Two-Step Rigid and Non-Rigid Image Registration for the Alignment of Multi-View Three-Dimensional Echocardiography Sequences

by

Srivathsan Shanmuganathan

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Medical Sciences - Radiology and Diagnostic Imaging
University of Alberta

Abstract

Ultrasound is a widely used imaging modality, which provides continuous real-time imaging of the human heart, brain, liver, and many other organs. Accurate cardiovascular evaluation plays an important role in early disease diagnosis and echocardiography plays an important role in the assessment of cardiovascular diseases. The lack of ionizing radiation and portability make echocardiography one of the safest imaging modalities. Although two-dimensional echocardiography is widely used to obtain the motion of the heart structures in real-time, real-time 3D echocardiography (RT3DE) imaging allows better 3D imaging by extracting spatial features along with temporal information, thus improving clinical decision making. There have been technological advances; however, the majority of acquired RT3DE images tend to be of low quality, characterized by the absence of anatomical information, decreased spatial and temporal resolution, speckle noise, and a limited field of view. By registering RT3DE images obtained from several sonography windows, it is possible to enhance the recognition of structures and achieve a substantial improvement in image quality as well as it is also useful in the fusion of echo images to image the entire heart.

This study proposes a fully automatic point-based rigid registration technique, followed by nonrigid B-spline registration, to align 4D echocardiogram images acquired from various sonographic windows. The methodology was evaluated using scans acquired from seven volunteers. The accuracy of registration was visually and quantitatively assessed by delineating the left ventricle in each scan and computing the Dice score overlap metric and the Hausdorff distance mutual proximity measure between the first scan and the rest. The overall findings demonstrate that the sug-

gested registration method improves image alignment compared to the initial scans, which might be helpful in the fusion of echocardiographic images.

Preface

This thesis was submitted as partial fulfillment for the degree of Master of Science (M.Sc.) in Radiology and Diagnostic Imaging at the University of Alberta. The thesis is an original work by Srivathsan Shanmuganathan and the presented work was accomplished between September 2021 and January 2024. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board.

Material for this thesis is based on the following articles:

Srivathsan Shanmuganathan, Michelle Noga, Bernadette Foster, Harald Becher, and Kumaradevan Punithakumar, "Image registration for multi-view three-dimensional echocardiography sequences," in 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE), 2023, pp. 260–264.

Srivathsan Shanmuganathan, Michelle Noga, Pierre Boulanger, Bernadette Foster, Harald Becher, and Kumaradevan Punithakumar, "Two-step rigid and non-rigid image registration for the alignment of three-dimensional echocardiography sequences from multiple views," Volume 12, pp. 53485–53496, IEEE Access, 2024, doi:10.1109/ACCESS.2024.3388293

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Kumaradevan Punithakumar, and the Director of the Servier Virtual Cardiac Center Dr. Michelle Noga for giving me the opportunity to pursue an M.Sc. in the Servier lab and providing tremendous support in my study and related research. None of my success could have been possible without their guidance and support.

I would also like to thank Dr. Harald Becher and Bernadette Foster from the Mazankowski Alberta Heart Institute, who have graciously provided me with echocardiography data, ground truth segmentations, and clinical guidance over the years.

Above all, I'd like to thank my parents and my brother for their constant love, support, and encouragement in all my pursuits.

Table of Contents

List of Tables

List of Figures

Abbreviations and Acronyms

4Q 4-beat Acquisition.

APL Apical.

CT Computed Tomography.

DRL Deep Reinforcement Learning.

ECG Echocardiogram.

ED End Diastolic.

ES End Systolic.

FOV Field of View.

HD Hausdorff Distance.

HMQ Heart Model Acquisition.

LV Left Ventricle.

MRI Magnetic Resonance Imaging.

PSL Parasternal.

RL Reinforcement Learning.

RT3DE Real-time 3D Echocardiography.

RV Right Ventricle.

US Ultrasound.

Chapter 1

Introduction

1.1 Motivation

Image registration is the process of aligning two or more images taken from different modalities or different windows into a common coordinate system [1]. In medical imaging, multiple scans of the same patient are often taken over time using single or multiple imaging modalities such as ultrasound, magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and single photon emission computed tomography (SPECT). The registration of these images acquired at different times or with different sensors can reveal minute changes or patterns that can indicate the progression of the disease or the effectiveness of treatment. Ultrasound images tend to be noisy and subject to unique artifacts, and it is possible to enhance the recognition of structures as well as achieve a substantial improvement in image quality using image registration to align images acquired at multiple sonography windows. In addition, it is also useful in the fusion of echo scans to image the entire heart. The definition of the window in echocardiography image acquisition refers to the specific location of the transducer on the patient. The images used in this study are acquired from the parasternal and apical windows. Figure 1.1 shows the scanning planes of the heart. In order to find the appropriate alignment between the two images, it is necessary to find a transformation to be able to link the points of one image with the corresponding points of the other. Over the years, many tech-

niques have been developed to register images to fix variations that cause images to be misaligned. Depending on the technique used, the registration algorithms can be divided into two types as *feature-based* and *intensity-based* methods [2].

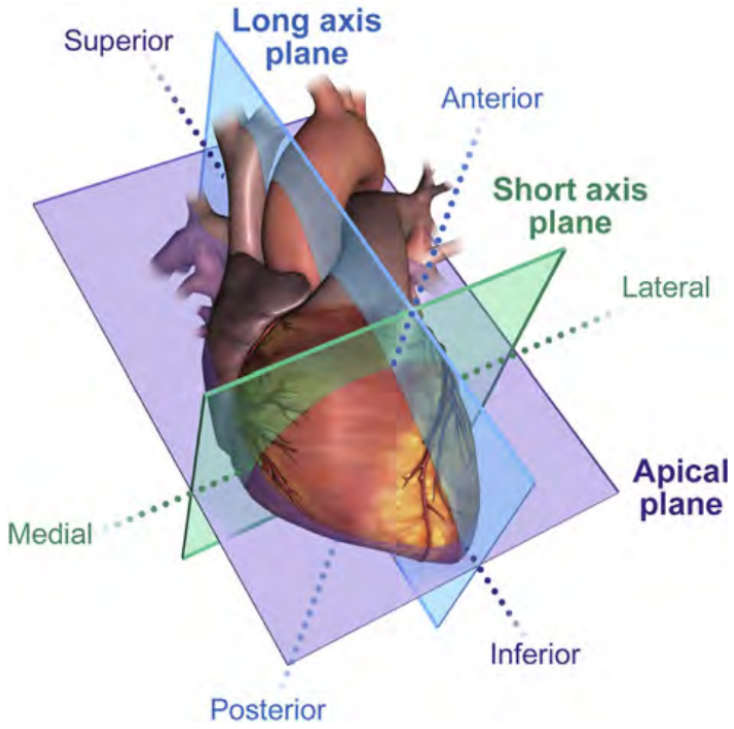


Figure 1.1: Scanning planes of the heart. The long and short axis planes correspond to images acquired in the parasternal window and the apical plane corresponds to images acquired from the apical window [3].

Feature-based registration relies on identifying distinctive features in the images, such as corners and edges, and matching them to perform image alignment, whereas intensity-based registration relies only on image pixel intensities. A transformation that maximizes the similarity between the pixel intensities of the images to be registered is calculated to find the optimal alignment. Mean squares, normalized correlation, and mutual information are some of the common intensity-based similarity metrics used in image registration. Based on the image coordinate transformation, we can categorize different types of image registration. *Rigid* registration is one of the simplest methods used to align images when only translations and rotations are needed, whereas the *affine* transform allows for shearing and scaling in addition to ro-

tation and translation. *Non-rigid or deformable* registration methods are used when two images are related through non-rigid geometric transformations, where the correspondence between images cannot be achieved without localized deformations. For example, the structures of the heart can deform nonlinearly with the cardiac motion. *Point-based or landmark-based* registration allows the embedding of expert knowledge into the registration procedure when dealing with challenging registration tasks that cannot be achieved with the usual registration techniques. In point-based registration, the transformation is computed by using predefined sets of landmark points or coordinates identified on the fixed and moving images. The images are aligned in such a way that the total Euclidean distance between the points on the fixed image and the corresponding points on the moving image is reduced.

1.2 Thesis contribution

An overview is provided of this thesis' contribution to the field of medical image registration.

Two-step rigid and non-rigid image registration for the alignment of images acquired from standard and nonstandard apical (Apl) windows:

This study proposes a fully automatic point-based rigid registration technique, followed by nonrigid B-spline registration, to align 4D echocardiogram images acquired at different apical windows.

Two-step rigid and non-rigid image registration for the alignment of images acquired from apical and parasternal (Psl) windows:

This study proposes a fully automatic point-based rigid registration technique, followed by nonrigid B-spline registration, to align 4D echocardiogram images acquired at apical and parasternal windows.

Visual and quantitative evaluation of the accuracy of registration:

The accuracy of registration was visually and quantitatively assessed by delineating the left ventricle in each scan and computing the Dice score overlap metric and the

Hausdorff distance mutual proximity measure between the first scan and the rest.

1.3 Thesis overview

The thesis is structured into 5 main chapters as follows:

In Chapter 2, the basics of image registration, registration algorithms for use between ultrasound images, image registration approaches in the literature, and details on how deep reinforcement learning is used for anatomical landmark detection are presented.

Chapter 3 presents the dataset, the methodology of the proposed algorithm, a point-based rigid registration followed by B-spline non-rigid registration to register multiview 3D echocardiography sequences. It discusses how the LV annotations were delineated to evaluate the accuracy of registration and provides details on how the Dice similarity coefficient and the Hausdorff distance measures were used to quantitatively evaluate the alignment between pairs of 3D and 4D scans.

Chapter 4 presents the training results of the landmark detection models and the visual and quantitative accuracy of the proposed algorithm using the registered images.

Finally, in Chapter 5, the work performed for this thesis is summarized, and future work, and limitations are discussed

Chapter 2

Background

2.1 Problem definiton

2.1.1 Heart anatomy and function

Heart anatomy

The cardiovascular system of the human body is made up of the heart and a network of blood vessels. The heart is the main organ of the cardiovascular system that pumps blood throughout the human body. The heart circulates the oxygenated blood from the lungs to all the body parts and works with other body systems to control the heart rate and blood pressure. The right side of the heart is responsible for receiving deoxygenated blood from the entire body and sending it to the lungs. The lungs oxygenate the blood and circulates to the left side of the heart. The left side of the heart is responsible for pumping the oxygenated blood to the rest of the body.

As shown in figure 2.1, the heart is divided into four chambers. Each side of the heart consists of two chambers, the upper chambers are called the atria, and the lower chambers, are the ventricles. The left atrium, the upper left chamber of the heart, receives oxygenated blood from the lungs and pumps it down into the left ventricle which circulates it to the body. The right atrium, the right upper chamber of the heart, receives deoxygenated blood from the body and pumps it into the right ventricle which then sends it to the lungs to be oxygenated. The left lower

chamber of the heart called the left ventricle (LV), receives blood from the left atrium and pumps it out under high pressure through the aorta to the body. Finally, the lower right chamber of the heart called the right ventricle (RV) receives deoxygenated blood from the right atrium and pumps it under low pressure into the lungs via the pulmonary artery. The heart itself is comprised of three layers of tissue. The outermost layer is the epicardium, the middle is the myocardium and the innermost layer is the endocardium. The myocardium is the thickest muscular layer, responsible for pumping the blood.

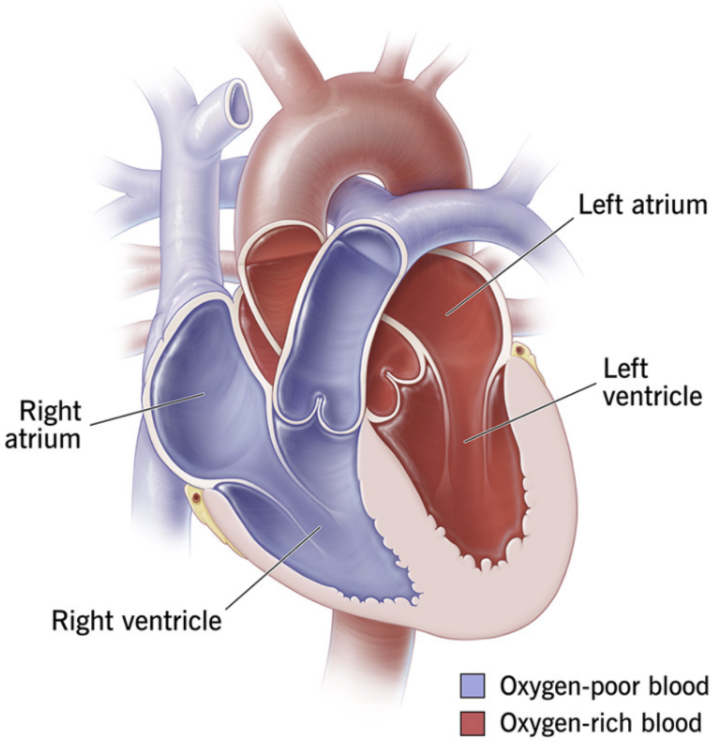


Figure 2.1: Basic anatomy of the heart [4].

Cardiac function and cycle

An electrocardiogram (ECG) is used to record and assess the heart's electrical activity over cardiac cycles. The use of the ECG offers a quick, non-invasive method of determining the heart's rhythm. Figure 2.2 shows the basic components of an ECG. The three main components are P wave, QRS complex, and T wave. The P represents the electrical activity of the upper heart chambers, the QRS complex

denotes the electrical movement of the ventricles and finally, the T wave represents the ventricles resetting electrically to prepare for the next muscle contraction [5]. Figure 2.3 shows an example of end-systolic and end-diastolic frames and the corresponding electrocardiogram tracings. With expertise, it is possible for clinicians to manually identify the ED and ES phases of an echocardiography image by visually inspecting each frame of the echo sequence for changes in the LV dimension and left-sided valves with relation to the ECG tracing [6].

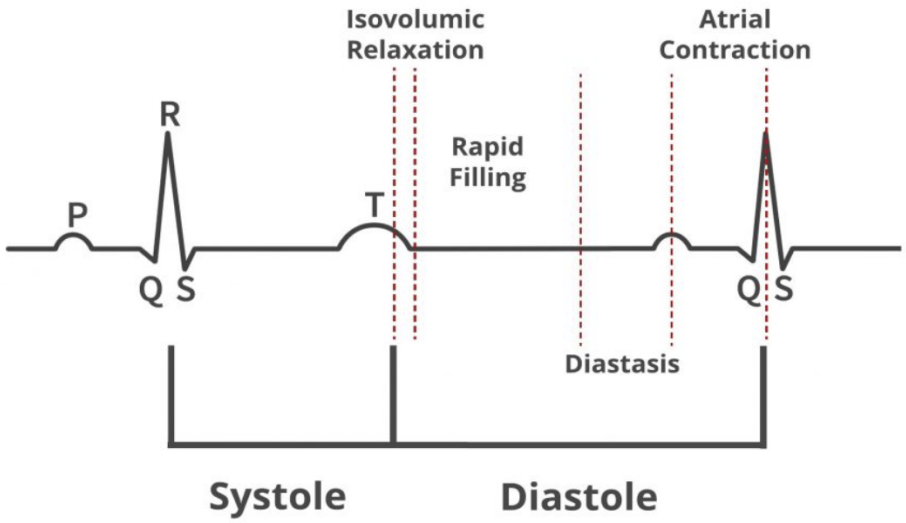


Figure 2.2: An example of an ECG of a heart [6].

2.1.2 Ultrasound imaging

Ultrasound is one of the most widely used imaging modalities that provides continuous real-time imaging. It is considered one of the safest modalities due to its lack of ionizing radiation. It is also portable and less expensive compared to other modalities, allowing it to be used in various fields of medicine [8]. Real-time three-dimensional (3D) acquisition is a recent advancement in ultrasound imaging. Conventional ultrasound transducers provide two-dimensional images, usually at 30 to 60 frames per second. Before the introduction of real-time 3D ultrasound imaging, 2D planar images were obtained by systematically moving a standard ultrasound transducer across

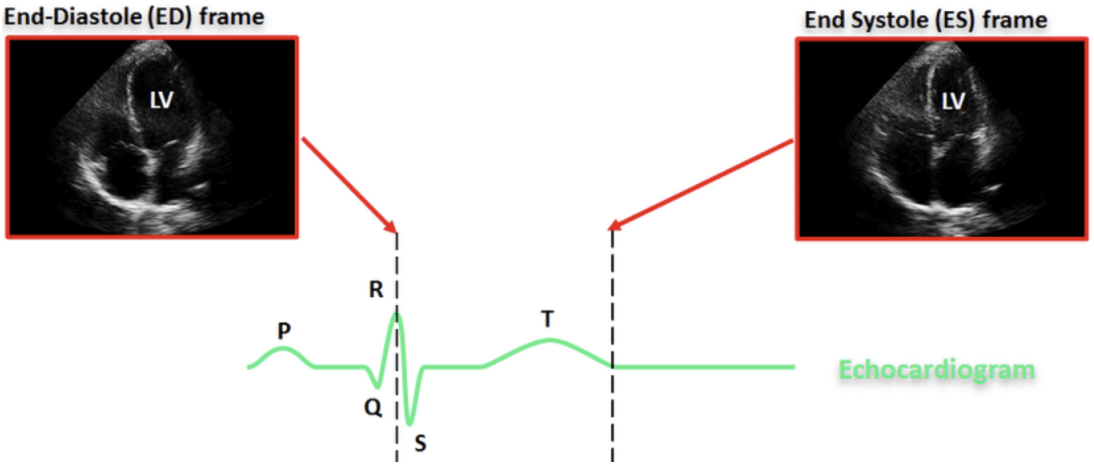


Figure 2.3: Sample end-diastolic (ED) and end-systolic (ES) frames with corresponding electrocardiogram (ECG) tracings [7].

a 3D area, and then the images were stacked to create 3D images [9]. However, this method was sensitive to patient motion and had difficulties in localizing the transducer. Furthermore, it is not sufficiently swift to capture three-dimensional images of a dynamic organ like the heart. On the other hand, real-time 3D ultrasound acquisition has the ability to generate around 20-30 volumes per second and accurately capture the movement of the heart with a remarkably high temporal resolution. Real-time 3D ultrasound imaging can be described as a form of four-dimensional (4D) imaging, where time is considered as the fourth dimension.

2.1.3 Evaluating cardiac function using ultrasound imaging

US imaging of the heart consists of both two-dimensional (2D) imaging and 3D imaging over time. There are several advantages of using the 3D modality compared to the traditional 2D approach. Regional wall motion analysis and inspecting the wall thickness are crucial parts of diagnosing disease in a patient. It can provide insight into specific areas of the heart that are functionally abnormally. With 2D imaging, the sonographer may have to modify the orientation of the transducer to observe a particular segment of the myocardium. The use of 3D US allows for a much larger

volume to be captured [8]. Apart from regional analysis, 3D US enables the ability for valves to be analyzed and characterized [8]. By using image registering to register RT3DE images obtained from several sonography windows, it is possible to enhance the recognition of structures and achieve a substantial improvement in image quality compared to the original scans.

2.2 Image registration overview and methods in literature

2.2.1 Basics of image registration

Image registration is the process of aligning two or more images or volumes and transforming them into the same coordinate system. The technique is used in a wide variety of medical imaging applications. The images can be obtained from various points in time, using different sensors and/or from different viewpoints. Typically, when we align two images, one is considered stationary and referred to as the target, reference, or fixed image. The other image is subjected to various transformations such as translation, rotation, warping, etc. This image is referred to as the source or moving image. The process of mapping features from one image to another is referred to as the geometrical transformation, deformation field, or displacement field. In order to register two or more images, it is necessary to estimate a transformation. Typically, these changes are categorized as *rigid*, which involves only rotation and translation to align images, *affine*, which includes scale factors and shears in addition to rigid registration, and *non-rigid/deformable*, where the correspondence between structures in two images cannot be achieved without some localized stretching or deformation. Figure 2.4 illustrates these classifications. Soft tissues, such as those found in the human body, generally do not conform to a rigid or affine approximation[10].

Respiration induces motion in the patient's body during cardiac imaging acquisition. Utilizing gating or synchronization with the heartbeat and/or respiratory cycle

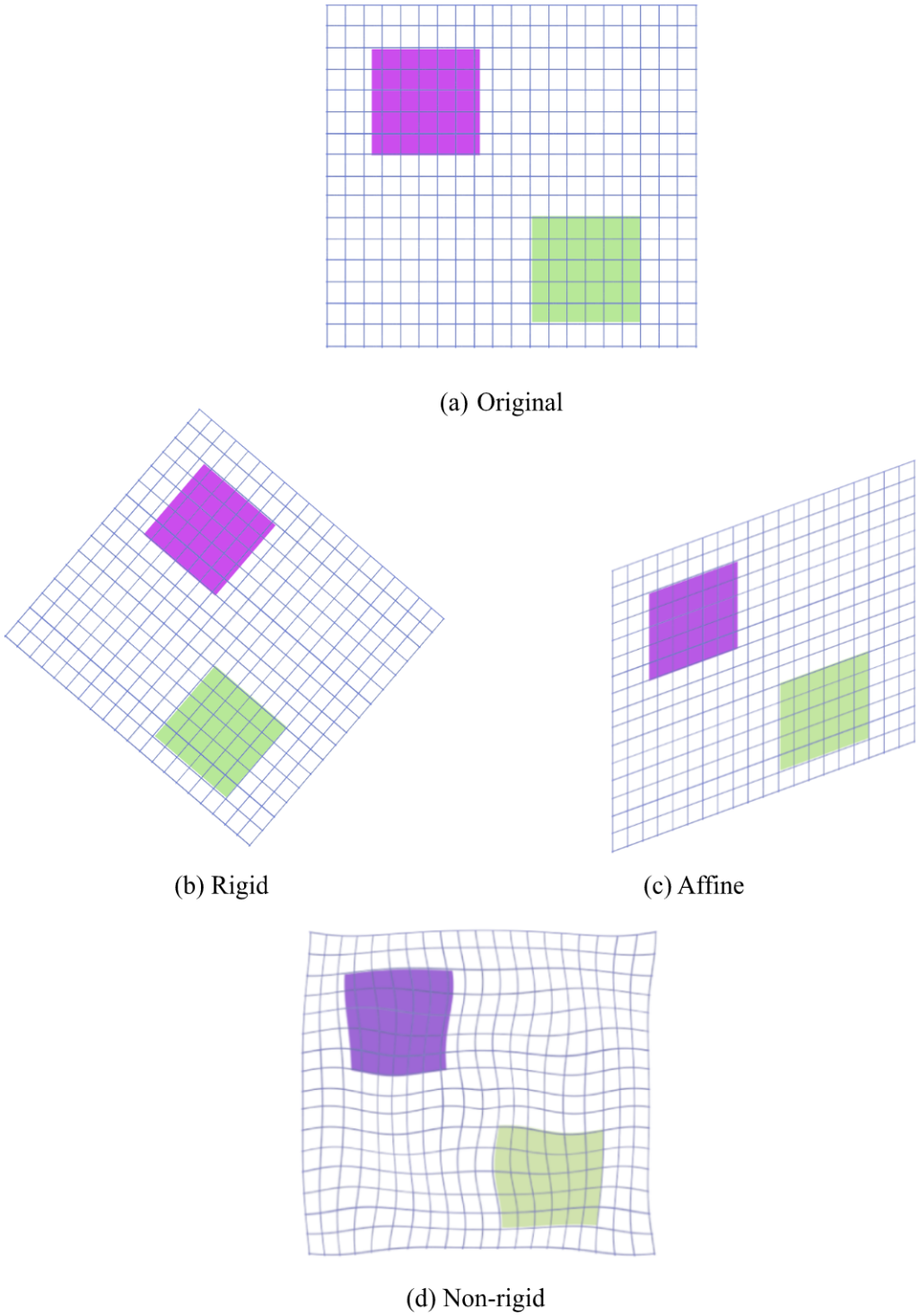


Figure 2.4: Three transformation models.

is a common approach to acquiring a series of volumes. However, the volumes may still be misaligned. Undesired patient movement can potentially lead to the occur-

rence of motion artifacts. In order to conduct an accurate analysis of the patient's progress over a period of time, it is necessary to align the volumes to a common reference frame. Hence, image registration is used to carry out this alignment. Image fusion involves merging data from numerous scans to expand the field of view (FOV) of the heart. This procedure is commonly conducted when the complete heart is not clearly seen in the acquired scan. Image registration is often used to align different volumes and facilitate the observation of key structures.

2.2.2 Landmark based image registration

The localization of anatomical landmarks or fiducial markers in medical imaging is used to initialize intensity-based registration between two images and to register between image spaces. In difficult registration problems, registering images using landmark points provides accurate results. During the registration process, the transformation that aligns the fixed and moving images given a set of pair landmarks is computed. SimpleITK's landmark-based transform initializer computes the best-fit transform that maps the fixed and moving images in the least squares sense (sum of the squares of the residuals). Point 1 in the fixed image will get mapped close to point 1 in the moving image, etc. Using 3D rigid transform alongside a set of landmarks would allow computing the translation and the rotation components to align the images minimizing the distance between given landmark points on two images.

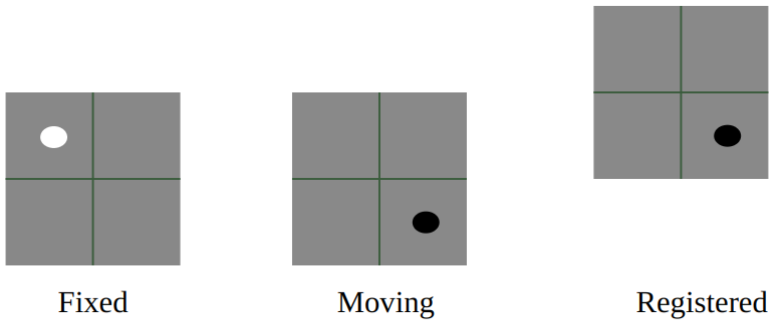


Figure 2.5: Registration using manually placed landmarks.

2.2.3 Deformable image registration

Deformable image registration is the process of finding correspondence between structures in two images that cannot be achieved without some localized stretching or deformation. The three primary components of deformable image registration are 1) the similarity measure, 2) the transformation model, and 3) regularisation. The similarity measure is typically specified as an objective function that obtains its optimal value when two images have a specific relationship. The similarity measure in intensity-based registration methods is defined directly by image intensities. The mean squares image similarity metric is calculated by obtaining the square of the difference between matching pixels in each image and then calculating the average of these squared differences. The mean squared error S_{MSE} could be defined as,

$$S_{MSE}(I_F(P), I_M(T(P,\mu))) = \frac{1}{n}\sum_{i=1}^{n}(I_F(P_i) - I_M(T(P_i,\mu)))^2 \tag{2.1}$$

where $I_F(P)$ is a fixed image, $I_M(P)$ is a moving image, transformation with parameters μ is $T(P,\mu)$ and n is the number of pixels in the image. The optimal value of this metric is zero. Poor matches between two images result in large values of the metric.

The Mutual Information (MI) [11] metric calculates the mutual information between two images. MI is a measure of how much information one voxel intensity in one image tells about the corresponding voxel in the second image. This involves measuring the probability density distributions (PDF) of the intensities of the fixed and moving images, without the need to explicitly define the exact form of the relationship. Mattes Mutual Information [12] is an implementation that samples the same pixels in each cycle. Using a predetermined collection of discrete positions to assess the marginal and joint probability density functions enhances the smoothness of the trajectory in the search space. Given a fixed image $I_F(P)$, a moving image $I_M(P)$, and a transformation with parameters μ is $T(P,\mu)$ the mutual information is

defined as,

$$S_{MI}(I_F(P), I_M(T(P,\mu))) = H(I_F(P)) + H(I_M(T(P,\mu))) - H(I_F(P), I_M(T(P,\mu))) \tag{2.2}$$

where $H(I_F(P))$ and $H(I_M(T(P,\mu)))$ are the marginal entropies, and $H(I_F(P), I_M(T(P,\mu)))$ is the joint entropy of fixed and moving images.

The transformation model is the second component of an image registration process. The transformation model specifies the collection of possible transformations necessary for aligning two images. In rigid parametrization, rotation and shifting are sufficient transformations to align two images with different orientations. Deformable transformations can be classified into parametric and non-parametric types. In non-parametric approaches, the estimation of a transformation is done by an unknown function. Parametric approaches rely on a specific parameterization of the transformation, such as spline-based methods [13]. The B-spline method uses B-spline curves to define a continuous deformation field to map voxels of fixed and moving images [14]. There is no unique solution to estimate a non-rigid transformation based on only a set of given control points/landmarks. There can be an infinite number of transformations that have different behavior in the remaining parts of the image but will match the corresponding landmarks. Usually, a regularization term is used to enforce some constraints to transformation, to solve the problem with a unique solution.

2.2.4 Multiresolution image registration

Multiresolution registration is an approach to reduce the computational complexity of image registration problems. It allows algorithms to work on both fine and coarse scales, rather than waiting for local pixel-level operations to converge at large scales. Multi-resolution methods have also been called coarse-to-fine, hierarchical, and pyramidal in the literature. The pyramid model for a multiresolution image registration method is shown in Figure 2.6. Level n has the original images, with each upper level featuring images reduced in size by a certain factor (typically 2) to create new images.

To acquire the intended low-resolution images, the process is repeated. The minimum resolution can be automatically determined or can be provided as input by the user. Initially, the images at the top level are aligned. Reducing image size simplifies the correspondence procedure by minimizing local geometric discrepancies between the images. Registration parameters at level n are estimated using the registration result at level $n-1$. The process is repeated until registration at the highest resolution (level n) is achieved. This subdividing process enables us to deal with smaller images and reduces the search area. In addition, knowing the approximate registration parameters makes finding the ultimate registration parameters faster.

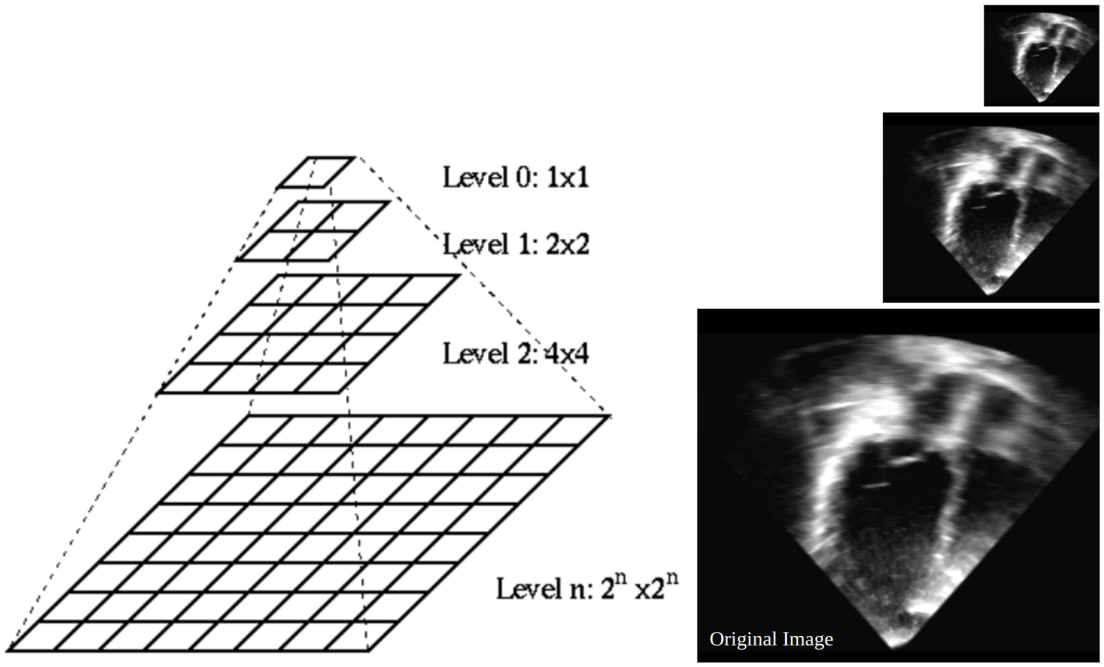


Figure 2.6: Coarse to fine image representation for multiresolution image registration.

2.3 Methods in literature

2.3.1 Ultrasound image registration

Ultrasound is often used to acquire images of the human heart, brain, liver, and other organs [15],[16],[17] and is increasingly being used in animal studies [8]. Since

US imaging is used to acquire images of different organs of the human body, over the years many registration algorithms have been developed to address monomodal ultrasound-to-ultrasound registration. In [18], a fully automatic, feature-based, non-rigid registration algorithm was proposed to register 3D brain ultrasound images. Their feature-based registration reduced the mean target registration error of images acquired before and after opening the dura mater from an initial value of 3.3 to 1.5 mm. In [19], a method was proposed using Gabor filters to extract features to register 3D ultrasound images of the fetal head. Among the different metrics the authors used, they achieved 92% of the successful registration rate for the correlation ratio metric. In [20], a hybrid procedure was proposed using global statistics and local textural features to register envelope-detected radio frequency ultrasound data of the human neck and was shown to outperform standard measures such as SSD, NCC, and Hellinger distance. In [21], a fully automated deformable registration algorithm was developed using the Bayesian regularization framework to register free-hand ultrasound volumes of female breasts. In [22], a semiautomated deformable approach was used using manually placed control points along with the Nelder simplex algorithm as an optimizer to register breast images. In [23], a registration framework with mutual information was proposed as a similarity metric to rigidly register 2D liver images. An affine registration method was proposed in [24], and a rigid registration approach was proposed in [25] to register real-time 4D liver ultrasound images. Compared to brain or head image registration, cardiac image registration is more complex because the heart is a moving organ and presents fewer accurate anatomical landmarks that can be used for registration [26]. Although several articles have been published in the field of medical image registration, only a few focus on cardiac image registration. In [27], a multiresolution deformable algorithm was proposed to automatically track the mitral annulus on 3D ultrasound images. To calculate performance, the root mean square difference between manually traced landmarks and automatic tracking was calculated, and the algorithm produced 1.96 ± 0.46 mm as the average error,

which is within the acceptable error range according to the authors. In [9], a mutual information-based registration framework was proposed to register real-time 3D ultrasound data. A pair of volume frames at identical cardiac phases were selected from two frame sets to perform the registration. The authors have evaluated the accuracy of the registration by artificially introducing a misalignment.

Echocardiography, also known as cardiac ultrasound, can acquire moving images of the heart in real time and is an important tool in the diagnosis of cardiovascular disease, a leading cause of death worldwide [28]. Accurate cardiovascular evaluation plays an important role in early diagnosis of the disease. Most often, two-dimensional (2D) echocardiography image sequences are acquired to see the motion of the heart; however, the imaging is limited to only one plane in each scan. The 3D echo technique is more efficient and useful and captures real-time three-dimensional views of the heart's structure and shows how well the heart pumps blood. Using 3D allows clinicians to study parts of a heart from different angles with greater accuracy, enabling a more accurate assessment of heart function. The real-time 3D echocardiography (RT3DE) imaging technique allows better 3D imaging by extracting spatial features along with temporal information, thus improving clinical decision making [29]. In addition, RT3DE allows the compilation of information on cardiac anatomy and function using single-beat or multi-beat acquisitions. However, the practicality of employing multiple pulses in multibeat modality is constrained by the need for precise control of breathing and maintaining normal heart rate during the image acquisition process [30].

Although there have been technological advancements, many acquired RT3DE images tend to be of poor quality, characterized by the absence of anatomical information, decreased spatial and temporal resolution, speckle noise, and a limited field of view (FOV) compared to 2D images [31], [32]. By registering RT3DE images obtained from several windows, it is possible to enhance the recognition of structures and achieve a substantial improvement in image quality [33]. Considering all of the

aforementioned drawbacks, the authors of [34] have introduced a method that registers apical and parasternal echocardiographic images by utilizing local orientation and phase differences. An initial rigid transformation was performed by identifying three manual landmark points on the end-diastolic frames of both the fixed and moving images. This was then followed by registration using a unique voxel similarity function. An inherent limitation of the algorithm being discussed is the requirement for manual involvement in order to obtain the first alignment. Our study introduces an entirely automated technique for aligning multiview 3D echocardiogram images. Additionally, we employ a non-rigid registration as a second step to rectify minor misalignments. The alignment accuracy of the proposed method is assessed by measuring the overlap between the left ventricular regions. As the starting point of the proposed method, we utilize Leroy et al.'s multi-agent communicative reinforcement learning system [35] to train a model. This model is used to automatically detect landmarks in 3D echocardiography images. Subsequently, we employ SimpleITK's landmark-based registration technique to rigidly align the images. Next, we employ a B-spline-based multiresolution deformable registration technique utilizing the Simple Elastix package to rectify slight discrepancies in the rigidly aligned images mentioned beforehand.

2.4 Deep reinforcement learning for anatomical landmark detection

2.4.1 Reinforcement learning

Reinforcement learning (RL) is a subset of Machine Learning (ML) that falls within the broader domain of Artificial Intelligence (AI) and is influenced by behavioral psychology and neuroscience [36]. In RL, an artificial agent learns by taking actions in an environment and receiving a new state along with a reward, similar to how animals learn via trial and error. Environmental policies are acquired by direct learning from

inputs with high dimensions. This problem is often modeled mathematically as a Markov decision process (MDP). Figure 2.7 shows a reinforcement learning loop. An agent at every timestep t is in a state S_t takes action A_t, receives a scalar reward R_t, and transitions to the next state S_{t+1} according to environment dynamics. The agent attempts to learn a policy or map from observations to actions, which would increase the total cumulative reward. This learning technique allows the agent to automatically execute a sequence of decisions that optimize a reward metric for the task, without any human interaction or explicit programming to carry out the job [37].

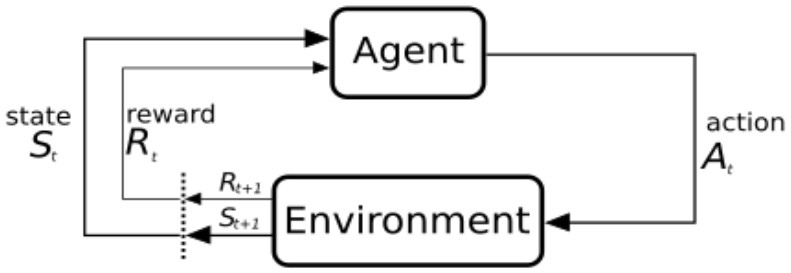


Figure 2.7: Reinforcement learning loop, the agent takes an action in the environment and receives the updated state and corresponding reward [38].

As mentioned above, the agent comprises two elements: a policy and a learning algorithm. A policy is a function that determines actions depending on observations obtained from the environment. Usually, the policy is a mathematical model that can approximate functions and has adjustable parameters, such as a complex neural network. The learning algorithm iteratively adjusts the policy settings using information from the actions, observations, and rewards. The objective of the learning algorithm is to identify an optimal policy that maximizes the total reward obtained throughout the learning.

2.4.2 Medical Applications

Reinforcement learning agents can be used in clinical applications to take advantage of their high accuracy and fast navigation abilities while working with 3D medical images. RL has been used in various clinical applications such as locating landmarks in fetal head ultrasound, adult brain, and cardiac MRI scans [39]. It has also been used to find standard view planes in brain MRI images [40]. Supervised machine learning solutions for anatomical landmark detection learn from large annotated image databases. These solutions for anatomy detection are subject to several limitations such as suboptimal feature engineering techniques and computationally expensive search strategies. To overcome these issues the detection problem has been reformulated as a deep reinforcement learning problem in [41]. In scenarios where a 3D CNN in a supervised context is used to predict landmarks in medical images, it would take the whole image as an input and outputs the predicted landmark's position. Whereas the power of the RL approach is that it takes much less memory and processing power while still producing state-of-the-art results [42].

2.4.3 Anatomical landmark detection

Deep reinforcement learning (DRL) combines elements of classical reinforcement learning with deep learning. The utilization of deep Q-networks (DQN) emerged due to the progressively complex state and action spaces, hence the inclusion of the term 'deep' in deep reinforcement learning. The goal is to evaluate the expected Q-value of a given state s when taking action a. Q-learning [43] aims to determine a Q-value that evaluates the effectiveness of doing a certain action in a particular state s by acquiring knowledge through a policy that maximizes the cumulative reward during the training process. Mnih et al. [36] suggested employing a deep neural network called DQN to estimate these q-values.

In this work we use Leroy et al.'s [42] proposed communicative DQN-based RL agents for the detection of anatomical landmarks in echocardiography images. These

agents are programmed to acquire knowledge through learning while exploring various locations in 3D medical imaging and tested on brain MRI and fetal head ultrasound images. Figure 2.8 depicts a schematic representation of the navigating agents within a 3D volume or environment *E*. The example shows three agents looking for landmark A. Similarly, in the proposed registration approach, three agents each, a total of six, were looking for landmarks B and C.

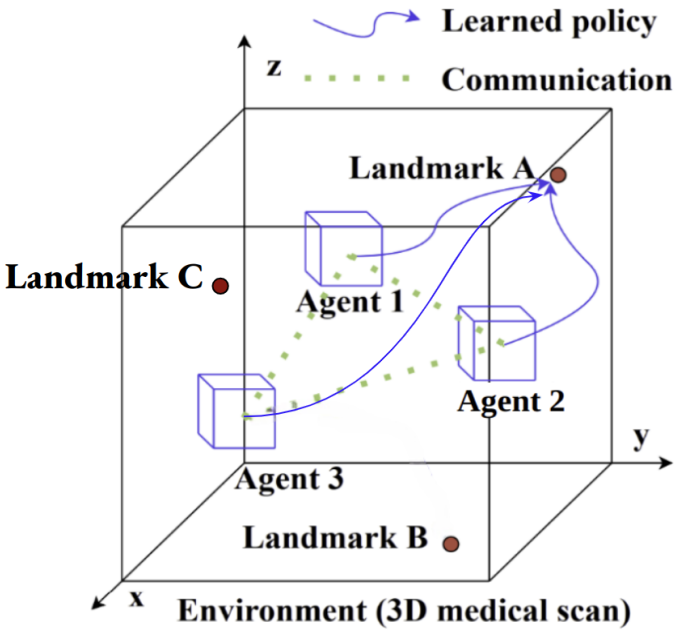


Figure 2.8: A schematic diagram of multi-agents interacting with the 3D image environment *E*. All three agents will take action toward a target landmark A during each step. After taking the sequential action, the learned policy is formed by the path between the starting points and the target landmarks [42].

For the agents, the state represents a region of interest (ROI) that is 45×45×45 voxels in size and it is centered on each agent. In order to enhance the stability and convergence of the network, the history of the last 4 states is provided as an input. At the start of each episode, every agent is positioned randomly within the 80% of the inner part of the image. The agent stops upon reaching the designated landmark. During the process of inference, the terminal state is activated when the agent oscillates around a target location. An agent can take action in the six directions

in the 3D Cartesian coordinates, namely left, right, up, down, forward, or backward till it reaches the terminal state. The reward is calculated using the difference between the Euclidean distance of the current point of interest and the target landmark and the point of interest of the previous step and the target landmark. The calculated reward is clipped between -1 to 1 to make sure positive rewards are given if the agent moves towards the target. The architecture of the communicative multi-agent reinforcement learning for anatomical landmark detection is shown in Figure 2.9. It shows an example of two agents sharing the same convolutional layers. The model takes a tensor of size $number\ agents \times 4 \times 45 \times 45 \times 45$ as an input. The convolutional layers are shared between all the agents. The agents learn to communicate by averaging the output of the fully connected layer of each agent, which is then concatenated to the input of the next fully connected layer. The last fully connected layer is a size of 6, which is the size of the action space (six directions) mentioned above. Finally, the model is trained using the DQN loss function [44].

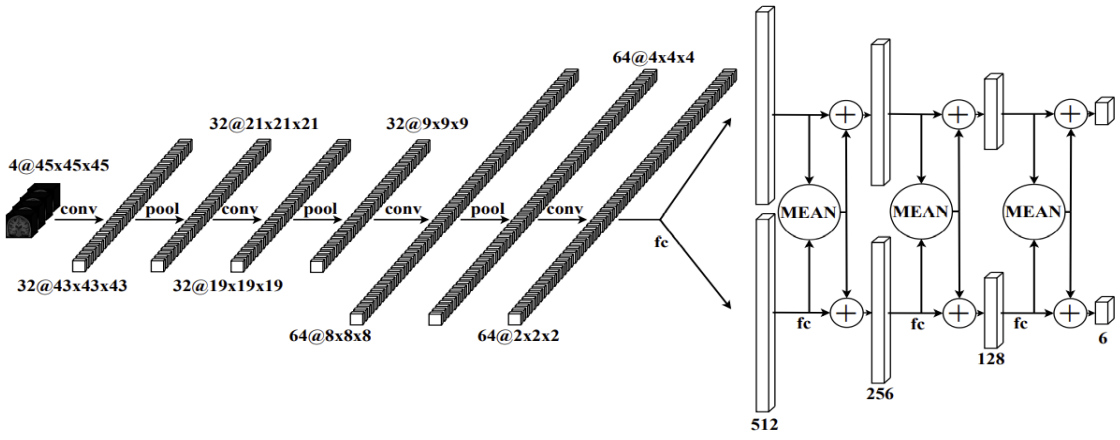


Figure 2.9: The communicative multi-agent reinforcement learning architecture for anatomical landmark detection [42].

Chapter 3

Two-Step Rigid and Non-Rigid Image Registration

A flow diagram that describes the entire registration process is shown in Figure 3.1. The individual steps of the procedure are explained below.

3.1 Participant population and data

Seven volunteers with no known cardiac disease, stable sinus rhythm, and 18 years or older were scanned. The Health Research Ethics Board of the University of Alberta granted approval for the study, and written informed consent was obtained from all participants. In this study, a total of 124 4D (3D+time) echocardiography scans of dimensions $n \times 272 \times 176 \times 208$ and $n \times 224 \times 176 \times 208$ were used where n represents the number of volumes or frames within a sequence and ranges between $25 - 30$ for each scan. All images were resampled to an isotropic voxel size with spacing $1 \times 1 \times 1$ mm along the x, y and z coordinate directions for training the anatomical landmark detection models.

3.2 Image acquisition and preprocessing

Three-dimensional echocardiogram images of the left ventricle (LV) were acquired in real-time from multiple windows, mostly apical and parasternal windows, using an X5 probe on a commercially accessible Philips EPIQ 7C scanner. The ultrasonic

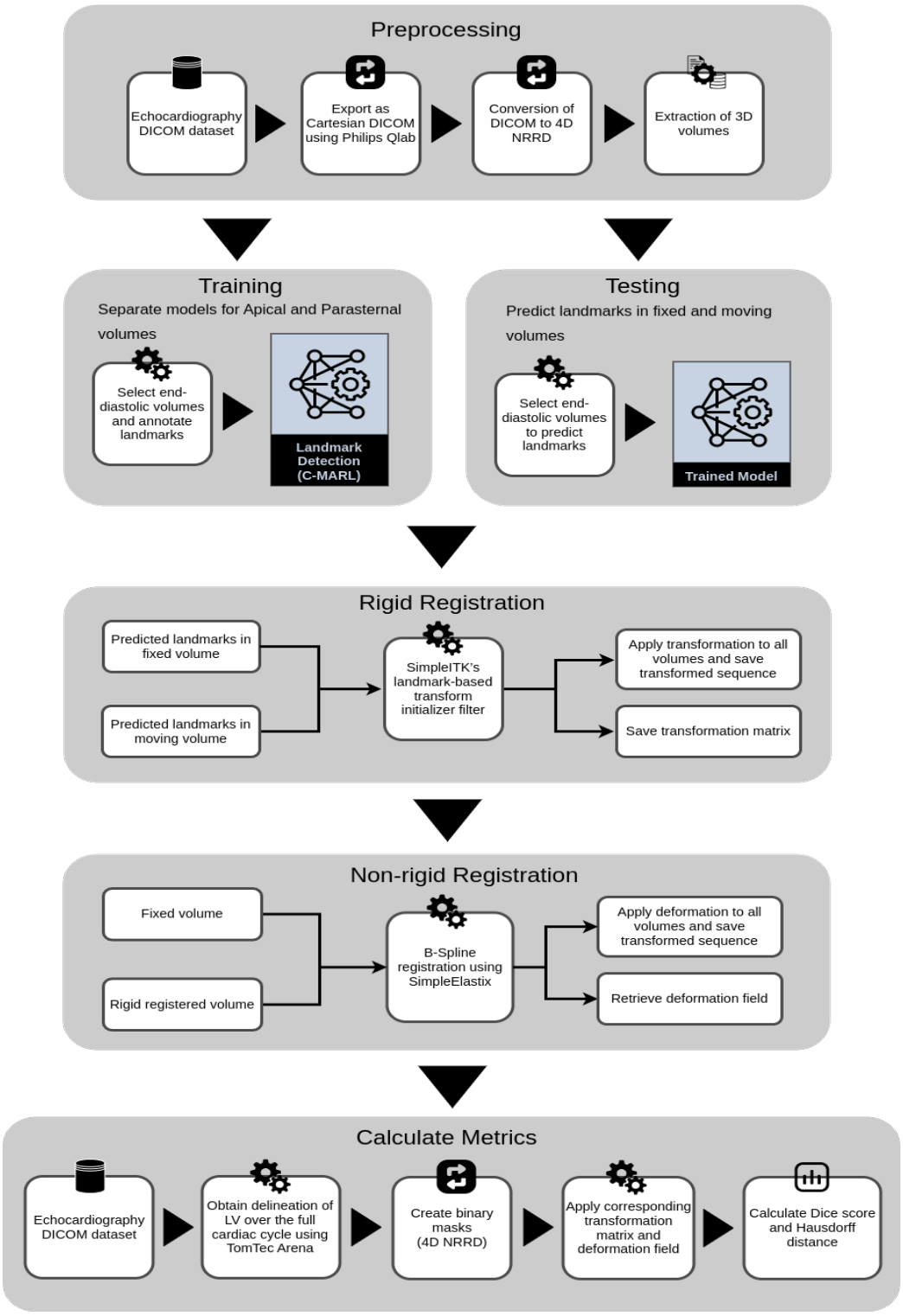


Figure 3.1: Chart showing the key steps in the registration procedure. (icons: Flaticon.com)

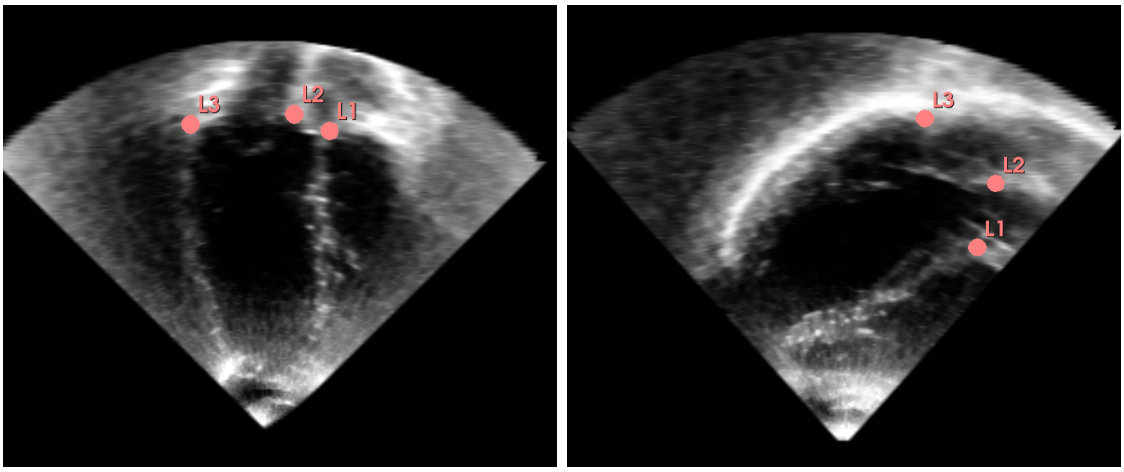
transducer was attached to a Universal Robots UR10e arm, which was used to control the transducer during the scanning procedure. ECG-synchronized multi-beat full-volume datasets were obtained, comprising around three to four images captured from the same acquisition window. Each volunteer had a single session in which all datasets were acquired. Participants were instructed to hold their breath to try to prevent the occurrence of stitching artifacts. The scans were acquired with a starting average frame rate of 25 ± 3 Hz employing 4-beat (4Q) and heart model acquisition (HMQ) techniques. DICOM images were exported using QLab software in Cartesian DICOM format to be compatible with third-party software. The proposed landmark detection model does not work directly on 4D images and is applied to 3D volumes (spatial). Therefore, the exported cartesian DICOM images were converted into NRRD sequences using the Pynrrd Python module and then 3D volumes were extracted from the NRRD sequence. In total, 213 apical and 149 parasternal 3D volumes at the end-diastolic phase were selected based on image quality among 7 volunteer participants.

3.3 Training

The first step of the proposed registration method was a landmark-based rigid registration using the SimpleITK library. The SimpleITK's landmark-based transform initializer calculates the best fit that maps the fixed and moving images in the least squares sense using an equal number of landmarks on the fixed and moving images. The landmarks were predicted using a neural network approach to remove the manual overhead of landmark annotation on the fixed and moving images. The proposed neural net models were trained using communicative deep Q-network-based reinforcement learning agents implemented by Leroy et al. [35] and used to automatically detect three landmarks, given a new 3D echocardiography volume. Previous research carried out using the proposed landmark detection method to identify landmarks on adult MRI and fetal head ultrasound showed improved accuracy in locating the de-

sired landmark through the use of communicative agents. To the best of the authors' knowledge, this is the first application of the communicative deep Q-network-based reinforcement learning technique to detect landmarks in 3D echocardiography images. Furthermore, the findings demonstrated that the collective efforts of multiple agents in looking for a single landmark are more effective than assigning a single agent to each landmark. Therefore, in our study, we employed three agents to search for each of our landmarks. Two models were trained: one specifically for parasternal 3D volumes and another specifically for apical landmark identification. We proceeded with different models on the basis of the following rationale. 1) There is a difference in the structures observed in the 3D apical and parasternal volumes; specifically, the apex part of the heart is usually not captured in the parasternal volumes. 2) There is an unequal distribution of data across the two datasets.

Figure 3.2 shows the chosen landmark points used for point-based registration.



(a) Apical (b) Parasternal

Figure 3.2: End-diastolic frames of apical and parasternal 3D volumes with manually annotated landmark points. The first point (L1) was picked where the right coronary cusp attaches to the aortic root annulus in PSLAX. The second (L2) and third (L3) points were selected at the anterior and posterior mitral valve leaflet attachment on the mitral valve annulus in the PSLAX, respectively.

We used multiplanar reconstruction on the collected 3D volumes to identify landmarks, specifically focusing on the end-diastolic frames. This was done using a 3-

chamber view. The initial point (L1) was selected where the right coronary cusp attaches to the aortic root annulus in the parasternal long axis view (PSLAX). The second (L2) and third (L3) locations were chosen at the anterior and posterior attachments of the mitral valve leaflet on the mitral valve annulus in the PSLAX view. The clinical expert, a diagnostic cardiac sonographer, manually labeled all the chosen landmarks using three orthogonal views in the open source program 3D Slicer [45]. 3D Slicer utilizes the RAS coordinate system as its default. The multi-agent landmark identification approach requires the use of real-world physical coordinates. To do this, the RAS coordinates of the annotated landmarks were transformed into physical x, y, and z coordinates by using the origin and spacing data. The 3D apical and parasternal datasets were separated to form the training set comprising 166 apical and 113 parasternal volumes, the validation set comprising 23 apical and 16 parasternal volumes, and the test set comprising 33 apical and 20 parasternal volumes.

The accuracy of the landmark detection model is assessed by computing the Euclidean distance error between the detected landmarks and the target landmarks. Therefore, similar to [35], the most suitable model was selected during the training process considering the highest accuracy achieved in the validation data set. A total of nine agents, three assigned to each landmark, performed a search within a 3D echo volume to locate three separate landmarks. The batch size was set to 128 and 64 for the apical and parasternal data, respectively, and the number of episodes was set to 75,000. Other training hyperparameters remained unchanged and used their default values. We used an AMD CPU and an Nvidia A100 graphics card equipped with 40GB of RAM for training. More details on how the agents learn are explained in section 2.4.3.

3.4 Image registration

3.4.1 Rigid registration

From the extracted 3D volumes, we chose the standard apical end-diastolic 3D volume of the highest quality from each volunteer as a fixed image. All remaining images were deemed to be the moving set. Image registration often involves applying a transformation to the set of moving images. We applied the model that had the highest validation accuracy to make predictions on both fixed and moving images for landmark identification. As stated in Section 3.3, the training process and prediction are based on the physical x, y, and z coordinates of the landmarks. SimpleITK uses LPS coordinates. Consequently, we transformed physical coordinates back into RAS coordinates. LPS coordinates were derived by negating the RAS coordinates. The following figure displays three spaces together with their respective axes.

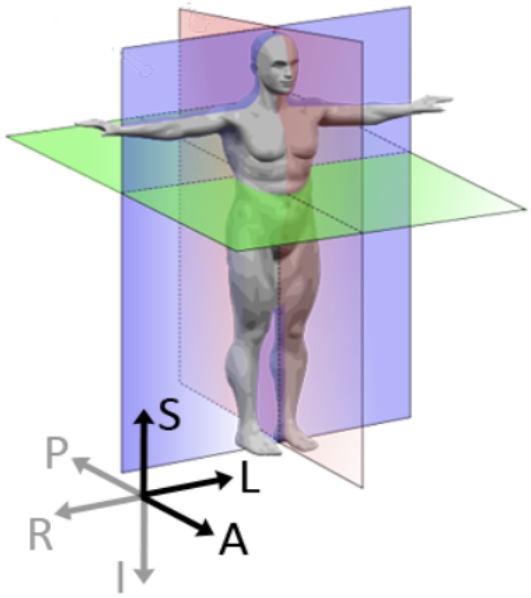


Figure 3.3: Figure illustrating the three spaces and their corresponding axes. (Source: slicer.org)

Finally, the predicted sets of coordinates were utilized to implement a point-based

rigid registration, which resulted in the transformation of the moving image to align with the fixed image. After the registration procedure was performed, the transformation was extended to the remaining 3D volumes in the same sequence, and the whole sequence of echo data was saved as an NRRD file. Furthermore, the resultant transformation matrix for each moving sequence with respect to the fixed image was stored in a file.

3.4.2 Non-rigid registration

SimpleElastix [46] is a freely available image registration library that extends SimpleITK and incorporates a Python wrapper for image registration techniques. For the second part of our two-step registration procedure, we used a multimetric, multiresolution B-spline deformable registration. This algorithm was implemented using the SimpleElastix library and applied to images that had already undergone rigid registration. The fixed image remained unchanged, whereas nonrigid registration was performed exclusively on the end-diastolic frame. B-spline registration involves mapping every voxel in a moving image to a corresponding voxel in a fixed image using a continuous deformation field defined by B-spline curves [14]. To enhance the accuracy of the registration process and reduce the time required for computation, a multiresolution framework was employed to apply the B-spline transform. This approach involved two consecutive phases for the apical images and a single phase for the parasternal images, resulting in an improved registration process. Furthermore, during the registration procedure, the similarity metric between the corresponding voxels of the fixed and moving pictures was computed using two metrics, namely, the mean squares and the Mattes mutual information. The mean squares similarity metric calculates the total squared differences between corresponding pixels in the fixed image and the moving image. The Mattes mutual information measure calculates mutual information using the approach proposed by Mattes et al. [12], comparing fixed and moving images. Using the parameters described above, a B-spline transform

was computed to determine the spatial correlation between the two 3D volumes. Subsequently, the deformation field obtained from the transform was utilized to modify the remaining 3D volumes in the sequence, and the modified sequence was stored.

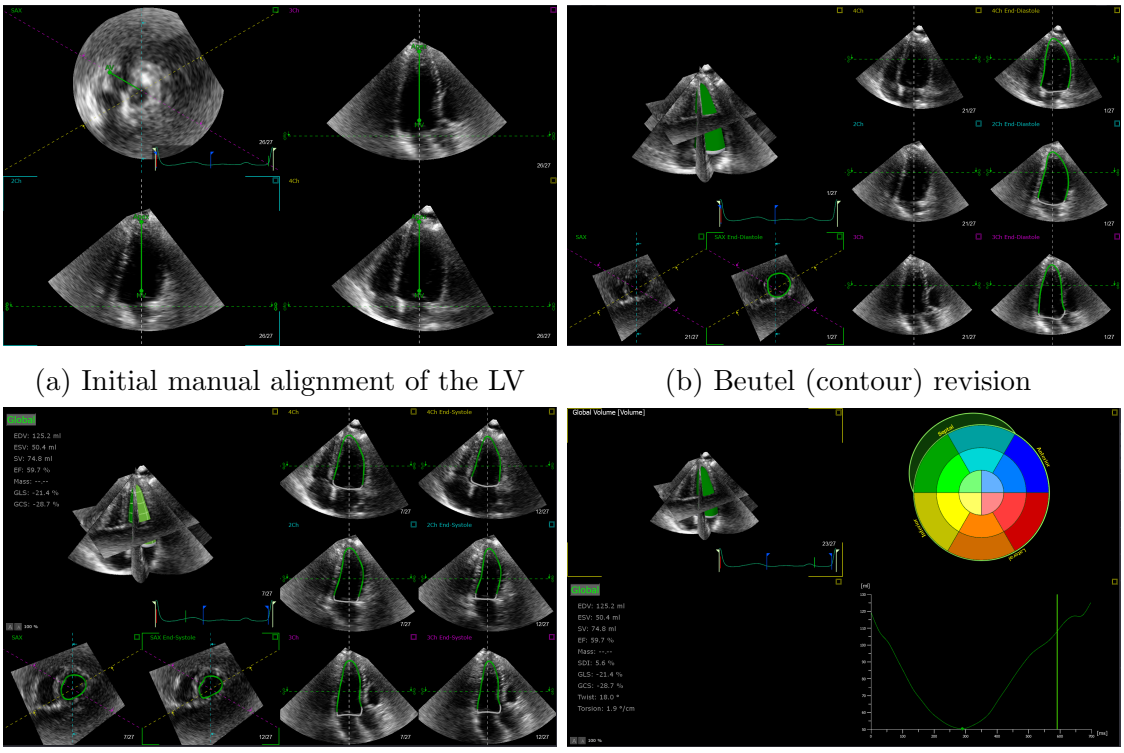
3.5 Evaluation

3.5.1 Delineation of left ventricle

In order to assess the accuracy of the registration, the degree of overlap between the regions of the left ventricle (LV) was measured. Usually, there is poor alignment between the LV areas in two distinct scans, specifically between the apical and parasternal scans. Once the images are registered, the alignment should be improved. Therefore, the Dice score, which measures the overlap between two areas of LV before and after registration, was computed to assess the accuracy of the registration. The Hausdorff distance was used to calculate the mutual proximity between the LV areas. The LV annotations were delineated by a diagnostic cardiac sonographer and a cardiologist using TomTec Arena[1], a commercially available software. TomTec Arena uses *speckle tracking* method for obtaining a delineation of the LV over the full cardiac cycle. Speckle tracking is a method that examines the tissue motion by analyzing speckle patterns. The individual steps are displayed in Figure 3.4. The clinician first manually identifies the apex and the mitral valve, and the LV is aligned vertically according to these two points. In the second step, a rough estimation of the endocardial borders is automatically produced for the ED and ES frames, where the clinician has the opportunity to adjust them. Then tracking (using a speckle tracking method) is initiated and revised if necessary. The final results include volumetric measurements, regional analysis, and metrics concerning the strain [47].

The annotations were performed as mentioned above on the echocardiogram pictures obtained immediately from the scanner. The LV annotations were exported as UCD segments in the original image coordinate system and subsequently transformed

[1]https://www.tomtec.de/products/tomtec-arena



(a) Initial manual alignment of the LV

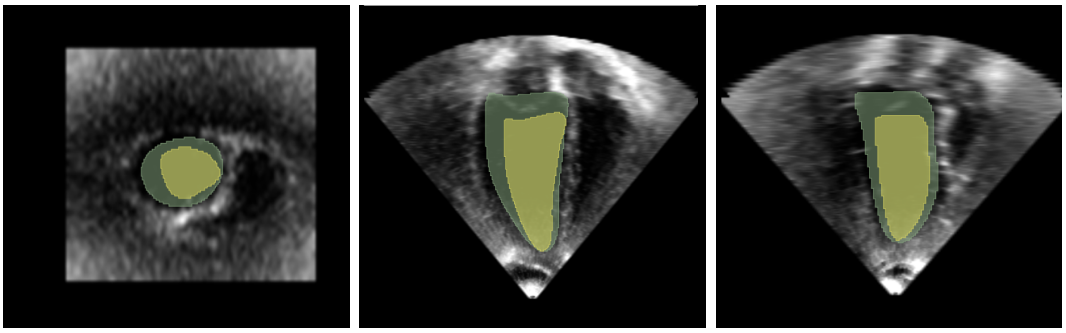
(b) Beutel (contour) revision

(c) Manual tracking revision

(d) Final analysis

Figure 3.4: The process for semi-automated delineation of the LV across the cardiac cycle using the TomTec Arena software. The annotations were exported as UCD segments after the analysis.

into NRRD sequences. Figure 3.5 shows a sample apical echocardiography image with the corresponding exported ED and ES annotations in different planes. The ED phase is rendered in green and the ES phase is rendered in yellow. To calculate the metrics, each LV annotation underwent a transformation using identical rigid and non-rigid transformations that were computed to align the respective echocardiography images. To assess the reliability of the extracted volume annotations of the left ventricular regions, we calculated the interrater agreement by comparing the annotations made by two raters on the same dataset. Interrater agreement yielded a mean Dice score value of 0.8413 across 9 apical and parasternal datasets taken from three volunteers.



(a) Axial (b) Coronal (c) Sagittal

Figure 3.5: Sample apical echocardiography image with the corresponding ED and ES annotations in different planes. ED is rendered in green and ES is rendered in yellow.

Figure 3.6: Sample 3D binary mask of delineated LV at the ED phase.

3.6 Metrics

3.6.1 Dice coefficient

The Dice coefficient is a statistical tool that measures the similarity of two sets of data. The overlap between each pair of annotations was measured using the metric. The Dice metric between a pair of 3D annotation volumes U, V was calculated as follows.

$$Dice(U, V) = \frac{2 * (U \cap V)}{U + V} \tag{3.1}$$

Since the rigid and non-rigid transforms were applied to all the 3D volumes within the sequence, the average Dice score between a pair of 4D annotation sequences

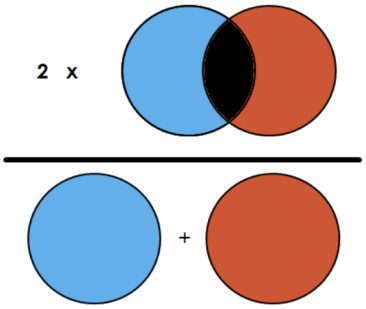


Figure 3.7: Illustration of Dice metric.

U_{seq}, V_{seq} was calculated as,

$$Dice(U_{seq}, V_{seq}) = \frac{1}{n} \sum_{i=1}^{n} Dice(U_i, V_i) \tag{3.2}$$

where n was the number of 3D volumes within a sequence. For cases where one sequence has more 3D volumes than the other, the extra 3D volumes were not included in the calculation. A high Dice coefficient value indicates a high level of similarity between the two masks, meaning that there is proper alignment. Conversely, a low Dice coefficient value indicates poor alignment. The Dice metric ranges from 0 to 1, with 1 indicating complete overlap.

3.6.2 Hausdorff distance

The Hausdorff distance (HD) [48] is used to measure the maximum distance between two objects, measuring how far two point sets are from each other. It calculates the distance of a set to the nearest point in the other set. A more general definition of Hausdorff distance between two finite point sets U and V would,

$$HD(U, V) = \max_{u \in U} \min_{v \in V} ||u - v|| \tag{3.3}$$

where $||\ ||$ is the Euclidean norm on the points U and V.

The average HD between a pair of 4D annotation sequences U_{seq}, V_{seq} was calculated as follows,

$$HD(U_{seq}, V_{seq}) = \frac{1}{n} \sum_{i=1}^{n} HD(U_i, V_i) \tag{3.4}$$

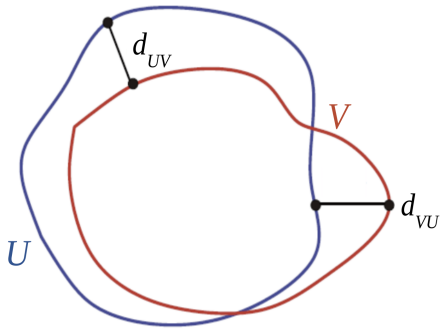


Figure 3.8: Computing the Hausdorff distance between the blue line U and the red line V.

where $HD(U_i, V_i)$ was the HD between two 3D volumes U_i, V_i and n was the number of 3D volumes within a sequence. For cases where one sequence has more 3D volumes than the other, the extra 3D volumes were not included in the calculation. A high Hausdorff distance value indicates the points are farthest from one set to another, meaning a poor alignment between LV annotations. Conversely, a low Hausdorff Distance value indicates better alignment.

Chapter 4

Results

In this section, we present the training results of the landmark detection models and evaluate the visual and quantitative accuracy of the proposed algorithm using the registered images. The proposed method is evaluated quantitatively using two metrics, namely, Dice metric (DM) and Hausdorff distance (HD in mm). Separate evaluations are done for registering images acquired at various apical windows (Apical-to-apical) and registering images acquired at apical and parasternal windows (Apical-to-parasternal). In the end, we provide a comprehensive analysis combining both results. We also compare the alignment of the registered images with the corresponding original images.

4.1 Model training and validation

In our proposed method we used three agents to search for each of the landmarks. For every agent, once they reach their final positions, we take the average as the final position for each landmark. As we have mentioned earlier, two models were trained, one for the identification of landmarks in apical 3D volumes and another for parasternal 3D volumes. Figure 4.1 shows the collab-DQN agents during testing. The agents are simultaneously looking for the anterior mitral valve leaflet attachment on the mitral valve annulus landmark (L2) in echo scans. The screenshot was taken on the agents right after they were spawned within the 80% starting zone in the environment. The

agent is denoted by a blue dot, and its field of view is shown as a yellow square surrounding it. The red dot is the landmark, with its red circle indicating its distance to the agent in the z dimension. At the bottom, if the preceding step brought the agent closer to the objective, the text is displayed in green. Conversely, if the agent moved further away, the text is shown in red. Similarly, we trained 3 agents each to identify the other two landmark points.

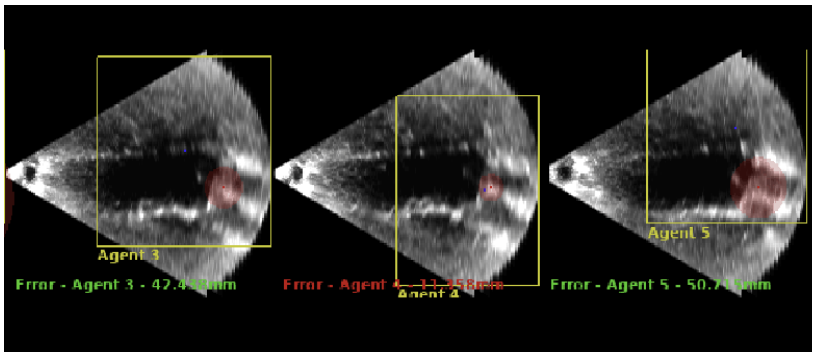


Figure 4.1: Collab-DQN agents during training. They are simultaneously looking for the anterior mitral valve leaflet attachment on the mitral valve annulus landmark (L2).

Table 4.1 shows the mean and standard deviation of the distance errors of three agents looking for the landmarks on both apical and parasternal validation datasets. A total of 23 apical and 16 parasternal volumes were used for this purpose. The second landmark (L2), for anterior mitral valve leaflet attachment on the mitral valve annulus, shows slightly better accuracy scores for both apical and parasternal models. The distance errors for L2 indicate, on average agents' position may deviate by 2.54 mm for apical and 2.71 mm for parasternal 3D volumes from the actual position.

We provide the related graphs obtained during the agent's training on the anterior mitral valve leaflet attachment on the mitral valve annulus landmark (L2). The initial 50,000 episodes of the training are shown. The remaining portion of the training process is omitted, as the agent has reached convergence and is relatively constant. In figures 4.2, 4.3, and 4.4 we can see the maximum distance error, minimum distance error, and mean distance error on the validation set over the number of episodes

Table 4.1: Mean and standard deviation of distance errors (in mm) of three agents for trained apical and parasternal models on the validation dataset.

Landmark	**Distance errors of 3 agents (mm)**	
	Apl	**Psl**
L1	3.08 ± 0.77	2.99 ± 0.76
L2	2.52 ± 0.06	2.71 ± 0.37
L3	3.36 ± 1.22	3.29 ± 0.69

respectively. The orange line corresponds to the training results on the apical images and the blue on the parasternal images. At the 50k episode mark, the average distance error is around 2.5 mm for apical and 2.99 mm for parasternal on the validation set. This is closer to the mean distance errors reported in Table 4.1 for the full 75k episodes. The mean of maximum distance error for all 9 agents looking for 3 landmarks towards the end of the training for the model that had the best performance on the validation set is around 17.7 and 9.11 for apical and parasternal respectively. This indicates that, in the worst-case situation, the agent's position may deviate by 17 mm for detecting landmarks on apical images and by 9 mm for detecting landmarks on parasternal images.

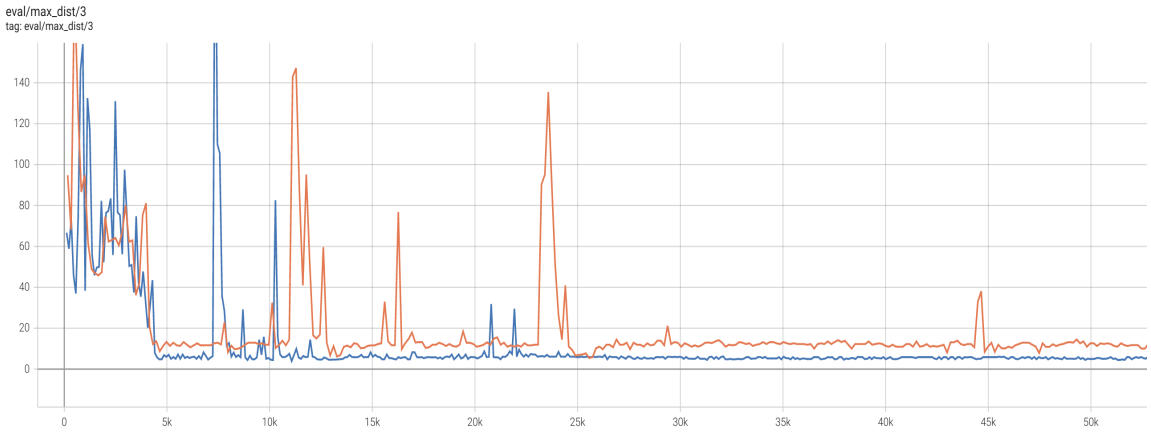


Figure 4.2: Maximum distance error on the validation set over the number of episodes. This shows worst cases are getting closer and closer to the correct landmark

In Figures 4.5 and 4.6 we can see the training loss and score. The loss ranges

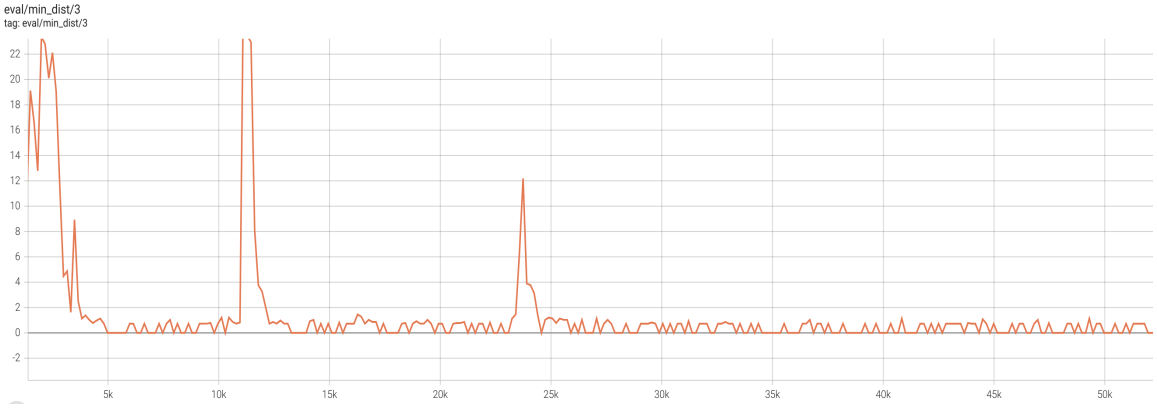


Figure 4.3: Minimum distance error on the validation set over the number of episodes.

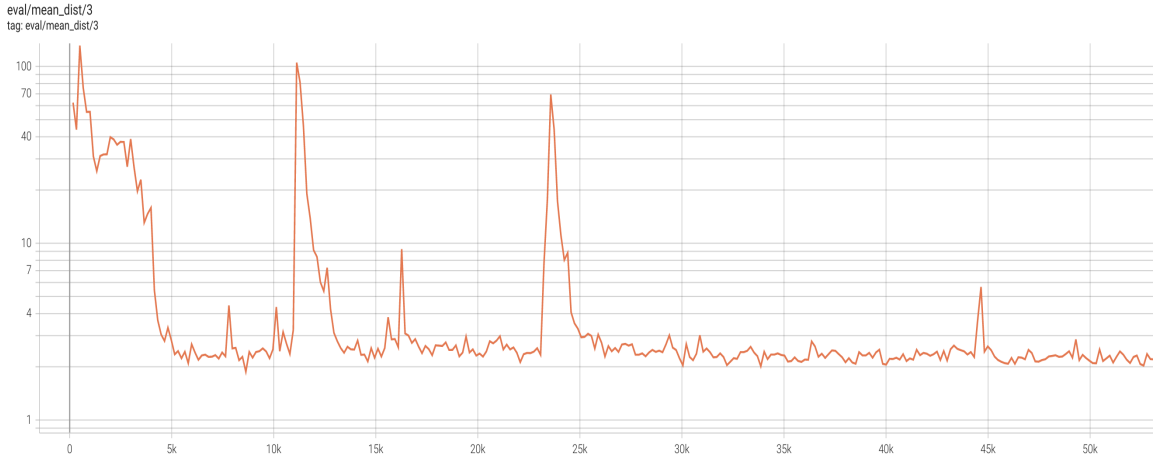


Figure 4.4: Average distance error on the validation set over the number of episodes (log-scale). Single agent trains for an epoch on the anterior mitral valve leaflet attachment on the mitral valve annulus landmark (L2) and is then evaluated on the validation set.

from 0.09 to 1.0 throughout training and is stable after convergence. The score is calculated by summing the rewards received throughout an episode. An ideal agent would possess a finite score that would vary because of its arbitrary starting position in the environment. The agent's score gradually increases during training and levels off between 50 and 100. An agent moving randomly would often receive a somewhat negative score since, on average, a random step would result in the agent moving further away from the landmark in a Euclidean distance scenario. The graphs validate that the agent is effectively learning during stable training.

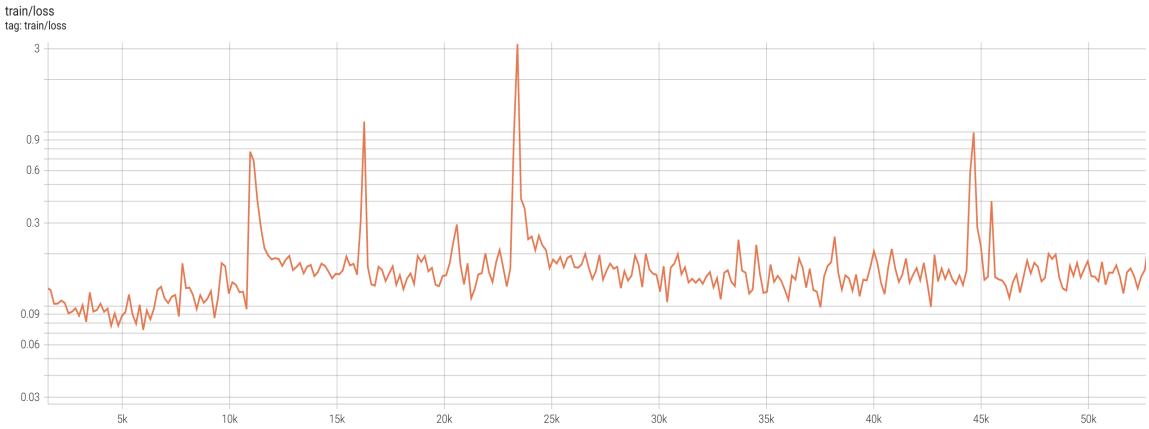


Figure 4.5: Huber loss of the single agent's DQN over the number of episodes (log-scale).

Figure 4.6: Score for each episode on the training set over the number of episodes (smoothened).

4.2 Apical-to-apical image registration

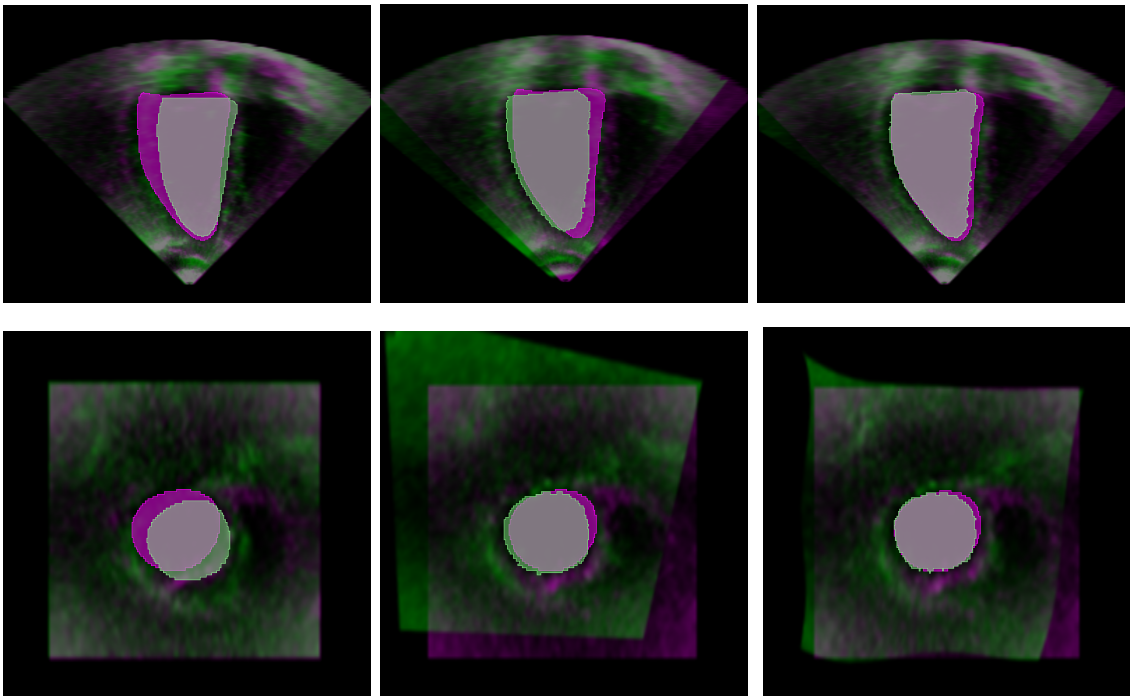
We evaluated the alignment accuracy of registering images acquired at apical windows. The results include registering two apical 4-chamber images, registering an apical 4-chamber image with an apical 3-chamber view, and registering a standard apical image with a non-standard apical image. In total, we used 33 (7 fixed, 26 moving) 4D apical echo sequences acquired from 7 volunteers for the evaluation.

4.2.1 Visual assessment

We used the 3D Slicer[1], a free open-source software used for medical image analysis and visualization, to visually evaluate the accuracy of the alignments of the registered 4D sequences and the corresponding LV annotations. Two 4D sequences, fixed and moving images, were loaded into the program, and by gradually adjusting the transparency value to switch between the foreground (moving apical) and background (fixed apical) images, an assessment for possible misalignments between the image pairs was done by the expert. Example registration results for a pair of apical slices with corresponding LV annotations are shown in Figure 4.7.

The fixed volume and the corresponding LV annotation are rendered in magenta, whereas the moving volume and annotation are rendered in green. The top row corresponds to long-axis slices, whereas the bottom row shows short-axis slices. The first column shows the original apical fixed and moving slices before any alignment. The second and third columns show rigid and non-rigid registered results, respectively. From the short axis view, one could clearly see the midlevel LV structures of the LV masks after alignment. Figure 4.8 shows original, rigid, and non-rigid registration results of an example pair of apical 3D echocardiography annotations. The first two rows show the long axis and the bottom two rows show short axis views. To assess the alignment accuracy through a cardiac cycle, slices taken at the end-diastolic state are visualized in the first and third rows, and slices taken at the end-systolic are visualized

[1]https://www.slicer.org/

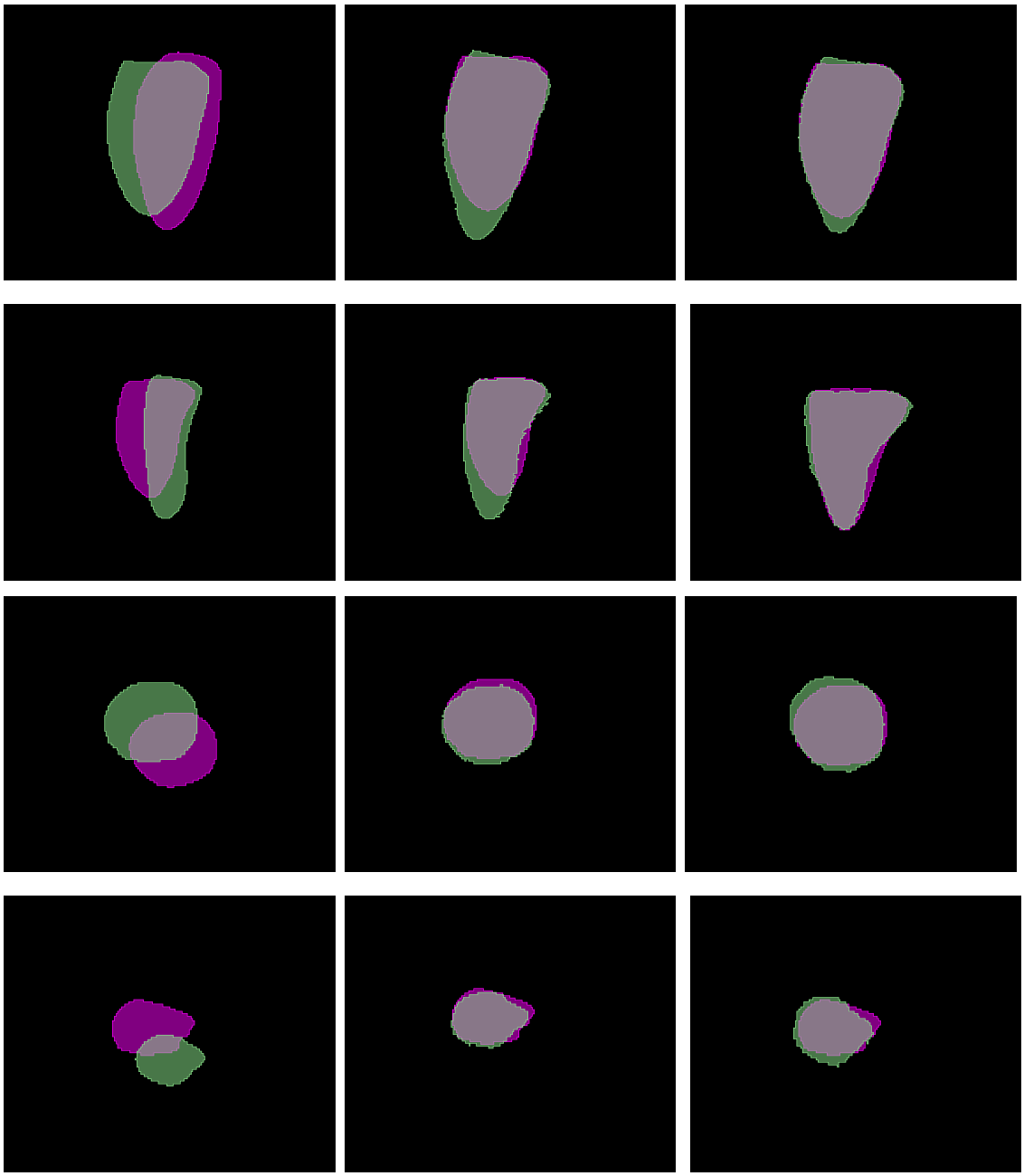


(a) Before registration (b) Rigid registration (c) Non-rigid registration

Figure 4.7: Example results for alignment of echo images with corresponding annotations. Original volumes, rigid and non-rigid registration results of an example pair of apical 3D echocardiography images in the long-axis (first row) and short-axis (bottom row) views. The fixed volume and the corresponding LV annotation are rendered in magenta, whereas the moving volume and annotation are rendered in green. The results indicate a significant alignment improvement for the apical and parasternal volumes and the corresponding LV annotations after image registration.

in the second and fourth rows. The fixed volume's LV annotation is rendered in magenta, and the moving volume's annotation is rendered in green. The results demonstrate clear alignment at end-diastolic and end-systolic states after performing the proposed two-step registration process compared to the original volumes.

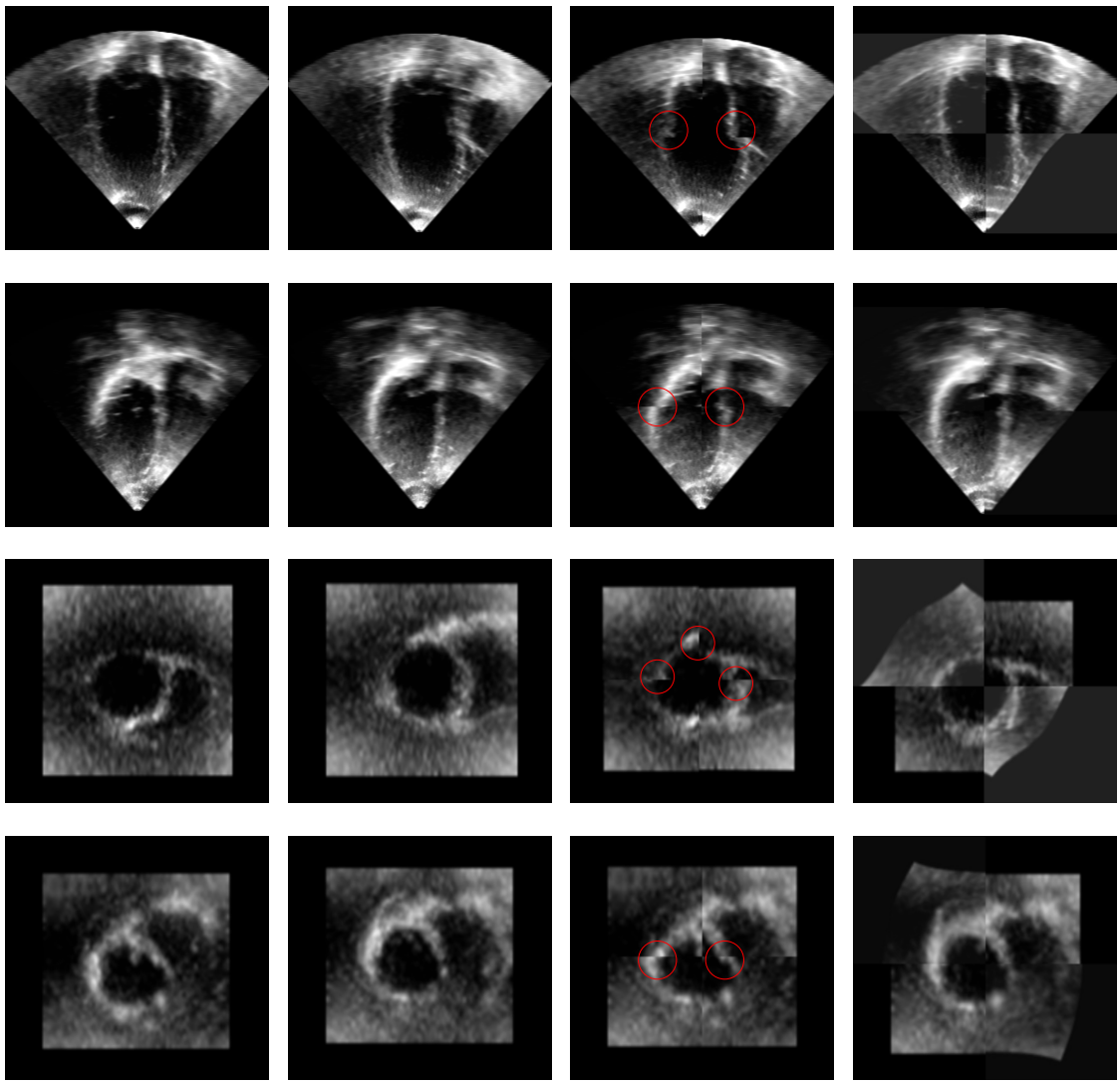
The checkerboard pattern is often used to combine two images and then compare the registration results. Sample results of registered images acquired at apical windows are shown in Figure 4.9. To assess the accuracy of the registration, the images are superimposed in a 2×2 checkerboard pattern. The first two columns show the original fixed apical and moving apical images respectively and the third column shows the superimposed checkerboard image before registration. The last column



(a) Before registration (b) Rigid registration (c) Non-rigid registration

Figure 4.8: Example results of annotations before and after registration. Original, rigid, and non-rigid registration results of an example pair of apical 3D echocardiography annotations in the long-axis (first two rows) and short-axis (bottom two rows) views. The first and third rows show end-diastolic and the second and fourth row shows end-systolic slices. The fixed volume's LV annotation is rendered in magenta, whereas the moving volume's annotation is rendered in green.

shows the checkerboard image after applying the proposed two step registration. For the checkerboard pattern, the top-left and bottom-right sections are taken from the fixed apical dataset, while the top-right and bottom-left sections correspond to the moving apical acquisition. Assessed visually, the continuity of the left ventricle walls is particularly visible in the images. Also, we could see a clear alignment of the LV from the short-axis papillary muscle view. It is clear that the algorithm succeeds in keeping the continuity of the main structures.



(a) Original apical (fixed image) (b) Original apical (moving image) (c) Before registration (d) After registration

Figure 4.9: Sample slices taken from two volunteers before and after alignment. The first two columns show the original fixed apical (a) and moving apical (b) images. The third (c) and fourth (d) columns show the two apical datasets that have been superimposed on a 2×2 checkerboard pattern before and after alignment, respectively. The upper two rows correspond to long-axis slices and the lower two show short-axis slices. Misalignments are highlighted in red circles. We can notice the continuity of the ventricular walls and better alignment of the LV after registration from the short-axis papillary muscle view.

4.2.2 Quantitative assessment

The proposed method was quantitatively evaluated by computing the Dice coefficients and the Hausdorff distance between the pair of annotations. The metrics were calculated for each 3D annotation volume within a sequence and the average values for entire annotation sequences were computed according to equations (3.2) and (3.4) explained in section 3.6, methodology. Both metrics were calculated for each sequence with rigid, non-rigid registration and before applying any transformations.

The bar plots in Figure 4.10 show the Dice score results for each pair of sequences acquired from 7 volunteers. Images taken from volunteer 1 are denoted as V1, volunteer 2 are denoted as V2, and so on for all 7 volunteers. A total of 33 apical scans were used in the evaluation. A significant improvement in the Dice coefficient was observed for images with major misalignment (V2-2, V2-3, V2-4, V3-4, V3-6). Additionally, we could also see an improvement in the Dice score for images with minor misalignments after applying the proposed two-step registration process. Overall, a significant increase in the Dice coefficient after rigid registration and further improvement after non-rigid registration were observed.

To further confirm the alignment accuracy, in Figure 4.11 we have compared the Dice score and Hausdorff distance for each 4D sequence. The scatter plots on the top figure represent the Dice score for each image and the box plot gives us a good indication of how the Dice scores were distributed before and after registration. The higher the value, the better the alignments are. We could see an improvement in Dice score after rigid registration and it is further improved after applying the non-rigid registration. The mean and standard deviation of the Dice score for the original images was 0.6490 ± 0.1846 and it was improved to 0.8222 ± 0.0581 after registration. The bottom plot shows the Hausdorff distance, which measures the maximum distance between any point in the LV annotations of the fixed to the moving image. The lower the value, the closer the LV annotations are. The distributions of the computed Haus-

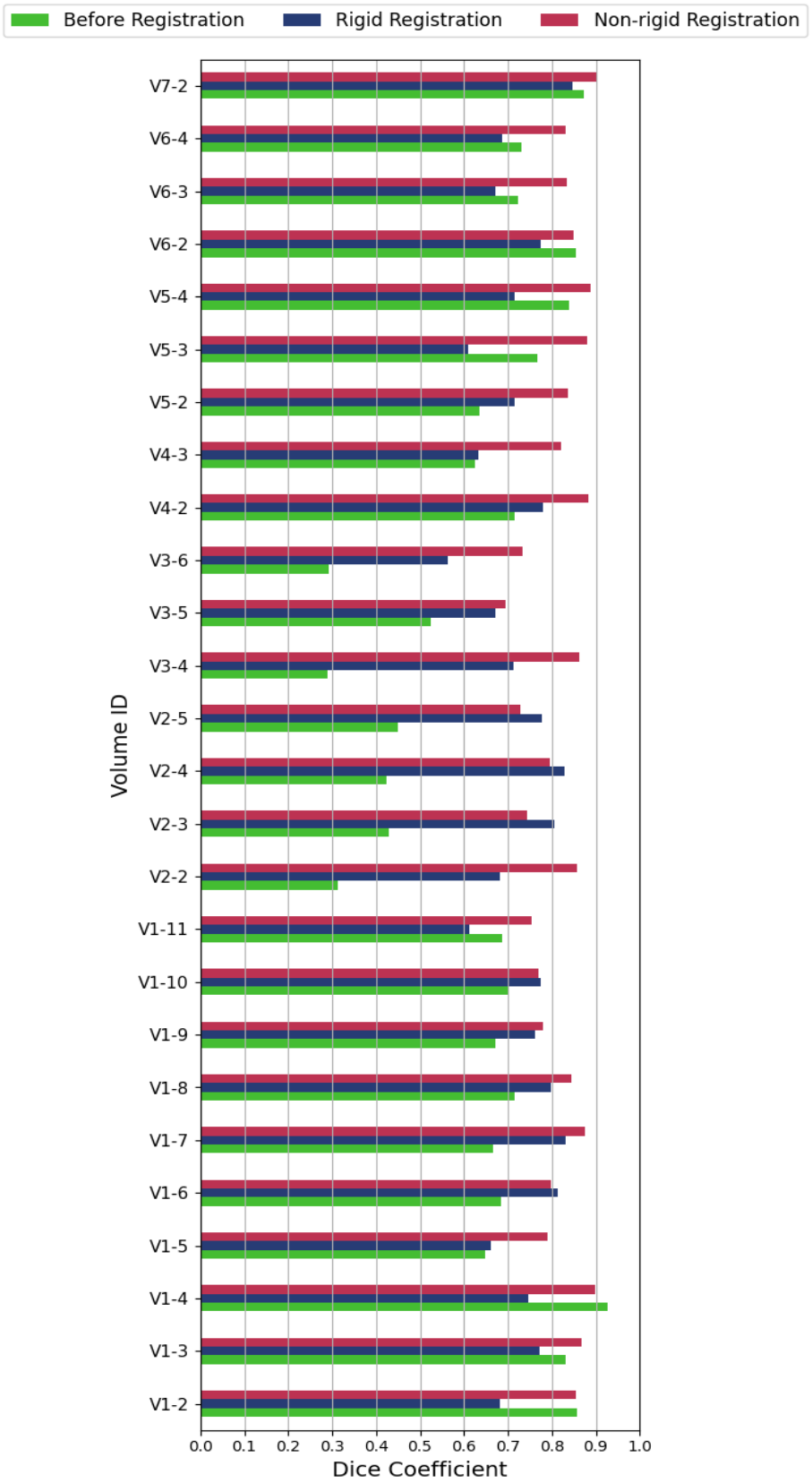


Figure 4.10: The Dice coefficient between pair of apical sequences with rigid, non-rigid registration and without any alignment for scans acquired from seven volunteers.

dorff distance are summarized using boxplots indicating a lower maximum distance between two registered annotations, a mean distance of 9.2 mm, compared to the echocardiography images without alignment where the calculated mean distance was 13.9 mm. The results indicate that the LV annotations of fixed and moving images move closer after registration.

Figure 4.12 shows the Dice coefficient and Hausdorff distance values for each 3D volume within a sequence with rigid, non-rigid registration and without registration. Each row corresponds to a pair of apical images taken from 4 different volunteers. The results show a significant improvement in the Dice coefficient values after registration at the end-diastolic and end-systolic phases. At the same time, the Hausdorff distance between the LV annotations becomes small, meaning that the annotations are getting closer to each other. The same was observed for other registered apical sequences.

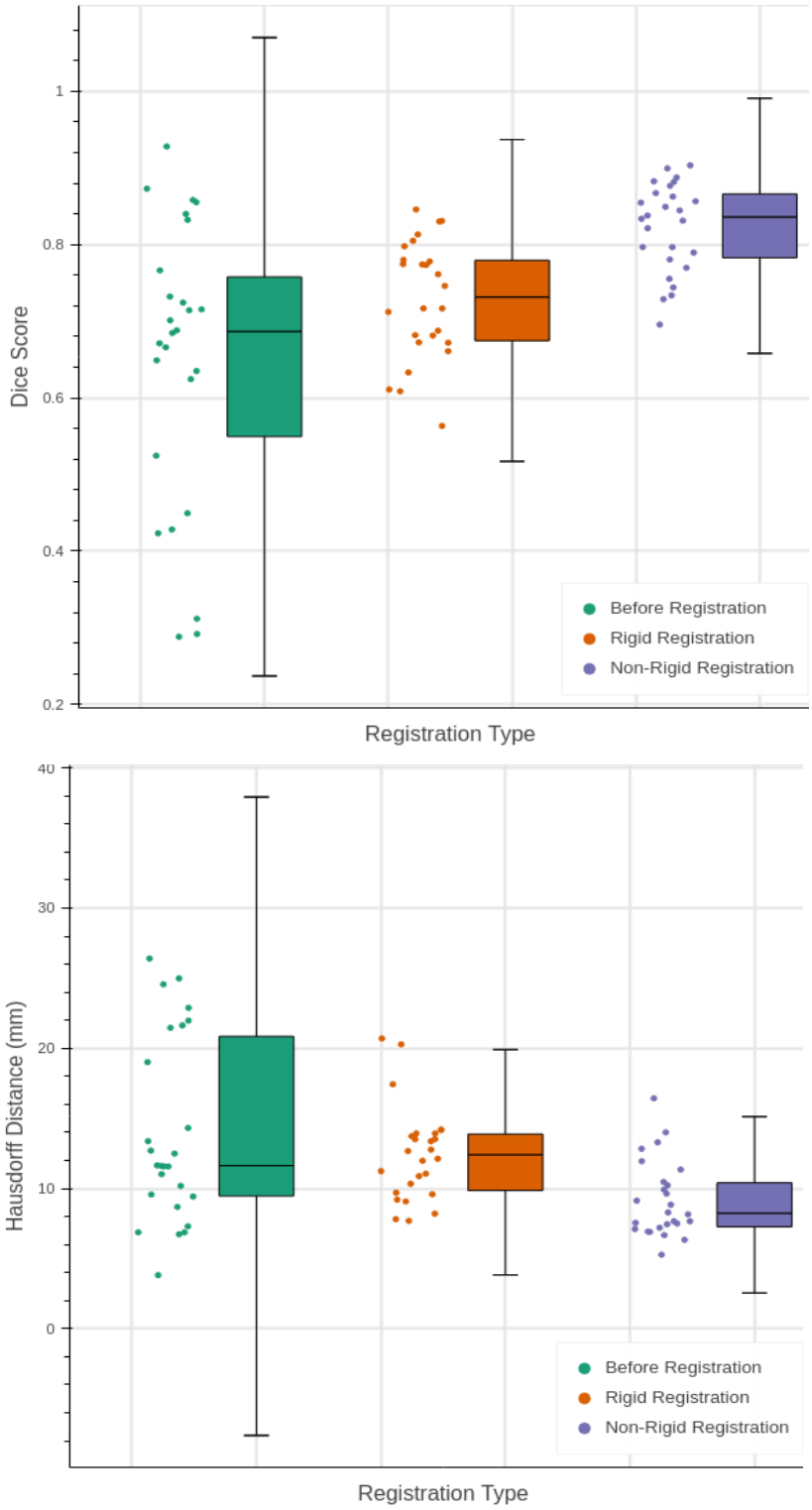


Figure 4.11: The Dice score and Hausdorff distance between pair of apical sequences with rigid and non-rigid registration and without any alignment for scans acquired from seven volunteers. The results show significant improvement in Dice score and close mutual proximity between registered echocardiography images.

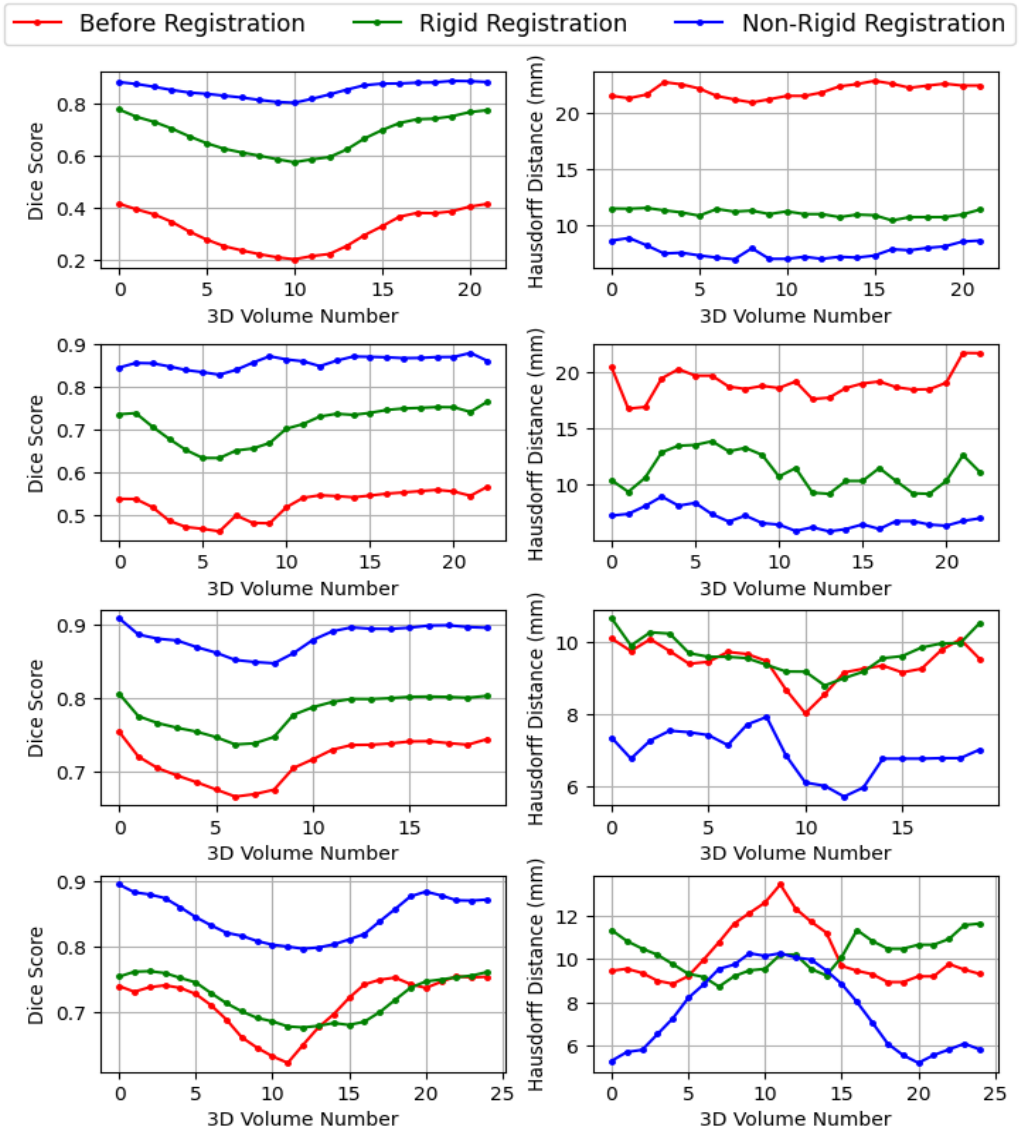


Figure 4.12: Sample Dice coefficient and Hausdorff distance values for each 3D volume within a sequence with rigid, non-rigid registration and without registration. Each row corresponds to a pair of apical sequences from four different volunteers. The results show significant improvement in the Dice coefficient and Hausdorff distance values after registration for both end-diastolic and end-systolic phases of echocardiography images.

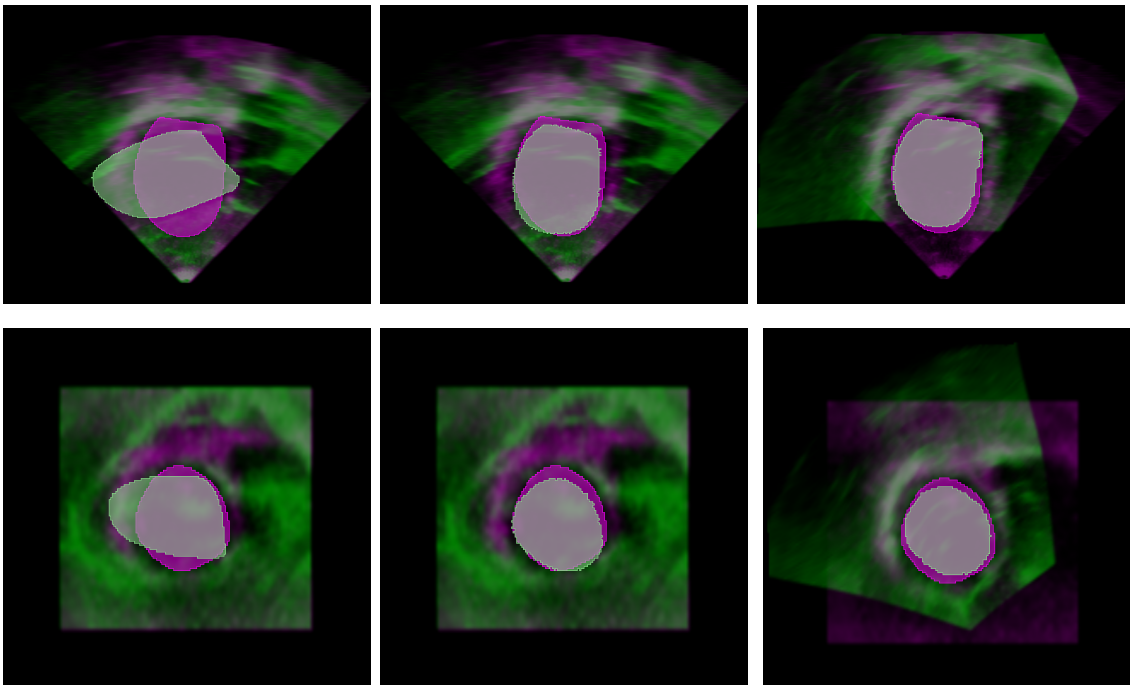
4.3 Apical-to-parasternal image registration

In this section, we evaluate the alignment accuracy of registering images acquired at apical and parasternal windows. The results include registering pairs of standard apical and standard parasternal images and registering pairs of standard apical and non-standard parasternal images. In total, we used 27 (7 fixed apical, 20 moving parasternal) 4D apical and parasternal echo sequences acquired from 7 volunteers for the evaluation.

4.3.1 Visual assessment

Similar to how we evaluated the alignment accuracy of apical scans, we used 3D Slicer to evaluate the alignment accuracy of registering apical and parasternal images. Two 4D sequences, fixed and moving images, were loaded into the program, and by gradually adjusting the transparency value to switch between the foreground (moving parasternal) and background (fixed apical) images, an assessment for possible misalignments between the image pairs was done. Example registration results for a pair of apical and parasternal slices with corresponding LV annotations are shown in Figure 4.13. The fixed volume and the corresponding LV annotation are rendered in magenta, whereas the moving volume and annotation are rendered in green. The top row corresponds to long-axis slices, whereas the bottom row shows short-axis slices. The first column shows the original apical fixed and parasternal moving slices before any alignment. The second and third columns show rigid and non-rigid registered results, respectively. From the short axis view, one could clearly see the midlevel LV structures of the LV masks after alignment. Figure 4.14 shows original, rigid, and non-rigid registration results of an example pair of apical and parasternal 3D echocardiography annotations. The first two rows show the long axis and the bottom two rows show short axis views. Slices taken at the end-diastolic state are visualized in the first and third rows. The second and fourth row shows

slices taken at the end-systolic state. The fixed volume's LV annotation is rendered in magenta, and the moving volume's annotation is rendered in green. The results demonstrate clear alignment at end-diastolic and end-systolic states after performing the proposed two-step registration process.

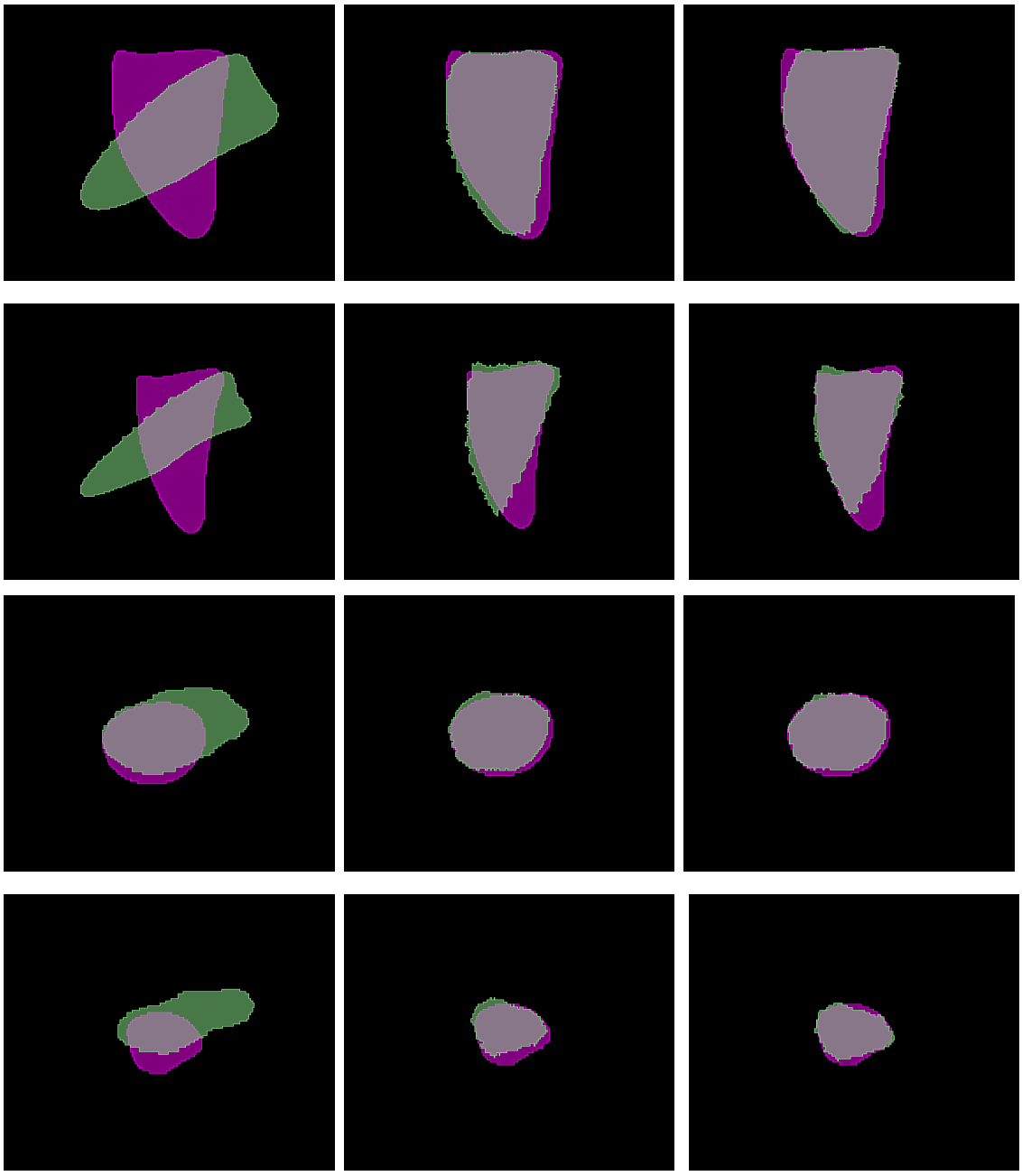


(a) Before registration (b) Rigid registration (c) Non-rigid registration

Figure 4.13: Example results for alignment of echo images with corresponding annotations. Original volumes, rigid and non-rigid registration results of an example pair of apical and parasternal 3D echocardiography images in the long-axis (first row) and short-axis (bottom row) views. The fixed volume and the corresponding LV annotation are rendered in magenta, whereas the moving volume and annotation are rendered in green. The results indicate a significant alignment improvement for the apical and parasternal volumes and the corresponding LV annotations after image registration.

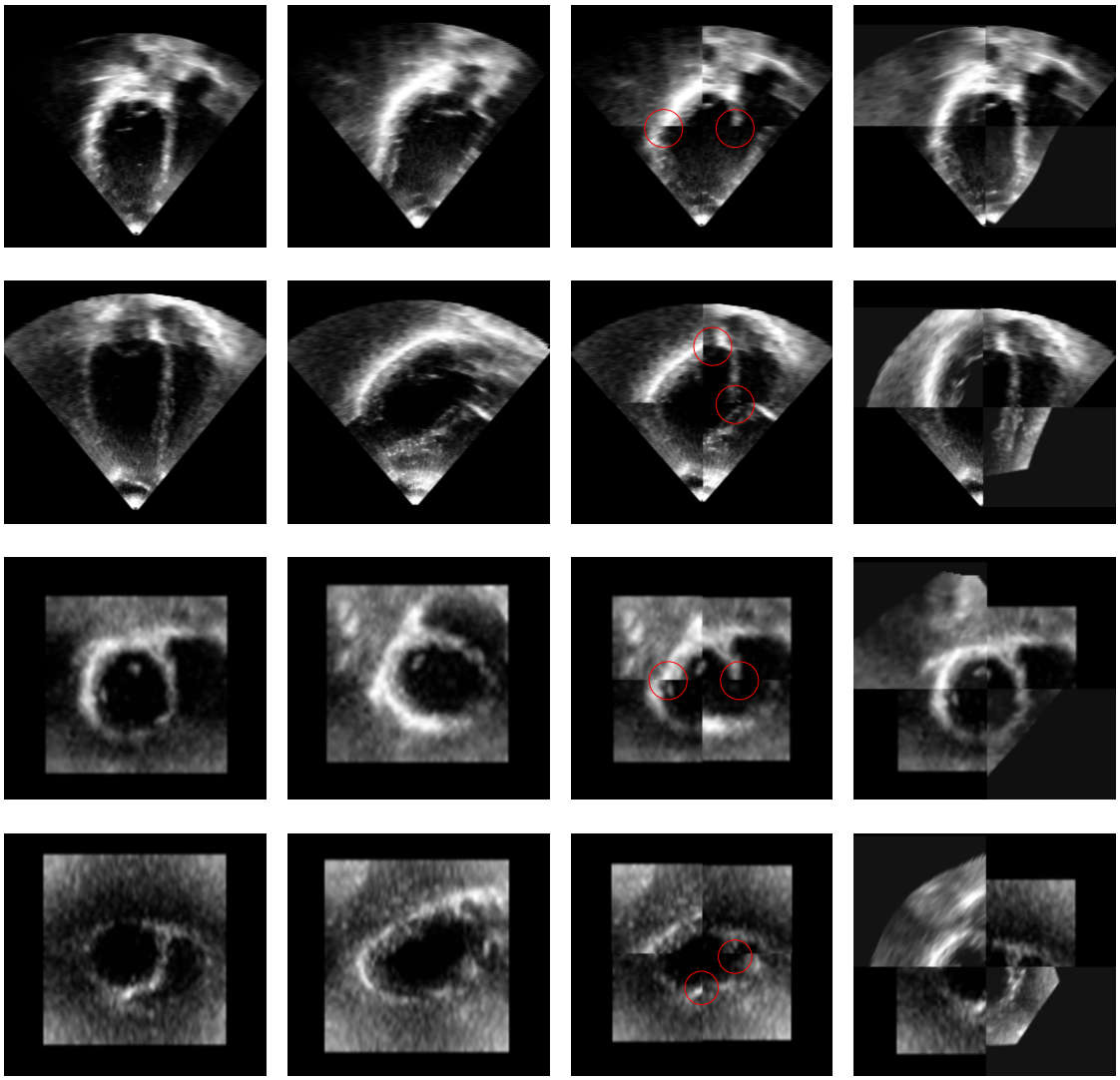
Figure 4.15 shows sample checkerboard pattern results of registered images acquired at apical and parasternal windows. The first two columns show the apical, fixed image, and parasternal, moving image respectively and the third column shows the superimposed 2×2 checkerboard image before registration. The last column shows the 2×2 checkerboard image after registration. For the checkerboard pattern,

the top-left and bottom-right sections are taken from the apical dataset, while the top-right and bottom-left sections correspond to the parasternal acquisition. Assessed visually, the continuity of the left ventricle walls is particularly visible in the images. Also, we could see a clear alignment of the LV from the short-axis papillary muscle view. It is clear that the algorithm succeeds in keeping the continuity of the main structures.



(a) Before registration (b) Rigid registration (c) Non-rigid registration

Figure 4.14: Example results of annotations before and after registration. Original, rigid and non-rigid registration results of an example pair of apical and parasternal 3D echocardiography annotations in the long-axis (first two rows) and short-axis (bottom two rows) views. The fixed volume and the corresponding LV annotation are rendered in magenta, whereas the moving volume and annotation are rendered in green.



(a) Original apical (fixed image) (b) Original parasternal (moving image) (c) Before registration (d) After registration

Figure 4.15: Sample slices taken from two volunteers before and after alignment. The first two columns show the original apical (a) and parasternal (b) images. The third (c) and fourth (d) columns show apical and parasternal datasets that have been superimposed on a 2×2 checkerboard pattern before and after alignment, respectively. The upper two rows correspond to long-axis slices and the lower two show short-axis slices. Misalignments are highlighted in red circles. We can notice the continuity of the ventricular walls and better alignment of the LV after registration from the short-axis papillary muscle view.

4.3.2 Quantitative assessment

Similar to apical sequences the metrics were calculated for each 3D annotation volume within a sequence and the average values for entire annotation sequences were computed according to equations (3.2) and (3.4) for each sequence with rigid, non-rigid registration and before applying any transformations.

The bar plots in Figure 4.16 show the Dice score results for each pair of sequences acquired from 7 volunteers. A total of 27 apical scans (7 apical, 26 parasternal) were used in the evaluation. Since apical and parasternal images are acquired at different windows, compared to apical sequences, we could see a lower Dice score for original images without any alignment. Once registered we could see significant improvements in the Dice coefficients for registering a pair of apical and parasternal images. Also, we could also see an improvement in the Dice score after applying non-rigid registration. Overall, a significant increase in the Dice coefficient after rigid registration and further improvement after non-rigid registration were observed.

Figure 4.17 shows the Dice score and Hausdorff distance for each 4D sequence. We could see an improvement in the Dice score after rigid registration which is further improved after applying non-rigid registration. The mean and standard deviation of the Dice score for the original images was 0.4706 ± 0.0977 and it was improved to 0.7616 ± 0.0713 after registration. Computed Hausdorff distances summarized using boxplots for rigid and non-rigid registration indicate a lower maximum distance between two registered annotations compared to the echocardiography images without alignment. Since the images were acquired at two different windows, apical and parasternal, the mean Hausdorff distance of the original images was 32.4 mm, higher than images acquired only at apical windows. After registration, this value was brought down to 12.9 mm. The results indicate that the LV annotations of fixed and moving images move closer after registration compared to the initial scans. Figure 4.18 shows the Dice coefficient and Hausdorff distance values for each 3D volume

within a sequence with rigid, non-rigid registration and without registration. Each row corresponds to a pair of apical and parasternal images taken from 4 different volunteers. The results show a significant improvement in the Dice coefficient values after registration at the end-diastolic and end-systolic phases. At the same time, the Hausdorff distance between the LV annotations becomes small, meaning that the annotations are getting closer to each other.

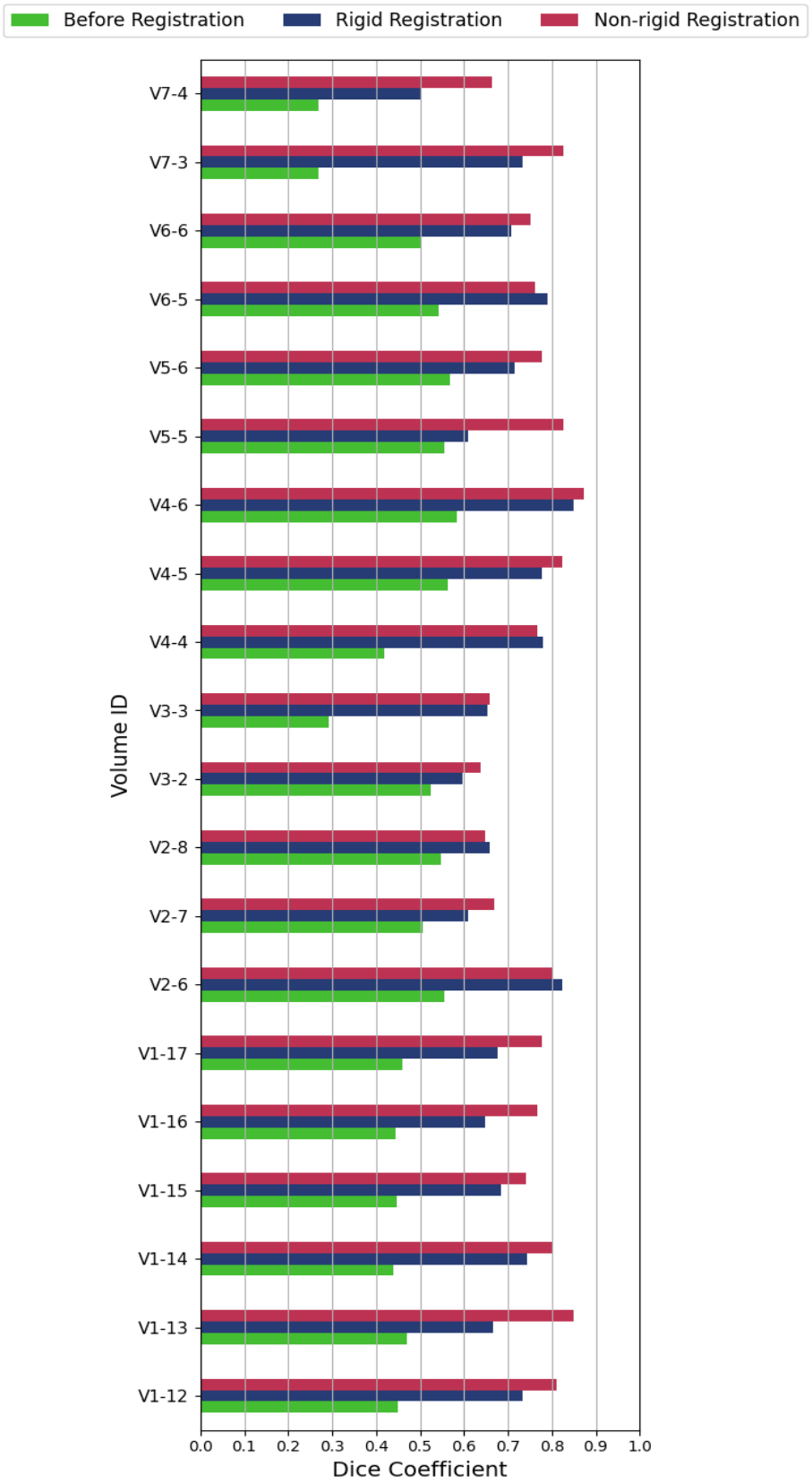


Figure 4.16: The Dice coefficient between pair of apical and parasternal sequences with rigid, non-rigid registration and without any alignment for scans acquired from seven volunteers.

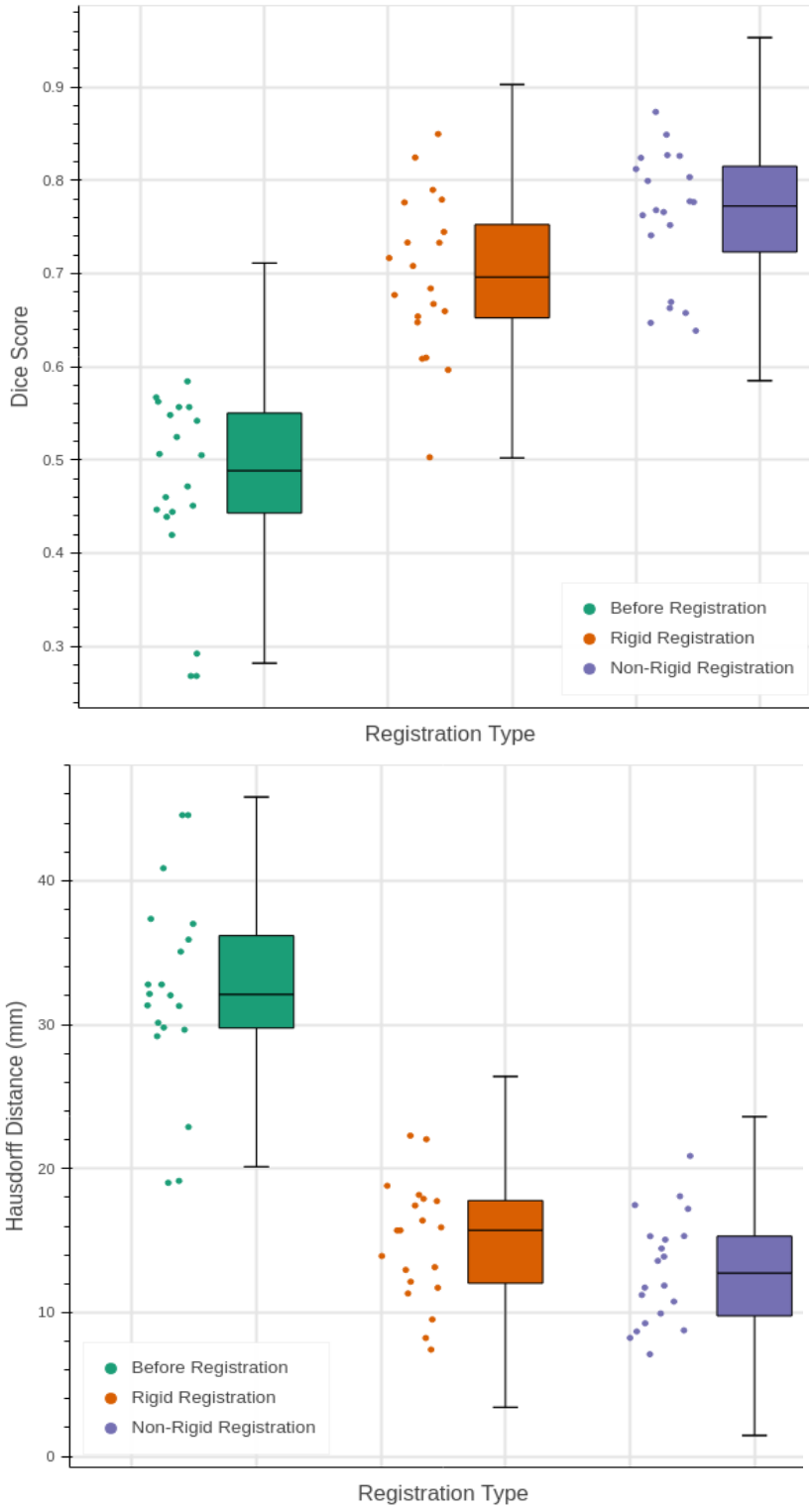


Figure 4.17: The Dice score and Hausdorff distance between pair of apical and parasternal sequences with rigid and non-rigid registration and without any alignment for scans acquired from seven volunteers. The results show significant improvement in Dice score and close mutual proximity between registered echocardiography images.

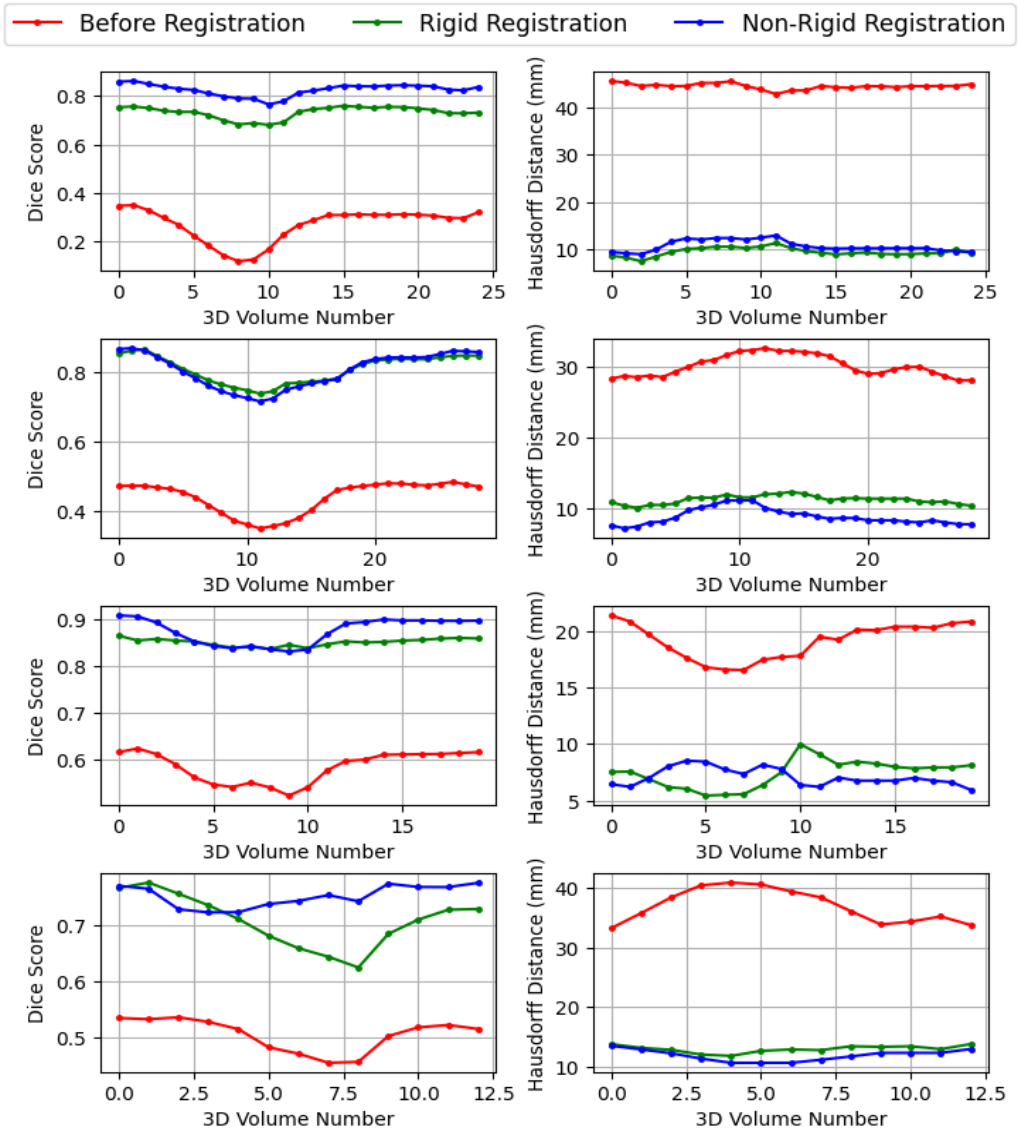


Figure 4.18: Sample Dice coefficient and Hausdorff distance values for each 3D volume within a sequence with rigid, non-rigid registration and without registration. Each row corresponds to a pair of apical and parasternal sequences from four different volunteers. The results show significant improvement in the Dice coefficient and Hausdorff distance values after registration for both end-diastolic and end-systolic phases of echocardiography images.

4.4 Combined results

In this section, we provide comprehensive results of registering images acquired at various apical and parasternal windows. A total of 53 (7 fixed, 46 moving) 4D sequences from 7 volunteers comprising both apical and parasternal windows were used to compute the values.

4.4.1 Visual assessment

The alignment accuracy of registering images acquired at apical and parasternal windows was visually evaluated using the 3D Slicer software. The fixed, moving images and the corresponding LV annotations were used for this purpose. In addition to that the fixed and moving images were superimposed in a 2×2 checkerboard pattern before and after registration in order to appreciate the quality of the alignment. Assessed visually, the results of registered images acquired at both apical and parasternal windows show significant alignment compared to the original scans. From the registered images presented in sections 4.2.1 and 4.3.1, the continuity of the left ventricle walls is particularly visible and a clear alignment of the LV can be seen from the short-axis papillary muscle view. This shows that the algorithm succeeds in aligning the images while keeping the continuity of the structures.

4.4.2 Quantitative assessment

The mean and standard deviation of the Dice coefficient values are given in Table 4.2. The first row contains the results for the registration of two images acquired from apical windows (Apl-to-Apl). The second row shows the results for the registration of the apical with the parasternal (Apl-to-Psl) window images. Since these images are acquired at two different windows, we could see a much lower Dice score (compared to apical images) for images without any alignment. The last row shows the combined results of apical image registration and apical to parasternal image registration. The overall results yielded an average Dice score of **0.7959**, after applying the two-step

registration process which indicates that there was an overall significant improvement in the alignment after registration.

Table 4.2: The mean and standard deviation of Dice score values measuring the overlap of entire 4D left ventricular annotations. A total of 53 4D sequences were used to compute the values.

Image View	**Dice Coefficient**		
	Before Registration	**Rigid Registration**	**Non-Rigid Registration**
Apl-to-Apl	0.6490 ± 0.1846	0.7279 ± 0.0767	0.8222 ± 0.0581
Apl-to-Psl	0.4706 ± 0.0977	0.6981 ± 0.0849	0.7616 ± 0.0713
Combined	0.5714 ± 0.1760	0.7149 ± 0.0809	0.7959 ± 0.0703

The mean and standard deviation of the Hausdorff distance values are given in Table 4.3. After applying the proposed registration method, the average distance decreased from 21.9 mm to 10.8 mm for registered images comprising both apical and parasternal windows. The overall results show a high Dice score and low Hausdorff distance for the registered images, indicating a better alignment compared to the original images.

Table 4.3: The mean and standard deviation of Hausdorff distance values measuring the overlap of entire 4D left ventricular annotations. A total of 53 4D sequences were used to compute the values.

Image View	**Hausdorff Distance (mm)**		
	Before Registration	**Rigid Registration**	**Non-Rigid Registration**
Apl-to-Apl	13.9 ± 6.6	12.4 ± 3.3	9.2 ± 2.7
Apl-to-Psl	32.4 ± 6.9	14.9 ± 4.2	12.9 ± 3.8
Combined	21.9 ± 11.4	13.5 ± 3.9	10.8 ± 3.7

For the same data, Figure 4.19 shows the Dice score and Hausdorff distance for each

4D sequence. In the top plot, the scatter plots represent the Dice score and the bottom plot represents the Hausdorff distance for each sequence. The box plot gives us a good indication of how the Dice scores and Hausdorff distances were distributed before and after registration. The distributions of the computed Hausdorff distance summarized using boxplots indicate a lower maximum distance between two registered annotations compared to the echocardiography images without alignment. The results indicate that the LV annotations of fixed and moving images move closer after registration.

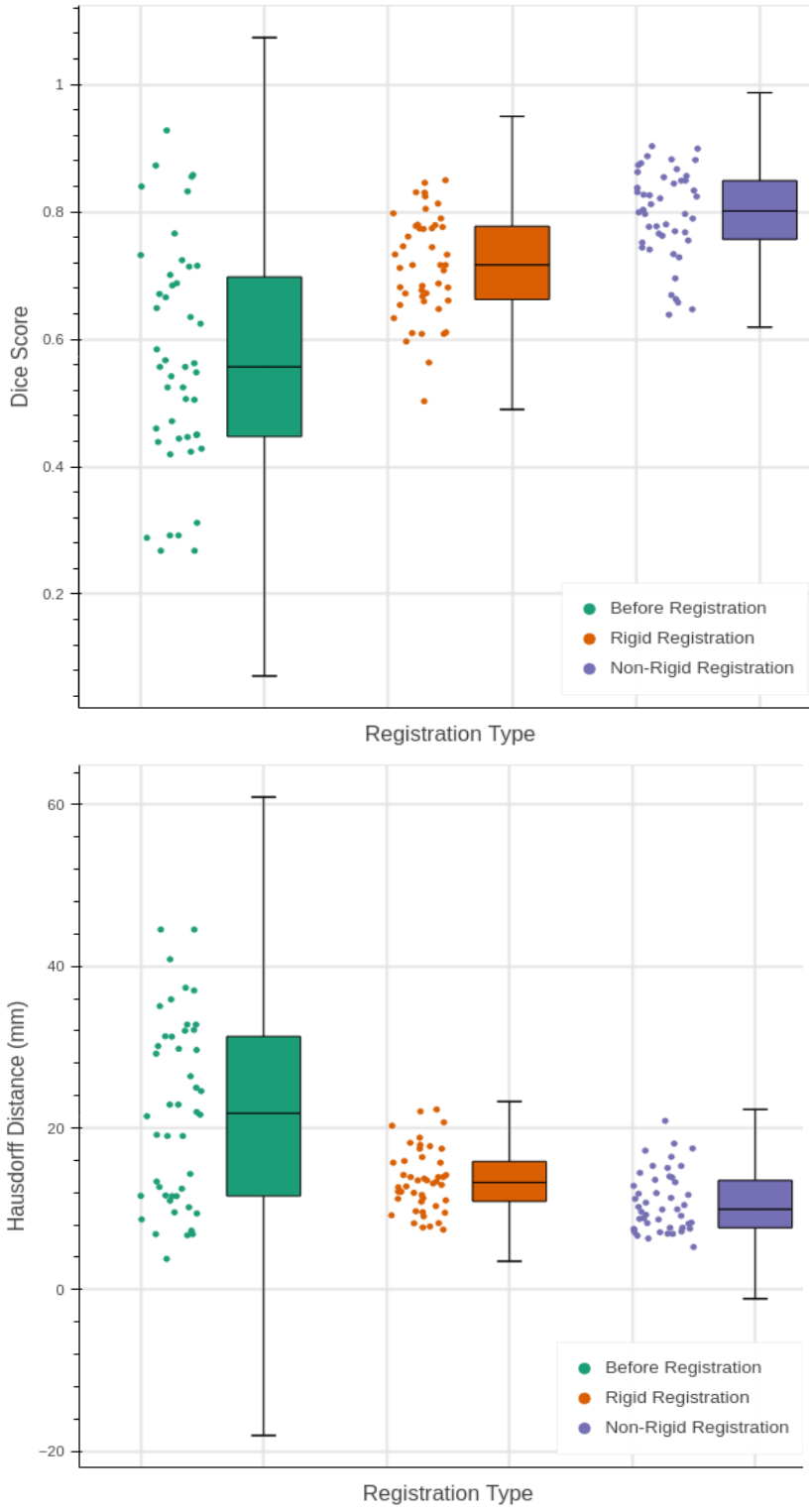


Figure 4.19: The Dice score and Hausdorff distance between pair of sequences with rigid and non-rigid registration and without any alignment for scans acquired from seven volunteers. The results show significant improvement in Dice score and close mutual proximity between registered echocardiography images.

4.4.3 Comparison with ANTs library

Advanced Normalization Tools (ANTs)[2] is a medical image registration and segmentation toolkit and is available as a Python library. ANTs registration was done using the same multiresolution framework settings as of SimpleElastix's; two consecutive phases for the apical images and a single phase for the parasternal images were used along with Mattes mutual information as the similarity metric. Tables 4.4 and 4.5 provide a comparison of the mean and the standard deviation of the Dice score and Hausdorff distance values between SimpleElastix and ANTs measuring the overlap of entire 4D left ventricular annotations for a total of 30 sequences (20 apical, 10 parasternal) acquired from 7 volunteers. We got better results for both Dice score and Hausdorff distance with Elastix library for registering images at both Apl-to-Apl and Apl-to-Psl windows.

Table 4.4: A comparison of the mean and standard deviation of the Dice score values between SimpleElastix and ANTs measuring the overlap of entire 4D left ventricular annotations.

Image View	**Dice Coefficient**		
	Rigid Registration	**Elastix Registration**	**ANTs Registration**
Apl-to-Apl	0.7343 ± 0.0733	**0.8453 ± 0.0412**	0.7192 ± 0.1526
Apl-to-Psl	0.7224 ± 0.0854	**0.7844 ± 0.0738**	0.5239 ± 0.2157
Combined	0.7303 ± 0.0763	**0.8250 ± 0.0605**	0.6541 ± 0.1961

[2]https://github.com/ANTsX/ANTs

Table 4.5: A comparison of the mean and standard deviation of the Hausdorff distance values between SimpleElastix and ANTs measuring the overlap of entire 4D left ventricular annotations.

Image View	**Hausdorff Distance (mm)**		
	Rigid Registration	**Elastix Registration**	**ANTs Registration**
Apl-to-Apl	10.99 ± 5.89	**8.33 ± 1.99**	12.52 ± 4.76
Apl-to-Psl	14.66 ± 4.11	**12.60 ± 3.56**	19.74 ± 5.89
Combined	12.21 ± 5.58	**9.75 ± 3.28**	14.93 ± 6.13

4.5 Conclusion

We proposed an algorithm using point-based rigid registration followed by B-spline non-rigid registration to register multiview 3D echocardiography sequences. The proposed method has yielded promising results in registering echocardiography images taken in different sonographic windows. Experiments were conducted to register pairs of apical images and pairs of apical and parasternal images. The accuracy of the registration was visually and quantitatively assessed using the Dice similarity coefficient and the Hausdorff distance measure between pairs of 3D and 4D scans by delineating the left ventricular region. The proposed approach was tested on 53 4D echocardiography sequences acquired from seven participants. The results yielded an average Dice score of 0.7959 and a Hausdorff distance of 10.8 mm for the registered echocardiography sequences. The high Dice score and lower Hausdorff distance for registered images show that the accuracy of alignment improves significantly after registration compared to the original scans, which could be useful in the fusion of echo images leading to a significant increase in image quality and structure definition.

4.5.1 Acknowledgement

The authors thank Alberta Innovates for the AICE Concepts funding and the Natural Sciences and Engineering Research Council of Canada (NSERC) for an RTI grant that supported this research work.

Chapter 5

Conclusion

5.0.1 Summary

Medical image registration techniques serve as the fundamental basis for procedures such as image-guided radiation therapy, image-guided radiation surgery, and computer-assisted diagnosis. Though medical images are acquired using multiple modalities, ultrasound is becoming one of the widely used modalities, because of its lack of ionization, portable, and inexpensive over other modalities. However, ultrasound images tend to be notoriously noisy and subject to unique artifacts. Registering real-time 3D echocardiography images acquired at different windows has the potential to improve image quality, is useful to image the entire heart, and also allows for better 3D imaging by extracting spatial features along with temporal information, thus improving clinical decision making.

The main goal of this thesis was to propose an algorithm to automatically register echocardiography images acquired at different sonography windows. The contributions of this study are (1) point-based rigid registration followed by a nonrigid registration to align images acquired from standard and nonstandard apical (Apl) windows, (2) point-based rigid registration followed by a nonrigid registration to align images acquired from apical and parasternal (Psl) windows, and 3) visual and quantitative evaluation of the accuracy of registration by measuring the overlap between the left ventricle regions. A total of 53 4D echocardiography sequences acquired from 7 vol-

unteers were used for this purpose. The overall results indicate that the proposed registration approach improves the alignment of the images compared to the original scans, which could be useful in the fusion of echo images.

5.0.2 Limitations and future works

However, it is essential to note that this research is limited because the images have been taken from volunteers without any heart abnormalities and the number of 3D volumes that have been used to train the landmark detection model is relatively small compared to real-world applications. Further research should focus on increasing the diversity of the dataset and incorporating more diverse volunteers to improve the model's generalizability. Future works could be extended to validate the proposed technique on a larger and more diverse dataset. This involves patient data collected from various ultrasound scanners and institutions. Testing this approach on individuals with major anomalies in the left ventricle, such as those with a history of cardiac attack, cardiomyopathy, and congenital disorders, would be intriguing.

Bibliography

[1] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, 325–376, 1992, ISSN: 0360-0300. DOI: 10.1145/146370.146374. [Online]. Available: https://doi.org/10.1145/146370.146374.

[2] P. M. Patel and V. M. Shah, "Image registration techniques: A comprehensive survey," *International journal of innovative research and development*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:61445208.

[3] C. Mitchell *et al.*, "Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: Recommendations from the american society of echocardiography," *Journal of the American Society of Echocardiography*, vol. 32, no. 1, pp. 1–64, 2019.

[4] *Chambers of the Heart*, en. [Online]. Available: https://my.clevelandclinic.org/health/body/23074-heart-chambers (visited on 01/21/2024).

[5] *Electrocardiogram (EKG) Components and Intervals*, en-US. [Online]. Available: https://myhealth.alberta.ca:443/Health/Pages/conditions.aspx?hwid=zm2308 (visited on 01/25/2024).

[6] Cardioserv, *Understanding the Basics: Physiology of Diastole*, en-US, Nov. 2017. [Online]. Available: https://www.cardioserv.net/echo-physiology-diastole/ (visited on 01/24/2024).

[7] F. Dezaki *et al.*, "Cardiac phase detection in echocardiograms with densely gated recurrent neural networks and global extrema loss," *IEEE Transactions on Medical Imaging*, vol. PP, pp. 1–1, Dec. 2018. DOI: 10.1109/TMI.2018.2888807.

[8] C. Che, T. S. Mathai, and J. Galeotti, "Ultrasound registration: A review," *Methods*, vol. 115, pp. 128–143, 2017, Image Processing for Biologists, ISSN: 1046-2023. DOI: https://doi.org/10.1016/j.ymeth.2016.12.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1046202316304789.

[9] V. Zagrodsky, R. Shekhar, and J. F. Cornhill, "Mutual information-based registration of cardiac ultrasound volumes," in *Medical Imaging 2000: Image Processing*, SPIE, vol. 3979, 2000, pp. 1605–1614.

[10] W. Crum, T. Hartkens, and D. Hill, "Non-rigid image registration: Theory and practice," *The British journal of radiology*, vol. 77 Spec No 2, S140–53, Feb. 2004. DOI: 10.1259/bjr/25329214.

[11] P. Viola and W. Wells, "Alignment by maximization of mutual information," vol. 24, Jan. 1995, pp. 16–23, ISBN: 0-8186-7042-8. DOI: 10.1109/ICCV.1995.466930.

[12] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "Nonrigid multimodality image registration," in *SPIE Medical Imaging*, 2001. [Online]. Available: https://api.semanticscholar.org/CorpusID:44108676.

[13] M. Holden, "A review of geometric transformations for nonrigid body registration," *IEEE Transactions on Medical Imaging*, vol. 27, no. 1, pp. 111–128, 2008. DOI: 10.1109/TMI.2007.904691.

[14] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid registration using free-form deformations: Application to breast mr images," *Medical Imaging, IEEE Transactions on*, vol. 18, pp. 712 –721, Sep. 1999. DOI: 10.1109/42.796284.

[15] T. Selbekk *et al.*, "Ultrasound imaging in neurosurgery: Approaches to minimize surgically induced image artefacts for improved resection control," *Acta Neurochirurgica*, vol. 155, pp. 973 –980, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:17540394.

[16] D. V. Sahani *et al.*, "Intraoperative us in patients undergoing surgery for liver neoplasms: Comparison with mr imaging," *Radiology*, vol. 232, no. 3, pp. 810–814, 2004, PMID: 15273336. DOI: 10.1148/radiol.2323030896. eprint: https://doi.org/10.1148/radiol.2323030896. [Online]. Available: https://doi.org/10.1148/radiol.2323030896.

[17] K. L. Hansen *et al.*, "Intraoperative vector flow imaging of the heart," in *2013 IEEE International Ultrasonics Symposium (IUS)*, 2013, pp. 1745–1748. DOI: 10.1109/ULTSYM.2013.0445.

[18] M. Letteboer, P. Willems, M. Viergever, and W. Niessen, "Non-rigid registration of 3d ultrasound images of brain tumours acquired during neurosurgery," Nov. 2003, pp. 408–415, ISBN: 978-3-540-20464-0. DOI: 10.1007/978-3-540-39903-2_50.

[19] F. Cen, Y. Jiang, Z. Zhang, H.-T. Tsui, T. Lau, and H. Xie, "Robust registration of 3-d ultrasound images based on gabor filter and mean-shift method," vol. 3117, Jan. 2004, pp. 304–316, ISBN: 978-3-540-22675-8. DOI: 10.1007/978-3-540-27816-0_26.

[20] T. Klein, M. Hansson, A. Karamalis, and N. Navab, "Registration of rf ultrasound data using hybrid local binary patterns," *Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro. IEEE International Symposium on Biomedical Imaging*, pp. 1072–1075, May 2012. DOI: 10.1109/ISBI.2012.6235744.

[21] G. Xiao, M. Brady, J. Noble, M. Burcher, and R. English, "Nonrigid registration of 3-d free-hand ultrasound images of the breast," *IEEE transactions on medical imaging*, vol. 21, May 2002. DOI: 10.1109/TMI.2002.1000264.

[22] C. R. Meyer *et al.*, "Semiautomatic registration of volumetric ultrasound scans," *Ultrasound in medicine & biology*, vol. 25, no. 3, pp. 339–347, 1999.

[23] S. Ramamoorthy, K. R., and S. Rajaram, "Registration of ultrasound liver images using mutual information technique," *Advances in Intelligent Systems and Computing*, vol. 246, pp. 147–153, Nov. 2014. DOI: 10.1007/978-81-322-1680-3_17.

[24] J. Banerjee, C. Klink, E. D. Peters, W. J. Niessen, A. Moelker, and T. van Walsum, "4d liver ultrasound registration," in *Biomedical Image Registration: 6th International Workshop, WBIR 2014, London, UK, July 7-8, 2014. Proceedings 6*, Springer, 2014, pp. 194–202.

[25] O. K. Øye, W. Wein, D. M. Ulvang, K. Matre, and I. Viola, "Real time image-based tracking of 4d ultrasound data," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I 15*, Springer, 2012, pp. 447–454.

[26] T. Mäkelä *et al.*, "A review of cardiac image registration methods," *IEEE transactions on medical imaging*, vol. 21, pp. 1011–21, Oct. 2002. DOI: 10.1109/TMI.2002.804441.

[27] H. Veene *et al.*, "Automatic mitral annulus tracking in volumetric ultrasound using non-rigid image registration," vol. 2015, Aug. 2015, pp. 1985–1988. DOI: 10.1109/EMBC.2015.7318774.

[28] W. Zhang, J. Noble, and J. Brady, "Real time 3-d ultrasound to mr cardiovascular image registration using a phase-based approach," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, 2006, pp. 666–669. DOI: 10.1109/ISBI.2006.1625004.

[29] O. Velasco *et al.*, "Real-time three-dimensional echocardiography: Characterization of cardiac anatomy and function—current clinical applications and literature review update," *BioResearch Open Access*, vol. 6, no. 1, pp. 15–18, 2017, PMID: 28303211. DOI: 10.1089/biores.2016.0033. eprint: https://doi.org/10.1089/biores.2016.0033. [Online]. Available: https://doi.org/10.1089/biores.2016.0033.

[30] L. Macron *et al.*, "Single-beat versus multibeat real-time 3d echocardiography for assessing left ventricular volumes and ejection fraction," *Circulation: Cardiovascular Imaging*, vol. 3, no. 4, pp. 450–455, 2010. DOI: 10.1161/CIRCIMAGING.109.925966. eprint: https://www.ahajournals.org/doi/pdf/10.1161/CIRCIMAGING.109.925966. [Online]. Available: https://www.ahajournals.org/doi/abs/10.1161/CIRCIMAGING.109.925966.

[31] K. Rajpoot, V. Grau, J. Alison Noble, H. Becher, and C. Szmigielski, "The evaluation of single-view and multi-view fusion 3d echocardiography using image-driven segmentation and tracking," *Medical Image Analysis*, vol. 15, no. 4, pp. 514–528, 2011, Special section on IPMI 2009, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2011.02.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841511000296.

[32] T. Lamb *et al.*, "Multi-view 3-d fusion echocardiography: Enhancing clinical feasibility with a novel processing technique," *Ultrasound in Medicine Biology*, vol. 47, no. 11, pp. 3090–3100, 2021, ISSN: 0301-5629. DOI: https://doi.org/10.1016/j.ultrasmedbio.2021.06.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301562921002891.

[33] V. Grau and J. A. Noble, "Adaptive multiscale ultrasound compounding using phase information," in *International conference on medical image computing and computer-assisted intervention*, Springer, 2005, pp. 589–596.

[34] V. Grau, H. Becher, and J. A. Noble, "Registration of multiview real-time 3-d echocardiographic sequences," *IEEE Transactions on Medical Imaging*, vol. 26, no. 9, pp. 1154–1165, 2007. DOI: 10.1109/TMI.2007.903568.

[35] G. Leroy, D. Rueckert, and A. Alansary, "Communicative reinforcement learning agents for landmark detection in brain images," in *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology*, S. M. Kia *et al.*, Eds., Cham: Springer International Publishing, 2020, pp. 177–186, ISBN: 978-3-030-66843-3.

[36] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:205242740.

[37] *What Is Reinforcement Learning? - MATLAB & Simulink*. [Online]. Available: https://www.mathworks.com/help/reinforcement-learning/ug/what-is-reinforcement-learning.html (visited on 02/12/2024).

[38] *Deep reinforcement learning*, en, Page Version ID: 1199844804, Jan. 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Deep_reinforcement_learning&oldid=1199844804 (visited on 02/09/2024).

[39] A. Alansary *et al.*, "Evaluating reinforcement learning agents for anatomical landmark detection," *Medical Image Analysis*, vol. 53, pp. 156–164, 2019, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2019.02.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841518306121.

[40] A. Alansary *et al.*, *Automatic view planning with multi-scale deep reinforcement learning agents*, 2018. arXiv: 1806.03228 [cs.CV].

[41] F.-C. Ghesu *et al.*, "Multi-scale deep reinforcement learning for real-time 3d-landmark detection in ct scans," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 176–189, 2019. DOI: 10.1109/TPAMI.2017.2782687.

[42] G. Leroy, D. Rueckert, and A. Alansary, "Communicative reinforcement learning agents for landmark detection in brain images," in *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology: Third International Workshop, MLCN 2020, and Second International Workshop, RNO-AI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 3*, Springer, 2020, pp. 177–186.

[43] C. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, pp. 279–292, May 1992. DOI: 10.1007/BF00992698.

[44] H. v. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, ser. AAAI'16, Phoenix, Arizona: AAAI Press, 2016, 2094–2100.

[45] A. Fedorov *et al.*, "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012, Quantitative Imaging in Cancer, ISSN: 0730-725X. DOI: https://doi.org/10.1016/j.mri.2012.05.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X12001816.

[46] K. Marstal, F. Berendsen, M. Staring, and S. Klein, "Simpleelastix: A user-friendly, multi-lingual library for medical image registration," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016, pp. 574–582. DOI: 10.1109/CVPRW.2016.78.

[47] *4D LV-ANALYSIS*. [Online]. Available: https://www.tomtec.de/products/application-finder/4d-lv-analysis/ (visited on 01/21/2024).

[48] M. Tang, M. Lee, and Y. J. Kim, "Interactive hausdorff distance computation for general polygonal models," *ACM Trans. Graph.*, vol. 28, no. 3, 2009, ISSN: 0730-0301. DOI: 10.1145/1531326.1531380. [Online]. Available: https://doi.org/10.1145/1531326.1531380.