh regularization (e.g., Laplacian and/or edge length loss constraints [68]). This is important since mesh regularization schemes can complicate the optimization process [42], whereas we observed our approach to converge significantly faster than MeshDeformNet (18 vs 32 hrs on a GTX2080Ti GPU).

For CFD simulations requiring the time-dependent motion of the heart over the cardiac cycle, our method has the advantage of deforming the template mesh in a temporally consistent manner, enabling automated construction of dynamic cardiac meshes within minutes on a standard desktop computer. Registration-based approaches, in contrast, often require test time optimizations that are computationally expensive and prone to local minimums, which often lead to inaccurate registration results such as underestimation of large deformation. Although deep-learning approaches have been proposed to speed-up the registration process [8, 96], large-deformation registration on cardiac images remains challenging for learning-based approaches. The establishment of temporal feature correspondence of our method is due to similar features of time-series images naturally being encoded into similar feature vectors by the CNN encoders and does not require explicit training. Nevertheless, ground truth data of anatomical landmarks could be incorporated in the future during training to further improve the accuracy of feature correspondence across different time frames or patients.

Blood flow simulations developed from our automated mesh generation process demonstrated clockwise circulatory flow patterns during diastole and converging flow patterns during systole in the ventricular cavity consistent with prior studies [156, 94]. However, we observed an average of 15-25% of errors in the simulated mean velocity and kinetic energy density, despite a promising mean LV Dice scores of 93% and an mean volume error of 6%.

Such amount of volume error is consistent with the inter- and intra-observer variations of manual LV segmentation[182, 68]. Indeed, simulation of blood flow is sensitive to uncertainties in geometry, such as inflow directions and vessel and/or LV wall smoothness [156, 16]. We plan to conduct intra- and inter-observer studies on the ground truth meshes to further understand the relationship between prediction uncertainties and the accuracy of CFD simulations. Nevertheless, our approach is among the first to enable creation of simulation-suitable meshes from patient images. And our design of using template meshes and control handles could support shape editing and analysis to study the effect of geometric variations on CFD simulations.

A limitation of the proposed method is that the used a single whole-heart template may not capture the full geometric variations observed clinically. In particular, the current template assumes four separate and distinct pulmonary vein ostia and thus may not fully capture pulmonary veins with alternate branching patterns, which can be important for preoperative planning of pulmonary and cardiac surgery [55]. Similarly, the template used would not be suitable for cardiac malformations such as single-ventricle patients with congenital heart diseases since the structures of the heart are significantly different from our current training template. Nonetheless, this framework could still be utilized if sufficient training data of, say, single-ventricle patients were available and a corresponding single-ventricle mesh template were used. To better handle the above applications, in future work we aim to add a template retrieval module to automatically select a template that best suits the input. Furthermore, implicit shape representation [27] can be combined with our learning-based shape deformation approach to predict cardiac structures with different anatomies.

5.5 Conclusion

We proposed a novel deep-learning approach that directly constructs simulation-ready whole heart meshes from image data and allows switching of template meshes to accommodate different modeling requirements. Our method leverages a graph convolutional network to predict the translations of a small set of control handles to smoothly deform a whole heart template using biharmonic coordinates. Our method consistently outperformed prior state-of-the-art methods in constructing simulation-ready meshes of the heart, and was able to produce geometries that better satisfy modeling requirements for cardiac flow simulations. We demonstrated application of our method on constructing dynamic whole heart meshes from time-series CT image data to simulate the cardiac flow driven by the cardiac motion. The presented approach is able to automatically construct whole heart meshes within seconds on a modern desktop computer and has the potential in facilitating high-throughput, large-cohort validation of patient-specific cardiac modeling, as well as its future clinical applications.

Chapter 6

Conclusion

In this thesis, we developed methods for automated mesh reconstruction from volumetric medical images to support computational simulations of cardiac functions. In chapter 2, we developed a streamlined and automated approach combining convolutional neural networks (CNN) to generate segmentations of cardiac tissues and surface post-processing techniques to generate meshes of the left ventricle for computational fluid dynamics simulations.

Direct construction of meshes can circumvent the intermediate CNN-based segmentation steps that are prone to topological artifacts due to a lack of awareness of the overall anatomy and topology of the target organs. In chapter 3, we developed a method to directly construct surface meshes of the whole heart from volumetric image data by learning to deform mesh templates of spheres, which are topologically equivalent to the cardiac structures modeled. This approach demonstrates the advantages of generating accurate, high-resolution meshes regardless of input image resolution, and the meshes are also anatomically consistent as they are topologically equivalent to the template. Nevertheless, learning deformation directly on decoupled sphere mesh templates still requires surface processing to join cardiac structures together for computational simulations involving multiple cardiac structures. In chapters 4 and 5, we further developed a learning-based mesh deformation method that deforms the space enclosing a realist whole heart mesh template, and once trained, the developed approaches can readily generate meshes of arbitrary subsets of cardiac structures by swapping the training template. In chapter 4, we combined free-form deformation with deep learning to predict the deformation of a control point grid from medical images. In chapter 5, we further improved this approach by using control handles and their biharmonic coordinates to parameterize the deformation of the whole heart. This approach uses fewer control handles while enabling more flexible and accurate deformation of the whole heart surface mesh. Our methods can efficiently construct 4D simulation-ready meshes within a minute on a standard desktop computer whereas prior methods may take hours of time and human efforts. It thus may potentially be used for large-cohort validation of computational simulations and clinical translations.

6.1 Future Directions

While these are encouraging progress towards the goal of efficient, automatic, and robust reconstruction of whole heart meshes for computational simulations of cardiac function, a number of challenges remain. We discuss some of these below and highlight interesting future directions.

Heart valves, thickness, and scarring of cardiac tissues: This thesis developed methods to construct the blood pools of four heart chambers, the left ventricle myocardium, and the blood pools of major vessels, namely aorta, pulmonary arteries, pulmonary veins, and vena cava. However, the heart valves and thickness of cardiac structures except for the left ventricle are not constructed. These structures could be important for computational simulations. For example, heart valves significantly affect the flow patterns and thus are necessary for accurate CFD simulation of the blood flow [34, 114, 134]. Tissue thicknesses are important for electromechanics simulations of cardiac contraction where the propagation of the electrical signals within the cardiac tissues needs to be modeled [152]. However, the heart valves and the tissues of the atrium and vessels in the heart are very thin compared to the image resolution of patient CT and MR scans. Identification of the thickness of these tissues is challenging even for radiology experts. A few approaches, however, have applied supervised deep learning methods to construct heart valves from high-resolution patient images [109, 110]. Furthermore, the scarring of cardiac tissues alters their material properties and affects the propagation of electrical signals [143, 22]. The shape and locations of the scars need to be incorporated into electrophysiology simulation to accurately model the propagation of electrical signals for individual patients [152, 26]. A few studies have applied deep learning to the segmentation of infractions in myocardium or atriums [153, 78]. It would be an interesting direction to combine these efforts with our methods to construct more detailed whole heart models.

Model construction for cardiac malformations: Our methods learn to deform a template mesh of the heart of a healthy subject and assume that the anatomies of the hearts of the target patients have the same topology as the initial heart template. This is typically true for many cardiac diseases but patients with cardiac malformation or congenital heart defects can have altered structures in their hearts, such as holes and missing parts that cannot be modeled by deforming a template of a normal heart. In contrast to explicitly representing shapes as surface meshes, a few recent studies represent shapes implicitly as signed distance functions and leveraged differentiable iso-surface extraction to construct meshes with varying topology [126, 41]. Furthermore, existing deep learning methods are usually trained on homogeneous datasets that primarily contain adult subjects with normal cardiac anatomies [58]. In the future, datasets containing pediatric subjects with congenital heart defects would be extremely valuable for developing learning-based methods to robustly construct cardiac geometries and computer models for those patients.

Generalization to unseen medical imaging domains: While learning-based approaches have demonstrated success in many medical image analysis tasks, these approaches have been often trained and validated using patient images from a single clinical center or acquired using homogeneous imaging protocols. Therefore, an important challenge for learning-based methods for medical image analysis is how well the proposed methods would generalize on unseen medical images from different scanner vendors or new imaging protocols. In this thesis, we validated our approaches on CT images acquired from a different source than the training dataset and demonstrated promising performance [68, 67]. However, our method performed worse on cine MR images that significantly differ from our training datasets [68]. Indeed, MR data are sensitive to a number of factors, including differences in vendors, magnetic coil types, and/or acquisition protocols. A few studies have published MR datasets across different clinical centers, imaging conditions, or scanner vendors, and attempted to develop domain-generalization approaches for medical image segmentation [15, 85, 84]. In the future, we plan to combine these approaches with our methods to robustly construct simulation-suitable meshes from unseen MR images. Furthermore, automatic uncertainty quantifications [39] and quality assessments [77] of the predicted whole heart surface meshes are both interesting directions to provide additional insights to clinicians and engineers when analyzing the outputs from deep-learning-based whole heart reconstruction methods.

Bibliography

[1] Martın Abadi et al. "Tensorflow: A system for large-scale machine learning". In: *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 2016, pp. 265–283.

[2] Hermenegild J Arevalo et al. "Arrhythmia risk stratification of patients after myocardial infarction using personalized heart models". In: *Nature Communications* 7.1 (2016), p. 11437. ISSN: 2041-1723.

[3] Rahman Attar et al. "3D Cardiac Shape Prediction with Deep Neural Networks: Simultaneous Use of Images and Patient Metadata". In: *MICCAI 2019*. Ed. by Dinggang Shen et al. Cham: Springer International Publishing, 2019, pp. 586–594. ISBN: 978-3-030-32245-8.

[4] Christoph Augustin et al. "Patient-specific modeling of left ventricular electromechanics as a driver for haemodynamic analysis". In: *Europace* 18 (Aug. 2016), pp. iv121–iv129. DOI: 10.1093/europace/euw369.

[5] Christoph M Augustin et al. "Anatomically accurate high resolution modeling of human whole heart electromechanics: A strongly scalable algebraic multigrid solver method for nonlinear deformation". eng. In: *Journal of computational physics* 305 (2016), pp. 622–646. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2015.10.045. URL: https://pubmed.ncbi.nlm.nih.gov/26819483%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4724941/.

[6] M. R. Avendi, Arash Kheradvar, and Hamid Jafarkhani. "A Combined Deep-Learning and Deformable-Model Approach to Fully Automatic Segmentation of the Left Ventricle in Cardiac MRI". In: *Medical image analysis* 30 (2015), pp. 108–119.

[7] Wenjia Bai et al. "Multi-atlas segmentation with augmented features for cardiac MR images". In: *Medical Image Analysis* 19.1 (2015), pp. 98–109. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2014.09.005. URL: http://www.sciencedirect.com/science/article/pii/S136184151400142X.

[8] Guha Balakrishnan et al. "VoxelMorph: A Learning Framework for Deformable Medical Image Registration". In: *IEEE Transactions on Medical Imaging* 38 (2019), pp. 1788–1800.

[9] Alessandra Bavo et al. "Patient-specific CFD simulation of intraventricular haemodynamics based on 3D ultrasound imaging". In: *Biomedical engineering online* 15 (Sept. 2016), p. 107. DOI: 10.1186/s12938-016-0231-9.

[10] O. Bernard et al. "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?" In: *IEEE Transactions on Medical Imaging* 37.11 (2018), pp. 2514–2525. ISSN: 1558-254X. DOI: 10.1109/TMI.2018.2837502.

[11] M. M. Bronstein et al. "Geometric Deep Learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42. DOI: 10.1109/MSP.2017.2693418.

[12] V. Bui et al. "Simultaneous Multi-Structure Segmentation of the Heart and Peripheral Tissues in Contrast Enhanced Cardiac Computed Tomography Angiography". In: *IEEE Access* 8 (2020), pp. 16187–16202. DOI: 10.1109/ACCESS.2020.2966985.

[13] Vy Bui et al. "Improving multi-atlas cardiac structure segmentation of computed tomography angiography: A performance evaluation based on a heterogeneous dataset". In: *Computers in Biology and Medicine* 125 (2020), p. 104019. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2020.104019. URL: http://www.sciencedirect.com/science/article/pii/S0010482520303504.

[14] Andrés Caballero et al. "Modeling Left Ventricular Blood Flow Using Smoothed Particle Hydrodynamics". In: *Cardiovascular Engineering and Technology* 8.4 (2017), pp. 465–479. ISSN: 1869-4098. DOI: 10.1007/s13239-017-0324-z. URL: https://doi.org/10.1007/s13239-017-0324-z.

[15] Víctor M. Campello et al. "Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge". In: *IEEE Transactions on Medical Imaging* 40 (2021), pp. 3543–3554.

[16] Simona Celi et al. "On the Role and Effects of Uncertainties in Cardiovascular in silico Analyses". In: *Frontiers in Medical Technology* 3 (2021).

[17] Liang-Chieh Chen et al. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (2018), pp. 834–848.

[18] Sumanth Chennupati et al. "MultiNet++: Multi-Stream Feature Aggregation and Geometric Loss Strategy for Multi-Task Learning". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 2019.

[19] Kavitha M. Chinnaiyan et al. "Rationale, design and goals of the HeartFlow assessing diagnostic value of non-invasive FFRCT in Coronary Care (ADVANCE) registry." In: *Journal of cardiovascular computed tomography* 11 1 (2017), pp. 62–67.

[20] Christophe Chnafa, Simon Mendez, and Nicoud Franck. "Image-Based Simulations Show Important Flow Fluctuations in a Normal Left Ventricle: What Could be the Implications?" In: *Annals of biomedical engineering* 44 (Apr. 2016). DOI: 10.1007/s10439-016-1614-6.

[21] Özgün Çiçek et al. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Ed. by Sebastien Ourselin et al. Cham: Springer International Publishing, 2016, pp. 424–432. ISBN: 978-3-319-46723-8.

[22] Caroline Mendonca Costa et al. "Modeling the Electrophysiological Properties of the Infarct Border Zone". In: *Frontiers in Physiology* 9 (2018).

[23] Sharon Cresci et al. "Heart Failure in the Era of Precision Medicine: A Scientific Statement From the American Heart Association." In: *Circulation: Genomic and Precision Medicine* (2019).

[24] George Crystal and Paul Pagel. "Right Ventricular Perfusion: Physiology and Clinical Implications". In: *Anesthesiology* 128 (Oct. 2017), p. 1. DOI: 10.1097/ALN.0000000000001891.

[25] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering". In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee et al. Vol. 29. Curran Associates, Inc., 2016, pp. 3844–3852. URL: https://proceedings.neurips.cc/paper/2016/file/04df4d434d481c5bb723be1b6df1ee65-Paper.pdf.

[26] Dongdong Deng et al. "A feasibility study of arrhythmia risk prediction in patients with myocardial infarction and preserved ejection fraction." In: *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology* 18 suppl 4 (2016), pp. iv60–iv66.

[27] Yu Deng, Jiaolong Yang, and Xin Tong. "Deformed Implicit Field: Modeling 3D Shapes with Learned Dense Correspondence". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 10281–10291.

[28] Chao Dong et al. "Image Super-Resolution Using Deep Convolutional Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2016), pp. 295–307.

[29] Jinming Duan et al. "Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach". In: *IEEE transactions on medical imaging* 38 (2019), pp. 2151–2164.

[30] O. Ecabert et al. "Automatic Model-Based Segmentation of the Heart in CT Images". In: *IEEE Transactions on Medical Imaging* 27.9 (2008), pp. 1189–1201. ISSN: 1558-254X. DOI: 10.1109/TMI.2008.918330.

[31] Olivier Ecabert et al. "Segmentation of the heart and great vessels in CT images using a model-based adaptation framework". In: *Medical Image Analysis* 15.6 (2011), pp. 863–876. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2011.06.004. URL: http://www.sciencedirect.com/science/article/pii/S1361841511000910.

[32] Mohammed S.M. Elbaz et al. "Assessment of viscous energy loss and the association with three-dimensional vortex ring formation in left ventricular inflow: In vivo evaluation using four-dimensional flow MRI". In: *Magnetic Resonance in Medicine* 77.2 (Feb. 2016), pp. 794–805. DOI: 10.1002/mrm.26129. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.26129. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.26129.

[33] Jonatan Eriksson et al. "Quantification of presystolic blood flow organization and energetics in the human left ventricle". In: *American journal of physiology. Heart and circulatory physiology* 300 (Mar. 2011), H2135–41. DOI: 10.1152/ajpheart.00993.2010.

[34] Réka Faludi et al. "Left ventricular flow patterns in healthy subjects and patients with prosthetic mitral valves. An". In: *The Journal of thoracic and cardiovascular surgery* 139 (Apr. 2010), pp. 1501–10. DOI: 10.1016/j.jtcvs.2009.07.060.

[35] Kunihiko Fukushima. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". In: *Biological Cybernetics* 36 (2004), pp. 193–202.

[36] Liuhao Ge et al. "3D Hand Shape and Pose Estimation From a Single RGB Image". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 10825–10834.

[37] M. Genet et al. "Modeling Pathologies of Diastolic and Systolic Heart Failure". In: *Ann Biomed Eng* 44.1 (2016), pp. 112–127.

[38] Morteza Gharib et al. "Optimal vortex formation as an index of cardiac health". In: *Proceedings of the National Academy of Sciences of the United States of America* 103 (May 2006), pp. 6305–8. DOI: 10.1073/pnas.0600520103.

[39] Biraja Ghoshal et al. "Estimating Uncertainty in Deep Learning for Reporting Confidence to Clinicians when Segmenting Nuclei Image Data". In: *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)* (2019), pp. 318–324.

[40] Ian J. Goodfellow et al. "Generative Adversarial Nets". In: *NIPS*. 2014.

[41] Benoit Guillard et al. "DeepMesh: Differentiable Iso-Surface Extraction". In: *ArXiv* abs/2106.11795 (2021).

[42] Kunal Gupta and Manmohan Chandraker. *Neural Mesh Flow: 3D Manifold Mesh Generationvia Diffeomorphic Flows*. 2020. arXiv: 2007.10973 [cs.CV].

[43] Marija Habijan et al. "Overview of the Whole Heart and Heart Chamber Segmentation Methods". In: *Cardiovascular Engineering and Technology* (2020). ISSN: 1869-4098. DOI: 10.1007/s13239-020-00494-8. URL: https://doi.org/10.1007/s13239-020-00494-8.

[44] William L. Hamilton, Zhitao Ying, and Jure Leskovec. "Inductive Representation Learning on Large Graphs". In: *NIPS*. 2017.

[45] Kaiming He et al. "Identity Mappings in Deep Residual Networks". In: *Computer Vision – ECCV 2016*. Ed. by Bastian Leibe et al. Cham: Springer International Publishing, 2016, pp. 630–645. ISBN: 978-3-319-46493-0.

[46] Nicholas Heller et al. "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge". In: *Medical Image Analysis* 67 (2021), p. 101821. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2020.101821. URL: https://www.sciencedirect.com/science/article/pii/S1361841520301857.

[47] Peijun Hu et al. "Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution." In: *Physics in medicine and biology* 61 24 (2016), pp. 8676–8698.

[48] Juan Iglesias and Mert Sabuncu. "Multi-Atlas Segmentation of Biomedical Images: A Survey". In: *Medical image analysis* 24 (Dec. 2014). DOI: 10.1016/j.media.2015.06.012.

[49] Fabian Isensee and Klaus Maier-Hein. "An attempt at beating the 3D U-Net". In: *ArXiv* abs/1908.02182 (2019).

[50] Fabian Isensee et al. "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi et al. Cham: Springer International Publishing, 2018, pp. 287–297. ISBN: 978-3-319-75238-9.

[51] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18.2 (2021), pp. 203–211. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01008-z. URL: https://doi.org/10.1038/s41592-020-01008-z.

[52] Alec Jacobson et al. "Bounded biharmonic weights for real-time deformation". In: *Communications of the ACM* 57 (2014), pp. 99–106.

[53] Cheng Ju, Aurélien Bibaut, and Mark Laan. "The Relative Performance of Ensemble Methods with Deep Convolutional Neural Networks for Image Classification". In: *Journal of Applied Statistics* 45 (Apr. 2017). DOI: 10.1080/02664763.2018.1441383.

[54] Angjoo Kanazawa et al. "Learning Category-Specific Mesh Reconstruction from Image Collections". In: *ECCV*. 2018.

[55] Asha Kandathil and Murthy R. Chamarthy. "Pulmonary vascular anatomy & anatomical variants." In: *Cardiovascular diagnosis and therapy* 8 3 (2018), pp. 201–207.

[56] Elias Karabelas et al. "Towards a Computational Framework for Modeling the Impact of Aortic Coarctations Upon Left Ventricular Load". In: *Frontiers in Physiology* 9 (May 2018), p. 538. DOI: 10.3389/fphys.2018.00538.

[57] Rashed Karim et al. "Algorithms for left atrial wall segmentation and thickness – Evaluation on an open-source CT and MRI image database". In: *Medical Image Analysis* 50 (2018), pp. 36–53. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2018.08.004. URL: http://www.sciencedirect.com/science/article/pii/S1361841518306431.

[58] Saeed Karimi-Bidhendi et al. "Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases". In: *Journal of Cardiovascular Magnetic Resonance* 22 (2020).

[59] Tero Karras, Samuli Laine, and Timo Aila. "A Style-Based Generator Architecture for Generative Adversarial Networks". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 4396–4405.

[60] Seyed Saeid Khalafvand et al. "Assessment of human left ventricle flow using statistical shape modelling and computational fluid dynamics". In: *Journal of Biomechanics* 74 (Apr. 2018). DOI: 10.1016/j.jbiomech.2018.04.030.

[61] Seyed Saeid Khalafvand et al. "Fluid-dynamics modelling of the human left ventricle with dynamic mesh for normal and myocardial infarction: Preliminary study". In: *Computers in biology and medicine* 42 (July 2012), pp. 863–70. DOI: 10.1016/j.compbiomed.2012.06.010.

[62] Diederik Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *International Conference on Learning Representations* (Dec. 2014).

[63] Alexander Kirillov et al. "PointRend: Image Segmentation as Rendering". In: 2019.

[64] Stefan Klein et al. "Elastix: A Toolbox for Intensity-Based Medical Image Registration". In: *IEEE transactions on medical imaging* 29 (Nov. 2009), pp. 196–205. DOI: 10.1109/TMI.2009.2035616.

[65] Fanwei Kong and Shawn C. Shadden. "Automatic Whole Heart Meshes Generation For Image-Based Computational Simulations By Learning Free-From Deformations". In: *International Conference on Medical Image Computing and Computer Assisted Intervention* (2021).

[66] Fanwei Kong and Shawn C. Shadden. "Automating Model Generation for Image-Based Cardiac Flow Simulation". In: *Journal of Biomechanical Engineering* 142.11 (Sept. 2020). 111011. ISSN: 0148-0731. DOI: 10.1115/1.4048032.

[67] Fanwei Kong and Shawn C. Shadden. "Learning Whole Heart Mesh Generation From Patient Images For Computational Simulations". In: *ArXiv* abs/2203.10517, Under review in IEEE Transactions on Medical Imaging (2022).

[68] Fanwei Kong, Nathan Wilson, and Shawn Shadden. "A Deep-Learning Approach For Direct Whole-Heart Mesh Reconstruction". In: *Medical Image Analysis* (2021), p. 102222. ISSN: 1361-8415. DOI: 10.1016/j.media.2021.102222.

[69] Fanwei Kong et al. "Finite element analysis of MitraClip procedure on a patient-specific model with functional mitral regurgitation". In: *Journal of Biomechanics* 104 (Feb. 2020), p. 109730. ISSN: 0021-9290.

[70] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks". In: *Communications of the ACM* 60 (2012), pp. 84–90.

[71] Ethan Kung et al. "Predictive modeling of the virtual Hemi-Fontan operation for second stage single ventricle palliation: Two patient-specific cases". In: *Journal of Biomechanics* 46.2 (2013), pp. 423–429. ISSN: 0021-9290.

[72] Andrey Kurenkov et al. "DeformNet: Free-Form Deformation Network for 3D Shape Reconstruction from a Single Image". In: *2018 IEEE Winter Conference on Applications of Computer Vision* (2018), pp. 858–866.

[73] Yann LeCun et al. "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation* 1 (1989), pp. 541–551.

[74] Christian Ledig et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 105–114.

[75] Lik Chuan Lee et al. "Applications of Computational Modeling in Cardiac Surgery". In: *Journal of Cardiac Surgery* 29 (2014).

[76] Jane A. Leopold and Joseph Loscalzo. "Emerging Role of Precision Medicine in Cardiovascular Disease." In: *Circulation Research* (2018).

[77] Kang Li, Lequan Yu, and Pheng-Ann Heng. "Towards reliable cardiac image segmentation: Assessing image-level and pixel-level segmentation quality via self-reflective references." In: *Medical image analysis* 78 (2022), p. 102426.

[78] Lei Li et al. "AtrialJSQnet: A New Framework for Joint Segmentation and Quantification of Left Atrium and Scars Incorporating Spatial and Shape Information". In: *Medical image analysis* 76 (2022), p. 102303.

[79] Shuailin Li, Chuyu Zhang, and Xuming He. "Shape-aware Semi-supervised 3D Semantic Segmentation for Medical Images". In: *MICCAI*. 2020.

[80] Liang Liang et al. "Machine Learning based 3D Geometry Reconstruction and Modeling of Aortic Valve Deformation Using 3D CT Images: Machine Learning based 3D Aortic Valve Modeling". In: *International Journal for Numerical Methods in Biomedical Engineering* 33 (Aug. 2016), e02827. DOI: 10.1002/cnm.2827.

[81] Jiangke Lin et al. "Towards High-Fidelity 3D Face Reconstruction From In-the-Wild Images Using Graph Convolutional Networks". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), pp. 5890–5899.

[82] Geert J. S. Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical image analysis* 42 (2017), pp. 60–88.

[83] Minghua Liu et al. "DeepMetaHandles: Learning Deformation Meta-Handles of 3D Meshes with Biharmonic Coordinates". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 12–21.

[84] Quande Liu et al. "FedDG: Federated Domain Generalization on Medical Image Segmentation via Episodic Learning in Continuous Frequency Space". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), pp. 1013–1023.

[85] Xiao Liu et al. "Semi-supervised Meta-learning with Disentanglement for Domain-generalised Medical Image Segmentation". In: *MICCAI*. 2021.

[86] William E. Lorensen and Harvey E. Cline. "Marching Cubes: A High Resolution 3D Surface Construction Algorithm". In: *SIGGRAPH Comput. Graph.* 21.4 (Aug. 1987), pp. 163–169. ISSN: 0097-8930. DOI: 10.1145/37402.37422. URL: https://doi.org/10.1145/37402.37422.

[87] Bingnan Luo et al. "Shape Constrained Network for Eye Segmentation in the Wild". In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2020), pp. 1941–1949.

[88] Gabriel Maher, Nathan Wilson, and Alison Marsden. "Accelerating cardiovascular model building with convolutional neural networks". In: *Medical & Biological Engineering & Computing* 57 (2019), pp. 2319–2335.

[89] Xiao-Jiao Mao, Chunhua Shen, and Yubin Yang. "Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections". In: *NIPS*. 2016.

[90] Pablo Martinez-Legazpi et al. "Contribution of the Diastolic Vortex Ring to Left Ventricular Filling". In: *Journal of the American College of Cardiology* 64 (Oct. 2014), pp. 1711–21. DOI: 10.1016/j.jacc.2014.06.1205.

[91] Pablo Martinez-Legazpi et al. "Contribution of the Diastolic Vortex Ring to Left Ventricular Filling". In: *Journal of the American College of Cardiology* 64 (Oct. 2014), pp. 1711–21. DOI: 10.1016/j.jacc.2014.06.1205.

[92] Kathleen S. McDowell et al. "Methodology for patient-specific modeling of atrial fibrosis as a substrate for atrial fibrillation". In: *Journal of Electrocardiology* 45.6 (2012), pp. 640–645. ISSN: 0022-0736.

[93] Zahra Mirikharaji and G. Hamarneh. "Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation". In: *MICCAI*. 2018.

[94] Rajat Mittal et al. "Computational modeling of cardiac hemodynamics: Current status and future outlook". In: *Journal of Computational Physics* 305 (Nov. 2015). DOI: 10.1016/j.jcp.2015.11.022.

[95] Pim Moeskops et al. "Deep Learning for Multi-task Medical Image Segmentation in Multiple Modalities". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Ed. by Sebastien Ourselin et al. Cham: Springer International Publishing, 2016, pp. 478–486. ISBN: 978-3-319-46723-8.

[96] Tony C. W. Mok and Albert C. S. Chung. "Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks". In: *ArXiv* abs/2006.16148 (2020).

[97] Siamak N Doost et al. "Heart blood flow simulation: A perspective review". In: *BioMedical Engineering OnLine* 15 (Dec. 2016). DOI: 10.1186/s12938-016-0224-8.

[98] Siamak N Doost et al. "The Numerical Analysis of non-Newtonian Blood Flow in Human Patient-Specific Left Ventricle". In: *Computer Methods and Programs in Biomedicine* 127 (Jan. 2016). DOI: 10.1016/j.cmpb.2015.12.020.

[99] Tuan Ngo, Zhi Lu, and Gustavo Carneiro. "Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine Magnetic Resonance". In: *Medical Image Analysis* 35 (May 2016). DOI: 10.1016/j.media.2016.05.009.

[100] Thanh Danh Nguyen, Olufemi Emmanuel Kadri, and Roman S. Voronov. "An Introductory Overview of Image-Based Computational Modeling in Personalized Cardiovascular Medicine". In: *Frontiers in Bioengineering and Biotechnology* 8 (2020).

[101] Vinh-Tan Nguyen et al. "A semi-automated method for patient-specific computational flow modelling of left ventricles". In: *Computer methods in biomechanics and biomedical engineering* 18 (Aug. 2013). DOI: 10.1080/10255842.2013.803534.

[102] Haibo Ni, Stefano Morotti, and Eleonora Grandi. "A Heart for Diversity: Simulating Variability in Cardiac Arrhythmia Research". In: *Frontiers in Physiology* 9 (2018).

[103] Steven A. Niederer, Kenneth S. Campbell, and Stuart G. Campbell. "A short history of the development of mathematical models of cardiac mechanics". In: *Journal of Molecular and Cellular Cardiology* 127 (2019), pp. 11–19.

[104] Steven A. Niederer, Joost Lumens, and Natalia A. Trayanova. "Computational models in cardiology". In: *Nature Reviews Cardiology* 16 (2018), pp. 100–111.

[105] Jesus R. Nieto and Antonio Susín. "Cage Based Deformations: A Survey". In: 2013.

[106] Anton Noordegraaf et al. "Pathophysiology of the right ventricle and of the pulmonary circulation in pulmonary hypertension: an update". In: *European Respiratory Journal* 53 (Dec. 2018), p. 1801900. DOI: 10.1183/13993003.01900-2018.

[107] Ujwani Nukala et al. "A Review of the Efforts and Hindrances of Modeling and Simulation of CAR T-cell Therapy". In: 2021.

[108] Sebastian Ordas et al. "A statistical shape model of the heart and its application to model-based segmentation". In: *Proceedings of SPIE - The International Society for Optical Engineering* 6511 (Mar. 2007). DOI: 10.1117/12.708879.

[109] D. H. Pak et al. "Efficient Aortic Valve Multilabel Segmentation Using a Spatial Transformer Network". In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. 2020, pp. 1738–1742.

[110] Daniel H. Pak et al. "Distortion Energy for Deep Learning-Based Volumetric Finite Element Mesh Generation for Aortic Valves". In: *MICCAI*. 2021.

[111] Junyi Pan et al. "Deep Mesh Reconstruction From Single RGB Images via Topology Modification Networks". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), pp. 9963–9972.

[112] Ares D Pasipoularides et al. "Diastolic right ventricular filling vortex in normal and volume overload states." In: *American journal of physiology. Heart and circulatory physiology* 284 4 (2003), H1064–72.

[113] Christian Payer et al. "Multi-label Whole Heart Segmentation Using CNNs and Anatomical Label Configurations". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Cham: Springer, 2018, pp. 190–198. ISBN: 978-3-319-75541-0.

[114] Gianni Pedrizzetti, Federico Domenichini, and G. Tonti. "On the Left Ventricular Vortex Reversal after Mitral Valve Replacement". In: *Annals of biomedical engineering* 38 (Mar. 2010), pp. 769–73. DOI: 10.1007/s10439-010-9928-2.

[115] Gianni Pedrizzetti et al. "The vortex - An early predictor of cardiovascular outcome?" In: *Nature reviews. Cardiology* 11 (June 2014). DOI: 10.1038/nrcardio.2014.75.

[116] Peng Peng et al. "A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging". In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29.2 (2016), pp. 155–195. ISSN: 1352-8661. DOI: 10.1007/s10334-015-0521-4. URL: https://doi.org/10.1007/s10334-015-0521-4.

[117] J. Peters et al. "Optimizing boundary detection via Simulated Search with applications to multi-modal heart segmentation". In: *Medical Image Analysis* 14.1 (2010), pp. 70–84. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2009.10.004. URL: http://www.sciencedirect.com/science/article/pii/S1361841509001194.

[118] Jhony K. Pontes et al. "Image2Mesh: A Learning Framework for Single Image 3D Reconstruction". In: *Computer Vision – ACCV 2018*. Ed. by C. V. Jawahar et al. Cham: Springer International Publishing, 2019, pp. 365–381. ISBN: 978-3-030-20887-5.

[119] Mark Potse et al. "Patient-specific modelling of cardiac electrophysiology in heart-failure patients". In: *EP Europace* 16 (Nov. 2014), pp. iv56–iv61. ISSN: 1099-5129.

[120] Alison Pouch et al. "Statistical Assessment of Normal Mitral Annular Geometry Using Automated Three-Dimensional Echocardiographic Analysis". In: *The Annals of thoracic surgery* 97 (Oct. 2013). DOI: 10.1016/j.athoracsur.2013.07.096.

[121] A. Prakosa et al. "Methodology for image-based reconstruction of ventricular geometry for patient-specific modeling of cardiac electrophysiology". In: *Progress in Biophysics and Molecular Biology* 115.2 (2014). Novel Technologies as Drivers of Progress in Cardiac Biophysics, pp. 226–234. ISSN: 0079-6107. DOI: https://doi.org/10.1016/j.pbiomolbio.2014.08.009. URL: http://www.sciencedirect.com/science/article/pii/S0079610714000844.

[122] Adityo Prakosa et al. "Personalized virtual-heart technology for guiding the ablation of infarct-related ventricular tachycardia". In: *Nature Biomedical Engineering* 2.10 (2018), pp. 732–740. ISSN: 2157-846X. DOI: 10.1038/s41551-018-0282-2. URL: https://doi.org/10.1038/s41551-018-0282-2.

[123] Alfio Quarteroni et al. "Integrated Heart—Coupling multiscale and multiphysics models for the simulation of the cardiac function". In: *Computer Methods in Applied Mechanics and Engineering* 314 (2017), pp. 345–407.

[124] Waseem Rawat and Zenghui Wang. "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review". In: *Neural Computation* 29 (2017), pp. 2352–2449.

[125] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 779–788.

[126] Edoardo Remelli et al. "MeshSDF: Differentiable Iso-Surface Extraction". In: *ArXiv* abs/2006.03997 (2020).

[127] Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2015), pp. 1137–1149.

[128] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24574-4.

[129] Federica Sacco et al. "Left Ventricular Trabeculations Decrease the Wall Shear Stress and Increase the Intra-Ventricular Pressure Drop in CFD Simulations". In: *Frontiers in Physiology* 9 (2018), p. 458. ISSN: 1664-042X. DOI: 10.3389/fphys.2018.00458. URL: https://www.frontiersin.org/article/10.3389/fphys.2018.00458.

[130] Alfonso Santiago et al. "Fully coupled fluid-electro-mechanical model of the human heart for supercomputers". In: *International Journal for Numerical Methods in Biomedical Engineering* 34 (Aug. 2018). DOI: 10.1002/cnm.3140.

[131] Torsten Schenkel et al. "MRI-Based CFD Analysis of Flow in a Human Left Ventricle: Methodology and Application to a Healthy Heart". In: *Annals of Biomedical Engineering* 37 (2008), pp. 503–515.

[132] Torsten Schenkel et al. "MRI-Based CFD Analysis of Flow in a Human Left Ventricle: Methodology and Application to a Healthy Heart". In: *Annals of biomedical engineering* 37 (Feb. 2009), pp. 503–15. DOI: 10.1007/s10439-008-9627-4.

[133] Thomas W. Sederberg and Scott R. Parry. "Free-form deformation of solid geometric models". In: *Proceedings of the 13th annual conference on Computer graphics and interactive techniques* (1986).

[134] Jung Hee Seo et al. "Effect of the mitral valve on diastolic flow patterns". In: *Physics of Fluids* 26 (Dec. 2014), p. 121901. DOI: 10.1063/1.4904094.

[135] M. Sermesant et al. "Patient-specific electromechanical models of the heart for the prediction of pacing acute effects in CRT: A preliminary clinical validation". In: *Medical Image Analysis* 16.1 (2012), pp. 201–215. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2011.07.003. URL: http://www.sciencedirect.com/science/article/pii/S1361841511001009.

[136] Denis Shamonin et al. "Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease". In: *Frontiers in neuroinformatics* 7 (Jan. 2013), p. 50. DOI: 10.3389/fninf.2013.00050.

[137] Florence Sheehan and Andrew Redington. "The right ventricle: Anatomy, physiology and clinical imaging". In: *Heart (British Cardiac Society)* 94 (Dec. 2008), pp. 1510–5. DOI: 10.1136/hrt.2007.132779.

[138] Evan Shelhamer, Jonathan Long, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2017), pp. 640–651.

[139] Hang Si. "TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator". In: *ACM Trans. Math. Softw.* 41.2 (Feb. 2015). ISSN: 0098-3500. DOI: 10.1145/2629697. URL: https://doi.org/10.1145/2629697.

[140] P. Simard, David Steinkraus, and John C. Platt. "Best practices for convolutional neural networks applied to visual document analysis". In: *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* (2003), pp. 958–963.

[141] SimVascular. *svFSI*. `https://github.com/SimVascular/svFSI`. 2020.

[142] Olga Sorkine-Hornung et al. "Laplacian surface editing". In: *SGP '04*. 2004.

[143] Joseph F. Spear et al. "Cellular electrophysiology of human myocardial infarction. 1. Abnormalities of cellular activation." In: *Circulation* 59 2 (1979), pp. 247–56.

[144] Marina Strocchi et al. "A publicly available virtual cohort of four-chamber heart meshes for cardiac electro-mechanics simulations". In: *PLOS ONE* 15.6 (June 2020), pp. 1–26.

[145] Marina Strocchi et al. "Simulating ventricular systolic motion in a four-chamber heart model with spatially varying robin boundary conditions to model the effect of the pericardium". In: *Journal of Biomechanics* 101 (2020), p. 109645. ISSN: 0021-9290. DOI: `https://doi.org/10.1016/j.jbiomech.2020.109645`. URL: `http://www.sciencedirect.com/science/article/pii/S002192902030052X`.

[146] Hugo Talbot et al. "Towards an interactive electromechanical model of the heart". In: *Interface Focus* 3.2 (2013), p. 20120091. DOI: `10.1098/rsfs.2012.0091`. eprint: `https://royalsocietypublishing.org/doi/pdf/10.1098/rsfs.2012.0091`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rsfs.2012.0091`.

[147] Dalin Tang et al. "Image-Based Patient-Specific Ventricle Models with Fluid-Structure Interaction for Cardiac Function Assessment and Surgical Design Optimization." In: *Progress in pediatric cardiology* 30 1-2 (2010), pp. 51–62.

[148] C. Tobon-Gomez et al. "Benchmark for Algorithms Segmenting the Left Atrium From 3D CT and MRI Datasets". In: *IEEE Transactions on Medical Imaging* 34.7 (2015), pp. 1460–1473. DOI: `10.1109/TMI.2015.2398818`.

[149] Catalina Tobon-Gomez et al. "Understanding the mechanisms amenable to CRT response: from pre-operative multimodal image data to patient-specific computational models". In: *Medical & Biological Engineering & Computing* 51 (2013), pp. 1235–1250.

[150] Qianqian Tong et al. "3D Deeply-Supervised U-Net Based Whole Heart Segmentation". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Cham: Springer, 2018, pp. 224–232. ISBN: 978-3-319-75541-0.

[151] Natalia A. Trayanova. "Computational Cardiology: The Heart of the Matter". In: *ISRN Cardiology* 2012 (2012).

[152] Natalia A. Trayanova, Jason Constantino, and Viatcheslav Gurev. "Electromechanical models of the ventricles". In: *American Journal of Physiology-Heart and Circulatory Physiology* 301.2 (2011), H279–H286.

[153] Eranga Ukwatta et al. "Myocardial Infarct Segmentation From Magnetic Resonance Images for Personalized Modeling of Cardiac Electrophysiology". In: *IEEE Transactions on Medical Imaging* 35 (2016), pp. 1408–1419.

[154] Adam Updegrove, Nathan M. Wilson, and Shawn C. Shadden. "Boolean and smoothing of discrete polygonal surfaces". In: *Advances in Engineering Software* 95 (2016), pp. 16–27. ISSN: 0965-9978. DOI: https://doi.org/10.1016/j.advengsoft.2016.01.015. URL: http://www.sciencedirect.com/science/article/pii/S0965997816300230.

[155] Adam Updegrove et al. "SimVascular: An Open Source Pipeline for Cardiovascular Simulation". In: *Annals of Biomedical Engineering* 45 (Dec. 2016). DOI: 10.1007/s10439-016-1762-8.

[156] Vijay Vedula et al. "Effect of trabeculae and papillary muscles on the hemodynamics of the left ventricle". In: *Theoretical and Computational Fluid Dynamics* 30 (May 2015). DOI: 10.1007/s00162-015-0349-6.

[157] Katharina Vellguth et al. "Development of a modeling pipeline for the prediction of hemodynamic outcome after virtual mitral valve repair using image-based CFD". In: *International Journal of Computer Assisted Radiology and Surgery* (July 2018). DOI: 10.1007/s11548-018-1821-8.

[158] Salim S. Virani et al. "Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association." In: *Circulation* (2021).

[159] N. Wang et al. "Pixel2Mesh: 3D Mesh Model Generation via Image Guided Deformation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020), pp. 1–1.

[160] Vicky Y. Wang, Poul M. F. Nielsen, and Martyn P. Nash. "Image-Based Predictive Modeling of Heart Mechanics." In: *Annual review of biomedical engineering* 17 (2015), pp. 351–83.

[161] Y. Wang, Z. Zhong, and J. Hua. "DeepOrganNet: On-the-Fly Reconstruction and Visualization of 3D / 4D Lung Models from Single-View Projections by Deep Deformation Network". In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), pp. 960–970. DOI: 10.1109/TVCG.2019.2934369.

[162] Yan Wang et al. "Deep Distance Transform for Tubular Structure Segmentation in CT Scans". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), pp. 3832–3841.

[163] Yifan Wang et al. "Neural Cages for Detail-Preserving 3D Deformations". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 72–80.

[164] Yu Wang et al. "Linear subspace design for real-time shape deformation". In: *ACM Transactions on Graphics* 34 (2015), pp. 1–11.

[165] Mingqiang Wei et al. "Feature-preserving optimization for noisy mesh using joint bilateral filter and constrained Laplacian smoothing". In: *Optics and Lasers in Engineering* 51.11 (2013), pp. 1223–1234. ISSN: 0143-8166. DOI: https://doi.org/10.1016/j.optlaseng.2013.04.018. URL: http://www.sciencedirect.com/science/article/pii/S0143816613001395.

[166] Mingqiang Wei et al. "Learning-based 3D surface optimization from medical image reconstruction". In: *Optics and Lasers in Engineering* 103.September 2017 (2018), pp. 110–118. ISSN: 01438166. DOI: 10.1016/j.optlaseng.2017.11.014. URL: https://doi.org/10.1016/j.optlaseng.2017.11.014.

[167] Chao Wen et al. "Pixel2Mesh++: Multi-View 3D Mesh Generation via Deformation". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), pp. 1042–1051.

[168] Udaranga Wickramasinghe et al. "Voxel2Mesh: 3D Mesh Model Generation from Volumetric Data". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by Anne L. Martel et al. Cham: Springer International Publishing, 2020, pp. 299–308. ISBN: 978-3-030-59719-1.

[169] Jelmer M. Wolterink et al. "An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework". In: *Medical Physics* 43.5 (2016), pp. 2361–2373. DOI: https://doi.org/10.1118/1.4945696. eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4945696. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4945696.

[170] Meng Ye et al. "PC-U Net: Learning to Jointly Reconstruct and Segment the Cardiac Walls in 3D from CT Data". In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Cham: Springer International Publishing, 2021, pp. 117–126. ISBN: 978-3-030-68107-4.

[171] Sohail Zahid et al. "Feasibility of using patient-specific models and the "minimum cut" algorithm to predict optimal ablation targets for left atrial flutter". In: *Heart Rhythm* 13.8 (2016), pp. 1687–1698. ISSN: 1547-5271.

[172] Matthew D. Zeiler and Rob Fergus. "Visualizing and Understanding Convolutional Networks". In: *ECCV*. 2014.

[173] Mo Zhang, Bin Dong, and Quanzheng Li. "Deep Active Contour Network for Medical Image Segmentation". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by Anne L. Martel et al. Cham: Springer International Publishing, 2020, pp. 321–331. ISBN: 978-3-030-59719-1.

[174] Si Zhang et al. "Graph convolutional networks: a comprehensive review". In: *Computational Social Networks* 6 (2019), pp. 1–23.

[175] Hang Zhao et al. "Loss Functions for Image Restoration With Neural Networks". In: *IEEE Transactions on Computational Imaging* 3 (2017), pp. 47–57.

[176] Hao Zheng et al. "A New Ensemble Learning Framework for 3D Biomedical Image Segmentation". In: *AAAI*. 2019.

[177] Liang Zhong et al. "Application of Patient-Specific Computational Fluid Dynamics in Coronary and Intra-Cardiac Flow Simulations: Challenges and Opportunities". In: *Frontiers in Physiology* 9 (2018).

[178] Jun-Yan Zhu et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks". In: *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2242–2251.

[179] X. Zhuang et al. "Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection". In: *Med Phys* 42.7 (2015), pp. 3822–3833.

[180] Xiahai Zhuang. "Challenges and Methodologies of Fully Automatic Whole Heart Segmentation: A Review". In: *Journal of Healthcare Engineering* 4 (2013), p. 981729. ISSN: 2040-2295. DOI: 10.1260/2040-2295.4.3.371. URL: https://doi.org/10.1260/2040-2295.4.3.371.

[181] Xiahai Zhuang and Juan Shen. "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI". In: *Medical Image Analysis* 31 (2016), pp. 77–87. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2016.02.006. URL: http://www.sciencedirect.com/science/article/pii/S1361841516000219.

[182] Xiahai Zhuang et al. "Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge". In: *Medical Image Analysis* 58 (2019), p. 101537. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2019.101537. URL: http://www.sciencedirect.com/science/article/pii/S1361841519300751.