Diffeomorphic Deformable Image Registration for Cardiac Magnetic Resonance Imaging (MRI)

by

Ameneh Sheikhjafari

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

Doctor of Philosophy

Department of Computing Science
University of Alberta

Abstract

Deformable registration of cardiac magnetic resonance imaging (MRI) is one of the crucial tasks in medical image analysis. It aims to find the unique transformation between images taken from the same scene at different times, from different views, and by different imaging modalities such as MRI and CT. The main goal of this thesis is to develop automated deformable registration methods, particularly improving the accuracy and robustness of image registration while preserving the topology and invertibility of the deformation.

Although deformation fields related to the point correspondence between a pair of images are high-dimensional, we propose a method that generates deformation fields from low-dimensional latent variables by minimizing a dissimilarity metric between a fixed and a warped moving image. This low-dimensional manifold formulation avoids the intractability associated with the high-dimensional search space that most other methods face during image registration.

Moreover, we propose an end-to-end learning-free multi-resolution framework. This method eliminates the need for a dedicated training set while exploiting the capabilities of neural networks to achieve accurate deformation fields. Since it is capable to share the parameters through the architecture, it can be used for Groupwise registration as well as pairwise registration. We integrated Gaussian filters with the GMCNet to impose a smoothness constraint which relaxes the need for an explicit regularization term and its corresponding weight in the cost function.

Additionally, we propose a learning-free diffeomorphic recursive framework, which models the changes in the deformation over multiple resolutions as opposed to the

deformation itself. The final deformation is estimated by a solution to an ordinary differential equation (ODE). Thus, the resulting algorithm is recursive. Following this recursion, the moving image is warped successively and enables the final prediction to be decomposed into smaller displacements.

Finally, we present an end-to-end unsupervised diffeomorphic framework based on moving mesh parameterization. This new parameterization of the deformation field has three significant advantages; firstly, it relaxes the need for an explicit regularization term and its corresponding weight in the cost function. Secondly, it guarantees diffeomorphism through explicit constraints applied to the transformation of the Jacobian determinant. Finally, it is suitable for cardiac data since it parameterizes the deformation using radial and rotational components.

Preface

This thesis was submitted as partial fulfillment of the degree Doctor of Philosophy (Ph.D.) in Computing Science at the University of Alberta. The thesis is an original work by Ameneh Sheikhjafari and the presented work was accomplished between September 2017 and September 2022. 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 papers:

Chapter 3:

Sheikhjafari Ameneh, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray. "Unsupervised deformable image registration with fully connected generative neural network." Medical Imaging with Deep Learning (MIDL) 2018.

Chapter 4:

Sheikhjafari Ameneh, Michelle Noga, Ahmed Ahmed, Kumaradevan Punithakumar, and Nilanjan Ray. "GMCNet: A generative multi-resolution framework for cardiac registration." Submitted to IEEE Open Access (2022) - Under Review.

Chapter 5:

Sheikhjafari Ameneh, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray. "A training-free recursive multiresolution framework for diffeomorphic deformable image registration." Applied Intelligence (2022): pp. 1-10.

Chapter 6:

Sheikhjafari Ameneh, Krishnaswamy Deepa, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray, "Unsupervised diffeomorphic cardiac image reg-

istration using parameterization of the deformation field." ArXive, 2022.

Sheikhjafari Ameneh, Krishnaswamy Deepa, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray, "Unsupervised diffeomorphic cardiac image registration using parameterization of the deformation field." Medical Image Analysis (2022) - Under Review.

Chapter 7:

Sheikhjafari Ameneh, Krishnaswamy Deepa, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray. "Deep Learning Based Parametrization of Diffeomorphic Image Registration for the Application of Cardiac Image Segmentation." Submitted to International Conference on Bioinformatics and Biomedicine BIBM (2022) - Under Review.

"Dreams are often most profound when they seem the most crazy."

- Sigmund Freud

I dedicate my thesis to my loving parents, Shahrbanoo Rafie and Davood Sheikhjafari whose words of encouragement and push for tenacity ring in my ears. A special feeling of gratitude for my mom, an independent, powerful woman who is my hero. I also dedicate this dissertation to my siblings, Rahele, Alireza, Adele, and Mohammadreza whose taught me to work hard for the things that I aspire to achieve and have never left my side.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisors Prof. Nilanjan Ray and Dr. Kumaradevan Punithakumar for providing tremendous support in my Ph.D. study and related research. Without their tireless guidance and unending faith in my potential, none of my success could have been possible. I would also like to thank Dr. Michelle Noga from the Servier lab, who has graciously provided me with clinical guidance over the years. I would like to extend my sincere thanks to my friend and colleague Deepa, one of the most diligent and kind individuals.

I would like to thank my family and all their encouragement. For my parents who raised me with love and supported me in all my pursuits. Thank you and love you.

Above all, I would like to thank my husband for his love and constant support, and for keeping me sane over the past few months. Thank you for being my muse, editor, proofreader, and sounding board.

Table of Contents

List of Tables

List of Figures

Abbreviations & Acronyms

BD Backward Deformation.

DetJ Determinant of Jacobian.

DM Dice Metric.

ED End diastolic.

EDV End Diastolic Volume.

EF Ejection Fraction.

ES End Systolic.

ESV End Systolic Volume.

FD Forward Deformation.

FFT Fast Fourier Transform.

HD Hausdorff Distance.

LV Left Ventricle.

MAD Mean Absolute Distance.

MRI Magnetic Resonance Imaging.

MYO Myocardium.

ROI Region of Interest.

RV Right Ventricle.

Chapter 1

Introduction

1.1 Motivation

Image registration is a method of aligning two or more images into the same coordinate system [1, 2]. It is required to find the geometrical transformation that aligns the images taken at different times, from different perspectives or using different modalities. Therefore the aligned images can be directly combined, compared and analyzed. Medical imaging is one of the challenging application domains of image registration [2, 3]. In different applications image registration can be called image alignment, matching stabilization and fusion. Based on the image coordinate transformation, we can have different types of registration *rigid*, when only translations and rotations are needed to align images, *affine* when the transformation maps parallel lines onto parallel lines, *projective* If it maps lines onto lines and finally *deformable/non-rigid*.

In *rigid* registration, rotation and shift are needed to align images, for instance, to correct head position between scans. Most of the early research in medical image registration was working on aligning brain images acquired with different modalities [4, 5]. For these applications, a rigid body approximation was sufficient as there is a relatively small difference in brain shape or position within the skull over the relatively short periods between scans.

Deformable/non-rigid registration often arises where two images are related through non-rigid geometric transformations. Deformable image registration is one of the key

technologies in medical image analysis. The term deformable is used to denote the fact that the mapping is associated with a nonlinear dense transformation. Most of the human body does not fit to a rigid or even affine approximation [6]. Therefore, the most challenging work in registration involves the development of non-rigid/deformable registration techniques for applications ranging from correcting soft-tissue deformation during imaging or surgery [2]. Application of deformable image registration includes motion compensation, multimodality fusion, monitoring of changes, segmentation, and atlas construction.

1.1.1 Deformable image registration

Deformable image registration is one of the crucial tasks in medical image analysis which aims to find the point-wise mapping between a pair of images. These techniques serve as the fundamental basis for procedures such as image-guided radiation therapy, surgery, tumour growth monitoring, minimally invasive treatments, and many other challenging problems [7, 8]. Studies have shown that deformable registration could also be used for the cardiac functional assessment and delineation of cardiac structures with magnetic resonance imaging (MRI) sequences.

Despite a large amount of work and relative success, deformable image registration is still not a solved problem [2, 9]. Many challenges remain in the definition of correspondences between the image elements and transformation models. Images can be corrupted by noise, illumination changes and blocking artifacts, blur artifacts due to body motion, or outliers, e.g. contrast agent, growing tumour or moving cells, which have no correspondences in the other image. That makes mathematically formulating the correspondence definition between images difficult. Furthermore, if two images are from different modalities, then the same anatomical object can have completely different intensities. The transformation model definition is the second main challenge in deformable image registration because the true underlying deformation is often unknown. The definition of proper similarity measure is another challenging

task in the presence of deformable registration.

Existing deformable image registration algorithms could be broadly categorized into 1) Optimization-based approaches, and 2) Learning-based approaches. Traditional methods for deformable image registration hardly use any training sets. Therefore, they are more generally and easily applicable. However, most similarity measures, especially in multimodal registrations, have lots of local optima around the global optimum which leads to early unanticipated convergence and stagnation. These problems are two common limiting dilemmas in the optimization field [10]. On the contrary, recently deep learning-based methods, which require a large training set, are routinely defeating traditional methods with respect to accuracy. However, learning-based methods include having to re-train for new images/modalities and the need to have access to training datasets that are quite substantial in size. More importantly, learning-based methods do not guarantee an optimal solution at the test time. In addition, [11] disclosed that manually crafted methods might outperform deep learning methods in unimodal registration cases. Thus, learning-based methods typically trade optimality for efficiency at deployment and it seems that they are trading generality and ease of use for computational time.

1.2 Thesis contribution

In the proposed research, we investigate learning and non-learning based on opposing views and bring the two ends closer together by using the advantages of both methods. For instance, CNNs have a couple of powerful properties: a) excellent function approximation capability and b) an appropriate regularization mechanism in the form of translational invariance. Additionally, modern tools, such as PyTorch and TensorFlow made back-propagation easy through any differentiable pipeline including deep networks. We would like to exploit these properties of deep neural nets vis-a-vis auto-grad to combine them with the classic features of traditional methods: multi-resolution structure, continuous nature, and learning-free nature of optimization.

An overview is provided of the contributions in this thesis in the field of medical image registration.

Unsupervised Deformable Image Registration with Fully Connected Generative Neural Network: The proposed method uses an FCNet as a generator to estimate deformation fields directly and at the same resolution as the images to be registered. Even though deformation fields related to the point correspondence between fixed and moving images are high-dimensional in nature, we assume that these deformation fields form a low-dimensional manifold in many real-world applications. Thus, in our method, a neural network generates an embedding of the deformation field from a low-dimensional vector. This low-dimensional manifold formulation avoids the intractability associated with the high-dimensional search space that most other methods face during image registration. As a result, while most methods rely on explicit and handcrafted regularization of the deformation fields, our algorithm relies on implicitly regularizing the network parameters. The proposed method generates deformation fields from low dimensional latent space by minimizing a dissimilarity metric between a fixed image and a warped moving image. Our method removes the need for a large dataset to optimize the proposed network. The results have demonstrated that a randomly-initialized neural network can be used as a handcraft before achieving satisfying performance for deformable medical image registration. It is remarkable because no aspect of the network is learned from data. Based on the author knowledge it is the first neural network-based manifold embedding method for unsupervised deformable image registration.

GMCNet: A generative Multi-resolution Framework for Cardiac Registration: As a second alternative for deformable image registration, we propose a novel approach to improve registration accuracy and obtain smooth transformations between image pairs using a generative multi-resolution training-free CNN-based architecture. Similar to the first proposed method the second framework is generative and capable of estimating desired deformation fields for pairs of images based on ran-

dom low dimensional vectors as inputs. In this method, we use a Gaussian filter to regularize deformation fields similar to the Demons algorithm. The results show that the second algorithm outperforms recent state-of-the-art learning and non-learning-based methods on three clinical cardiac MRI datasets with respect to accuracy. This framework is faster, more accurate and has fewer parameters than the first proposed method.

A Training-free Recursive Multi-resolution Framework for Diffeomorphic Deformable Image Registration: We have proposed a training-free approach for deformable image registration that is built upon the principle of an ordinary differential equation as a third solution. As [12] reveal that using ordinary differential equations lead to faster, more accurate result for time-series predictions and computing gradients with constant memory cost. This method has reasonable computational time and it is suitable for large deformation as well as small deformation. In this method, the rate of change of deformation over image resolution is parameterized by a CNN. Our formulation yields an Euler-integration type recursive scheme to estimate the spatial transformations between fixed and moving multi-resolution image pyramids. We optimize the parameters of the CNN by minimizing a loss between the fixed and the warped moving image pyramids. In contrast to traditional non-learning methods, the proposed approach is more flexible and expressive to model deformations because it integrates the outputs of a CNN starting from the coarsest to the finest image resolutions to approximate spatial transformations. In comparison to learning-based methods, the proposed method neither requires a dedicated training set nor suffers from any training bias. Evaluations were performed on three clinical cardiac MRI datasets, in terms of mean Dice coefficient and Hausdorff distance. The evaluation results demonstrate that the proposed registration approach yields state-of-the-art accuracy in a reasonable time.

Learning-based: Learning A Diffeomorphic Registration Model Using Moving Mesh Correspondence for Cardiac MRI: All the previous methods

are 2D registrations. Therefore, we propose a novel 2D-to-2D and 3D-to-3D diffeomorphic registration algorithm by investigating a new parameterization of the deformation field, which describes a deformation field with its transformation Jacobian determinant and curl of the end velocity field. The proposed representation of the deformation field has some advantages: 1) it relaxes the need for an explicit regularization to produce a physically plausible result because smoothness is implicitly embedded in the solution. 2) Parameterizing a deformation field in terms of radial and rotational components through unsupervised learning, making it especially suited for processing cardiac data. 3) the deformation can be ensured to be diffeomorphic by directly requiring the transformation Jacobian to be positive. 4) The other desirable constraints also can be enforced within the same framework using an explicit restriction on the transformation Jacobian such as incompressibility constraint.

Learning-based: Learning A Diffeomorphic Registration Model Using Moving Mesh Correspondence for Cardiac MRI: All the previous methods are 2D registrations. Therefore, we propose a novel 2D-to-2D and 3D-to-3D diffeomorphic registration algorithm by investigating a new parameterization of the deformation field, which describes a deformation field with its transformation Jacobian determinant and curl of the end velocity field. The proposed representation of the deformation field has some advantages: 1) it relaxes the need for an explicit regularization to produce a physically plausible result because smoothness is implicitly embedded in the solution. 2) Parameterizing a deformation field in terms of radial and rotational components through unsupervised learning, making it especially suited for processing cardiac data. 3) the deformation can be ensured to be diffeomorphic by directly requiring the transformation Jacobian to be positive. 4) The other desirable constraints also can be enforced within the same framework using an explicit restriction on the transformation Jacobian such as incompressibility constraint.

Diffeomorphic Image Registration for the Application of Cardiac Image Segmentation: In this chapter we propose a novel end-to-end supervised cardiac

MRI segmentation framework based on a diffeomorphic deformable registration that can segment the left ventricle from 2D and 3D images or volumes. In order to represent the actual cardiac deformation, in this methodology the transformation is parameterized using radial and rotational components, computed using a deep learning approach. The deep learning method is trained using a set of pair images along with their segmentation masks. The formulation guarantees transformations that are invertible and prevents mesh folding which is essential for preserving the topology of the segmentation results. A physically plausible transformation is achieved by employing diffeomorphism in computing the transformations and activation functions that constrain the range of the radial and rotational components.

1.3 Thesis overview

The thesis is structured into 8 main chapters as follows: In Chapter 2, previous works both traditional methods, as well as deep learning-based approaches in the literature, are presented. Chapter 3 presents an unsupervised dformable image registration with a fully connected generative neural network. The proposed approach is a sequential registration and is evaluated on a public set of MRI scans. In chapter 4 GMCNet: A generative multi-resolution framework for cardiac registration is presented. The method is evaluated on three datasets and compared with state-of-the-art learning and non-learning-based methods. Chapter 5 presents a training-free recursive multi-resolution framework for diffeomorphic deformable image registration. The proposed method is compared against learning and non-learning-based algorithms. Evaluations were performed on three different datasets. Chapter 6 discusses a 2D-to-2D and 3D-to-3D diffeomorphic learning-based algorithm with moving mesh correspondence for deformable image registration. Three datasets were employed for the evaluation. Chapter 7 presents a diffeomorphic image registration for the application of cardiac image segmentation. Finally, in Chapter 8, a summary of the significant results, future work, and limitations are discussed.

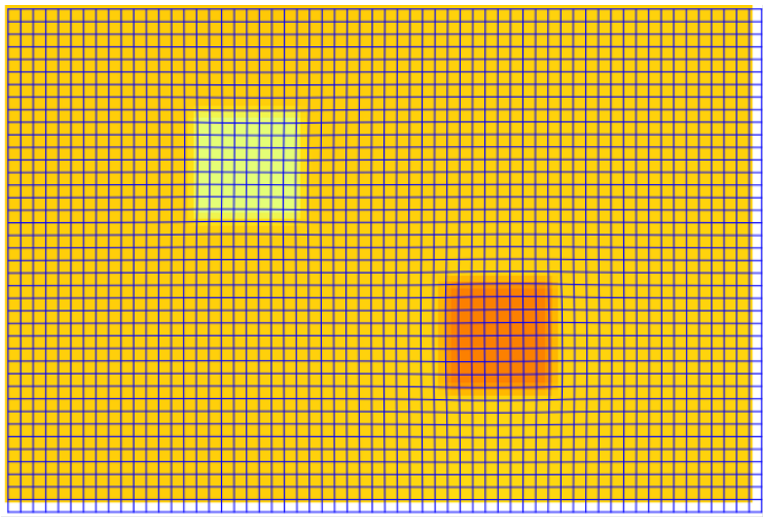
Chapter 2

Background

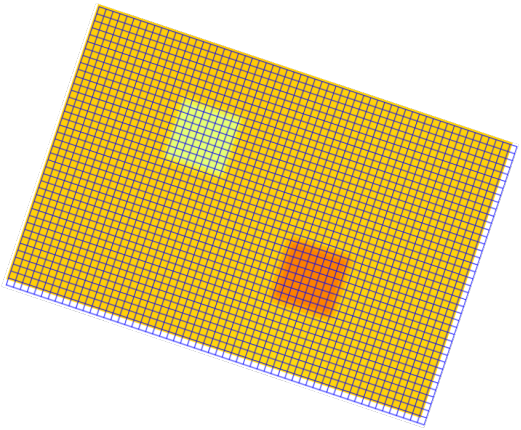
2.1 Basics of image registration

Registration is the process of matching different images and transforming them into the same coordinate system. The images can be acquired from different times, by different sensors and/or from different viewpoints. Commonly when we aligning two images, one is regarded as static and is known as the target, reference, or fixed image and the other image undergoing transformation (i.e. translated, rotated, warped, etc.) which is known as the source, or moving image. The geometrical transformation that maps features in one image to features in another is known as the transformation, deformation field, or displacement field. Which needs to be estimated/modelled in order to register two or more images. These transformations are usually classified as being rigid (where images simply need to be rotated and translated with respect to one another to achieve correspondence), affine (which is an extension of rigid registration, and includes scale factors and shears) or non-rigid/deformable (where the correspondence between structures in two images cannot be achieved without some localized stretching of the images) as illustrated in fig. 2.1. Most parts of the human body especially soft tissues do not conform to a rigid or even an affine approximation [2].

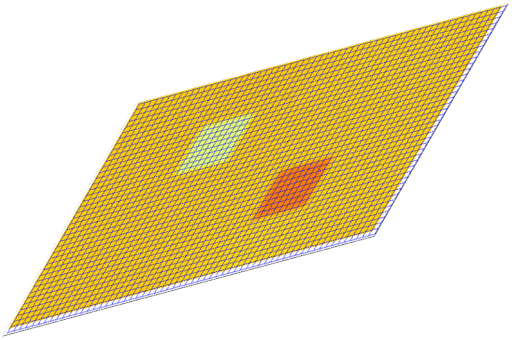
One of the challenges with the non-rigid registration problem is the large number of parameters that makes it time-consuming and hard to solve. Since Rigid and



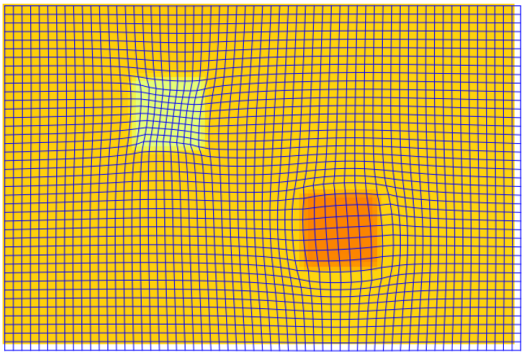
(a) Original



(b) Rigid



(c) Affine



(d) Non-rigid

Figure 2.1: Three transformation models.

affine registrations have fewer parameters (degree of freedom), which typically can be calculated in seconds or minutes. It is worth emphasizing that increasing flexibility needs more parameters to be determined, which requires more computational time. Therefore, most non-rigid optimization-based registration algorithms require minutes or hours [2] to be solved. Another challenge with non-rigid registration is asymmetric transformation. Asymmetric transformation is defined when there is a vector that describes how each point in the source image displaces the corresponding point in the target image, but there is no guarantee that each point in the target image can be related to a corresponding position in the source image. Which may make a gap in the target image. Some research has been done on symmetric schemes that guarantee the same result whether image A is matched to image B or vice versa [13]. This may be more appropriate for some applications (matching one normal brain to another) than others (monitoring the growth of a lesion). Finally, there is the question of redundancy. Since non-rigid registration is an ill-posed problem, there will be many different deformation fields (possible solutions) to align images which behave differently. For instance, where deformation field can be "folded" to improve the image match but in a non-physical way which is disallowed situations.

2.2 Deformable Image Registration

Deformable image registration has three main components include: 1) the similarity measure, 2) the transformation model, 3) regularization. The similarity measure usually is defined as an objective function that achieves its optimum when two images have a certain relationship. We discuss different similarity measures in section 2.2.1.

The second component of an image registration algorithm is the transformation model. The transformation model defines the set of plausible transformations required for aligning two images. For instance, in rigid parametrization, rotation and shifting are adequate transformations to align two images with different orientations. Deformable transformations are subdivided into parametric and non-parametric ones.

Parametric methods assume a particular parametrization of the transformation, e.g. spline-based [9]. In non-parametric methods, a transformation is estimated as an unknown function. We overview different transformation models in section 2.2.2.

As image registration is inherently an ill-posed problem; the existence and the uniqueness of the solution are not guaranteed and regularization is essential to make the problem well-posed. Regularization adds constraints to the transformations. In parametric methods regularization implicitly is applied to the transformation by using a low dimensional set of parameters. The theory of the regularization approaches in image registration is reviewed in section 2.2.3. The last component of an image registration algorithm is the optimization method that finds a set of parameters to optimize a given similarity measure. Some examples of optimization methods are employed including Gradient Descent, Quasi-Newton, and Stochastic Gradient Descent methods [14]

2.2.1 Similarity Measure

The similarity measure in intensity-based registration methods is defined directly by image intensities. Intensity-based similarity measures based on their application area are divided into two monomodal and multimodal classes.

Sum-of-squared-differences (SSD) is one of the simplest intensity-based measures

$$E_{SSD}(T) = \sum_{n=1}^{N} (I_F{}^n - (I_M{}^n \circ T))^2 \tag{2.1}$$

The optimal similarity measure of SSD occurs when two images only differ by Gaussian noise. Even though, SSD is simple and widely used in image registration, but is very sensitive to outliers and image artifacts. Sum-of-absolute-difference (SAD) is an approach to reducing the effect of outliers

$$E_{SAD}(T) = \sum_{n=1}^{N} |I_F{}^n - (I_M{}^n \circ T)| \tag{2.2}$$

Both SSD and SAD assume two images to have equal intensities at the correct alignment. On the contrary, the squared correlation coefficient (CC) is maximized when

the images are linearly related

$$E_{CC}(T) = \frac{(\sum_{n=1}^{N}({I_F}^n - \bar{I_F})({I_M}^n - \bar{{I_M}^T}))^2}{\sum_{n=1}^{N}({I_F}^n - \bar{I_F})^2 \sum_{n=1}^{N}({I_M}^n - \bar{{I_M}^T})^2} \tag{2.3}$$

where $\bar{I_F}, \bar{I_M}$ are the intensity mean of images I_F and I_M respectively. And ${I_M}^T$ is $I_M \circ T$.

The Structural Similarity (SSIM) [15] index is another method for measuring the similarity between two images. If we consider image I to have a perfect quality, then the similarity measure can serve as a quantitative measurement of the quality of the image J.

$$SSIM(I_F, I_M) = \frac{(2\mu_{I_F}\mu_{I_M} + c_1)(2\sigma_{I_F I_M} + c_2)}{(\mu_{I_F}^2 + \mu_{I_M}^2 + c_1)(\sigma_{I_F}^2 + \sigma_{I_M}^2 + c_2)} \tag{2.4}$$

where μ_{I_F} and μ_{I_M} are the average of the images and σ_{I_F} and σ_{I_M} are the variance of the images I_F and I_M respectively. σ_{II_M} is the co-variance I_F and I_M. Constant $c_1 = (K_1 L)^2$, $c_2 = (K_2 L)^2$, and L is the dynamic range of the pixel values (255 for grayscale image) and $K_1 << 1, K_2 << 1$ are small constant. The system separates the task of similarity measurement into three comparisons: luminance, contrast and structure. As the MI family assumes a statistical relation between the intensities of the images they are suited not only for monomodal but also for multimodal image pairs.

Mutual information (MI) [16] maximizes for the simultaneous low joint entropy and high marginal entropies.

$$E_{MI}(T) = H(I_F) + H(I_M \circ T) - H(I_F, I_M \circ T) = \sum_i \sum_j P^T(i,j) log \frac{P^T(i,j)}{P_{I_M \circ T}(j) P_{I_F}(i)} \tag{2.5}$$

where $P^T(i,j)$ is the joint probability of the intensities of the images, and $P_{I_M \circ T}(j)$, $P_{I_F}(i)$ are the marginal probabilities. MI measures how well one image explains the other. Maximized MI is obtained when the value of a voxel in the first image is a good predictor of the corresponding voxel in the second image.

2.2.2 Transformation Model

The Transformation model is the way to deform one image to match the other. Non-rigid transformation subdivide into parametric and non-parametric transformation [9]. Parametric transformation such as B-spline depends on a set of parameters. B-spline utilizes a mesh of control points (parameters) and interpolates in between with B-spline basis functions. The sparseness of control points limits the possible transformations. In non-parametric image registration methods, the transformation is estimated as an unknown function. With a non-parametric approach, the complex local deformation can be modelled. This approach is ill-posed and requires constraining the transformation by using regularization to be smooth or locally rigid.

Parametric Approach

In parametric image registration methods, the spatial transformation is usually defined as a function of a set of parameters. The popular non-rigid parametric transformations include radial basis functions (RBF) and B-splines. RBF approach defines the transformation as a linear combination of the basis functions

$$f(x) = \sum_{n=1}^{N} w_n U(\|x - x_n\|) \tag{2.6}$$

where w_n are unknown parameters. The basis function U depends on the Euclidean distance from x to the given control point x_n. Among different RBFs, TPS and Cubic B-spline are more popular in medical image analysis.

Non-Parametric Approach

Non-parametric image registration estimates the transformation as an unknown function without explicit parametrization. The solution to such a problem is not unique, it is called the ill-posed problem. To limit the problem a regularisation term can be used to penalize some undesirable transformation properties. The transformation is usually defined as

$$f(x) = x + u(x) \tag{2.7}$$

where $u(x)$ is an unknown displacement function that optimize the objective function:

$$E(u) = E_{sim}(I, J(x + u(x))) + \gamma E_{reg}(u(x)) \tag{2.8}$$

where E_{sim} is the similarity measure term, E_{reg} is the regularization term and γ is a trade-off parameter.

2.2.3 Regularization

Regularization is an essential key in image registration. It imposes certain properties such as sparsity, smoothness and rigidity on a function to make this ill-posed problem well-posed. For instance, estimating a non-rigid transformation based on only a set of given control points/landmarks. There is no unique solution to such a problem. There is an infinite number of transformations that have different behaviour in the remaining parts of the image but will match the corresponding landmarks. By defining a regularization term to enforce some constraints to transformation, the problem can be uniquely solved. Alternatively, the solution can be regularized by explicit parametrization, e.g. rigid. A standard regularization approach is to augment the existing optimization problem with a regularization term and minimize the following functional

$$E(f) = S(f) + \gamma R(f) \tag{2.9}$$

where $S(f)$ is objective function, $R(f)$ is the regularization term and γ is the trade-off parameter. There are two types of regularization implicit and explicit regularization.

Implicit Regularization

The regularization is imposed implicitly **1) Parametarizing the transformation model by lower degrees of freedom** by using a low dimensional set of parameters such as free form deformations (FFD), B-spline, radial basis function (RBF). For instance, B-spline parametrization limits the deformation by the control point spacing and the smoothness of the B-spline basis functions. **2) Restricting the space of**

the transformation by looking at special spaces. The main idea of changing the space is that instead of adding regularization to the transformation we can define the transformation in a regularized space.

Explicit Regularization

The regularization is imposed explicitly by **1) Deformation field filtering:** One of the explicit regularization methods is to use the filtering deformation field, like the Gaussian filter in the demons framework. The purpose of this smoothing is to suppress noise and preserve the geometric continuity of the deformed image. **2) Adding penalty/regularization term:** Standard regularisation terms to make the problem well-posed include the diffusion, curvature, elastic and fluid regularization.

2.3 Image Registration Overview and Methods in Literature

2.3.1 Non-Learning-Based Image Registration Methods

Traditionally, deformable registration is resolved by optimizing a similarity metric that measures the closeness between the fixed and the warped moving image. Several studies using optimization within the space of displacement vector field such as discrete methods [17–20] and Demons [21, 22] have been proposed for the registration problem. The basic concept behind the demons framework for deformable image registration is that the pixels in the fixed image act as local forces (applied by "demons") that can displace the pixels in the moving image to match the fixed image. The intuition behind the image force is that if the pixel value in the moving image is lower than the pixel value in the fixed image, demons push according to the image gradient. Demons push the opposite of the image gradient if the pixel value of the moving image is higher than the target value.

In [23] a combination of segmentation and registration is suggested based on nonlinear elasticity which uses a polyconvex for regularization. A numerical implementation

of the registration with a polyconvex regularization term is also presented by Burger et al. in [24]. Another study [25] proposed a method based on the log-domain for spatial transformation where a physical constraint is applied to estimate the myocardial strain from cine MRI in the registration process. Due to the properties of diffeomorphic transforms, namely folding-free and invertible [7], some methods rely on such transformations computed using an artificial velocity over time governed by the Lagrange transport equation to model deformations. The importance of these properties has led to the wide use of diffeomorphism in registration algorithms [26–28].

In [29, 30], a registration framework was proposed based on moving mesh (grid generation) to compute point-to-point correspondences. They used the L^2 norm as a dissimilarity measure. The authors modelled the deformation using radial (divergence) and rotational (curl) components that is better for analyzing the heart as this closely matches actual cardiac motion.

Despite the advancements in transformation computations, the deformable registration is an ill-posed problem that requires explicit regularization to obtain a unique solution [7]. Several regularization such as elastic-based [24] or diffusion-based methods [31, 32] have been suggested to overcome the issue. Besides, most of these methods require users to find parameters on hand. The process of finding the parameters that match the characteristics of certain data depends entirely on the users' intuition and their several tedious attempts.

2.3.2 Learning-based image registration methods

Recently, learning-based predictive approaches, notably deep learning techniques have been successfully used for image registration. Depending on the manner the networks are trained, they can be categorized into supervised and unsupervised learning. Deep-learning-based image registration can be divided into three main groups of the algorithm: 1) iterative, 2) supervised and 3) unsupervised registration [11].

Most learning-based algorithms use convolutional layers to learn the similarity/dissimilarity between the moving and fixed images. In iterative registration, the gradient descent is applied to iteratively update the parameters of the deformation field based on the given similarity/dissimilarity metric. When the solution space in registration is high dimensional, an iterative algorithm can lead to slow registration. This motivated the development of networks that could estimate the transformation in one step. However, fully supervised transformation estimation needs the exclusive use of ground truth data to define the loss function. The difficult nature of the acquisition of reliable ground truth remains a significant hindrance which leads to unsupervised approaches. One key innovation that has been useful to these works is the spatial transformer network (STN).

In supervised-learning methods, a convolutional neural network is trained using examples of medical images along with their ground truth transformation to predict the transformation directly on test images. Inspired by U-Net network [33], and using mesh segmentation to compute the reference transformation, [34] predicted the deformation field for 3D cardiac MR registration. [35] proposed a convolutional neural network (CNN) to estimate the displacement vector for 3D brain MR volumes. Where equalized active-points guided sampling and similarity between image patches were used to guide the learning process. Even though the accuracy of these approaches is considerable, their performance is dependent on the quality of the ground truth. The most significant drawback of the supervised methods is that the actual ground truth of a desired neural network output is not often available. There are only a limited number of individuals with the expertise to provide quality labels, and therefore, the supervised methods have limited ability to enhance the size and diversity of datasets.

Even though the supervised learning-based methods could rely on synthetic examples generated using random transformations or well-established traditional registration methods to obtain reference transformations, such training options could limit the robustness and accuracy of the algorithm in practice. Firstly, synthetic examples

may not be realistic and lead to poor generalization. Secondly, using the transformation from a traditional method as a reference inherently limits the prediction accuracy of the trained network to the accuracy of the original method [36]. The limitation of supervised methods has motivated different research groups to investigate dual or weakly supervised transformation estimation [37–39]. Nevertheless, these methods still require manually produced labels or segmentation.

To overcome some of the limitations imposed by supervised learning-based methods such as the need for expert annotated data, unsupervised image registration has received a lot of attention. In other words, the quality of the model is not dependent on the quality of the labels. Inspired by the spatial transformer network (STN) [40], [41, 42] proposed an unsupervised network for deformable registration which relied on normalized cross-correlation (NCC) and bending-energy regularization terms to train a fully convolutional neural network (FCN). [43, 44] used similarity to train a general framework for unsupervised image registration. Most unsupervised methods [41, 43, 45] use B-spline or STN as a differentiable warping of the moving images. However, these methods have not demonstrated that their deformations are sufficiently regular and plausible [46].

Generative models such as a generative adversarial network (GAN), stochastic variational autoencoder (VAE) and adversarial autoencoder (AAE) showed promising results in medical imaging application in learning data distribution from large image training set [7]. Recently, several research groups [45, 47] used unsupervised adversarial for image registration. In contrast to some generative approaches that used a GAN to ensure the predicted deformation is realistic, [48] proposed a GAN to perform deformable image registration of 3D MR volumes. [49] proposed simultaneous segmentation and registration of chest X-rays using a GAN framework. The network relies on three inputs: a reference image, a moving image, and the segmentation mask of the reference image and outputs the segmentation mask of the transformed image and the deformation field. Three discriminators are used to assess the quality of the

generated outputs, deformation field, warped image, and segmentation using cycle consistency and a Dice metric. The approach parameterizes deformable registration of 4D CT thoracic image volumes that uses the sum of squared differences (SSD) as a similarity metric. [7] proposed a generative and probabilistic model for diffeomorphic image registration which first trains an encoder-decoder neural network to estimate the deformation field by providing a large dataset of training images. By learning the low-dimensional global latent space, during the test time they use the decoder part of the network. However, the performance of these methods depends on the structure of a large training set [50].

2.3.3 Multi-Resolution Registration

A problem with non-linear medical image registration is the high complexity and multiple local minima in the optimization surface [51]. The multi-resolution approach is one of the solutions that help to overcome these issues. Multi-resolution methods have also been called coarse-to-fine, hierarchical, and pyramidal in the literature. These methods create from each image two or more images of increasingly smaller sizes with fewer image geometric differences. Furthermore, using multi-resolution techniques simplify the optimization and sapped up the registration process [52]. The structure of a multi-resolution image registration method is shown in Fig. 2.2. Level 0 is the original images, and in each level, the images are decreased in size by a factor of 2 (typically) to obtain new images. To obtain the desired lowest-resolution images, the process is repeated. The desired lowest resolution can be determined automatically or provided by the user. First, the images at the level n are registered. Smaller images simplify the correspondence process because local geometric differences between the images will be reduced sufficiently. To estimate registration parameters at level $n-1$ the registration result at level n is used.

By subdividing the images at one level higher resolution (level $i-1$) into correspondence subimages (level i) and registering two images at a given resolution (level

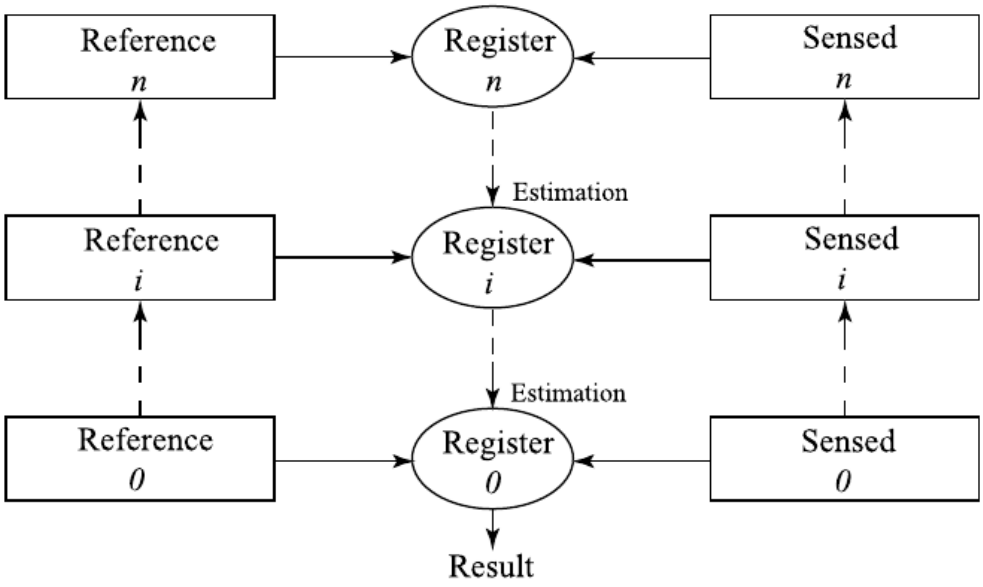


Figure 2.2: The structure of multiresolution image registration [53].

i), the registration parameters can be refined using information into corresponding subimages. The subdividing process causes us to deal with smaller images and reduces the search area. Besides, knowing the approximate registration parameters makes finding the ultimate registration parameters faster.

Information about the registration of images at level i is used to guide registration at level $i-1$, and the process is repeated until registration at the highest resolution (level 0) is achieved.

Gaussian Pyramid

It used a low-pass filter with Gaussian characteristics. Let G_0 be the original image. It becomes the bottom or zero levels of the Gaussian pyramid. Each pixel of the next pyramid level, image G_1, is obtained as a weighted average of the pixels in image G_0 within an $n \times n$ window. Each pixel of G_2 is then obtained from G_1 by applying the same pattern of weights. The window moves horizontally or vertically so that its centre is the second-next pixel of the current pixel, i.e., the sample distance at each

level is double that in the previous level. As a result, each image in the sequence is represented by an array which is half as large as its predecessor.

$$g_l(i,j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m,n) g_{l-1}(2i+m, 2j+n) \tag{2.10}$$

where w is Gaussian mask.

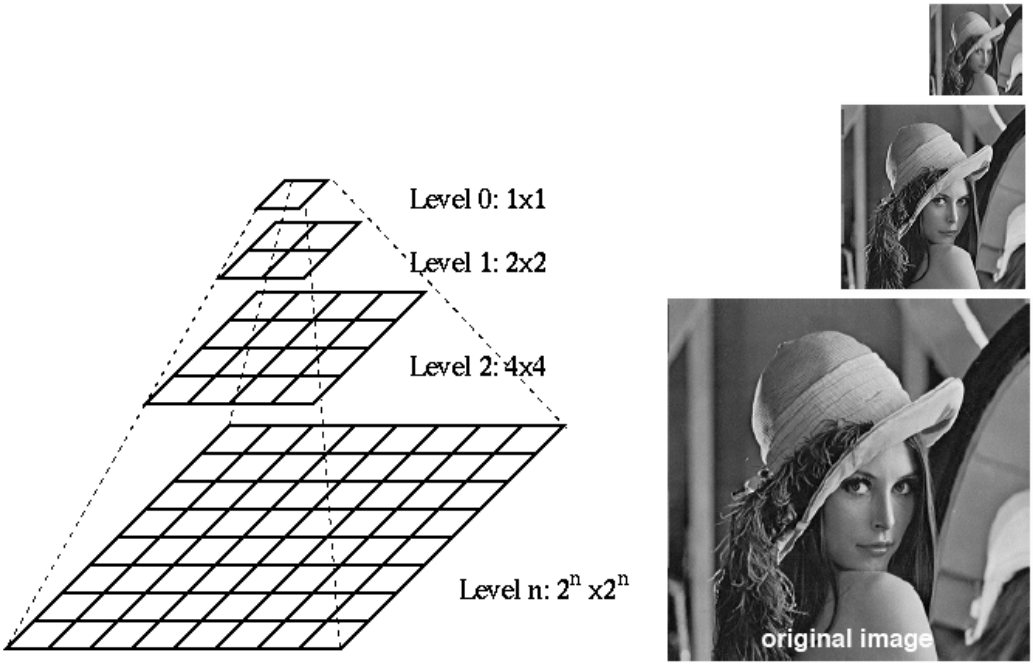


Figure 2.3: Image in multiresolution representation [54].

2.3.4 Groupwise Image Registration

In groupwise image registration methods, several images are registered to a common coordinate system and instead of a single geometric transformation, there is a group of transformations that has to be found [55] One method for groupwise registration

is selecting the first (or any other) image as the reference image, and then the rest of the images are registered to the reference using pairwise image registration. The choice of the reference image is important in such an approach. In sequential image registration, the next image is registered onto the previous and the deformation field is tracked over time. Often two neighbouring images in a sequence have a small deformation and distant images have large deformation. Another variation of groupwise image registration is to register the next image onto the average of previously aligned images which removes noise. Finally, the "true" groupwise non-rigid image registration registers all images simultaneously [55, 56]. Such a formulation eliminates the bias of choosing any particular reference image.

2.4 Problem Definition

The main goal of image registration is to find the optimal spatial deformation field T_μ parameterized by μ that warps a moving image I_M to align it with a fixed image I_F. The optimal values of μ could be obtained by minimizing an objective function of the form:

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}}\, L(I_F, I_M \circ T_\mu)) + \lambda R(T_\mu), \tag{2.11}$$

where L is a metric that measures how well I_F and the warped moving image $I_M \circ T_\mu$ are aligned. R is a penalty term that imposes smoothness on spatial deformation T_μ and λ controls the relative importance of the term. Eq.2.11 is a problem of finding deformation field T for defined similarity metric L and regularization R, but the image registration problem is not simply optimization, because the optimization of solving the problem greatly changes according to L and R. If L or R is too simple, the optimization of 2.11 is also too easy and the resulting T is likely to be trivial. Therefore, depending on the given source and target images, a different L and a different R should be given.

In our proposed approaches, the transformation is defined by the parameters of

the network and by optimizing these parameters we estimate the deformation field. In learning-based, unsupervised methods, the objective function is defined as:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{(I_F, I_M)} L(I_F, I_M \circ f_\theta(I_F, I_M)) + \lambda R(f_\theta(I_F, I_M)). \tag{2.12}$$

These methods use a neural network, such as CNN [41, 43], f with parameters θ. Eq. (2.12) minimizes a loss L between the fixed, I_F and the warped moving image, $I_M \circ f_\theta$ from a set of training data.

2.4.1 Heart anatomy and function

Heart anatomy

The heart which is almost located in the center of the chest is the hardest working muscle in the human body. The cardiovascular system is made up of the heart and blood vessels. Its job is to circulate blood throughout the body. The blood brings oxygen and nutrients to the tissues.

As it shown in Figure 2.4, the heart is divided into four chambers. The upper chambers are termed atria, and the lower chambers are termed ventricles. The upper right chamber of the heart is Left atrium which receives oxygenated blood from the lungs and pumps it down into the left ventricle which delivers it to the body. The right upper chamber of the heart is right atrium which 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 is Left ventricle. 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 that receives deoxygenated blood from the right atrium and pumps it under low pressure into the lungs via the pulmonary artery is Right ventricle (RV). 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.

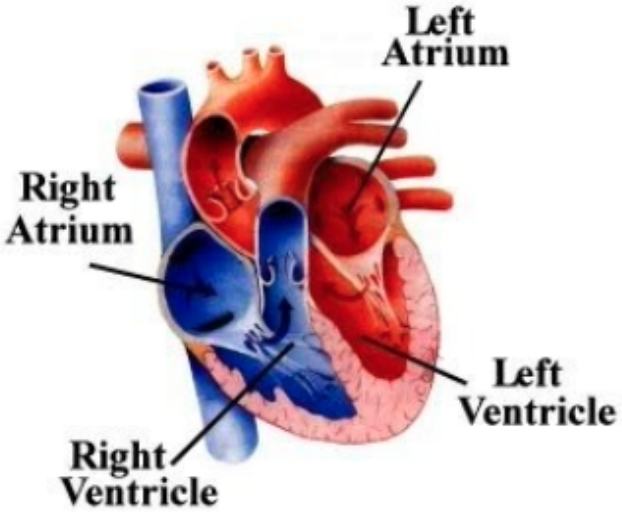


Figure 2.4: Basic anatomy of the heart [57].

2.4.2 MRI

A strong magnetic field is utilized to capture MR images and produce maps of hydrogen atoms in water or fat molecules in the body. By considering the spin of atomic nuclei as a magnetic vector, the proton behaves like a magnet. The image acquisition consists of an initial sequence of exciting pulses and recording of the emitted signal. In the next step, the amplitude of the signal is used to generate maps showing the heart structures. A 3D cross-sectional images of the heart can be created using cardiac MRI. MRI produces images with high resolution and high tissue contrast useful for different assessments such as heart valves, size and blood flow through the major vessels, etc. MRI is also utilized in diagnosing a variety of cardiovascular disorders such as tumors and inflammatory conditions, and for the preoperative treatment planning procedure and monitoring the progression of certain disorders over time.

2.5 Summary

In this section, we overviewed the general framework for image registration, and the key components of image registration, including popular similarity measures and transformation models, and showed the role of the regularization theory in image registration.

Chapter 3

Unsupervised Deformable Image Registration with Fully Connected Generative Neural Network

3.1 Overview

This chapter introduces a novel fully connected network (FCNet) to solve the optimization problem for image registration. The proposed method uses an FCNet as a generator to estimate deformation fields directly and at the same resolution as the images to be registered. Even though deformation fields related to the point correspondence between fixed and moving images are high-dimensional in nature, we assume that these deformation fields form a low-dimensional manifold in many real-world applications. Thus, in our method, a neural network generates an embedding of the deformation field from a low-dimensional vector. This low-dimensional manifold formulation avoids the intractability associated with the high-dimensional search space that most other methods face during image registration. As a result, while most methods rely on explicit and handcrafted regularization of the deformation fields, our algorithm relies on implicitly regularizing the network parameters. The proposed method generates deformation fields from low dimensional latent space by minimizing a dissimilarity metric between a fixed image and a warped moving image. Our method removes the need for a large dataset to optimize the proposed

network. The proposed method is quantitatively evaluated on a publicly available dataset. The results demonstrate that the proposed method improves performance in comparison to a moving mesh registration algorithm, and also it correlates well while is independent of manual segmentations which are created by any experts. Figure 3.1 displays a flowchart of the 2D registration method. Most deep learning-based image registration methods learn spatial transformations from training data with known deformation fields [58–61]. On the other hand, other methods estimate spatial transformations by using the pairs of images (fixed and moving images) [41, 62]. On the contrary, in the proposed method for each pair of images a low random dimensional vector (also known as a latent vector) is defined and it passes to the network as an input. Suppose $\{I_i\}_{i=1}^n$ is a sequence of images that we would like to register. We have a neural network with parameters θ that computes a deformation field $f_\theta(t_i) : R^d \rightarrow R^{N\times N}$, where N is the number of pixels and d is a number much smaller than N (in our work d is 25). We can call f_θ as an embedding function. Thus, for the i^{th} image in the sequence, the neural network takes in a d-dimensional vector t_i and outputs a deformation field $\phi(t_i)$. We can warp a moving image I_i by this deformation field to get the warped image $I(f_\theta(t_i))$. Therefore, we can minimize the following cost function for registering image sequence $\{I_i\}_{i=1}^n$:

$$E_{data}(\theta, \{t_i\}_{i=1}^n) = \sum_i |I_i - I_{mov(i)}(\phi(t_i))| \tag{3.1}$$

The minimization is performed jointly over the parameters θ of the neural network and latent vectors $\{t_i\}_{i=1}^n$, where I_i is a fixed image and the corresponding moving image is $I_{mov(i)}$.

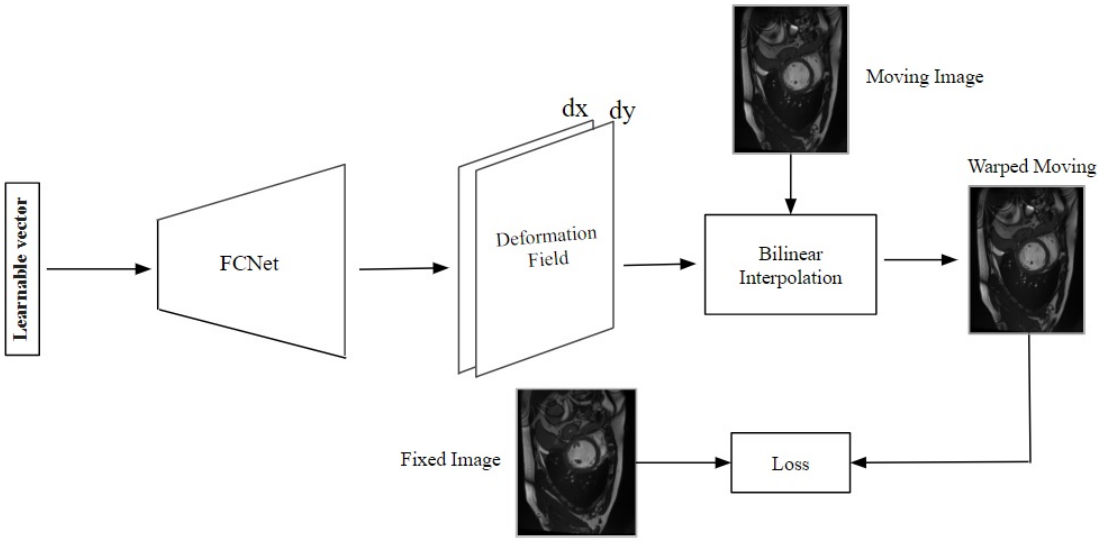


Figure 3.1: Schematic of the proposed image registration approach based on a fully connected network. The network uses one or more learnable vector which is initialized by a random vector as an input.

3.2 Methodology

Image registration is inherently ill-posed so that the existence and the uniqueness of the solution are not guaranteed [63]. Thus, regularization is essential to avoid both physically implausible displacement fields and local minimum during optimization [64]. As most optimization methods based on regularization image registration are typically computationally expensive and time-consuming, an alternative is to regularize network parameters θ. This regularization considers the mean of the sum of squares of the network weights (MSW):

$$MSW(\theta) = \frac{1}{N_w} \sum_{n=1}^{N_w} w_n^2, \tag{3.2}$$

where N_w represents the number of network weight parameters and w_n is an element of the parameter matrix in a vector expression W. The weights and biases of the network are initialized as random variables drawn from a Gaussian distribution. Finally, our optimization problem can be formulated as follows:

$$E(\theta, \{t_i\}_{i=1}^n) = E_{data}(\theta, \{t_i\}_{i=1}^n) + \lambda MSW(\theta), \tag{3.3}$$

The above optimization problem can be solved by backpropagation with stochastic gradient descent.

We propose a deformable registration framework using FCNet (fully connected network). In this method, in order to register each pair of images in a sequence, a low dimensional vector (also known as latent vector) is passed to the network as an input. One technique to find the latent vectors is using Auto-encoder (AE). The auto-encoder is made of two parts, encoder and decoder. The encoder brings the data from a high dimensional input to low dimensional output. Since using an auto-encoder network to find latent vector add time and computational cost to our framework, we initialized input vectors randomly and update it along with the parameters of the network. the FCNet applies 8 fully connected layers to generate a two dimensional deformation field dx, dy. The weights are initialized using a normal distribution with zero mean and standard deviation that is a function of the filter kernel dimensions.

Our model is implemented using Tensorflow. Adam optimization technique [65] is used with learning rate 1×10^{-4}, image batch size 10 and $\lambda = 0.1$. The results were obtained with NVIDIA GTX 1080 Ti GPU, and 2000 iterations were adopted for the optimization that takes 4–6 minutes per image. The optimization is stopped when the maximum iteration number is reached or the loss function is not improving after three iterations. A grid search algorithm is used to chose the number of layers and other hyper-parameters.

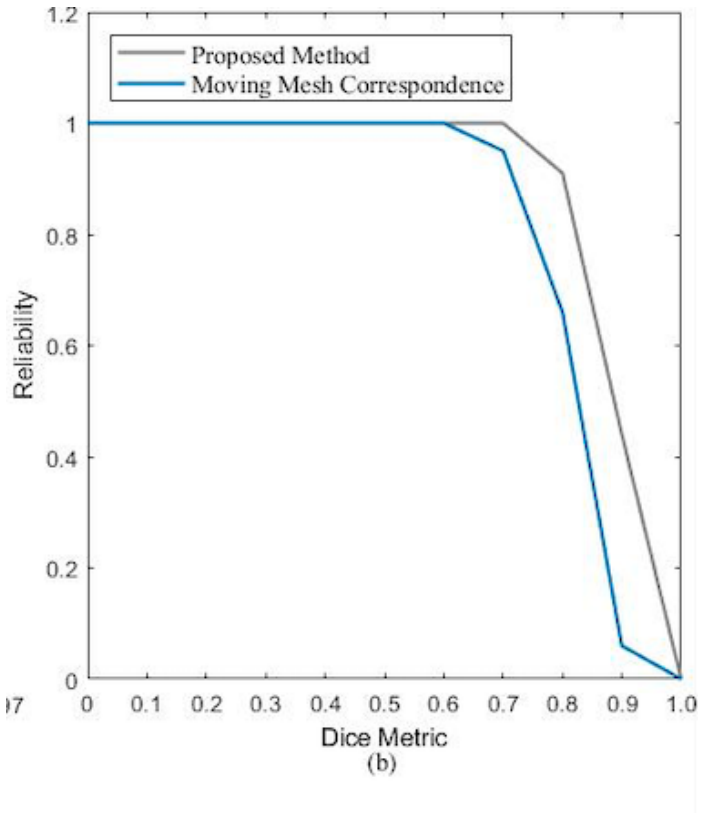


Figure 3.2: Comparison between proposed method and the methods by [30]. Reliability $R(d) = Pr(DM > d)$) for the proposed method and [30]. The proposed method led to a higher reliability curve.

.

Table 3.1: The mean, standard deviation of Dice score and Reliability function ($R(d) = P(DM > d)$). The higher the DM and R, the better the performance.

	Dice	$R(0.80)$	$R(0.85)$	$R(0.90)$
MM[30]	0.85 ± 0.03	0.95	0.66	0.06
our method	0.89 ± 0.03	1	0.91	0.44

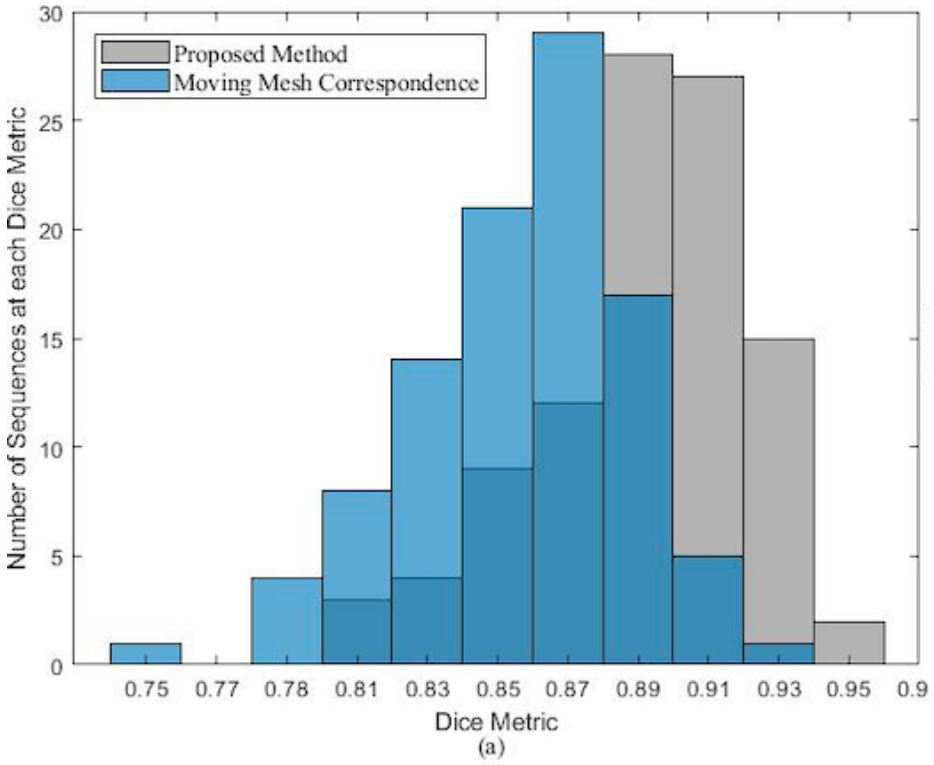


Figure 3.3: Comparison between proposed method and the methods by [30] based on Dice metric (DM).

.

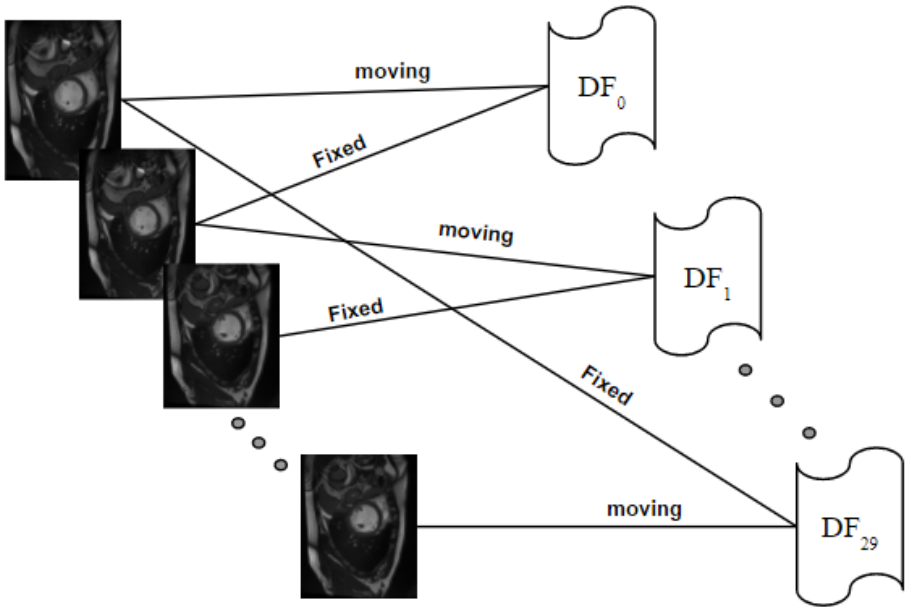


Figure 3.4: Selection of fixed and moving images in one sequence of the medical image. DF_i $(i = 0, 1, 2, ..., K)$ are deformation fields generated by the network for each pair of fixed and moving images.

3.3 Experiments

The proposed method is evaluated on Automated Cardiac Diagnosis Challenge (ACDC) dataset cardiac cine MR sequences 3.3.1 and $slice = 5$, a total of 30000 images. In each sequence I_i is selected as a moving image and I_{i+1} is selected as a fixed image Figure 3.4.

3.3.1 Data

Automated Cardiac Diagnosis Challenge (ACDC) dataset [66] which contains MRI scans including 100 stack of short-axis cardiac cine MRI sequences, each consisting of 12 to 35 temporal frames. The spatial resolution varies from 1.37 to 1.68 mm^2/pixel with a slice thickness of 5 mm to 8 mm (in general 5mm) and sometimes an inter-slice gap of 5mm. Each sequence consists of 28 to 40 images that cover the cardiac cycle

completely or partially.

3.3.2 Quantitative Evaluation

Table 3.1 shows the accuracy for FCNet in comparison with moving mesh correspondence method and it shows that our approach led to a significant improvement in average the accuracy. In addition, Table 3.1 reports the reliability of both methods in different accuracy levels and plot $R(d)$ as a function of d that is shown in Figure 3.2. Our algorithm led to a higher reliability curve and improvement in reliabilities.

3.3.3 Visual Assessment

By using a grid mesh, we show the displacement fields obtained by FCNet Figure 3.5. The comparison of the result of the FCNet and moving mesh correspondence method [30] are given in Figure 3.6. In Figure 3.7, we give a representative sample of borders obtained by our method. The FCNet approach yielded more accurate results than the moving mesh method.

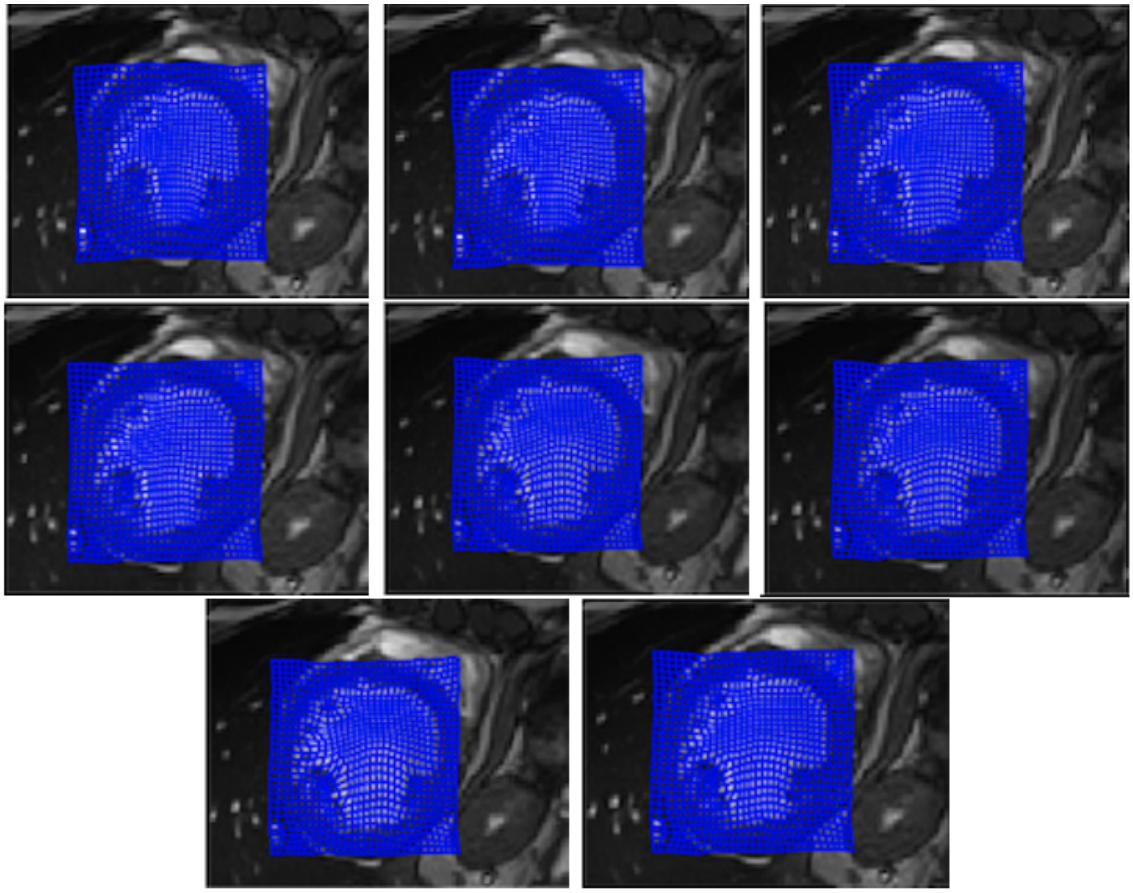


Figure 3.5: Representative examples of the displacement field obtained by the proposed method.

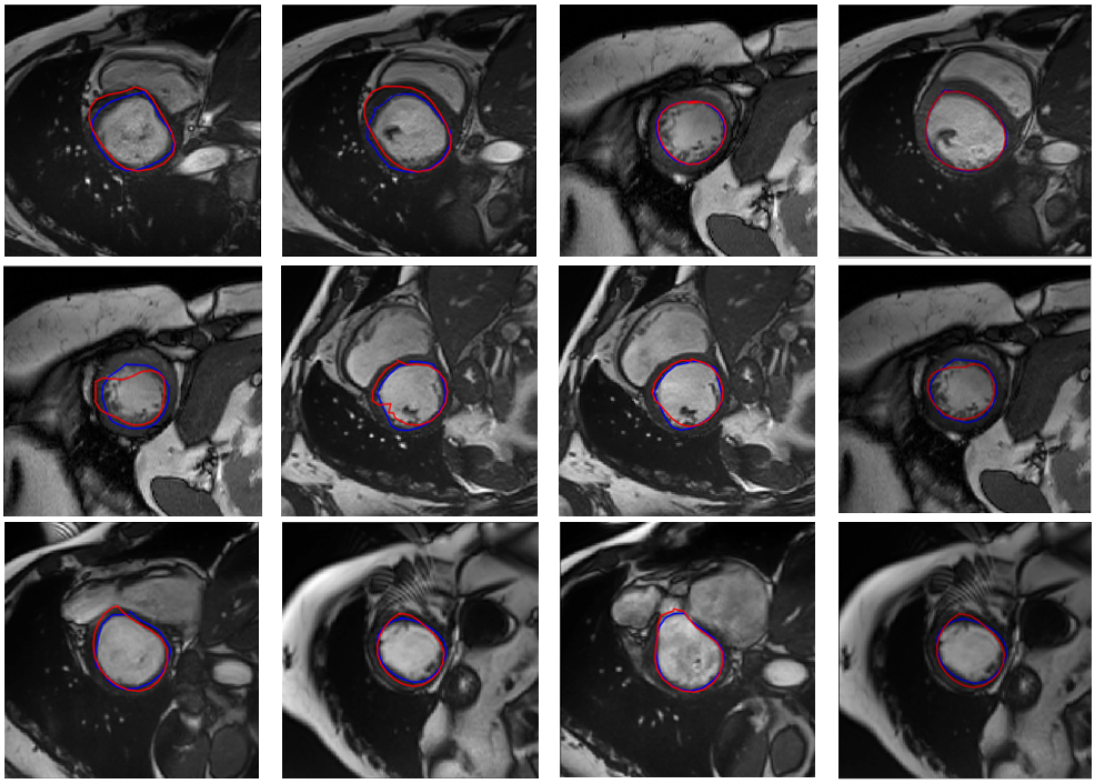


Figure 3.6: Representative examples of obtained borders of the LV with FCNet (blue) and moving mesh correspondence [30] (red) methods where FCNet provided significantly more accurate results than [30].

3.4 Conclusion

We proposed a deformable image registration algorithm based on the deep fully connected network to generate spatial transformations. Our network predicted the deformation field at the same resolution of fixed and moving images based on low dimensional learnable vectors initialized by random values, independent from moving and fixed images. The results have demonstrated that unsupervised deep learning models built upon generative fully connected networks can achieve satisfying performance for deformable medical image registration. This study shows that the proposed approach improves the performance over recent state-of-the-art image registration with respect to accuracy.

The drawback of our proposed method is that the fully connected network has lots of parameters to optimize, so that it is time-consuming.

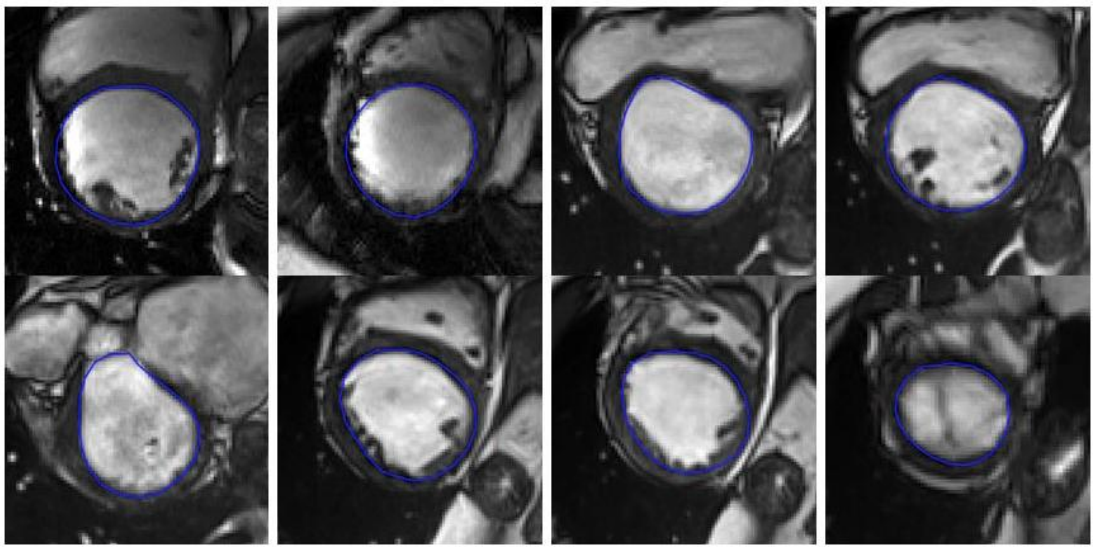


Figure 3.7: Sample of the boundary results with the proposed FCNet.

Chapter 4

GMCNet: A generative multi-resolution framework for cardiac registration:

4.1 Overview

Cardiac deformable registration is a fundamental step in quantitatively assessing cardiac function and generating diagnostic measurements. Recently, learning-based methods have yielded high performance for image registration by training on a large dataset and using a regularized optimization framework to minimize dissimilarities between an image pair. These methods have yielded well-suited approaches for image registration by training on a large dataset. The performance of learning-based methods is related to their ability to learn information from a large number of samples of medical images which is 1) difficult to obtain and 2) also makes the framework biased to the specific domain of data.

In this chapter, inspired by the chapter 3 we propose a novel end-to-end learning-free generative multi-resolution convolutional neural network (GMCNet) for image registration. Even though learning-based methods have yielded high performance for image registration, their performance depends on their ability to learn information from a large number of samples of medical images which is 1) difficult to obtain and 2) also makes the framework biased to the specific domain of data.

The proposed framework was quantitatively evaluated on clinical cardiac magnetic resonance images (MRI) over three different datasets and compared against the performance of nine state-of-the-art learning-based and optimization-based algorithms. To examine GMCNet from different aspects, we assess (1) robustness; (2) performance on pairwise registration; (3) influence of spatial transformation in a controlled environment; (4) impact of different multi-resolution structures.

The proposed learning-free method eliminates the need for a dedicated training set while exploiting the capabilities of neural networks to achieve accurate deformation fields. Due to its capability of parameter sharing through the architecture, the GMCNet can be used as a groupwise registration as well as pairwise registration. Additionally, inspired by the Demons algorithm Gaussian filters integrated with the GMCNet to impose a smoothness constraint which relaxes the need for an explicit regularization term and the corresponding weight in the cost function. The proposed framework yielded promising results on clinical cardiac magnetic resonance images over three different datasets when quantitatively compared against nine state-of-the-art learning and optimization-based algorithms. Different aspects of the GMCNet are explored by assessing 1) the robustness; 2) performance on pairwise registration; 3) the influence of spatial transformation in a controlled environment; and 4) the impact of different multi-resolution structures. The results provide confirmatory evidence that using temporal information from the 2D cardiac sequence to estimate the deformation field leads to registration with higher accuracy and also a highly robust performance under different noise levels. Moreover, since the proposed method is iterative and does not need a sample of images for training, its prediction is not domain-specific and can be applied to any sequence of images.

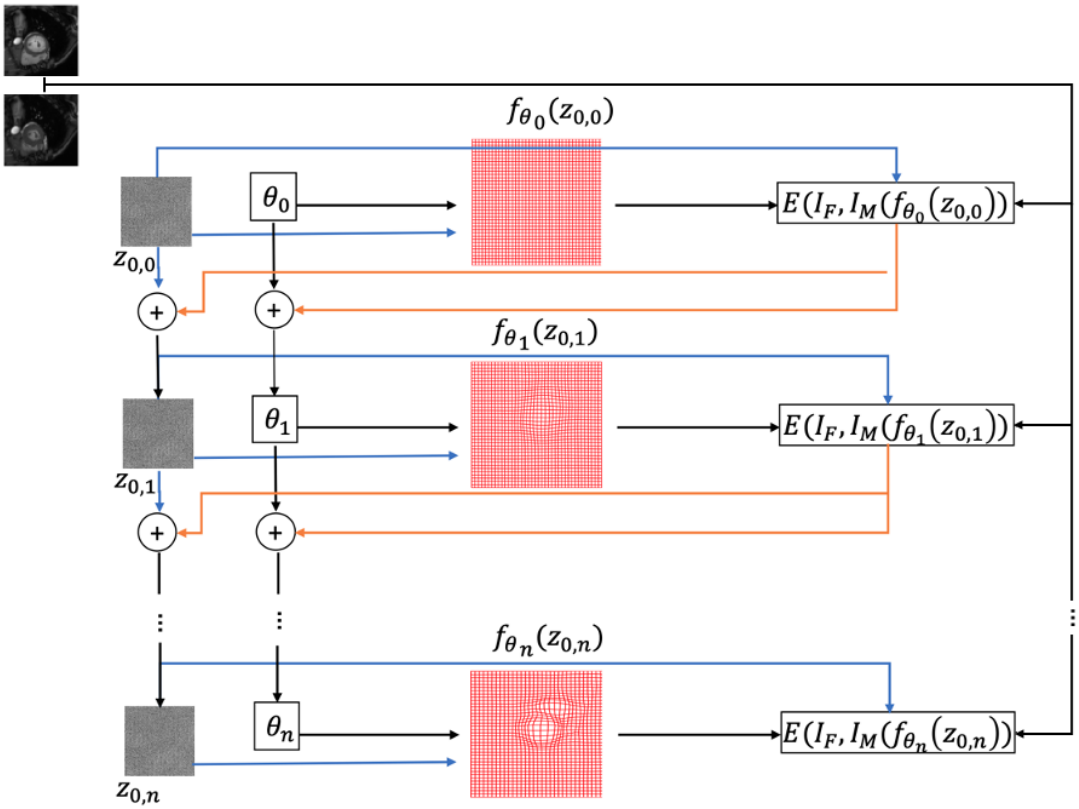


Figure 4.1: Iterations in the proposed GMCNet are shown for a pair of images for one resolution. The parameter set θ of the network f and latent variable z_0 for an image pair are iteratively updated through the back-propagation to minimize the registration cost (4.3) between the fixed image I_F and warped moving image I_M.

4.2 Methodology

The main goal of deformable image registration is to find the optimal spatial deformation ϕ_θ parameterized by θ that warps a moving image I_M to align it with a fixed image I_F. The optimal values of θ could be obtained by minimizing an objective function of the form:

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, L(I_F, I_M \circ \phi_\theta)) + \lambda R(\phi_\theta) \tag{4.1}$$

where L is a metric that measures the dissimilarity between I_F and the warped moving image $I_M \circ \phi_\theta$. R is a regularization term that imposes smoothness on spatial

deformation. λ is a weight coefficient corresponding to the regularization term. In unsupervised deep learning-based methods, a CNN is trained on a set of data to minimize a dissimilarity metric and a spatial transformer layer is used to warp the moving image I_M.

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{(I_F, I_M)} L(I_F, I_M \circ f_\theta(I_F, I_M)) + \lambda R(g_\theta(I_F, I_M)) \tag{4.2}$$

where f is the network and θ is the parameters of the network.

4.2.1 Generative Multi-resolution Convolutional Network (GMCNet)

Inspired by 3 for each pair of moving and fixed image a latent variable z is defined which is sampled from a random normal distribution and has the same dimensions as the input images $z \in \mathbb{R}^{H' \times W'}$. Where H and W are the height and width of images to be registered, respectively. The GMCNet maps the input latent variables to a desire deformation fields and uses the corresponding batch of fixed and moving images for the loss function. Thus, during the optimization process the network's parameters θ and z_i are optimized through back-propagation simultaneously.

So the new registration objective function can be formulated as follows:

$$\theta^*, z^* = \underset{\theta, z}{\operatorname{argmin}}\, L(I_F, I_M \circ f_\theta(z)), \tag{4.3}$$

where the θ^* and z^* are obtained using a neural network optimizer such as stochastic gradient descent. As it can be seen in the new formulation there is no need for an external regularization term and a corresponding weight. Fig. 4.1 shows the optimization of an input latent variable z and the parameters of the network θ during the iterative processing for a sample pair of images. For the sake of simplicity, the optimization process is showed for one resolution framework.

The proposed architecture implements the general principle of multiresolution framework, where computations follow coarse-to-fine resolutions[67]. It has three

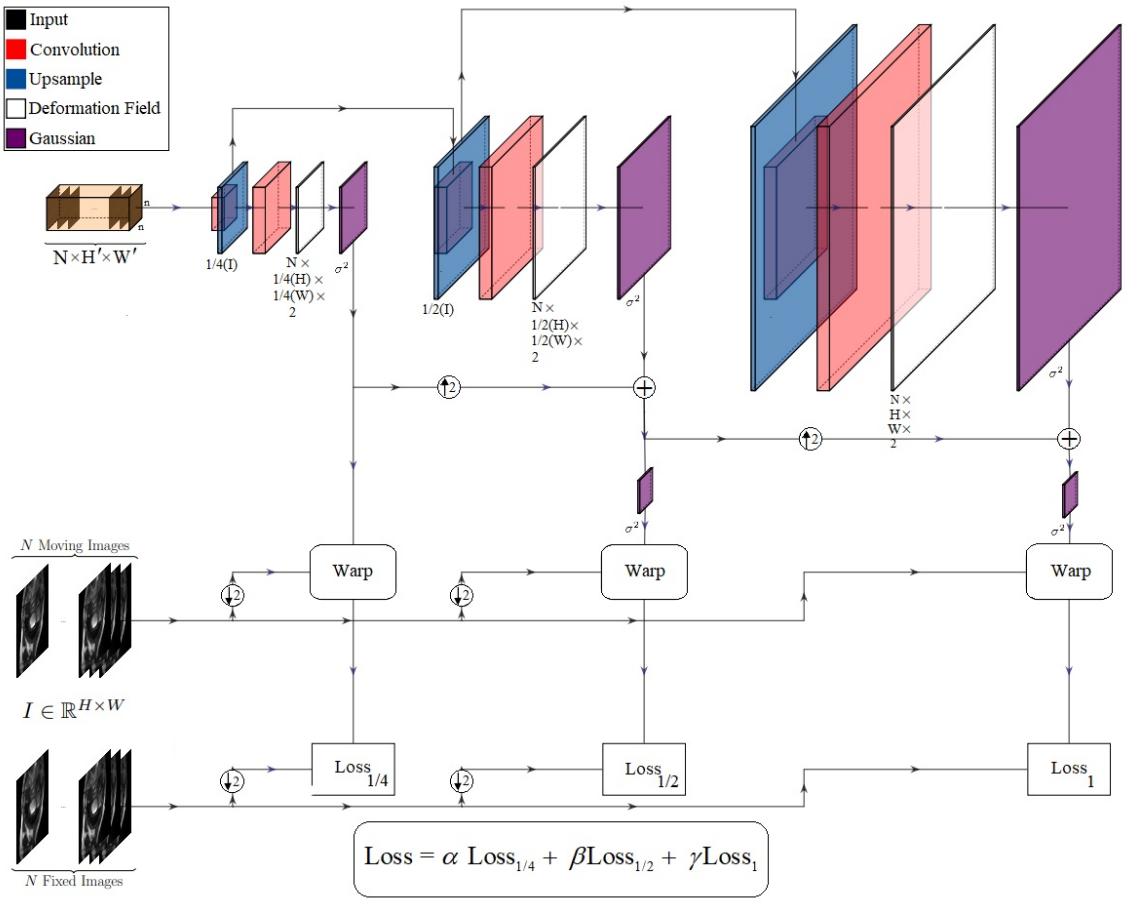


Figure 4.2: The proposed generative multiresolution convolutional network (GMCNet) architecture for deformable image registration. The network generates a deformation field based on low dimensional random inputs.

resolutions/scales denoted by a quarter $g_{\theta_1}^{1/4}$, a half $g_{\theta_1,\theta_2}^{1/2}$ and the original resolution, $g_{\theta_1,\theta_2,\theta_3}^{1}$. The $\theta_1, \theta_2, \theta_3$ are used to denote parameters of each resolution of GMCNet.

A spatial Gaussian smoothing kernel is integrated with the GMCNet to yield sufficiently smooth deformation fields. Such strategy has been adopted in the Demons algorithm, where unconstrained optimization is followed by Gaussian filtering to impose a smoothness constraint. So that, at each resolution a Gaussian kernel is applied on the deformation field and then the smooth scaled deformation field is upsampled

and added to the next scale. The components of the proposed GMCNet architecture are shown in Fig. 4.2.

4.2.2 Groupwise Registration

Because of parameters sharing property, the GMCNet can be used as a groupwise registration framework when we have a sequence of images to be registered $\{I_i\}_{i=1}^{N}$. In groupwise image registration methods, several images are registered to a common coordinate system and therefore, a group of transformations has to be computed instead of a single geometric transformation[55]. A common practice for groupwise registration is to select the first (or any other) image as the reference image, and then register the rest of the images to the reference using pairwise image registration. The choice of the reference image is important in such an approach. In the proposed

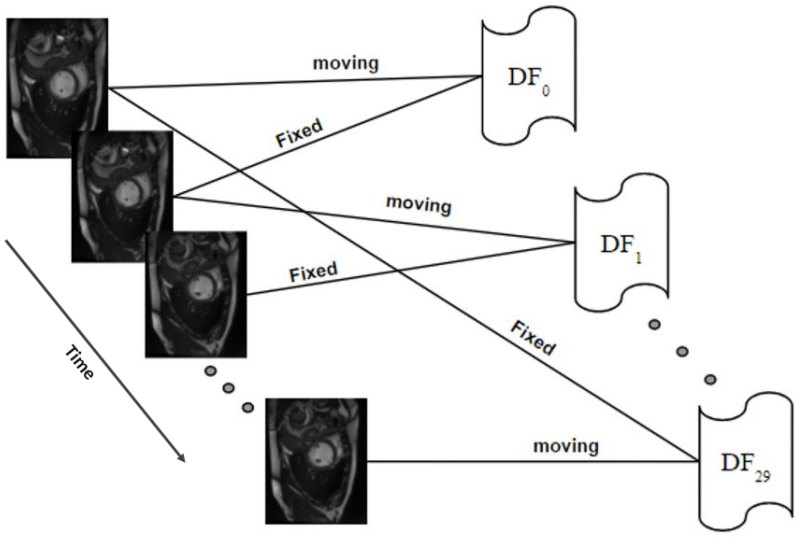


Figure 4.3: Illustration of choosing fixed and moving images in a sequence. The frame i is selected as a moving image and the next frame $i+1$ is selected as a fixed image.

method each image in a cardiac sequence is registered to the next image in the sequence. To register a set of images $\{I_i\}_{i=1}^N$, the i^{th} image in a sequence is set as a moving image I_M^i and $i+1^{th}$ image is set as a fixed image I_F^i, for $i = 1, ..., N-1$ as shown in Fig. 4.3. Such a formulation eliminates any potential bias introduced by choosing a particular reference image. In our formulation, the weights θ are shared through the entire sequence. Such consistency may not be obtained using traditional pairwise approaches where the registration is performed by splitting the sequence into a set of image pairs. For a set of N image pairs $\{I_F^i, I_M^i\}_{i=1}^N$, N latent variables are defined $\{z_i\}_{i=1}^N$. Where N is the number of images in a sequence.

In the case of groupwise registration (4.3) is changed as follows:

$$\theta^*, \{z_i^*\}_{i=1}^N = \underset{\theta, \{z_i\}_{i=1}^N}{\operatorname{argmin}} \sum_{i=1}^{N} L(I_F^i, I_M^i(g_\theta(z_i))) \tag{4.4}$$

4.2.3 Loss Function and Optimization

In each resolution a separate loss is defined to measure the dissimilarities between the fixed and warped moving images at that resolution as follow:

The loss function for the coarsest resolution:

$$Loss_{1/4} = \sum_{i=1}^{N} L(I_F^{1/4,i}, I_M^{1/4,i} \circ \phi_{\theta_1}^{1/4}(z_i)) \tag{4.5}$$

where $I_F^{1/4,i}$ and $I_M^{1/4,i}$ are respectively down-sampled fixed and moving images at the quarter resolution.

The loss function for the second (half) resolution:

$$Loss_{1/2} = \sum_{i=1}^{N} L(I_F^{1/2,i}, I_M^{1/2,i} \circ \phi_{\theta_1,\theta_2}^{1/2}(z_i)) \tag{4.6}$$

Similarly, the loss function for the full resolution is defined as:

$$Loss_1 = \sum_{i=1}^{N} L(I_F^{1,i}, I_M^{1,i} \circ \phi_{\theta_1,\theta_2,\theta_3}^{1}(z_i))) \tag{4.7}$$

The final loss function is a linear combination of loss functions in three resolutions:

$$Loss = \alpha Loss_{1/4} + \beta Loss_{1/2} + \gamma Loss_1, \tag{4.8}$$

We optimize the whole network as one stage to reduce the computational time. All the parameters of the network $\theta_{1,2,3}$ are initialized only once and they are shared for all images in the sequence. It is possible to optimize the network in three phases so that the first stage optimizes it with respect to θ_1 and $\{z_i\}_{i=1}^N$, the second stage optimizes θ_1, θ_2 and $\{z_i\}_{i=1}^N$ and finally, the third stage optimizes loss (4.8) with regard to all parameters θ_1, θ_2, θ_3 and $\{z_i\}_{i=1}^N$.

4.2.4 Learning-free framework

Since the propose method is learning-free there is no need to obtain a training set. It means the GMCNet is not trained on a given data set. On this basis, to register any sequence/pair of images the z and the parameters of the network θ are optimized iteratively from scratch. The optimization is terminated based on two criteria, the step size and iteration number. In each iteration, loss function is evaluated, and if it does not improve, the step size will be reduced. The process is terminated once the maximum number of iterations has been reached, or the step size is below the defined threshold.

4.2.5 Implementation details

The proposed GMCNet consists of six convolutional layers with strides $(1, 1, 1, 1)$ and three upsample layers. The inputs fit in a convolutional Gaussian layer with $\sigma = 2$ and kernel size 7. Each scale contains two convolutional layers, two upsample layers and a convolutional Gaussian layer with $\sigma = 3$ and kernel size 15 in front of the estimated deformation field and after adding a correction. Exponential linear unit (ELU) is selected as an activation function and NCC is used as a loss function. The values of α, β, and γ are chosen as $0.5, 0.25$, and 0.25 respectively. We initialized a learning rate of 5×10^{-4} with the Adam optimizer and a batch size of 10. The maximum iteration number for is 800 and the minimum threshold of the step size is 0.5×10^{-7}. The framework has been implemented using Python programming language with the

Tensorflow machine learning module and it is tested with an NVIDIA GTX 1080 Ti graphics processing unit with 11GB memory. The weights are initialized using a normal distribution with zero mean and standard deviation that is a function of the filter kernel dimensions. The optimization is stopped when the maximum iteration number is reached or the loss function is not improving after three iterations. A grid search algorithm is used to chose the number of layers and other hyper-parameters.

4.3 Experiments

The proposed framework is evaluated on clinical cardiac MRI images over three datasets.

4.3.1 Datasets

The following three datasets are considered in this study:

Automated Cardiac Diagnosis Challenge (ACDC)[66] This dataset contains multiple temporal 2D short-axis cardiac cine MRI sequences acquired from 100 patients and one of the publicly available datasets for cardiac MRI assessment. The spatial resolution varies from 1.37 to 1.68 mm^2/pixel with a slice thickness of 5 mm to 8 mm (in general 5mm) and sometimes an inter-slice gap of 5mm. Each sequence consists of 28 to 40 images that cover the cardiac cycle completely or partially.

The Sunnybrook Cardiac Challenge data (SCD)[68] This dataset contains multiple temporal 2D short-axis cardiac cine MRI scans acquired from 45 patients. Each cine sequence includes 20 frames to cover the cardiac cycle. The image resolution is 256×256, with a pixel spacing of 1.25 mm and slice thickness of 8 mm.

Left Atrium (LA) This dataset includes 100 temporal 2D long-axis cine MRI steady-state sequences from the 2, 3 and 4-chamber views. It was acquired from the University Alberta Hospital. Each cycle includes 25 or 30 frames with image

resolutions 176×189 – 256×208 and image spacing $1.445 - 1.795$ mm. The ground truth manual segmentation was initially performed by a medical student and edited by an experienced radiologist. The 2ch, 3ch and 4ch are used in the rest of the paper to denote 2, 3 and 4-chamber sequences, respectively.

4.3.2 Evaluation

In this section, we evaluate and compare the performance of the proposed framework with both optimization-based state-of-the-art algorithms, SimpleElastix (Elastix)[69], Moving Mesh (MM)[30], Real-Time Image-based Tracker (RRT)[70], Fast Symmetric Forces Demons (Demons)[71], LCC-Demons[32], Symmetric Normalization[27] and also learning-based state-of-the-art algorithms VoxelMorph (VM)[45], learning probability model for diffeomorphic registration (LPM)[7] and DIRNet[41], a CNN-based, end-to-end unsupervised deformable image registration. We denote GMCNet as GMCNet_s and GMCNet_p to identify the sequential registration and pairwise registration.

Quantitative Evaluation Metrics

The proposed method was evaluated quantitatively using four metrics, namely, Dice metric (DM), Hausdorff distance (HD in mm), determinant of Jacobian of the deformation field $\det(J)$, and reliability $R(d)$.

Dice Metric The DM [73] is a well-known segmentation based metric to measure the similarity (overlap) between two regions, warped moving and fixed image. The DM of two regions A and B is formulated as:

$$DM(A, B) = \frac{2|A \cap B|}{A + B} \tag{4.9}$$

Hausdorff Distance The HD [74] is another well-known metric which measures the maximum deviation between two regions' contours. The HD between two contours

Table 4.1: Quantitative evaluation of the results for cardiac MRI registration on the LA dataset. The evaluation is performed in terms of Dice (mean± standard deviation) and HD (mean). The 2ch, 3ch and 4ch stand for the 2, 3 and 4-chamber. Values in bold indicate the best performance. Undef stands for Undeformed.

Methods	Dice			HD		
	2ch	3ch	4ch	2ch	3ch	4ch
Undef	0.79 ± 0.07	0.78 ± 0.08	0.78 ± 0.09	7.37	7.70	8.66
FCG[72]	0.81 ± 0.11	0.84 ± 0.13	0.80 ± 0.10	7.5	7.67	8.2
RRT[70]	0.81 ± 0.13	0.85 ± 0.08	0.80 ± 0.11	7.3	6.99	8.01
Elastix[69]	0.82 ± 0.11	0.86 ± 0.10	0.82 ± 0.10	7.28	6.82	7.56
MM[30]	0.84 ± 0.06	0.83 ± 0.06	0.83 ± 0.08	6.58	6.48	6.77
Demons[71]	0.84 ± 0.08	0.85 ± 0.06	0.82 ± 0.10	7.41	7.33	7.84
SyN[27]	0.87 ± 0.06	0.86 ± 0.13	0.84 ± 0.11	6.92	7.52	7.51
GMCNet_p	0.81 ± 0.12	0.84 ± 0.09	0.81 ± 0.11	7.1	5.1	6.3
GMCNet_s	**0.88 ± 0.05**	**0.87 ± 0.05**	**0.84 ± 0.07**	**6.2**	**4.9**	**6.2**

(C_A) and C_B is formulated as:

$$\begin{aligned} \mathrm{HD}(C_A, C_B) = &\max(\max_i(\min_j(d(p_A^i, p_B^j))), \\ &\max_j(\min_i(d(p_A^i, p_B^j)))) \end{aligned} \tag{4.10}$$

where p_A^i, p_B^j denote the set of all the points in C_A and C_B respectively. The term $d(\cdot)$ denotes the Euclidean distance. The DM and HD were computed by comparing the delineations obtained using the registration methods with expert manual contours. Table 4.1, 4.2 and 4.3 show the mean and standard deviations of DM and HD for all algorithms evaluated on LA, ACDC and SCD datasets, respectively. The reported values for methods indicated with $*$ were taken from previous publications [7, 41]. It should be noted that the presented method (GMCNet approach) has a good performance on all mentioned datasets (DM=$0.85 - 0.87$) and the reported values in

Table 4.2: Quantitative evaluation of the results for cardiac MRI registration on the ACDC dataset. The evaluation is performed over the left ventricle in terms of Dice (mean± standard deviation) and HD (mean). Values in bold indicate the best performance.

Method	Dice	HD
Undeformed	0.71 ± 0.145	10.1
LCC-Demon*[32]	0.79 ± 0.096	9.21
VM*[45]	0.79 ± 0.096	8.46
LPM S1*[7]	0.79 ± 0.091	7.58
FCG[72]	0.80 ± 0.150	8.80
SyN*[27]	0.80 ± 0.091	8.24
LPM S3*[7]	0.81 ± 0.085	6.88
RRT[70]	0.83 ± 0.161	5.75
MM[30]	0.83 ± 0.153	5.64
Elastix[69]	0.84 ± 0.162	5.51
GMCNet_p	0.84 ± 0.170	7.43
GMCNet_s	**0.86 ± 0.141**	**5.35**

* Results are reported in[7] in which standard deviation could not be estimated. N/A refers to the value that are not reported in the original paper.

Table 4.1, 4.2 and 4.3 demonstrate that our approach outperforms other registration methods in terms of both DM and HD.

Reliability: We also evaluated the performance of the proposed algorithm using a reliability function computed based on DMs for each dataset. The complementary cumulative distribution function is defined for each $d \in [0, 1]$ as the probability of

Table 4.3: Quantitative evaluation of the results for cardiac MRI registration on the SCD dataset. The evaluation is performed in terms of Dice (mean± standard deviation) and HD (mean). Values in bold indicate the best performance.

Method	Dice	HD
Undeformed	0.62 ± 0.15	16.02
RRT[70]	0.71 ± 0.19	13.10
FCG[72]	0.71 ± 0.2	12.82
Demons[71]	0.71 ± 0.18	12.46
MM[30]	0.72 ± 0.12	12.53
Elastix[69]	0.79 ± 0.08	11.12
DIRNet*[41]	0.80 ± 0.08	N/A
SyN[27]	0.81 ± 0.16	8.9
GMCNet_p	0.71 ± 0.16	7.95
GMCNet_s	**0.87 ± 0.09**	**4.58**

* Results are reported in[41]. N/A refers to the value that are not reported in the original paper.

obtaining DM higher than d over all volumes.

$$\begin{aligned} R(d) &= P_r(Dice > d) \\ &= \frac{\text{\# Images segmented with DM higher than d}}{\text{total number of images}}. \end{aligned} \tag{4.11}$$

$R(d)$ measures how reliable the algorithm is in yielding accuracy d. The corresponding reliability $R(d)$ is plotted as a function of Dice in Fig. 4.4. The reliability values at $d = 0.80, 0.85$, 0.90 and 0.95 are reported in Table 4.4, 4.5, and 4.6 for ACDC, SCD and LA datasets, respectively. Our algorithm led to a higher reliability curve on all ACDC, SCD, and LA datasets. We have 2%, 17% and 2% improvement in $R(0.80)$, respectively on ACDC, SCD and LA. For instance, on ACDC, we obtained $R(0.85) = 0.65$, i.e, an excellent agreement ($DM > 0.85$) in 65% of the cases, whereas the Elastix[69] achieved 63% of the cases with a similar accuracy. Also, we obtained

Table 4.4: Reliability function ($R(d) = Pr(DM > d)$) for ACDC dataset. The higher the R, the better the performance.

Method	R(0.80)	R(0.85)	R(0.90)	R(0.95)
RRT[70]	0.73	0.60	0.40	0.13
MM[30]	0.73	0.61	0.42	0.16
Elastix[69]	0.74	0.63	0.45	0.20
GMCNet_s	**0.74**	**0.65**	**0.48**	**0.23**

Table 4.5: Reliability function ($R(d) = Pr(DM > d)$) for SCD. The higher the R, the better the performance.

Method	R(0.80)	R(0.85)	R(0.90)	R(0.95)
RRT[70]	0.45	0.31	0.16	0.04
Demons[71]	0.22	0.16	0.08	0.02
MM[30]	0.46	0.30	0.20	0.09
Elastix[69]	0.46	0.33	0.19	0.04
SyN[27]	0.55	0.48	0.20	0.09
GMCNet_s	**0.75**	**0.61**	**0.37**	**0.13**

$R(0.85) = 0.60$ on SCD, whereas the SyN[27] method, which has higher accuracy among other methods, achieved 43%.

det(J): To analyze deformation regularity in different algorithms, we show the determinant of the Jacobian $\det(J)$[75]. If the value of $\det(J)$ equals to 1, the area remains constant after the transformation, whereas the value smaller or larger than 1 indicates the local area shrinkage or expansion, respectively. The negative values of $\det(J)$ imply that local folding and twisting have occurred, which is physically not realizable and mathematically not invertible[45]. The minimum of $\det(J)$ of deformation for each pair of images is reported. We use the $\det(J)$ computed by SimpleITK[76–78] to quantify deformation regularity. Fig. 4.5 reports the minimum $\det(J)$ values observed in each method. No negative minimum values of $\det(J)$ were observed for test cases which indicates that the proposed method does not lead to local mesh folding or twisting. This implies that the estimated deformations are

Table 4.6: Reliability function ($R(d) = Pr(DM > d)$) for LA. The higher the R, the better the performance.

Method	R(0.80)	R(0.85)	R(0.90)	R(0.95)
LA-2ch				
RRT[70]	0.64	0.44	0.26	0.06
Demons[71]	0.75	0.51	0.24	0.07
MM[30]	0.78	0.50	0.17	0.04
Elastix[69]	0.60	0.44	0.22	0.08
SyN[27]	0.88	0.72	0.44	0.14
GMCNet_s	**0.90**	**0.74**	**0.45**	**0.18**
LA-3ch				
RRT[70]	0.81	0.62	0.37	0.12
Demons[71]	0.82	0.54	0.26	0.04
MM[30]	0.68	0.44	0.14	0.03
Elastix[69]	0.75	0.59	0.36	0.08
SyN[27]	0.87	0.66	**0.41**	0.11
GMCNet_s	**0.87**	**0.68**	0.40	**0.13**
LA-4ch				
RRT[70]	0.56	0.33	0.15	0.03
Demons[71]	0.59	0.39	0.26	0.09
MM[30]	0.45	0.31	0.16	0.04
Elastix[69]	0.50	0.28	0.13	0.02
SyN[27]	0.79	0.58	**0.35**	0.09
GMCNet_s	**0.79**	**0.59**	0.34	**0.10**

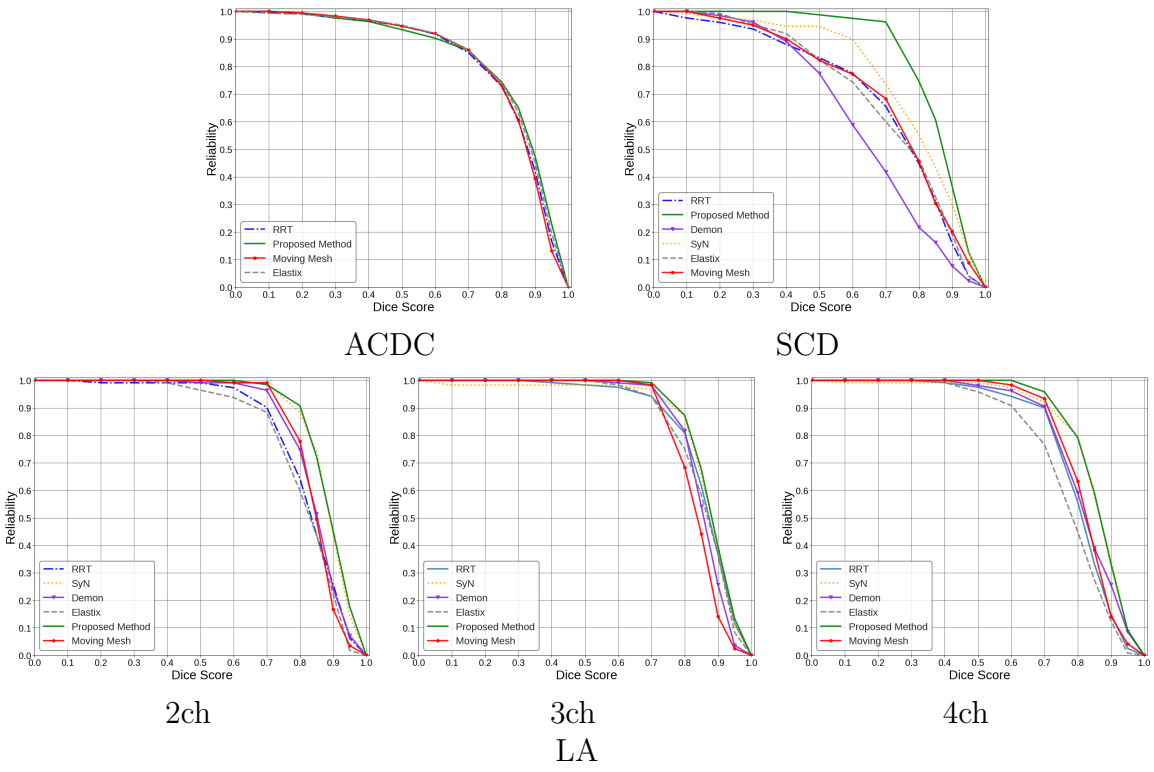
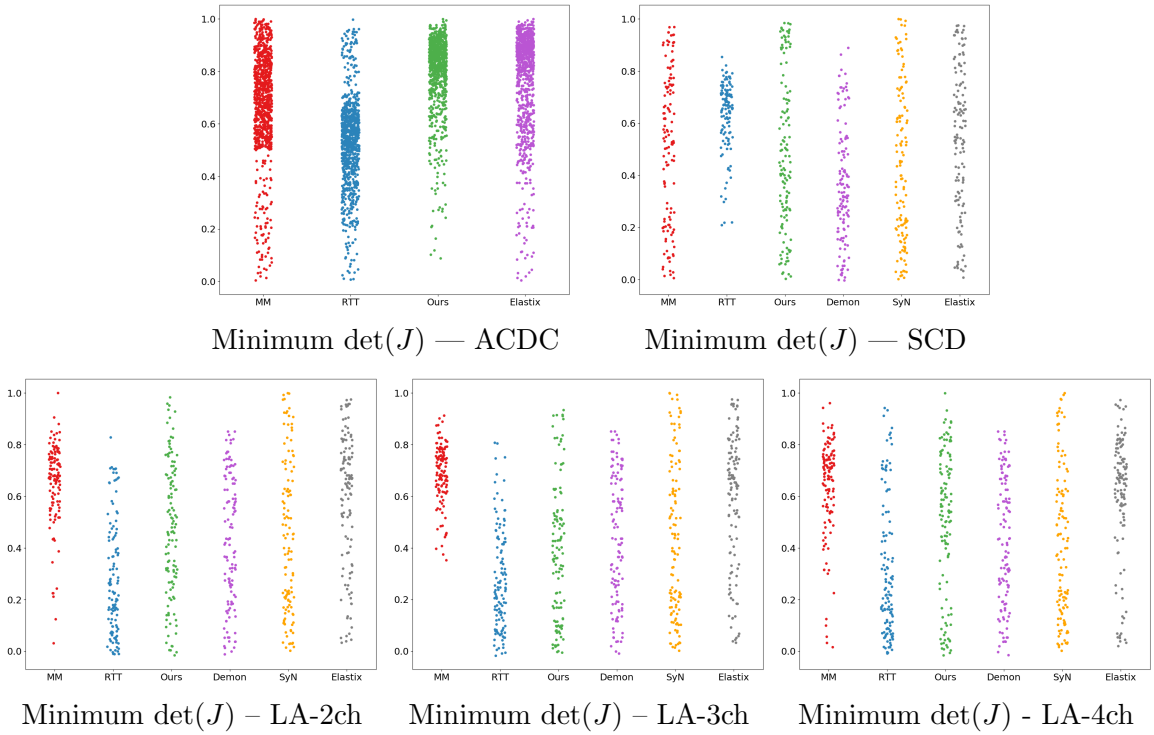


Figure 4.4: Reliability versus Dice metric of different algorithms: MM[30], RRT[70], Demons[45], SyN[27], SimpleElastix (Elastix)[69] and the proposed algorithm. The figures omit results for methods with no public implementation available.

physically realizable. Sample of warped moving images and corresponding Jacobian determinant with grid overlay is shown in Fig. 4.8.

Impact of Multi-resolution

We assess the impact of using different multi-resolution structures with GMCNet on the performance, and reported the corresponding evaluations in terms of DM and HD in Table 5.5. The results indicate that the high performance was obtained with the use of three resolutions.



Minimum det(J) — ACDC Minimum det(J) — SCD

Minimum det(J) – LA-2ch Minimum det(J) – LA-3ch Minimum det(J) - LA-4ch

Figure 4.5: Minimum of determinant of Jacobian results for different algorithms: MM[30], RRT[70], Demons[45], SyN[27], SimpleElastix (Elastix)[69] and the proposed algorithm. The figures omit results for methods with no public implementation available.

Assessment of robustness against different noise levels

To assess the robustness of proposed algorithm against different noise levels, we selected a random sequence from SCD and applied the same level of noise (speckle noise with $\mu = 1$ and $\sigma_s = [0, 0.5, \ldots, 4]$) to every frame in the sequence except for two randomly selected frames which were corrupted with twice as much noise level than other frames (standard deviation = 2σ). The DM of each algorithm was measured and plotted in Fig. 4.6. The results show that our method is almost independent from the noise level and at all noise level, we have the highest DM. This happens because the proposed method uses shared weights for all frames in a sequence, which capture

Table 4.7: Q
uantitative cardiac MRI registration results on the ACDC, SCD and LA based on different number of resolution of GMCNet.]Quantitative cardiac MRI registration results on the ACDC, SCD and LA based on different number of resolution of GMCNet. The evaluation is performed in terms of Dice (mean± standard deviation) and HD (mean). The 2ch, 3ch and 4ch stand for the 2, 3 and 4-chamber. Values in bold indicate the best performance.

	Dataset	Dice	HD
One Resolution	ACDC	0.79 ± 0.15	8.01
	SCD	0.73 ± 0.09	9.43
	LA-2ch	0.80 ± 0.11	7.42
	LA-3ch	0.82 ± 0.08	7.27
	LA-4ch	0.79 ± 0.10	7.53
Two Resolutions	ACDC	0.81 ± 0.13	6.81
	SCD	0.77 ± 0.11	7.31
	LA-2ch	0.81 ± 0.09	7.11
	LA-3ch	0.84 ± 0.10	6.98
	LA-4ch	0.80 ± 0.09	7.01
Three Resolutions	ACDC	**0.86 ± 0.14**	**5.35**
	SCD	**0.87 ± 0.09**	**4.58**
	LA-2ch	**0.88 ± 0.06**	**6.2**
	LA-3ch	**0.87 ± 0.05**	**4.9**
	LA-4ch	**0.84 ± 0.07**	**6.2**
Four Resolutions	ACDC	0.75 ± 0.18	8.45
	SCD	0.85 ± 0.08	5.83
	LA-2ch	0.82 ± 0.09	7.3
	LA-3ch	0.85 ± 0.08	7.03
	LA-4ch	0.82 ± 0.10	6.83

the correlation between frames. This, per se, leads to a more robust performance under different levels of noise than non-learning based methods. Also, the drop rate

for the other methods is higher than the proposed GMCNet method.

Evaluation against ground truth deformation by thin-plate spline

To analyze the spatial transformation in a controlled environment, we used the thin-plate spline algorithm to generate a true deformation field[79]. The algorithm requires an initial contour at end-diastole which was obtained by randomly selecting patients from ACDC. First, the ground truth contour was obtained for the end-diastolic frame. Then, by using a diffeomorphic registration method a set of control points were generated. The contours include 20 equally spaced points which were subsampled from the generated sequence of control points. The thin-plate spline method was then used to produce a smooth interpolation between these sets of points. We defined the least bent surface that fits through the control points as follows:

$$f(x,y) = a_1 + a_2x + a_3y + \sum_{i=1}^{n}(w_iU(|P_i - (x,y)|), \tag{4.12}$$

where the first three terms $a_1 + a_2x + a_3y$ defines the best fitted plane through the control points and the $\sum_{i=1}^{n}(w_iU(|P_i - (x,y)|)$ term is correspondence to the bending forces provided by n control points.

The deformation fields generated by the thin plate spline approach were used to verify and compare the performance of different parameterization approaches: RRT[70], Demons[71] and MM[30], Elastix[69], SyN[27], and the proposed GMCNet method.

The root mean squared error (RMSE) was calculated between the deformation fields for both directions (x, y) and summed. The RMSE is defined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2} \tag{4.13}$$

where n is the total number of pixels in the deformation field and (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ denote the true and estimated deformed points, respectively.

Fig. 4.7 displays the performance of the various registration algorithms compared to the ground truth thin plate spline method. These algorithms were evaluated by the

difference in the deformation grids, RMSE, where images were corrupted with different degrees of noise. It can be seen that when there is no added noise, most of the methods yield similar RMSE in the range of $0.52-0.53$. However, the algorithm performance is highly dependent on the noise level and even small increase in the value of noise would highly increase the RMSE. On the contrary, the presented method demonstrates less sensitive to noise levels with the standard deviation values of $0.0-1.5$. Even for the greater noise levels with standard deviation values of $1.5-3.5$, our method still yields the least RMSE.

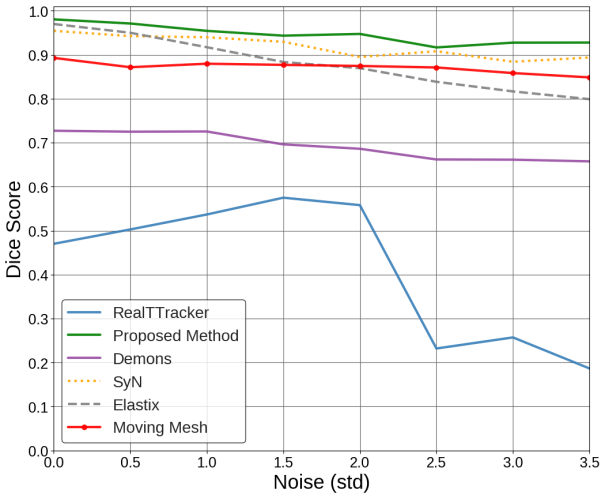


Figure 4.6: Dice scores (y-axis) for a randomly selected sequence from SCD corrupted with speckle noise with increasing standard deviations (x-axis). See details in the text.

Pairwise registration

To assess the effects of groupwise registration, we also applied our method as pairwise registration on images acquired at end-systolic and end-diastolic phases which has large deformation compares to neighbour frames in a sequence that has small

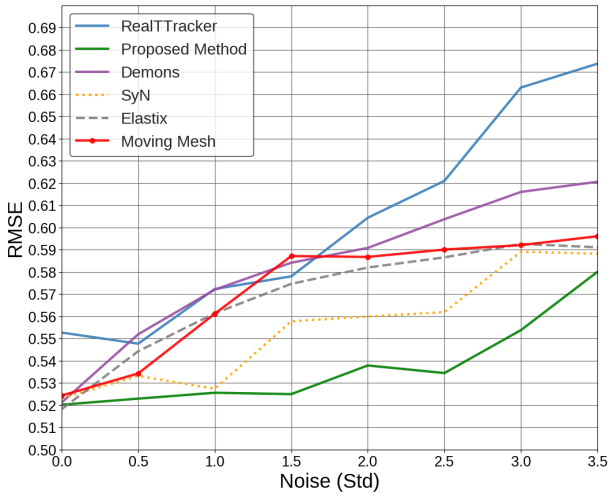


Figure 4.7: Root mean square error (RMSE) for quantitative evaluation of the accuracy between the ground truth deformation field and the six registration methods. The lower the values of RMSE, the more accurate the registration.

difference. Tables 4.1, 4.2 and 4.3 show the mean and standard deviations of DM and HD for pairwise GMCNet_p, which is evaluated on ACDC, SCD and LA datasets, respectively. It can be seen from these tables that the proposed sequential GMCNet_s outperforms GMCNet_p in each tested case in terms of DM and HD.

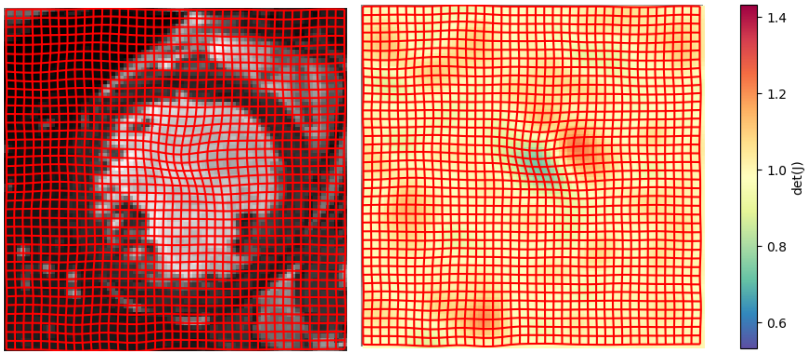


Figure 4.8: Registration results by the proposed GMCNet method showing a warped moving image with grid overlay and determinant of Jacobian.

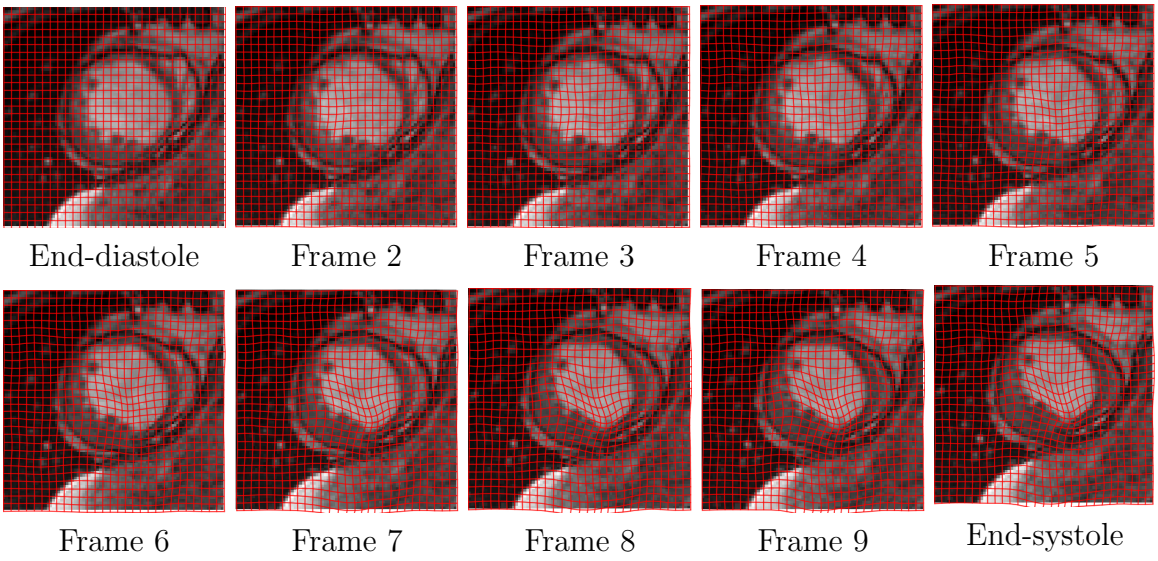


Figure 4.9: An example showing the predicted deformation results over the systolic phase of the cardiac cycle. It start from End-diastole (Frame 1) to End-systole which is 10^{th} in this sequence. The grid deformation over the sequence shows the impact of the application of the smoothness. Each frame is registered to the next frame with a smooth transformation field computed using the proposed method.

4.4 Discussion

In this study, we proposed a CNN based registration approach to obtain accurate results by exploiting temporal information from 2D image sequences. Registering a sequence of images plays an important role in many applications including cardiac functional assessment for MRI sequences. For instance, the left ventricle dysfunction is a significant condition for adults and often require the assessment of the regional function[80]. A point-to-point registration approach could be used for detecting regional left ventricular function abnormality[81]. In addition to the functional assessment of left ventricle from short-axis MRI sequences, there are several other applications that could benefit from image registration applied to 2D sequences. This includes functional assessment of left ventricle from short-axis MRI sequences[81], the functional assessments of left and right atria from long-axis MRI sequences[82, 83].

An example of the prediction results of the GMCNet method can be visualized in

Fig. 4.9 where the images were acquired from a patient's MRI sequence. This example shows that the proposed method leads to deformation fields with no local folding or twisting. We evaluated the approach on cine-MRI registration and compared registration performance in terms of DICE and Hausdorff distances to nine popular learning and non-learning based algorithms[32],[69], [7], [27],[30],[41],[45],[70],[71], and our conference version method[72]. The performance of our approach showed significant improvements in terms of registration accuracy.

The proposed method is applied as a sequential registration as well as pairwise registration for the datasets tested in this study. In sequential registration, all images were processed together, and therefore, we computed the equivalent time for pairwise registration by dividing the total time with the number of image pairs. In this case, the equivalent mean computational times were around 13, 10, and 17 seconds per image pair for ACDC, SCD, and LA datasets, respectively. In the case of pairwise registration, the mean of computational times were around 25, 30, and 47 seconds for ACDC, SCD and LA datasets, respectively. Thus the proposed method is suitable for medical applications that are not time-sensitive and the priority is given to robustness and accuracy.

4.5 Conclusion

We proposed a learning-free fully automated approach using the structure of an untrained generative multi-resolution convolutional neural network for deformable medical image registration. Optimizing the latent variables during the registration eliminates the need for regularization and tuning. The proposed method has yielded promising results on the cardiac MRI images in comparison to the learning and non-learning based methods. Moreover, the proposed method's prediction is not domain-specific and could be applied to any medical image sequence without resorting for annotated training data. Our algorithm captures the correlation between frames in a sequence which leads to a more robust performance under different levels of noise.

Currently, the presented methodology is applied to temporal 2D registration problem. In the future, we will extend the methodology to temporal 3D sequences.

Chapter 5

Recursive multiresolution framework for diffeomorphic deformable image registration

5.1 Overview

Diffeomorphic deformable image registration is one of the crucial tasks in medical image analysis, which aims to find a unique transformation while preserving the topology and invertibility of the transformation. Deep convolutional neural networks (CNNs) have yielded well-suited approaches for image registration by learning the transformation priors from a large dataset. The improvement in the performance of these methods is related to their ability to learn information from several sample medical images that are difficult to obtain and bias the framework to the specific domain of data. In this paper, we propose a novel diffeomorphic training-free approach; this is built upon the principle of an ordinary differential equation. Our formulation yields an Euler integration type recursive scheme to estimate the changes of spatial transformations between the fixed and the moving image pyramids at different resolutions. The proposed architecture is simple in design. The moving image is warped successively at each resolution and finally aligned to the fixed image; this procedure is recursive in a way that at each resolution, a fully convolutional network (FCN) models a progressive change of deformation for the current warped image. The entire system is end-to-end

and optimized for each pair of images from scratch. In comparison to learning-based methods, the proposed method neither requires a dedicated training set nor suffers from any training bias. We evaluate our method on three cardiac image datasets. The evaluation results demonstrate that the proposed method achieves state-of-the-art registration accuracy while maintaining desirable diffeomorphic properties. The main elements of this work are:

1. In the proposed method, an FCN models *the changes* in the deformation over multiple resolutions as opposed to the deformation itself that most existing methods do. The final deformation is estimated by a solution to an ordinary differential equation (ODE). Thus the resulting algorithm is recursive in nature.

2. Following this recursion, the moving image is warped successively, enabling the final prediction, which might consist of large displacements, to be decomposed into smaller displacements.

3. Within the recursive algorithmic framework, we introduce diffeomorphism, which guarantees the inverse consistency of deformations.

4. The proposed method is learning-free, i.e., it optimizes the parameters of the FCN from scratch for every new pair of images and eliminates the need for a dedicated training set.

5. The parameters of the FCN are shared across all resolutions and the neural net loss function values at all resolutions are optimized simultaneously.

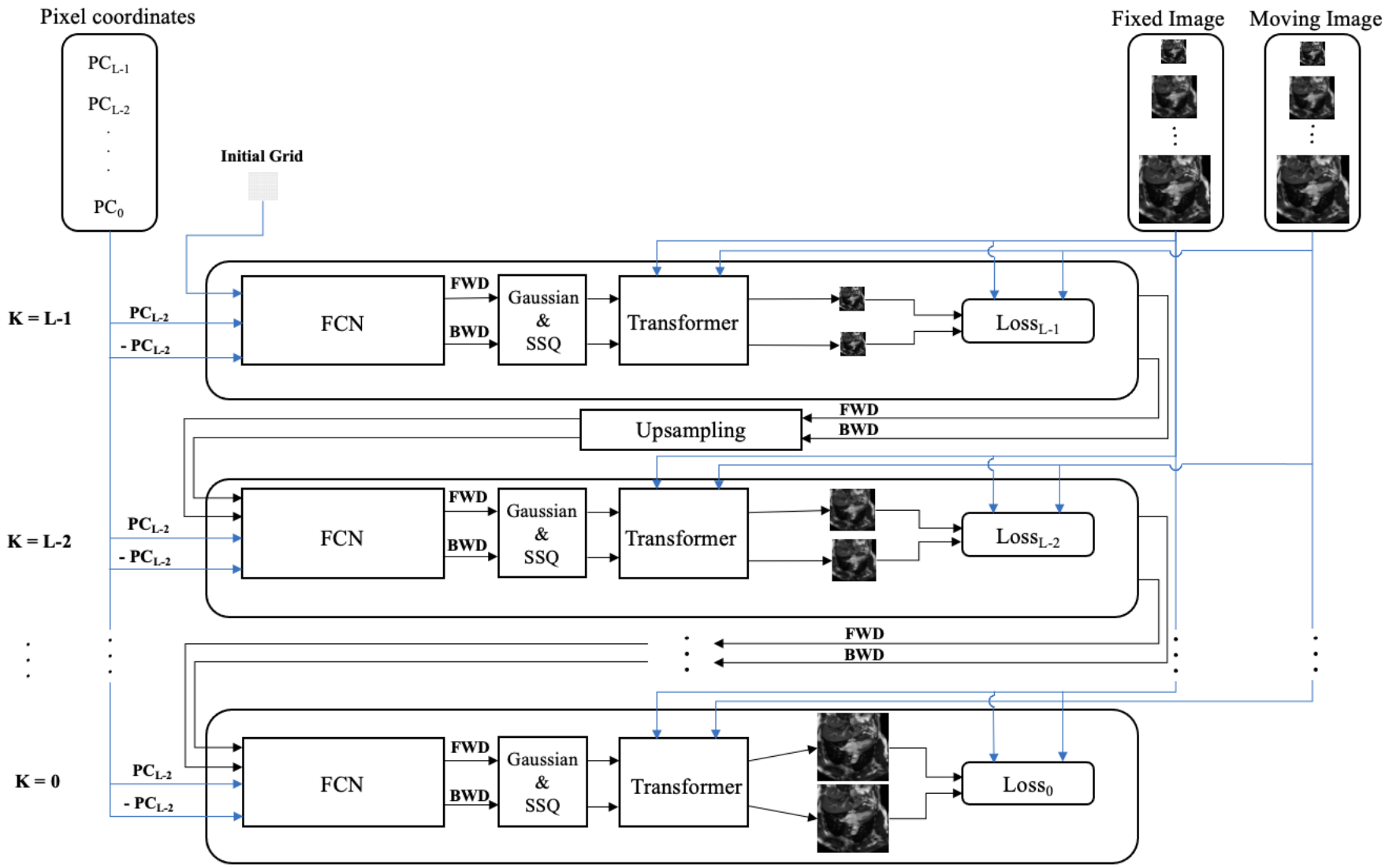


Figure 5.1: Schematic of the proposed recursive algorithm. $PC_{K-1}, PC_{K-2}, ..., PC_0$ are pixel coordinates correspond to the image pyramids. PC and $-PC$ are passed as the inputs to the network for forward and backward directions. An initial grid with zero displacement is used as a correction for the resolution $t = K - 1$. The outputs of the FCN, forward deformation field FWD and backward deformation field BWD will be smooth by a Gaussian filter and then passed to the scaling and squaring method SSQ. The final FWD and BWD generated at each resolution will be passed to an upsampler and then the upsampled results will be used as a correction for the next resolution. For each pair of images we repeat the whole process (e.g. 800 iterations) till we find the optimum value of θ

5.2 Methodology

Let $I_m \in \Omega$, $I_f \in \Omega$ denote the moving image and the fixed image respectively which are both defined over d-dimensional space Ω. ϕ is the deformation field which maps $\phi : \Omega \rightarrow \Omega$. Commonly, deformable registration methods construct a deformation prediction function g_θ which takes moving and fixed images as inputs and predicts a dense deformation field that aligns I_m to I_f. In contrast, in the proposed approach, we construct g_θ as a function of coordinate grid (not images) to predict the the change in the deformation field from one resolution to the next resolution. We cascade this procedure by recursively performing registration on the multi-resolution levels. Following this recursion, the change of deformation field is predicted at each resolution, enabling the final deformation field probably with large displacement to be decomposed into cascade with small displacement.

In multi-resolution pyramids, a Gaussian pyramid of images is constructed where the original image lies at the bottom level and subsequently higher levels have a downscaled Gaussian blurred version of the image. Multi-resolution pyramids often serve to accelerate the optimization and yield better accuracy. Using a multi-resolution recipe, two image pyramids are built: I_F^t and I_M^t for $t = K, \ldots, 1$, where K is the maximum level in the pyramid. Here, $I_F^1 = I_F$ and $I_M^1 = I_M$ are the original fixed and moving images, respectively.

Considering resolution t as the continuous variable, the change in deformation over resolution t is conceived by a FCN g_θ:

$$\frac{d}{dt}\phi_t = g_\theta(X_t, \phi_t), \tag{5.1}$$

where X_t, a multi-dimensional array, is the pixel coordinate grid at resolution t and D_t is the deformation field, another multi-dimensional array at resolution t. Thus, we model the change in deformation as a function of both the coordinate grid and the deformation. Discretization of the resolution (i,e., pyramid levels) leads to a solution

Algorithm 1: ODE-based Multi-resolution Diffeomorphism

Input: Pixel coordinates $\{X_t\}_{t=1}^{K}$ for all resolutions
Output: Deformation fields $\{D_t\}_{t=1}^{K}$ for all resolutions
-Initialize a spatial transformation
$V_K = f_\theta(X_K, 0)$
$\phi_K = EXP(V_K)$
for $t \leftarrow K-1$ **to** 0 **do**
$\phi_{t+1}^{up} = UpSample(\phi_{t+1})$
$V_t = g_\theta(X_t, \phi_{t+1}^{up})$
$V_t \leftarrow LowPass(V_t)$
$\phi_t = \phi_{t+1}^{up} \circ EXP(V_t)$
end

by a Euler integration with an initial value ϕ_K:

$$\phi_t = \phi_{t+1} + g_\theta(X_{t+1}, \phi_{t+1}), \quad \text{for } t = K-1, K-2, ..., 1, \tag{5.2}$$

where $t = K$ is the coarsest and $t = 1$ is the finest/original resolution. ϕ_{t+1} is the upsampled deformation field from a lower level ϕ_t and $g_\theta(D_{t+1})$ is the changes of the deformation field at resolution $t+1$ from resolution t.

Additionally, to have the same canonical range of pixel coordinates at every resolution, in our implementation, we use the range $[-1,1] \times [-1,1]$ for pixel coordinates. With this view, a multi-resolution pyramid adds more samples in the space $[-1,1] \times [-1,1]$ as we go from lower to higher resolutions.

A reasonable deformation field should prevent folding and be continuously varying (diffeomorphism). By taking the composite form in the Euler integration instead of the additive form, we introduce diffeomprhism to the framework. Further, if we treat the output of the FCN as the velocity field, then exponential of the velocity field will yield the multi-resolution diffeomorphic deformation:

$$\phi_t = \phi_{t+1} \circ EXP(f_\theta(X_{t+1}, \phi_{t+1})), \quad \text{for } t = K-1, K-2, ..., 1. \tag{5.3}$$

The computational scheme for ((5.3)) appears in Algorithm 1, where we introduced two more functionalities: upsampling and smoothing. Upsampling interpolates a deformation from a lower to the next upper level of resolution and smoothing of the velocity field acts as an implicit regularizer for image registration.

In this work, we use geometric optimization that uses the local canonical coordinates [31]. Based on this method, to estimate the current deformation, a compositional update rule is used between the previous estimate and the exponential map EXP of the displacement field. The exponential map is efficiently calculated by using the scaling and squaring method [84] and the composition of displacement fields. The exponentiation of the displacement field ensures the diffeomorphism of the mapping. Therefore, our recursive diffeomorphic image registration is obtained with Algorithm 1. Also the schematic of the proposed recursive algorithm is shown in 5.1. To optimize the FCN parameters θ, we use two Gaussian pyramids, one each for the moving I_M^t and the fixed I_F^t images, where $t = 1, ..., K$. Then, the FCN's parameters will be optimized based on the difference between warped moving image and fixed image at the same resolution. In our approach, the neural network will be optimized from the scratch for each pair of images being registered, and therefore, the trainable parameter values of the network will at convergence will be different. Our approach is a classical learning-free setting which would not require a dedicated training set with annotations of any kind from experts.

The parameters θ of the FCN are found by solving the following optimization using gradient descent (we use PyTorch's autograd):

$$\min_{\theta} \sum_{t=1}^{K} \sum_{x \in \Omega} L(I_F^t(x), I_M^t(D_t(x))) + \gamma \left\| D_t(x) - x \right\|^2, \tag{5.4}$$

where Ω denotes the fixed and moving images domain, $I_F^t(x)$ is the fixed image pixel value at pixel location x and resolution level t. $I_M^t(D_t(x))$ in (5.4) refers to the pixel value on the moving image I_M^t at location x after being displaced by deformation field $D_t(x)$. $L(.,.)$ is a differentiable loss function. In order to use the same range of displacement for all levels of resolution, we normalized the deformation field and the index at all resolutions in a canonical range $[-1, 1]$. In most of the learning-based deformable image registration approaches, the inverse mapping is often ignored [85]. However, the accuracy of the nonrigid registration could be improved by computing

the mappings from moving to fixed image and vice versa as well as exploiting the invertible property of diffeomorphic transformations.

In the proposed registration framework, we employ a FCN with two dense velocity fields to generate two deformation fields, namely, forward ϕ^F and backward ϕ^B deformation fields, that register images I_M to I_F and I_F to I_M, respectively, using Algorithm 1. We then enforce the consistency constraint that these two transformations are inverse mappings of each other by adding a regularization term to the loss function. Our FCN architecture consists of 4 convolution layers, each contains twenty-four 5×5 filters with a stride of 1, followed by a rectified linear unit (ReLU) activation, except for the last layer.

Note that the output of the Algorithm 1 is $\phi_t = [\phi_t^F, \phi_t^B]$ in which ϕ_t^F and ϕ_t^B are forward and backward deformation fields, respectively, see Fig. 5.2 as an example. With this view the following forward-backward optimization is used:

$$\begin{aligned}\min_{\theta} \sum_{t=1}^{K} \sum_{x \in \Omega} [&L_1(I_F^t(x), I_M^t(\phi_t^F(x))) + L_1(I_M^t(x), I_F^t(\phi_t^B(x))) + \\ &\alpha L_2(\phi_t^B(\phi_t^F(x)), x) + \alpha L_2(\phi_t^F(\phi_t^B(x)), x) + \\ &\gamma \left\| \phi_t^F(x) - x \right\|^2 + \gamma \left\| \phi_t^B(x) - x \right\|^2]\end{aligned} \tag{5.5}$$

We use differentiable mutual information (MI) [86] and structural similarity index metric (SSIM) [86] as similarity metric $L_1 = MI + SSIM$ and mean squared error (MSE) as penalizes term to enforce the inverse consistency $L_2 = MSE$.

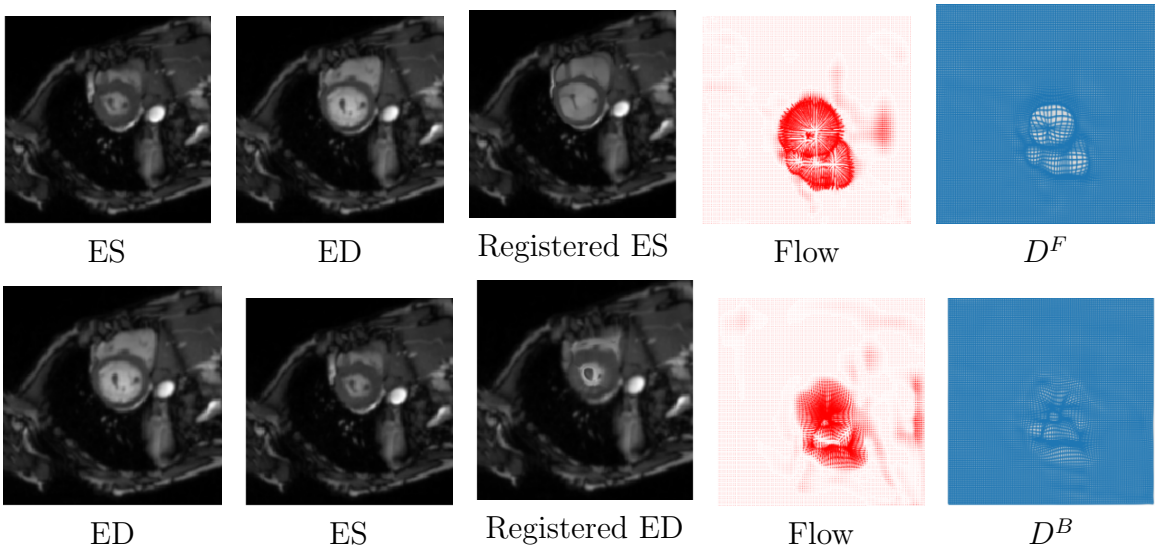


Figure 5.2: A sample result of forward and backward registrations with velocity fields and grid deformation fields from ACDC dataset.

Table 5.1: Quantitative evaluation of the results for cardiac MRI registration on the ACDC dataset. The evaluation was performed in terms of Dice (mean± standard deviation) and HD (mean) and the average number of pixels with non-positive Jacobian Determinant (lower is better). Values in bold indicate the best performance.

	Dice	HD	$\det(J_\phi) < 0$
Unregistered	0.65 ± 0.20	17.76	N/A
RTT(L2L2)	0.76 ± 0.18	7.19	0.10
RTT(L2L1)	0.78 ± 0.18	6.64	0.25
Elastix	0.79 ± 0.18	11.26	0.32
LCC-D	0.79 ± N/A	9.21	N/A
VM	0.79 ± N/A	8.46	N/A
LVM (S1)	0.79 ± N/A	7.58	N/A
SyN	0.80 ± N/A	8.24	N/A
MM	0.81 ± 0.18	8.59	0
LVM (S3)	0.81 ± N/A	6.88	N/A
Proposed Method	**0.92 ± 0.10**	**4.76**	**0**

*Results are are reported in [7]. **Results are reported in [41]. N/A refers to the value that are not reported in the original paper.

Table 5.2: Quantitative evaluation of the results for cardiac MRI registration on the SGD dataset. The evaluation was performed in terms of Dice (mean± standard deviation) and HD (mean) and the average number of pixels with non-positive Jacobian Determinant (lower is better). Values in bold indicate the best performance.

Methods	**Dice**	HD	$\det(J_\phi) < 0$
Unregistered	0.62 ± 0.15	N/A	N/A
RTT(L2L2)	0.70 ± 0.12	5.36	0.30
RTT(L2L1)	0.70 ± 0.12	5.36	0.30
Elastix	0.79 ± 0.08	N/A	0.37
SyN	0.81 ± 0.16	8.9	0.02
MM	0.72 ± 0.12	3.48	0
Demons	0.65 ± 0.18	15.46	0.42
DIRNet	0.80 ± 0.08	N/A	N/A
Proposed Method	**0.85 ± 0.15**	**5.06**	**0**

*Results are are reported in [7]. **Results are reported in [41]. N/A refers to the value that are not reported in the original paper.

Table 5.3: Quantitative evaluation of the results for cardiac MRI registration on the LA dataset. The evaluation was performed in terms of Dice (mean± standard deviation) and HD (mean) and the average number of pixels with non-positive Jacobian Determinant (lower is better). Values in bold indicate the best performance.

	LA								
Methods	Dice			HD			$\det(J_\phi) < 0$		
	2ch	3ch	4ch	2ch	3ch	4ch	2ch	3ch	4ch
Unregistered	0.79 ± 0.07	0.78 ± 0.08	0.78 ± 0.09	7.37	7.70	8.66	N/A	N/A	N/A
Elastix	0.81 ± 0.13	0.86 ± 0.08	0.80 ± 0.11	7.9	6.99	8.01	0.68	0.5	0.75
RTT	0.82 ± 0.11	0.85 ± 0.10	0.80 ± 0.10	7.28	6.82	7.56	0.53	0.45	0.7
MM	0.84 ± 0.06	0.83 ± 0.06	0.83 ± 0.08	6.58	6.48	6.77	0.01	0	0
Demons	0.84 ± 0.08	0.85 ± 0.06	0.82 ± 0.10	7.41	7.33	7.84	0.025	0.042	0
SyN	0.87 ± 0.06	0.86 ± 0.13	0.84 ± 0.11	6.92	7.52	7.51	0.033	0	0.01
Proposed Method	**0.89 ± 0.11**	**0.92 ± 0.06**	**0.86 ± 0.11**	**6.26**	**5.7**	**7.44**	0	0	0

* Results are are reported in [7]. **Results are reported in [41]. N/A refers to the value that are not reported in the original paper.

5.3 Experiments

5.3.1 Data:

The proposed framework is evaluated on three clinical cardiac MRI datasets:

Automated Cardiac Diagnosis Challenge (ACDC)[66] This dataset contains multiple temporal 2D short-axis cardiac cine MRI sequences acquired from 100 patients and one of the publicly available datasets for cardiac MRI assessment. The spatial resolution varies from 1.37 to 1.68 mm^2/pixel with a slice thickness of 5 mm to 8 mm (in general 5mm) and sometimes an inter-slice gap of 5mm. Each sequence consists of 28 to 40 images that cover the cardiac cycle completely or partially.

The Sunnybrook Cardiac Challenge data (SCD)[68] This dataset contains multiple temporal 2D short-axis cardiac cine MRI scans acquired from 45 patients. Each cine sequence includes 20 frames to cover the cardiac cycle. The data set is equally divided into 15 patient scans for training, 15 patient scans for validation, and 15 patient scans testing. The image resolution is 256×256, with a pixel spacing of 1.25 mm and slice thickness of 8 mm. To measure the accuracy of our method in comparison to other methods, registration is performed on the 15 test scans.

Left Atrium (LA) This dataset includes 100 temporal 2D long-axis cine MRI steady-state sequences from the 2, 3 and 4-chamber views. It is acquired from the University Alberta Hospital. Each cycle includes 25 or 30 frames with image resolutions 176×189 – 256×208 and image spacing $1.445 - 1.795$ mm. The ground truth manual segmentation is initially performed by a medical student and edited by an experienced radiologist. The 2ch, 3ch and 4ch are used in the rest of the paper to denote 2, 3 and 4-chamber sequences, respectively.

5.3.2 Implementation:

The proposed method is implemented in Python programming language using Pytorch module. In neural network design, the Adam optimization with 800 update iterations and a learning rate of 5×10^{-4} are used for all the three datasets. The level of pyramid K is considered 2 with $\alpha = 1/(K)$, $\gamma = \frac{\lambda \times 1}{K}$, $\lambda = 5$. The neural net framework is evaluated on an NVIDIA GeForce GTX 1080 Ti GPU.

5.3.3 Evaluation Metrics:

The Dice metric (DM), reliability, Hausdorff distance (HD) and the determinant of Jacobian are used as the evaluation metrics to compare the performance of the algorithms.

Dice Metric The DM [73] is a well-known segmentation based metric to measure the similarity (overlap) between two regions, warped moving and fixed image. The DM of two regions A and B is formulated as:

$$DM(A, B) = \frac{2|A \cap B|}{A + B} \tag{5.6}$$

Reliability The algorithms are examined by evaluating the reliability function [87] of the obtained DMs using (5.7). The complementary cumulative distribution function is defined for each $d \in [0, 1]$ as the probability of obtaining DM higher than d over the entire set. The $R(d)$ measures how reliable is the algorithm in yielding accuracy d.

$$R(d) = P_r(Dice > d) = \frac{\text{\# Images segmented with DM higher than d}}{\text{total number of images}} \tag{5.7}$$

Hausdorff Distance The HD [74] is another well-known metric which measures the maximum deviation between two regions' contours. The HD between two contours (C_A) and C_B is formulated as:

$$\text{HD}(C_A, C_B) = \max(\max_i(\min_j(d(p_A^i, p_B^j))), \max_j(\min_i(d(p_A^i, p_B^j)))) \tag{5.8}$$

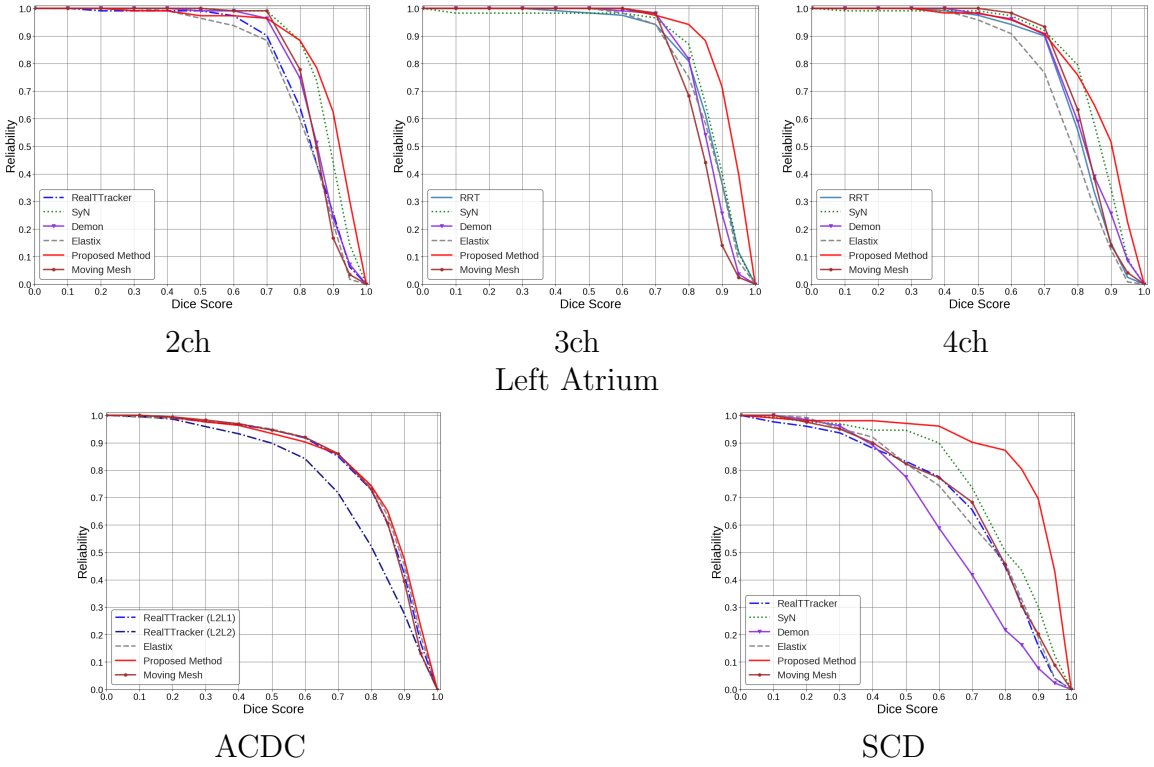
Table 5.4: Quantitative cardiac MRI registration results on 50% patients of ACDC, SCD and LA datasets based on different loss functions. The evaluation was performed in terms of average Dice, HD and the forward and backward registration time (per second).

SCD				ACDC			
Methods	Dice	HD	Time(s)	Methods	Dice	HD	Time(s)
MSE	0.83	9.1	15	MSE	0.87	5.73	15
SSIM	0.86	7.9	20	SSIM	0.92	4.31	21
SSIM + MI	0.89	6.01	26	SSIM + MI	0.93	4.31	35

LA							
Methods	Dice			HD			Time (s)
	2ch	3ch	4ch	2ch	3ch	4ch	
MSE	0.84	0.86	0.84	6.83	6.88	7.84	15
SSIM	0.91	0.93	0.89	6.26	6.11	6.42	22
SSIM + MI	0.90	0.93	0.87	6.89	5.22	6.40	35

where p_A^i, p_B^j denote the set of all the points in C_A and C_B respectively. The term $d(\cdot)$ denotes the Euclidean distance.

Determinant of Jacobian We quantify and analyze deformation regularity using the determinant of the Jacobian, $\det(J_\phi)$. Jacobian matrix captures the local behaviours of the deformation field. If the value of $\det(J)$ equals to 1, the volume remains constant after the transformation, the value smaller than 1 shows the local volume shrinkage and greater than 1 shows the local volume expansion. The negative values imply that local folding and twists have occurred, which is physically not realizable and mathematically not inevitable [45].



2ch 3ch 4ch

Left Atrium

ACDC SCD

Figure 5.3: Reliability versus Dice metric for different algorithms. The figures omit results for methods with no public implementation available.

5.3.4 Evaluation Results and Discussions

The performance of the proposed framework is compared with nine state-of-the-art learning-based and classic deformable registration algorithms including SimpleElastix (Elastix) [69], Moving Mesh (MM) [30], Real-Time Image-based Tracker (RTT) [70], Demons [71], LCC-Demons (LCC-D) [32], Symmetric Normalization (SYN) [27], [45], [7] and DIRNet [41].

Table **??**, shows the mean and standard deviations of Dice scores, HD and the average number of pixels with non-positive $\det(J_\phi)$ over all subjects ACDC, SCD, and left atrium datasets. Compared to the conventional and learning-based methods, our framework has the best performances in terms of Dice scores. In Fig. **??**, the highest accuracy, $50th$ percentile, and the lowest accuracy in terms of Dice scores achieved by the proposed method are shown visually, where the red, blue, violet, gray

Table 5.5: Quantitative cardiac MRI registration results on the ACDC, SCD and LA based on different number of resolution. The evaluation is performed in terms of Dice (mean± standard deviation) and HD (mean). The 2ch, 3ch and 4ch stand for the 2, 3 and 4-chamber. Values in bold indicate the best performance.

	Dataset	Dice	HD
One Resolution	ACDC	0.85 ± 0.20	6.03
	SCD	0.85 ± 0.17	7.70
	LA-2ch	0.87 ± 0.08	7.6
	LA-3ch	0.91 ± 0.05	6.59
	LA-	0.88 ± 0.10	6.63
Two Resolutions	ACDC	**0.87 ± 0.17**	**6.27**
	SCD	**0.86 ± 0.16**	**7.06**
	LA-2ch	**0.89 ± 0.09**	**7.59**
	LA-3ch	**0.93 ± 0.03**	**5.47**
	LA-4ch	**0.87 ± 0.11**	**6.67**
Three Resolutions	ACDC	0.83 ± 0.22	6.70
	SCD	0.85 ± 0.18	7.22
	LA-2ch	0.86 ± 0.09	7.03
	LA-3ch	0.90 ± 0.06	6.22
	LA-4ch	0.86 ± 0.11	6.42
Four Resolutions	ACDC	0.74 ± 0.24	8.96
	SCD	0.85 ± 0.19	7.50
	LA-2ch	0.82 ± 0.10	7.06
	LA-3ch	0.86 ± 0.07	6.93
	LA-4ch	0.81 ± 0.10	7.01

and orange contours depict the correspondence segmentation results by the proposed method, RRT1, RRT2, Elastix and MM, respectively and the green contours depict the ground truth.

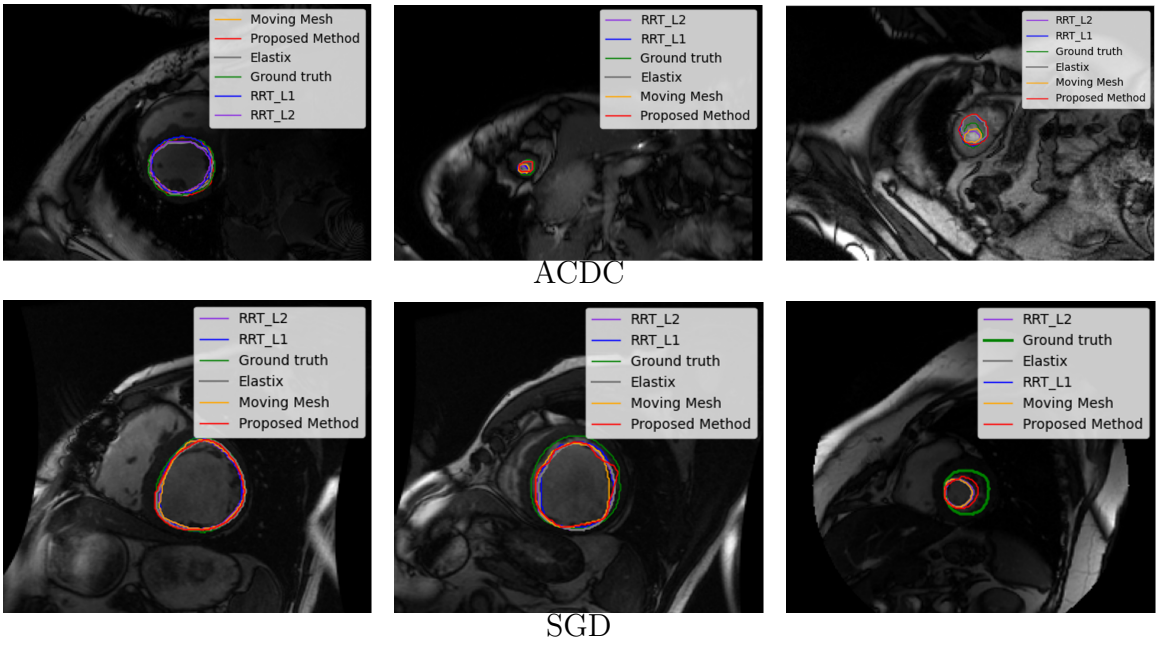


Figure 5.4: The segmentation results with the highest (first column), 50th percentile (2nd column), lowest (3rd column) Dice score values for (a) ACDC (1st row), (b) SGD (2nd row). The green represent the ground truth, also the red, blue, violet, gray and orange contours represent the boundary corresponds to the registration by proposed method, RRT1, RRT2, Elastix and MM, respectively.

To quantify and analyse deformation regularity, we computed the determinant of the Jacobian $\det(J_\phi)$ using SimpleITK [71]. No negative values of $\det(J)$ are observed for the proposed method as can be seen in Table **??**.

We also evaluated the performance of the proposed algorithm in terms of the reliability function $R(d)$ [87]. The corresponding $R(d)$ are plot as a function of Dice score in Fig. 5.3. Our algorithm led to a higher reliability curve on all ACDC, SCD, and left atrium datasets.

We examined different loss functions, namely, MSE, SSIM and SSIM + MI, in our framework and reported the average time to register one pair of images (Computing both forward and backward deformation fields) as well as the average of DM and HD over 50% of patient datasets in Table 5.4. It can be seen that the MSE loss function led to the best average time; however, it does not yield the best accuracy in terms of

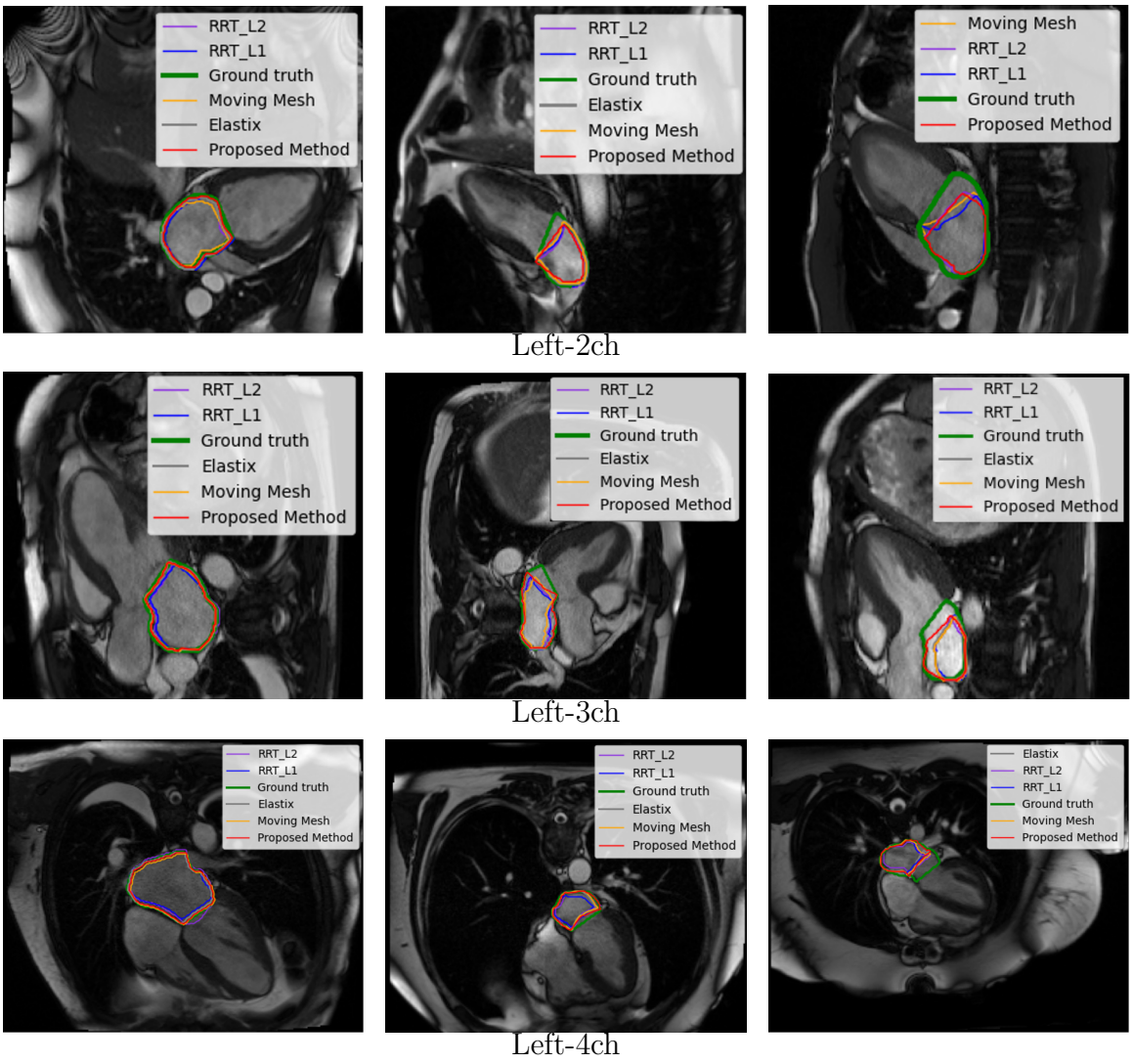


Figure 5.5: The segmentation results with the highest (first column), 50th percentile (2nd column), lowest (3rd column) Dice score values for left atrium 2-chamber (3rd row), (d) 3-chamber (4th row), 4-chamber (5th row) views. The green represent the ground truth, also the red, blue, violet, gray and orange contours represent the boundary corresponds to the registration by proposed method, RRT1, RRT2, Elastix and MM, respectively.

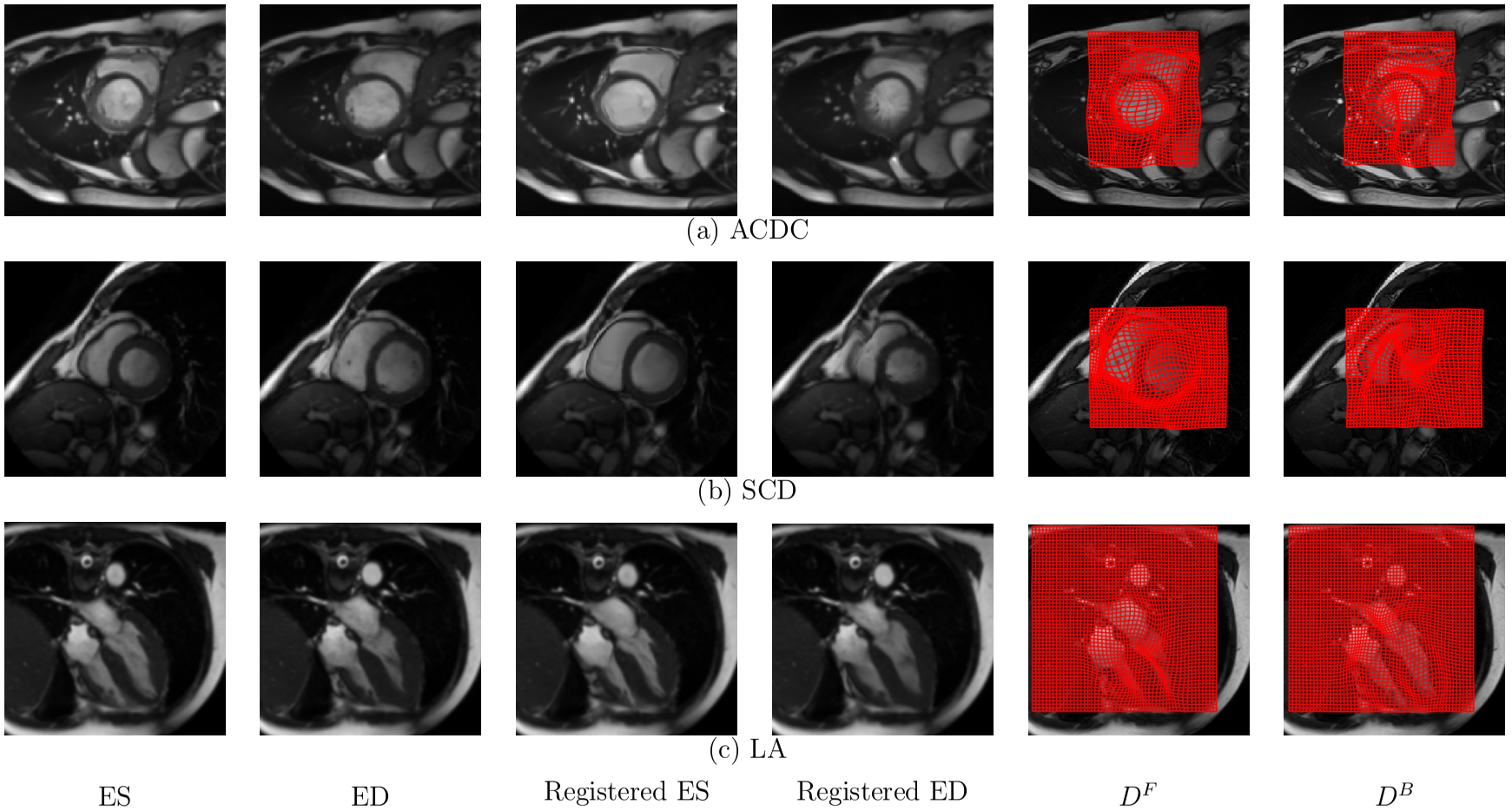


Figure 5.6: Examples of pair image registration End-systolic (ES) and End-diastolic (ED) from (a) ACDC, (b) SCD and (c) LA datasets. Each row shows forward and backward registered images and the corresponding grid deformations. The registrations are applied to the whole image, however, the grids are cropped to have a better resolution

the Dice score.

We assessed the impact of using different multi-resolution structures on the performance, and reported the corresponding evaluations in terms of DM and HD in Table 5.5. The results indicate that the high performance is obtained with the use of two resolutions.

Examples of registered images from ACDC, SCD, and left atrium datasets with their corresponding grid deformation are shown in Fig. 5.6 (a), (b), and (c), respectively. Fig. 5.6 shows the original ED, ES and their correspondence registered images and grid forward and backward deformations over images. The algorithm is applied on whole images; however, to display the deformation tracking on the part that has the most changes, we cropped the grid.

We have also evaluated the outputs from the forward deformation field by ignoring the bidirectional registration to present its effects on the Dice score and HD. The forward deformation field yields Dice score values of 0.88 ± 0.17 and 0.87 ± 0.14 and HD values of 6.08 and 5.19 for ACDC and SCD, respectively, while the proposed bidirectional approach yields Dice score values of 0.92 ± 0.10 and 0.90 ± 0.15 and HD values of 4.76 and 5.06 for ACDC and SCD, as reported in Table **??**. This clearly indicates the importance of including the bidirectional registration to obtain high performance.

5.4 Conclusion

We present an ODE-based diffeomorphic recursive framework for multi-resolution deformable registration using a FCN, to estimate the change velocity of forward and backward deformation fields. We have then utilized inverse consistency loss to further guarantee the desirable diffeomorphic properties of the resulting solutions. The proposed framework is learning-free and does not require a dedicated training set.

The proposed approach simultaneously estimates the forward and backward mapping at all levels of the multi-resolution pyramid. We evaluated and compared the

method using three different MRI datasets against several state-of-the-art traditional and learning-based registration methods. The results demonstrate that our method outperforms both traditional and learning-based methods.

Chapter 6

Unsupervised 2D-to-2D and 3D-to-3D Diffeomorphic Registration framework

6.1 Overview

In this chapter we propose a deformable registration algorithm based on unsupervised learning by using the moving mesh parameterization of deformation fields. The moving mesh parametrization is originally designed to generate a grid suitable for solving partial differential equations [88]. It naturally leads to a formulation of diffeomorphic image registration as a constrained optimization problem which we solved it by using deep learning. Such strategy has been adopted in the Demons algorithm, where unconstrained optimization is followed by a smoothing filter to impose a smoothness constraint 5. Moving mesh parameterization models a deformation field with its transformation Jacobian determinant and the curl of the end velocity field. The new model of the deformation field has three significant advantages; firstly, it relaxes the need for an explicit regularization term and the corresponding weight in the cost function. The smoothness is implicitly embedded in the solution which results in a physically plausible deformation field. Secondly, it guarantees diffeomorphism through explicit constraints applied to the transformation of the Jacobian determinant to keep it positive. Finally, it is suitable for cardiac data processing, since the

nature of this parameterization is to define the deformation field in terms of the radial and rotational components. [29, 30, 89]. The effectiveness of the algorithm is investigated by evaluating the proposed method on three different data sets including 2D and 3D cardiac MRI scans. The results demonstrate that the proposed framework outperforms existing learning-based and non-learning-based methods while generating diffeomorphic transformations. The proposed 2D-to-2D registration framework is tested on 2 publicly available datasets and 1 dataset which were scanned at the Mazankowski Alberta Heart Institute (Edmonton, Alberta, Canada). The 3D-to-3D version is tested on one publicly available dataset. The results show competitive performance in comparison to learning-based and non-learning-based methods.

6.2 Methodology

Most of the learning-based algorithms formulate the deformable registration problem as the minimization of the following equation:

$$\phi^* = \underset{\phi}{\operatorname{argmin}}\, L(I_F, I_M \circ \phi(\xi)) \tag{6.1}$$

where ξ denotes the pixel location in the image domain Ω, $\phi : \Omega \rightarrow \Omega$ denotes the transformation function, and the dissimilarity metric is denoted by $L(.)$. With the above formulation, introducing a regularization is necessarily to obtain a unique solution. Without regularization, this may result in multiple physically non-plausible solutions.

In our setting, we tackle these issues with the help of the moving mesh parameterization.

6.2.1 Moving Mesh Grid Generation

In moving mesh generation formulation, we need to define a monitor function μ and a vector field V. To avoid adding extra terms to the above formulation and having a

unique solution, more constraints are required to be added using a monitor function μ and curl of end velocity field γ.

First a continuous monitor function is defined and constrained by:

$$\int_{\Omega} \mu = |\Omega|. \tag{6.2}$$

The goal here is to find a transformation ϕ_1: $\Omega \rightarrow \Omega$, $\partial\Omega \rightarrow \partial\Omega$ such that the transformation Jacobian determinant $J_{\phi_1}(\xi)$ is equal to the monitor function μ :

$$J_{\phi}(\xi) = det\nabla\phi_1(\xi) = \mu(\xi). \tag{6.3}$$

To find the transformation ϕ_1 which satisfies 7.2, the following steps need to be taken,

Step 1: A vector field $V(\xi)$ is defined such that:

$$div\, V(\xi) = \mu(\xi) - 1. \tag{6.4}$$

Step 2: A velocity vector field based on artificial-time is then constructed from $V(\xi)$:

$$V_t(\xi) = \frac{V(\xi)}{t + (1-t)\mu(\xi)}, t \in [0,1] \tag{6.5}$$

The desire transformation ϕ_1 can be found by solving the following ordinary differential equation (ODE) at $t = 1$, $\phi_1(\xi) = \psi(\xi, t = 1)$ where $\psi(\xi$,t=0$) = \phi_0(\xi)$

$$\frac{\psi(\xi,t)}{dt} = V_t(\psi(\xi,t)), t \in [0,1], \tag{6.6}$$

Where $\phi_0(\xi)$ is the identity mapping and $det\nabla\phi_0(\xi) = 1$ and $\phi_0(\xi) = \xi$. Since the $\phi_1(\xi)$ is the desire transformation that we are looking for, we drop the subscript and use $\phi(\xi)$ for the rest of the paper. The main problem is how to find $V(\xi)$ such that $divV(\xi) = \mu(\xi) - 1$. There are different methods to solve this problem such as the div-curl system. To solve the problem with the div-curl system, we need to find the divergence and curl at each point and set up the div-curl system of equations for

Algorithm 2: Moving Mesh based deformable registration

Input: Given two 2D/3D pair of images, fixed image I_F and moving image I_M. The upper bound τ_{ub} and lower bound τ_{lb} of the transformation Jacobian determinant

Output: Deformation field ϕ

Step 1: Pass the input to the CNN to compute $\mu(\xi)$ and $V(\xi)$;

Step 2: Impose constraints from (7.7) for each pixel location $\xi \in \Omega$:

$$\mu(\xi) \leftarrow \frac{|\Omega|}{\sum_{\xi \subset \Omega} \mu(\xi)}$$

Step 3: Compute a curl of velocity field $V(\xi)$ that satisfies (7.3) and compute the deformation field ϕ

Step 4: Compute the loss function

Step 5: Update the μ and $V(\xi)$ using back-propagation

each point. By solving this system we can reconstruct a differentiable and invertible transformation.

$$\begin{cases} divV(\xi) = \mu(\xi) - 1 \\ curlV(\xi) = \gamma(\xi). \end{cases} \tag{6.7}$$

To have a unique ϕ a constraint need to be applied to the div of the vector field $V(\xi)$ 7.6. The generated transformation ϕ now can be parameterized with transformation Jacobian determinant and the curl of the end velocity field.

6.2.2 Diffeomorphic Image Registration

Using the above parameterization, the diffeomorphic image registration can be formulated as a constrained optimization problem. Let I_F and I_M be 2D/3D fixed and moving images/volumes, defined over $\Omega \rightarrow \mathbb{R}^2/\Omega \rightarrow \mathbb{R}^3$. We need to find $\mu(\xi)$ and $\gamma(\xi)$ $\forall \xi \in \Omega$, that optimize a similarity metric L_{Sim} between the warped moving image and fixed image, subject to the following constraints:

$$\begin{cases} \int \mu(\xi) d\xi = |\Omega| \\ \tau_{ub} > \mu(\xi) > \tau_{lb} \end{cases} \tag{6.8}$$

where the τ_{ub} is the upper bound and τ_{lb} is the lower bound of the transformation Jacobian determinant which were set by the user. The $\tau_{lb} > 0$ guarantees the diffeomorphism.

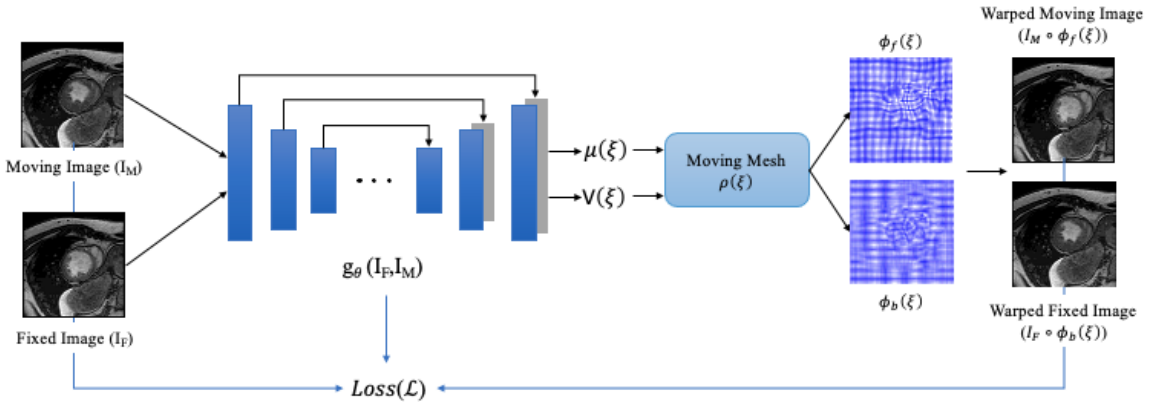


Figure 6.1: Overview of end-to-end unsupervised architecture. The ConvNet $g_\theta(I_F, I_M)$ takes the input fixed image(I_F) and moving image(I_M) and outputs the transformation Jacobian determinant $J_\phi(\xi) = \mu(\xi)$, and the vector field $V(\xi)$. Then the diffeomorphic forward and backward transformations ϕ_f and ϕ_b are computed using the moving mesh approach. Finally, the moving and fixed images are warped using ϕ_f and ϕ_b.

6.2.3 Numerical Methods

2D Div-curl solver

We represent the deformation field by divergence and curl (div-curl) system representation [90] (7.6). To find $V(\xi)$ under the null condition we converted the (7.6) into a set of Poisson equations as follows and used a Fast Fourier Transform (FFT) based Poisson solver. As shown in (6.9) the radial component is given by F^1 and the rotational components is given by F^2:

$$\begin{cases} \Delta V_x &= \mu_x - \gamma_y = F^1, \\ \Delta V_y &= \mu_y + \gamma_x = F^2, \end{cases} \tag{6.9}$$

3D Div-curl solver

The div-curl system for the 3D case is given in Equation (7.6). Where the divergence of the deformation field represents the radial motion while the curl operator represents the rotation of the media around every point. The 3D operator directly extends from the 2D curl, where each rotational component represents the rotational motion of the deformation field about each of the three axes. As it shown in (6.10) the radial component is given by f^1 and the three rotational components are given by f^2, f^3

and f^4. For the 3D version, there ate three unknowns (V_x, V_y, V_z) with four scalar equations which makes this system overdetermined. Furthermore, a dummy variable θ is introduced to solve the system. (please check [91] for more details.)

$$\begin{cases} divV = \dfrac{\partial V_x}{\partial x} + \dfrac{\partial V_y}{\partial y} + \dfrac{\partial V_z}{\partial z} = f^1 \\ curl_x V = \dfrac{\partial \theta}{\partial x} + \dfrac{\partial V_z}{\partial y} - \dfrac{\partial V_y}{\partial z} = f^2 \\ curl_y V = \dfrac{\partial \theta}{\partial y} + \dfrac{\partial V_x}{\partial z} - \dfrac{\partial V_z}{\partial x} = f^3 \\ curl_z V = \dfrac{\partial \theta}{\partial z} + \dfrac{\partial V_y}{\partial x} - \dfrac{\partial V_x}{\partial y} = f^4. \end{cases} \tag{6.10}$$

Similar to the 2D version, we converted the (6.10) into a set of Poisson equations as follows:

$$\begin{cases} \Delta V_x &= f_x^1 + f_z^3 - f_y^4 = F^1, \\ \Delta V_y &= f_y^1 + f_x^4 - f_z^2 = F^2, \\ \Delta V_z &= f_z^1 + f_y^2 - f_x^3 = F^3. \end{cases} \tag{6.11}$$

Then the Euler method with arbitrary time steps is used to compute the transformation ϕ from $V(\xi)$ via (7.4) and (7.5). For derivation and numerical implementation details, we refer the reader to [91].

6.2.4 Data driven parameter computation

Despite the traditional methods that iteratively and manually compute the parameters and update the gradient [30, 89] which are time-consuming, we use an unsupervised CNN and back-propagation Algorithm 2. In the proposed framework, the network parameters are learnt in an unsupervised fashion and a diffeomorphic deformation field is generated by moving mesh parameterization Figure 6.1.

As shown in Figure 6.1, the network takes I_F and I_M as input and outputs the monitor function $\mu(\xi)$ and the velocity vector filed $V(\xi)$. Then using the curl of end velocity and a div-cur system a diffeomorphic transformation ϕ is computed. To establish the uniqueness of the solution the Dirichlet boundary condition is used [92]. Additionally, a diffeomorphism, which is corresponded to a positive transformation Jacobian determinant, is enforced explicitly via the monitor function [91]. All of the

steps are designed to be differentiable and the network parameters are learnt using stochastic gradient descent optimization.

6.2.5 Registration

To train the framework a set of pair images (I_F, I_M) were given. Then using the monitor function and curl of end velocity, the desire ϕ was computed. Finally, the moving image was warped to have the minimum dissimilarity with fixed image I_F. For each pair of image, we simultaneously calculated the forward transformation which registers the fixed image I_F to moving image I_M and the backward transformation which registers the moving image I_M to fixed image I_F. A symmetric loss function is used as follows:

$$\phi^* = \underset{\theta,\mu,\gamma}{\operatorname{argmin}}\{w \times L(I_F, I_M \circ \phi_f) + w \times L(I_M, I_F \circ \phi_b)\} \tag{6.12}$$

where ϕ_f is the forward transformation and ϕ_b is the backward transformation.

The registration process is performed pairwise on both 2D images and 3D volumes. In the cardiac data sets, the end-diastolic and end-systolic images are passed to the proposed framework as input to compute the forward transformation ϕ_f and the reverse transformation ϕ_b. For the 2D version the mean squared error (MSE) and for the 3D version the normalise cross correlation (NCC) is used as dissimilarity metric.

6.3 Experiments

We perform a series of experiments to evaluate the registration accuracy of the proposed diffeomorphic CNN method against the state-of-the-art methods. The evaluations were performed over three data sets consisting of clinical 2D cardiac MR images to assess the performance of the 2D version of our method. We also evaluated the 3D version of the proposed framework using ACDC data set in 3D.

6.3.1 Data sets

The following three data sets are considered in this study:

Automated Cardiac Diagnosis Challenge (ACDC) [66] This data set contains multiple temporal 2D short-axis cardiac cine MRI sequences acquired from 100 patients and is one of the publicly available data sets for cardiac MRI assessment. The spatial resolution varies from 1.37 to 1.68 $mm^2/pixel$ with a slice thickness of 5 mm to 8 mm (in average 5mm). The testing set contained 20 cases of each of the following cardiac diseases: dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), previous myocardial infarction (MINF), abnormal right ventricle (RV) and healthy (Normal). The images are cropped to a size of 128×128, and padded the third dimension to 16 for the 3D voxels.

The Sunnybrook Cardiac Challenge data (SCD) [68] This data set contains multiple temporal 2D short-axis cardiac cine MRI scans acquired from 45 patients. Each cine sequence includes 20 frames to cover the cardiac cycle. The data set is equally divided into 15 patient scans for training, 15 patient scans for validation, and 15 patient scans for testing. The image resolution is 256×256, with a pixel spacing of 1.25 mm and slice thickness of 8 mm.

Left Atrium (LA) This data set includes 100 temporal 2D long-axis cine MRI steady-state sequences from the 2, 3 and 4-chamber views, acquired from the University Alberta Hospital. Each cycle includes 25 or 30 frames with image resolutions 176×189 – 256×208 and image spacing $1.445 - 1.795$ mm. The ground truth manual segmentation is initially performed by a medical student and edited by an experienced radiologist. The 2ch, 3ch and 4ch are used in the rest of the paper to denote 2, 3 and 4-chamber sequences, respectively. The results are compared on end-diastolic and end-systolic frames.

6.3.2 Quantitative Evaluation Metrics

The proposed method is evaluated quantitatively using four metrics, namely, Dice metric (DM), Hausdorff distance (HD in mm), determinant of Jacobian of the defor-

mation field $\det(J)$, and reliability $R(d)$.

Dice Metric: The DM [73] is a segmentation-based metric to measure the similarity (overlap) between two regions, warped moving and fixed images. Where the Dice score of 1 indicates complete overlap and Dice score of 0 indicates no overlap. The DM of two regions A and B is formulated as:

$$DM(A,B) = \frac{2|A \cap B|}{A+B} \tag{6.13}$$

Hausdorff Distance: The HD [74] is another metric which measures the maximum deviation between two regions' contours. The HD between two contours (C_A) and C_B is formulated as:

$$\begin{aligned} \mathrm{HD}(C_A, C_B) = &\max(\max_i(\min_j(d(p_A^i, p_B^j))), \\ &\max_j(\min_i(d(p_A^i, p_B^j)))) \end{aligned} \tag{6.14}$$

where p_A^i, p_B^j denote the set of all the points in C_A and C_B respectively. The term $d(\cdot)$ denotes the Euclidean distance.

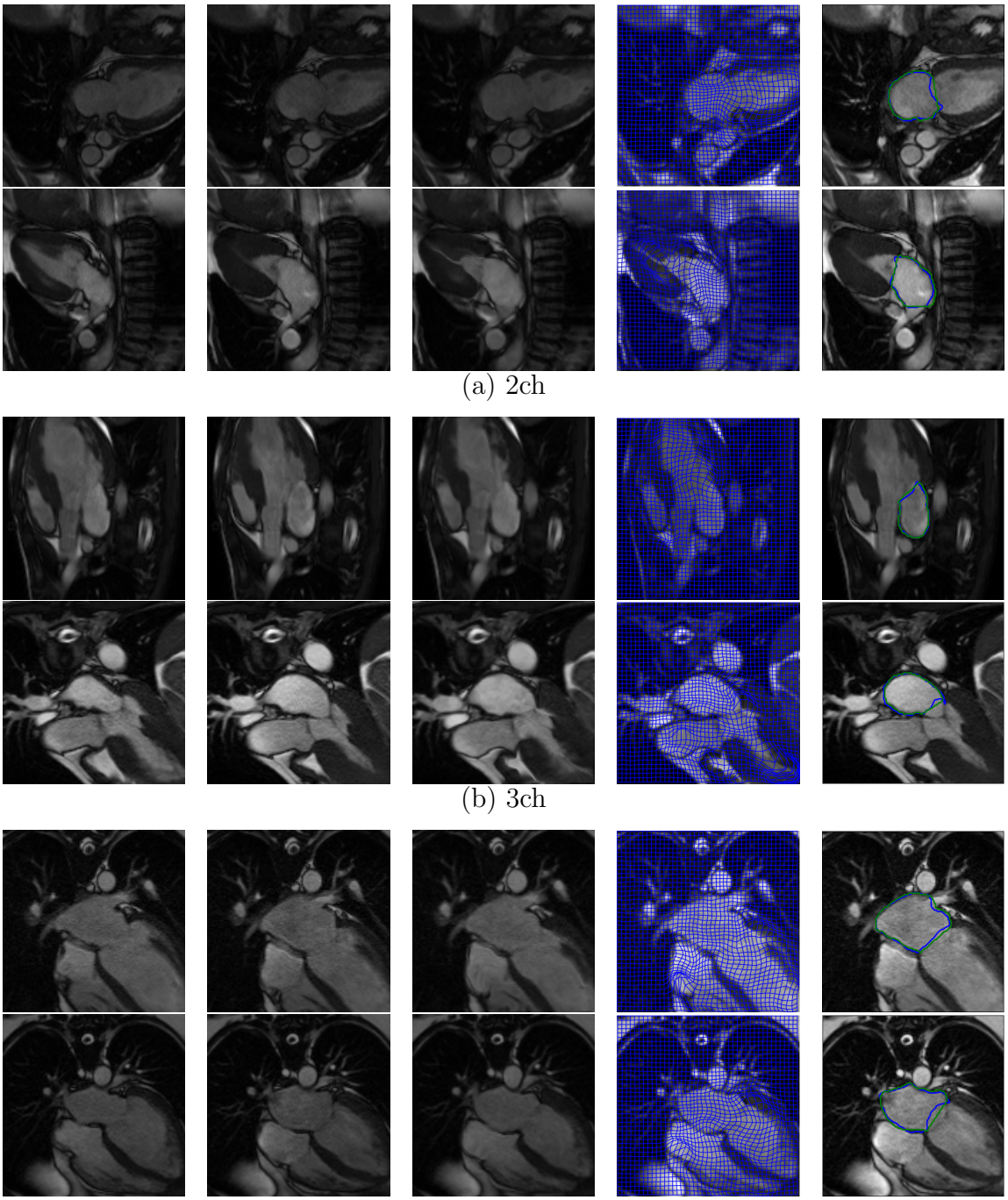
Reliability: We also evaluated the performance of the proposed algorithm using a reliability function computed based on DMs for each data set. The complementary cumulative distribution function is defined for each $d \in [0,1]$ as the probability of obtaining DM higher than d overall volumes.

$$\begin{aligned} R(d) &= P_r(Dice > d) \\ &= \frac{\#\text{ Images segmented with DM higher than d}}{\text{total number of images}}. \end{aligned} \tag{6.15}$$

$R(d)$ measures how reliable the algorithm is in yielding accuracy d.

det(J): To analyze deformation regularity in different algorithms, we calculate the determinant of the Jacobian $\det(J)$ [75]. If the value of $\det(J)$ equals 1, the area remains constant after the transformation, whereas the value smaller or larger than 1 indicates the local area shrinkage or expansion, respectively. The negative value of $\det(J)$ implies that local folding and twisting have occurred, which are physically not realizable and mathematically not invertible [45].

LA



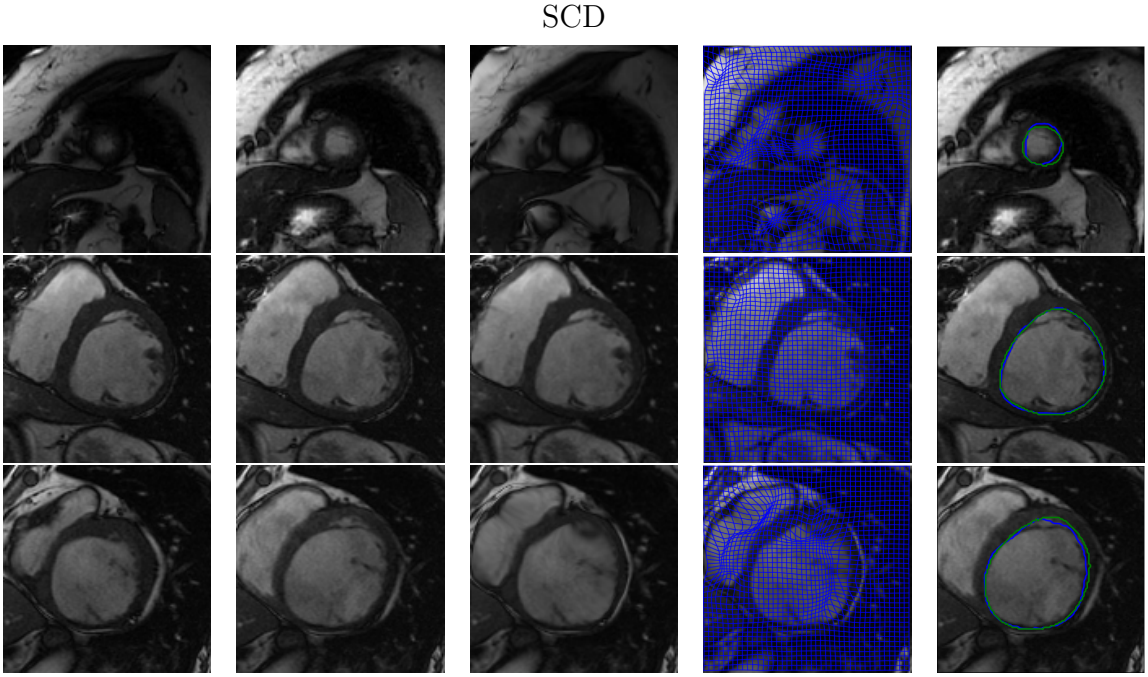
(a) 2ch

(b) 3ch

(a) ES (b) ED (c) Warped ES (d) DF (e) GT

(c) 4ch

Figure 6.2: Samples of registered images on the left atrium data set with the corresponding deformation field grid (DF). The end-systolic (ES) frame is the moving image and end-diastolic (ED) frame is the fixed image. The warped ES of each row is shown in the third column. The last column labeled ground truth (GT) displays the true segmentation and the predicted segmentation, which are shown by the green line and blue line respectively. The 2ch, 3ch and 4ch stand for the 2, 3 and 4-chamber.



(a) ES (b) ED (c) Warped ES (d) DF (e) GT

Figure 6.3: Samples of registered images on the SCD with the corresponding deformation filed grid (DF). End-systolic (ES) frame is the moving image and end-diastolic (ED) frame is the fixed image. The warped ES of each row is shown in the third column. The last column labeled ground truth (GT) displays the true segmentation and the predicted segmentation, which are shown by the green line and blue line respectively.

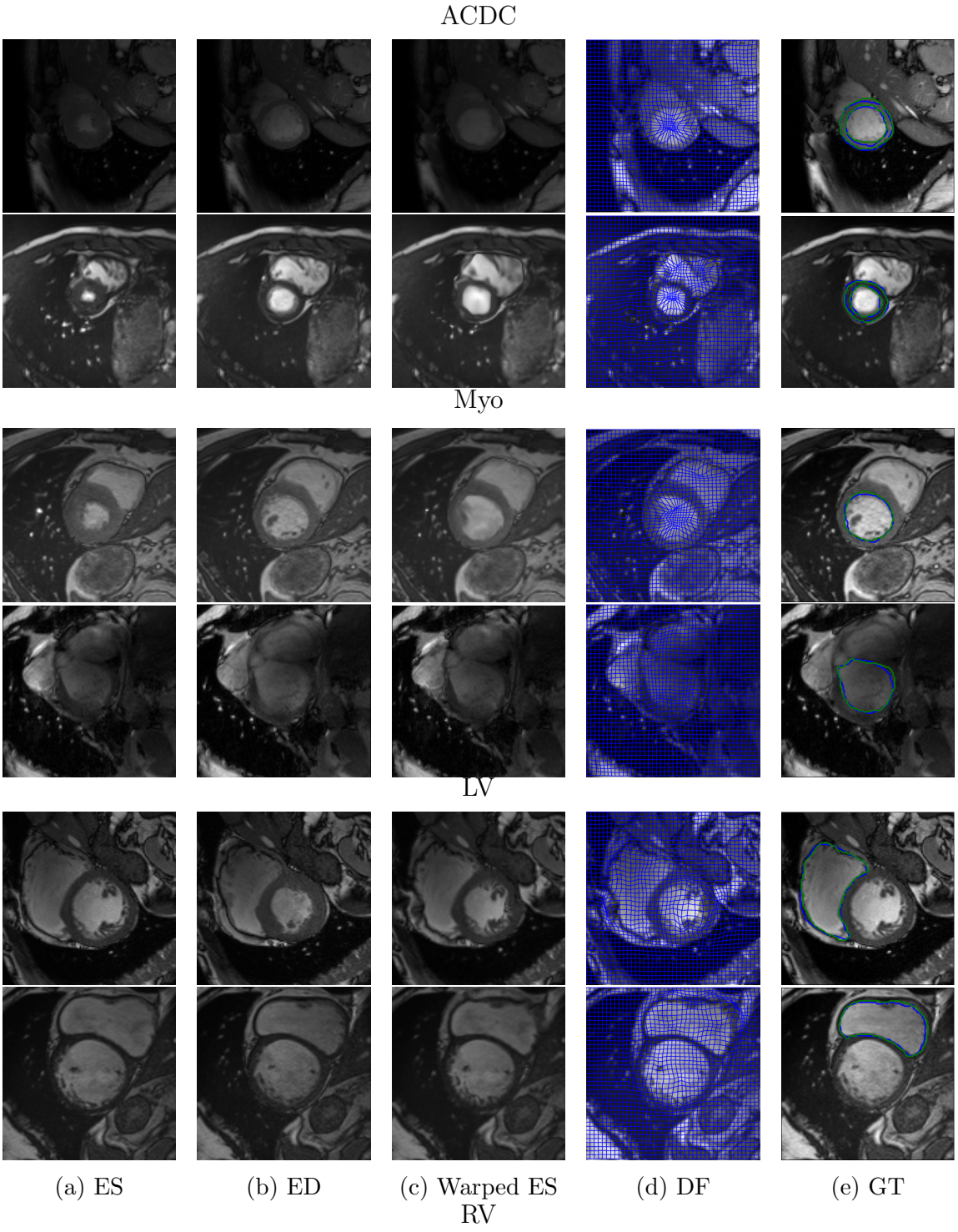


Figure 6.4: Samples of registered images on the ACDC on the myocardium (Myo), left ventricle (LV), and right ventricle (RV) anatomical structures with the correspondence deformation filed grid (DF). End-systolic (ES) frame is the moving image and end-diastolic (ED) frame is the fixed image. The warped ES of each row is shown in the third column. The last column labeled ground truth (GT) displays the true segmentation and the predicted segmentation, which are shown by the green line and blue line respectively.

Table 6.1: Quantitative evaluation of the results for cardiac MRI registration on the 2D ACDC data set. The following metrics are reported for each method: The Dice score *Dice* (mean± standard deviation), Hausdorff distance HD, the percentage of the number of pixels with negative Jacobian determinant $\%|J_\theta| < 0$, and reliability $R(0.75)$. Smaller values of HD and larger values of *Dice* indicate more accurate results. Also the smaller $\%|J_\theta| < 0$ indicates less mesh folding. The higher probability values of $R(0.75)$ show that more patients have the dice score higher or equal to %0.75. Values that are highlighted in bold indicate the metric that gave the best performance compared to the other algorithms.

Method	Dice	HD	$\%\|J_\theta\| < 0$	$R(0.75)$
Undeformed	0.71 ± 0.15	10.1	–	–
Demon [93]	0.76 ± 0.10	8.3	0.27	0.36
SyN [27]	0.80 ± 0.09	8.1	0.28	0.66
LPM [7]	0.79 ± 0.10	7.6	0.38	0.46
MM [30]	0.83 ± 0.15	5.64	0	0.81
Elastix [69]	0.84 ± 0.14	4.51	0.12	0.82
Proposed Method	**0.88 ± 0.11**	**3.85**	**0**	**0.89**

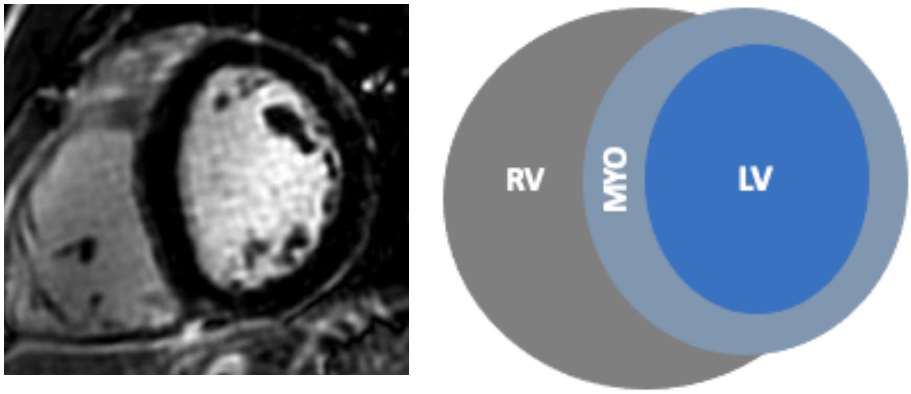


Figure 6.5: Example of cardiac anatomy: The left image demonstrates a sample MR short axis slice from the ACDC data set. The figure on the right displays the cardiac anatomy (LV = left ventricle, RV = right ventricle, MYO = myocardium) which are used for measuring the registration and segmentation accuracy.

6.3.3 Baseline Methods

We compared the performance of the proposed framework with state-of-the-art algorithms, SimpleElastix (Elastix) [69],(MM) [30, 94], Fast Symmetric Forces Demons

Table 6.2: Quantitative evaluation of the results for cardiac MRI registration on the 2D LA data set. The following metrics are reported for each method: The Dice score *Dice* (mean± standard deviation), Hausdorff distance HD, the percentage of the number of pixels with negative Jacobian determinant $\%|J_\theta| < 0$, and reliability $R(0.75)$. The 2ch, 3ch and 4ch stand for the 2, 3 and 4-chamber. Values in bold indicate the best performance.

(a) 2ch

Methods	Dice	HD	$\%\|J_\theta\| < 0$	$R(0.75)$
Undeformed	0.79 ± 0.07	7.37	–	–
Demons [71]	0.84 ± 0.08	7.41	0.38	0.89
SyN [27]	0.87 ± 0.06	**6.38**	0.18	0.95
MM [30]	0.84 ± 0.06	6.58	0	0.92
Elastix [69]	0.82 ± 0.11	7.28	0.28	0.74
Proposed Method	**0.88 ± 0.04**	6.54	**0**	**0.95**

(b) 3ch

Methods	Dice	HD	$\%\|J_\theta\| < 0$	$R(0.75)$
Undeformed	0.78 ± 0.08	7.70	–	–
Demons [71]	0.85 ± 0.06	7.33	0.36	0.94
SyN [27]	0.86 ± 0.13	7.52	0.21	0.93
MM [30]	0.83 ± 0.06	6.48	0	0.88
Elastix [69]	0.86 ± 0.10	6.82	0.26	0.9
Proposed Method	**0.87 ± 0.05**	**6.3**	**0**	**0.94**

(c) 4ch

Methods	Dice	HD	$\%\|J_\theta\| < 0$	$R(0.75)$
Undeformed	0.78 ± 0.09	8.66	–	–
Demons [71]	0.82 ± 0.10	7.84	0.43	0.77
SyN [27]	0.84 ± 0.11	7.51	0.20	0.86
MM [30]	0.83 ± 0.08	6.77	0	0.87
Elastix [69]	0.82 ± 0.10	7.56	0.38	0.64
Proposed Method	**0.87 ± 0.05**	**6.1**	**0**	**0.99**

Table 6.3: Quantitative evaluation of the results for cardiac MRI registration on the 2D SCD data set. The following metrics are reported for each method: The Dice score *Dice* (mean± standard deviation), Hausdorff distance HD, the percentage of the number of pixels with negative Jacobian determinant $\%|J_\theta| < 0$, and reliability $R(0.75)$. Smaller values of HD and larger values of *Dice* indicate more accurate results. Also the smaller $\%|J_\theta| < 0$ indicates less mesh folding. The higher probability values of $R(0.75)$ show that more patients have the dice score higher or equal to %0.75. Values that are highlighted in bold indicate the metric that gave the best performance compared to the other algorithms.

Method	Dice	HD	$\%\|J_\theta\| < 0$	$R(0.75)$
Undeformed	0.62 ± 0.15	16.02	–	–
Demons [71]	0.68 ± 0.18	12.46	0.4	0.36
SyN [27]	0.81 ± 0.16	8.9	0.02	0.70
LPM [7]	0.78 ± 0.08	7.6	0.38	0.63
MM [30]	0.72 ± 0.12	12.53	0	0.59
Elastix [69]	0.79 ± 0.08	11.12	0.37	0.62
Proposed Method	$\mathbf{0.88 \pm 0.09}$	**5.25**	**0**	**0.90**

(Demons) [71], Symmetric Normalization [27] which are optimization based methods and diffeomorphic learning-based methods LPM [7] and LapIRN [95].

2D Image Registration Results

Tables 7.1, 7.3 and 7.2 provide a summary of the results of the proposed method, the mean and standard deviations of DM, HD, the percentage of the number of pixels with negative Jacobian determinant $\%|J_\theta| < 0$, and reliability $R(0.75)$ on the held out test set on ACDC, LA, and SCD data sets, respectively. Figures 6.2, 6.3, 6.4 show samples of registered images on the LA, SCD, and ACDC data sets with the corresponding deformation field grid. End-systolic frame is the moving image and end-diastolic frame is the fixed image. The registered image of each row is shown in the third column. Also, the true and predicted segmentation maps are shown by the green and blue line respectively. For each new 2D pair of images, the registration process takes an average of 0.05 ± 0.03 seconds on a GPU.

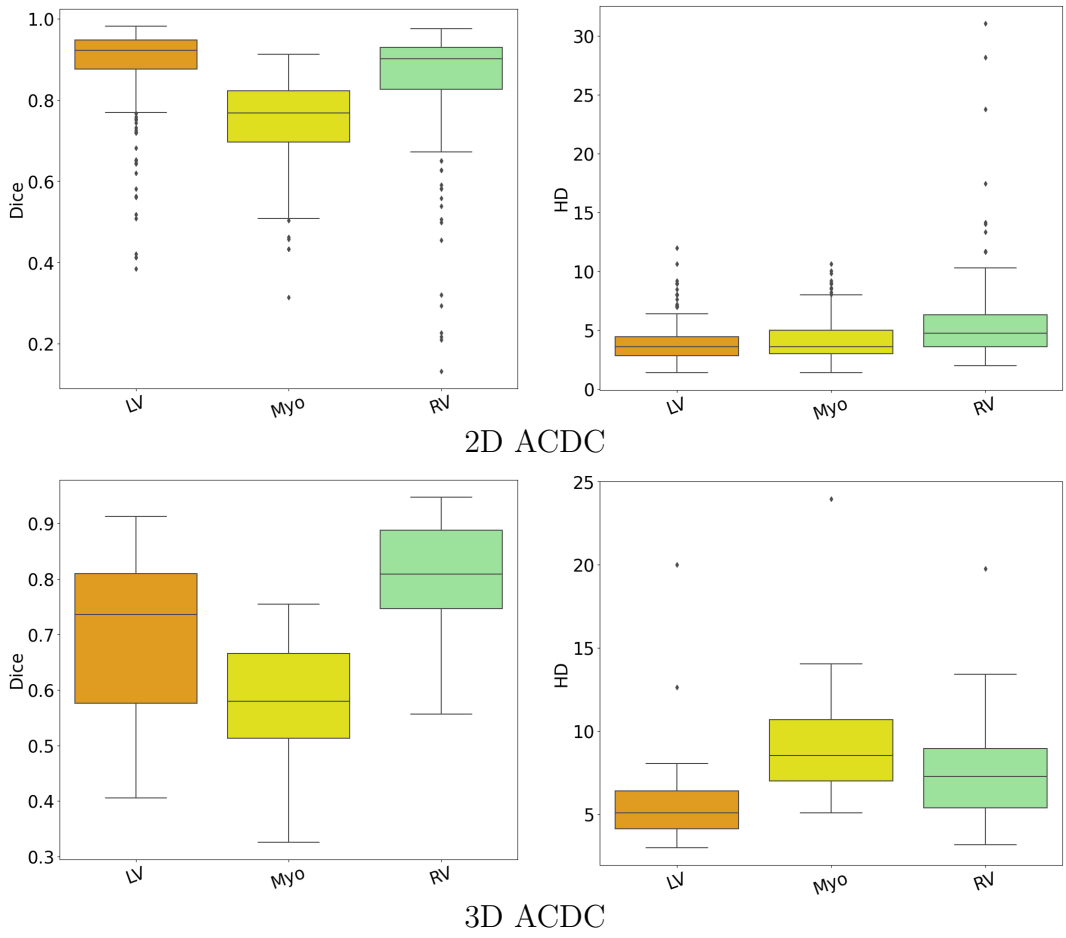


Figure 6.6: Boxplots indicating DM(mean± standard deviation) and HD (mean) for anatomical structures on the 2D and 3D ACDC dataset. The evaluation is performed over the myocardium (Myo), left ventricle (LV), and right ventricle (RV) anatomical structures.

The ACDC data set is originally a 3D data set where a set of 2D axial slices are stacked to form a 3D volume. To evaluate the 2D version of the proposed framework on ACDC, we computed 2D metrics on each slice separately and aggregated the results over all slices to obtain the final values reported in Table 7.1.

Since the ACDC data set offers different anatomical structures of the heart, myocardium (Myo), left ventricle (LV), and right ventricle (RV) Figure. 6.5. We also evaluated the proposed method on all three structures to show the reliability of the proposed method. Figure 6.6 illustrates the DM and HD results on different structures.

The presented method shows a better performance among the all compared methods in all aspects e.g., there is a noticeable difference between the obtain Dice score and Hausdorff distance. As can be seen, the improvement is not just limited to these two parameters, the Jacobian determinant is zero which means there is no folding or twisting in the transformation. This is in contrast to other methods where the determinant Jacobian is non-zero. Figure. 6.7 shows the end-diastolic and end-systolic images and the determinant of the Jacobian ($|J_\theta|$) with grid overlay for five example patients. As shown in all tables and Figure. 6.7, no negative values were observed on the test data for the proposed method which means our approach produced smooth and regular deformations.

Table 6.4: Quantitative evaluation of the results for cardiac MRI registration on the 3D ACDC data set. The following metrics are reported for each method: The Dice score *Dice* (mean± standard deviation), Hausdorff distance HD, the percentage of the number of pixels with negative Jacobian determinant $\%|J_\theta| < 0$, and reliability $R(0.75)$. Smaller values of HD and larger values of *Dice* indicate more accurate results. Also the smaller $\%|J_\theta| < 0$ indicates less mesh folding. The higher probability values of $R(0.75)$ show that more patients have the dice score higher or equal to %0.75. Values that are highlighted in bold indicate the metric that gave the best performance compared to the other algorithms.

Method	Dice	HD	$\%\|J_\theta\| < 0$	$R(0.75)$
Undeformed	0.71 ± 0.145	10.1	–	–
Demon [93]	0.80 ± 0.17	8.3	0.34	0.28
SyN [27]	0.80 ± 0.091	8.1	0.17	0.51
LPM [7]	0.81 ± 0.085	7.3	0.12	0.52
LapIRN [95]	0.72 ± 0.162	7.4	0	0.35
MM [94]	0.75 ± 0.156	7.03	0	0.56
Elastix [69]	0.83 ± 0.161	5.75	0.09	0.60
Proposed Method	**0.84 ± 0.06**	**5.3**	**0**	**0.78**

3D Image Registration results

The publicly available Automated Cardiac Diagnosis Challenge (ACDC) data set was employed for the evaluation of the proposed 3D-to-3D registration algorithm. Table 7.4 provides a summary of the results of the proposed method on the ACDC data set. The presented method displays a better performance among all the compared methods in all aspects e.g., there is a noticeable difference between the obtained Dice score and Hausdorff distance. Also, the higher probability values of $R(0.75)$ proves that the proposed method is more reliable than the other compared methods since more patients have the dice score higher or equal to %0.75. In addition, similar to 2D version, the Jacobian determinant is also zero in 3D version which means there is no mesh folding in the transformation. The registration process takes an average of 0.07 ± 0.005 seconds on a GPU to register an unseen 3D pair of images.

Figure 6.8 displays a correlation plot, where the ground truth volume in mL is plotted against the volume from the proposed method. The clustering of the dots to the reference yellow line indicates the high agreement between the proposed method to the ground truth. The analysis produced a Pearson correlation coefficient of 0.98.

Implementation and Parameters Analysis

The proposed method is implemented in Python programming language using Pytorch module. The network is designed based on a UNet-style architecture [33] which includes a convolutional layer with 16 filters, three downsampling layers with 32,64,64 convolutional filters and a stride of two, and upsampling convolutional layers with 64,64,32,32,32,16 filters. The Adam optimization with learning rate of 5×10^{-4} is used for all the three datasets. The proposed framework is evaluated on an NVIDIA GeForce GTX 1080 Ti GPU.

To guarantee the diffeomorphism and keep the transformation determinant Jacobian positive, different activation functions are used to apply constraints on μ and $V(\xi)$ and keep their range in $(0, 1)$ and $(-\lambda, +\lambda)$ respectively. Where λ can be any

value in range of $(1, \infty)$, we set $\lambda = 10$ in our experiment. Then using the two hyper-parameters, lower bound $\tau_{lb} \in (0, 1)$ and upper bound $\tau_{ub} \in (1, \lambda)$ of the transformation Jacobian determinant $|J_\theta|$, the user can control the amount of movement which directly affects the evaluation metrics. By increasing the values of τ_{lb} and τ_{ub}, each node in a grid (each pixel) can have a larger displacement; however, after a certain point, the results do not change significantly. We vary the precision τ_{lb}, τ_{ub} and set them to 0.2, 8.0 respectively. The chosen values resulted in the best Dice score and HD distance.

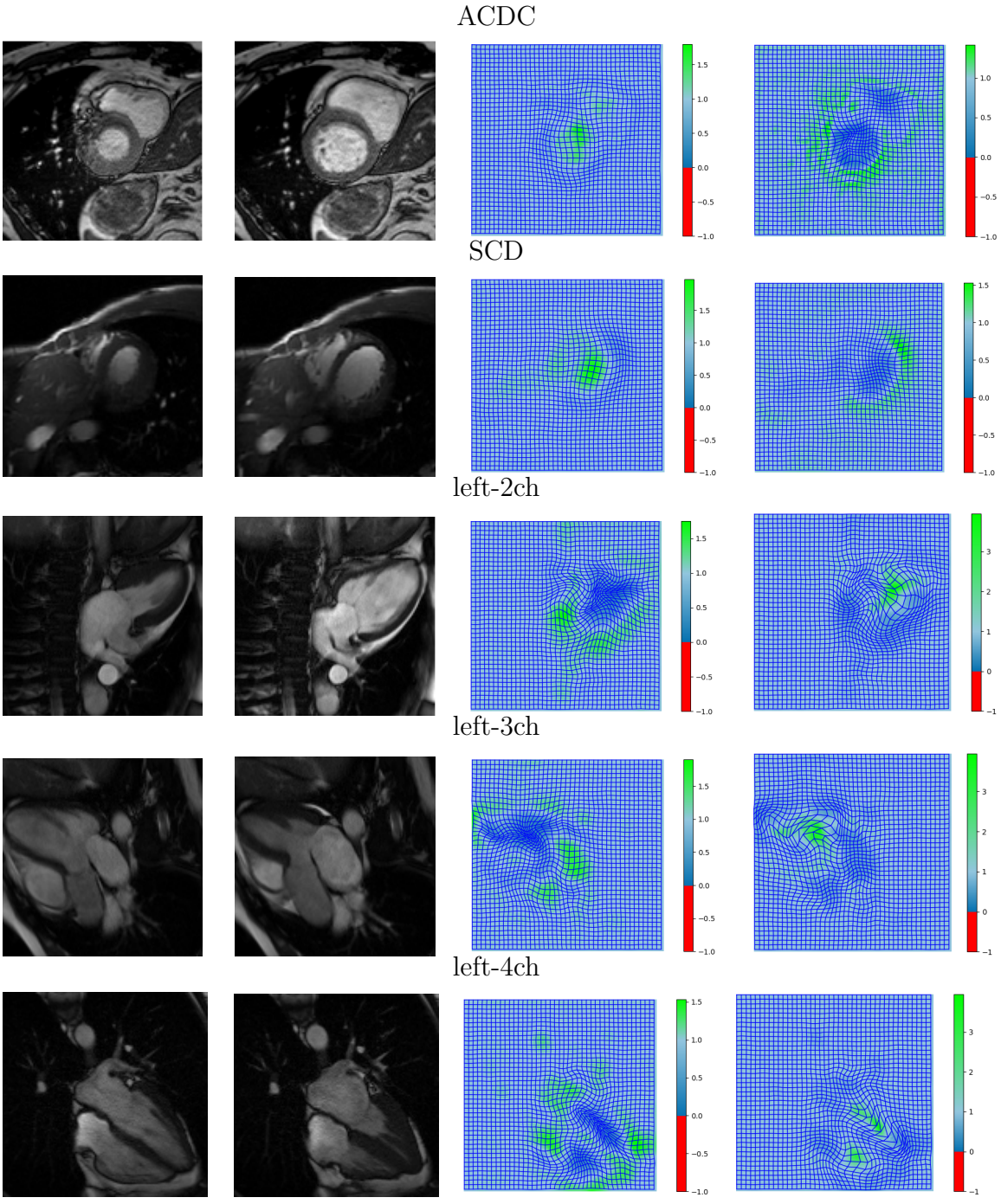


Figure 6.7: 2D registration results for five example patients, where the first column is the end-systolic image and the second column is the end-diastolic image. The grid deformations in the 3rd column displays the deformation from end-systole to end-diastole, while the last column displays the deformation from end-diastole to end-systole. The color represents the value of the Jacobian determinant, where red indicates values below 0, which is where mesh folding occurs. It can be seen that using the proposed method, no mesh folding occurs.

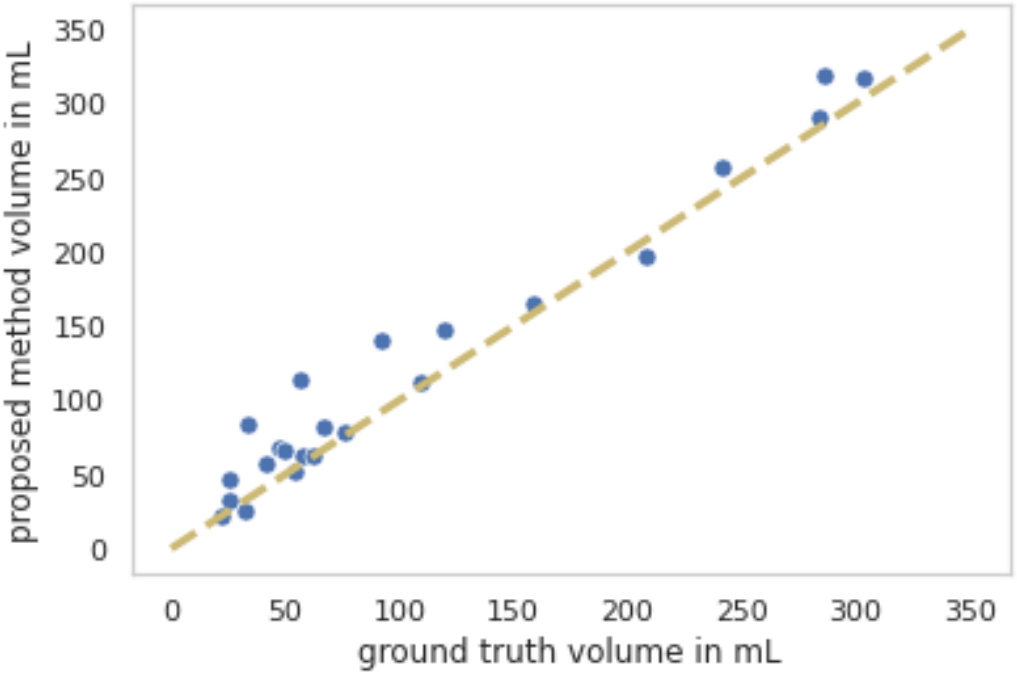


Figure 6.8: The ground truth volume in mL plotted against the volume from the proposed method, where each patient is represented by a blue dot. The yellow dotted line indicates the y=x line for reference. The Pearson correlation coefficient calculated is 0.98, revealing a high correlation of the proposed method to the ground truth.

Acknowledgment

The authors wish to thank Alberta Innovates for the AICE Concepts funding that supported this research work.

6.4 Conclusion

In this work, we build a principled connection between classical registration methods and recent learning-based approaches. We propose an end-to-end framework for diffeomorphic image registration and derive a learning algorithm that leverages a convolutional neural network and unsupervised learning for fast runtime. To achieve diffeomorphic transforms, we integrate a new parameterization of deformation fields for 2D-to-2D and 3D-to-3D diffeomorphic registration algorithm for the application of MRI cardiac registration.Which describes a deformation field with its transformation Jacobian determinant and curl of the end velocity field. It also relaxes the need for an explicit regularization to produce a physically plausible result because smoothness is implicitly embedded in the solution. Removing explicit regularization makes the need for an empirical trade-off between the similarity term and the regularization term, which may cause bias [96], unnecessary.

Furthermore, by directly requiring the transformation Jacobian to be positive, the deformation can be ensured to be diffeomorphic. The other desirable constraints also can be enforced within the same framework using an explicit restriction on the transformation Jacobian such as incompressibility constraint.

Additionally, the proposed parameterization naturally describes a deformation field in terms of radial and rotational components, making it especially suited for processing cardiac data [97]. We also provide an anatomical surface deformation model. If image segmentations are available for a particular anatomical structure, the model incorporates them naturally in the same joint framework.

Our algorithm can infer the registration of new image pairs in under a second.

Compared to traditional methods, our approach is significantly faster, and compared to recent learning-based methods, our method offers diffeomorphic guarantees. We demonstrate that the surface extension to our model can help improve registration while preserving properties such as low runtime and diffeomorphisms. The proposed algorithm was evaluated on end-diastole to end-systole cardiac cine-MRI registration on two publicly available ACDC Challenge [98] and Sunnybrook datasets (SCD)[68] as well as a set of Left Atrium obtained from the Mazankowski Alberta Heart Institute. We compared registration performance in terms of DICE and Hausdorff distances to six other registration methods, Symmetric Normalization diffeomorphic registration from the Dipy package [27], two versions of the Demons algorithm (classical and fast symmetric forces) from ITK [93], the Elastix software package [99, 100], a learning-based framework voxcelMorph [7] and optimization-based registration MM. The proposed algorithm is diffeomorphic, allowing it to capture the true deformation of the cardiac tissue. Observing the percentage of voxels with a Jacobian determinant less than zero, all of the other registration methods yielded mesh folding for either the MRI dataset. The presence of mesh folding may result in the inability of these methods to capture the true anatomical motion.

6.5 Conclusion

In this work, we build a principled connection between classical registration methods and recent learning-based approaches. We propose an end-to-end framework for diffeomorphic image registration and derive a learning algorithm that leverages a convolutional neural network and unsupervised learning for fast runtime. To achieve diffeomorphic transforms, we integrate a new parameterization of deformation fields for 2D-to-2D and 3D-to-3D diffeomorphic registration algorithm for the application of MRI cardiac registration, which describe a deformation field with its transformation Jacobian determinant and curl of the end velocity field. It also relaxes the need for an explicit regularization to produce a physically plausible result, as smoothness is

implicitly embedded in the solution. Removing explicit regularization makes the need for an empirical trade-off between the similarity term and the regularization term, which may cause bias [96], unnecessary.

Furthermore, by directly requiring the transformation Jacobian to be positive, the deformation can be ensured to be diffeomorphic. The other desirable constraints also can be enforced within the same framework using an explicit restriction on the transformation Jacobian such as incompressibility constraint. Additionally, the proposed parameterization naturally describes a deformation field in terms of radial and rotational components, making it especially suited for processing cardiac data [97]. Our algorithm can infer the registration of new image pairs in under a second, which is significantly faster than traditional iterative methods. Compared to recent learning-based methods, our method offers a guarantee of a diffeomorphic transform.

The proposed algorithm was evaluated on end-diastolic to end-systolic cardiac cine-MRI registration on two publicly available ACDC Challenge [98] and Sunnybrook data sets (SCD) [68] as well as a set of left atrium images obtained from the Mazankowski Alberta Heart Institute. The proposed algorithm is diffeomorphic, allowing it to capture the true deformation of the cardiac tissue. Observing the percentage of voxels with a Jacobian determinant less than zero, most of the other registration methods yielded mesh folding for either the MRI data sets. The presence of mesh folding may result in the inability of these methods to capture the true anatomical motion.

Chapter 7

Diffeomorphic Image Registration for the Application of Cardiac Image Segmentation

7.1 Overview

Cardiac segmentation from magnetic resonance imaging (MRI) is one of the essential tasks to analyze the anatomy and function of the heart for the assessment and diagnosis of cardiac diseases. However, cardiac MRI generates hundreds of images per scan and manual annotation of them is difficult and time consuming, and therefore processing these images automatically is of interest. This study proposes a novel end-to-end supervised cardiac MRI segmentation framework based on a diffeomorphic deformable registration that can segment the left ventricle from 2D and 3D images or volumes. In order to represent the actual cardiac deformation, the methodology parameterizes the transformation using radial and rotational components, computed using a deep learning approach. The deep learning method is trained using a set of pair images along with their segmentation masks. The formulation guarantees transformations that are invertible and prevents mesh folding which is essential for preserving the topology of the segmentation results. A physically plausible transformation is achieved by employing diffeomorphism in computing the transformations and activation functions that constrain the range of the radial and rotational components. The method was

evaluated over three different data sets and showed significant improvements compared to exacting learning and non-learning based methods in terms of the Dice score and Hausdorff distance metrics.

7.2 Methodology

The objective of this study is to enhance the accuracy of a supervised 2D and 3D segmentation framework using a diffeomorphic non-rigid image registration algorithm 6. The method computes a point-to-point mapping between two images or volumes using a convolutional neural network while preserving the topology.

Finding the optimum mapping between a pair of images can be formulated as follows:

$$\phi^* = \underset{\phi}{\operatorname{argmin}}\, L(I_2, I_1 \circ \phi(\xi)) \tag{7.1}$$

where $L(.)$ is dissimilarity metric, $\xi \in \Omega$ denotes the pixel location in the image domain Ω, $\phi : \Omega \rightarrow \Omega$ is the transformation function, I_1 is template image and I_2 is the image or volume that need to be segmented. To have a unique solution for the above formulation and more constrained, we define a deformation field using a monitor function μ where $\mu : \Omega \rightarrow \mathbb{R}$ and $\int_\Omega \mu = |\Omega|$ and curl of end velocity field $v : \Omega \rightarrow \mathbb{R}$.

The objective here is to find a transformation ϕ: $\Omega \rightarrow \Omega$, $\partial\Omega \rightarrow \partial\Omega$ so that the Jacobian determinant of the transformation $J_\phi(\xi)$ is equal to the monitor function μ :

$$J_\phi(\xi) = det\nabla\phi(\xi) = \mu(\xi). \tag{7.2}$$

To find the transformation ϕ which satisfies 7.2, we first compute a vector field $\gamma(\xi)$ such that:

$$div\, \gamma(\xi) = \mu(\xi) - 1. \tag{7.3}$$

and then a velocity vector field based on artificial-time is constructed from $\gamma(\xi)$:

$$\gamma_t(\xi) = \frac{\gamma(\xi)}{t + (1-t)\mu(\xi)}, t \in [0,1] \tag{7.4}$$

The desired transformation ϕ can be found by solving the following ordinary differential equation (ODE) at $t = 1$, $\phi(\xi) = \psi(\xi, t = 1)$ and $\psi(\xi$,t=0$) = \phi_0(\xi)$

$$\frac{\psi(\xi,t)}{dt} = \gamma_t(\psi(\xi,t)), t \in [0,1], \tag{7.5}$$

Where $\phi_0(\xi)$ is the identity mapping and $det\nabla\phi_0(\xi) = 1$ and $\phi_0(\xi) = \xi$.

The main problem is how to find $\gamma(\xi)$ such that $div\gamma(\xi) = \mu(\xi) - 1$. To solve this problem we use the div-curl system, by finding the divergence and curl at each point and set up the div-curl system of equations for each point which leads to reconstruct a differentiable and invertible transformation.

$$\begin{cases} div\gamma(\xi) = \mu(\xi) - 1 \\ curl\gamma(\xi) = v(\xi). \end{cases} \tag{7.6}$$

To have a unique ϕ the Dirichlet boundary condition is used [92]. Additionally, a diffeomorphism, which corresponds to a positive transformation Jacobian determinant, is enforced explicitly via the monitor function [91].

7.2.1 Segmentation Framework

Let I_1 and I_2 be 2D/3D pair images/volumes, defined over $\Omega \rightarrow \mathbb{R}^2/\Omega \rightarrow \mathbb{R}^3$ and S_1 and S_2 are their corresponding segmentation maps where each pixel/voxel is assigned to a desired anatomical label or background. As shown in Figure 7.1, the presented segmentation framework includes three steps as follows.

- First, the proposed network $g_\theta(I_1, I_2)$ which is designed based on the UNet-style architecture [33] takes I_1 and I_2 along with S_1 and S_2 as inputs and outputs $\mu(\xi)$, and $\gamma(\xi)$.

- Then, using the moving mesh approach, we compute v, which is the curl$\gamma(\xi)$, and the desire ϕ. Where the following constraints are imposed on $\mu(\xi)$ to guarantee the diffeomorphism:

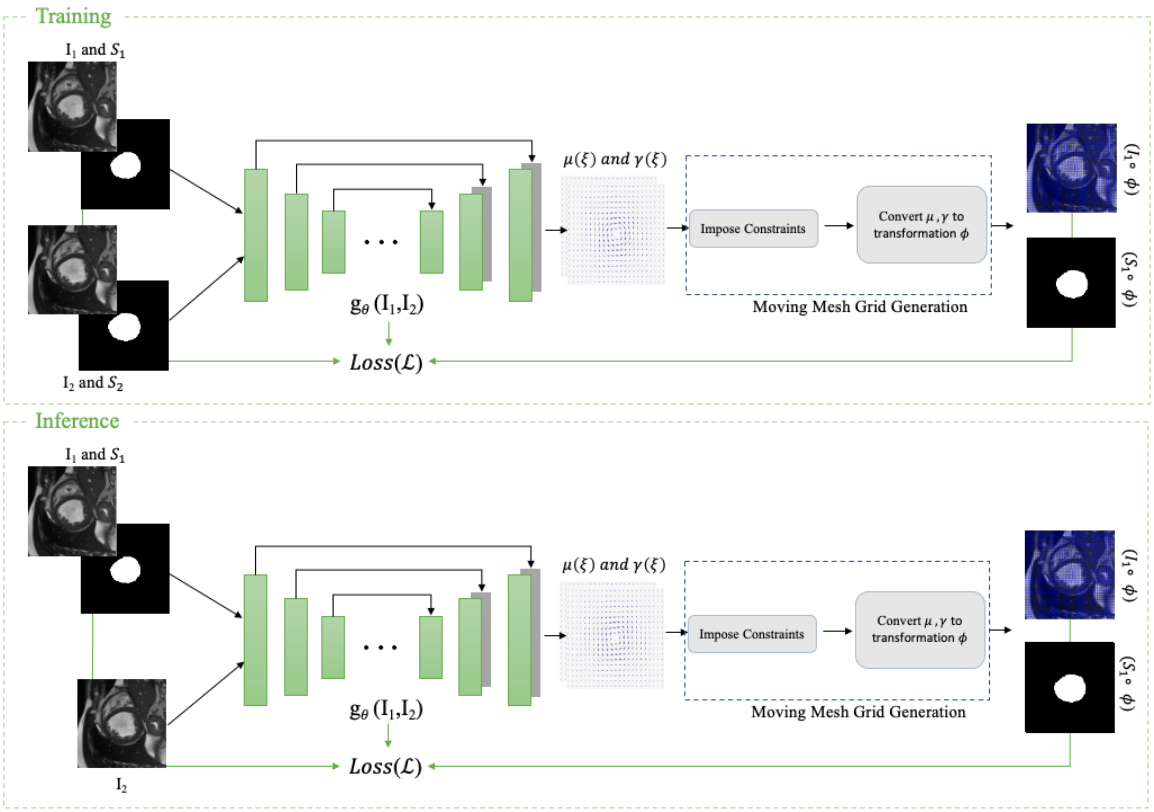


Figure 7.1: Overview of supervised architecture. During the training phase, a pair of images I_2 and I_1 with their correspondence segmentation maps (S_2) and (S_1) are passed to the CNN, $g_\theta(I_2, I_1)$ as inputs. The CNN then outputs the radial and rotational parameters. Then using the moving mesh grid generation method the transformation ϕ is computed. Finally, the segmentation map and source image are warped using the predicted ϕ and generate the desired segmentation $S_1 \circ \phi$. In the inference phase, only the segmentation map corresponding to the template image I_1 is passed to the framework and it outputs the desire segmentation map S_2^* for I_2 using $S_1 \circ \phi$.

$$\begin{cases} \int \mu(\xi) d\xi = |\Omega| \\ \tau_{ub} > \mu(\xi) > \tau_{lb} \end{cases} \tag{7.7}$$

where the τ_{ub} is the upper bound of the transformation Jacobian determinant, and τ_{lb} is the lower bound which were set by the user. The $\tau_{lb} > 0$ guarantees the diffeomorphism.

- Finally, the segmentation map S_1 is warped according to the computed transformation ϕ, $S_1 \circ \phi$ and desire S_2^* is generated.

In summary, during the training phase, the network takes I_1 and I_2 images/volumes, which are the end-diastolic and end-systolic frames of the cardiac sequence, along with their corresponding segmentation maps S_1 and S_2. The network is trained to obtain the point-to-point correspondence using the diffeomorphic transformation which is then used to compute the segmentation map at end-systole using the segmentation map at end-diastole. Since all the steps are designed to be differentiable, the network parameters are learnt using a stochastic gradient descent-based method with the following loss function:

$$L(\phi; I_2, S_2, I_1, S_1) = NCC(I_2, I_1 \circ \phi) + Dice(S_2, S_1 \circ \phi) \tag{7.8}$$

where NCC is normalized cross correlation metric and Dice is the Dice metric. As shown in Eq. 7.8, to improve the accuracy of the segmentation we measure the difference between images/volumes as well as segmentation maps.

In the inference phase we only pass the segmentation map corresponding to the template image I_1 to the framework and it outputs the desire segmentation map S_2^* for I_2 using $S_1 \circ \phi$.

7.3 Experiments

We perform a series of experiments to evaluate the accuracy of the proposed supervised segmentation framework against the state-of-the-art algorithms, SimpleElastix (Elastix)[69], Moving Mesh (MM)[30, 94], Fast Symmetric Forces Demons (Demons)[71], LCC-Demons[32], Symmetric Normalization[27] which are well known optimization based algorithms and diffeomorphic learning-based methods LPM [7], LapIRN [95], and Unsupervised diffeomorphic cardiac registration (UDCR) 6. To assess the performance of the 2D version of our method, the evaluations were performed over three data sets consisting of clinical 2D cardiac MR images. We also evaluated the 3D version of the proposed framework using ACDC data set in 3D.

7.3.1 Data sets

The following three data sets are considered in this study:

Automated Cardiac Diagnosis Challenge (ACDC)[66] This data set contains multiple temporal short-axis cardiac cine MRI sequences acquired from 100 patients and is one of the publicly available data sets for cardiac MRI assessment. The spatial resolution varies from $1.37mm^2/pixel$ to $1.68mm^2/pixel$ with a slice thickness of 5 mm to 8 mm. The images are cropped to the size of 128×128 as 2D images and $128 \times 128 \times 16$ as 3D volume. These dimensions were chosen to save computation time and are not a limitation of the framework.

The Sunnybrook Cardiac Challenge data (SCD)[68] This data set contains multiple temporal 2D short-axis cardiac cine MRI scans acquired from 45 patients. The data set is equally divided into 15 patient scans for training, 15 patient scans for validation, and 15 patient scans for testing. The image resolution is 256×256, with a pixel spacing of 1.25 mm and slice thickness of 8 mm.

Left Atrium (LA) This data set includes 100 temporal 2D long-axis cine MRI steady-state sequences from the 2, 3 and 4-chamber views. It is acquired from the University Alberta Hospital. Image resolutions 176×189 – 256×208 and image spacing $1.445 - 1.795$ mm. The 2ch, 3ch and 4ch are used in the rest of the paper to denote 2, 3 and 4-chamber sequences, respectively.

7.3.2 Quantitative Evaluation Metrics

The proposed method is evaluated quantitatively using four metrics, namely, Dice score, Hausdorff distance (HD in mm), determinant of Jacobian of the deformation field $\det(J)$, and reliability $R(d)$.

Dice Score The Dice Score [73] is a well-known segmentation-based metric to measure the similarity (overlap) between two regions, warped moving and fixed images.

The Dice score of two regions A and B is formulated as:

$$Dice(A, B) = \frac{2|A \cap B|}{A + B} \tag{7.9}$$

Hausdorff Distance The HD [74] is another well-known metric which measures the maximum deviation between two regions' contours. The HD between two contours (C_A) and C_B is formulated as:

$$\begin{aligned} \mathrm{HD}(C_A, C_B) = &\max(\max_i(\min_j(d(p_A^i, p_B^j))), \\ &\max_j(\min_i(d(p_A^i, p_B^j)))) \end{aligned} \tag{7.10}$$

where p_A^i, p_B^j denote the set of all the points in C_A and C_B respectively. The term $d(\cdot)$ denotes the Euclidean distance.

Reliability: We also evaluated the performance of the proposed algorithm using a reliability function computed based on Dice scores for each data set. The complementary cumulative distribution function is defined for each $d \in [0, 1]$ as the probability of obtaining Dice score higher than d overall volumes.

$$\begin{aligned} R(d) &= P_r(Dice > d) \\ &= \frac{\#\text{ Images segmented with Dice score higher than d}}{\text{total number of images}}. \end{aligned} \tag{7.11}$$

$R(d)$ measures how reliable the algorithm is in yielding accuracy d.

7.3.3 Segmentation Results

2D Segmentation Results

The summary of the results, the mean and standard deviations of Dice score, HD and reliability $R(0.75)$ on ACDC, LA, and SCD data sets are reported in tables 7.1, 7.2, and 7.3 respectively. Figures 7.2, 7.3, and 7.4 respectively show samples of segmentation results on the ACDC, SCD, and LA data sets. The end-diastolic frame is I_1 and the end-systolic frame is I_2. The true and predicted segmentation maps are shown by the green and red lines respectively. The time for segmenting an unseen 2D image is 0.03 ± 0.007 seconds on a GPU.

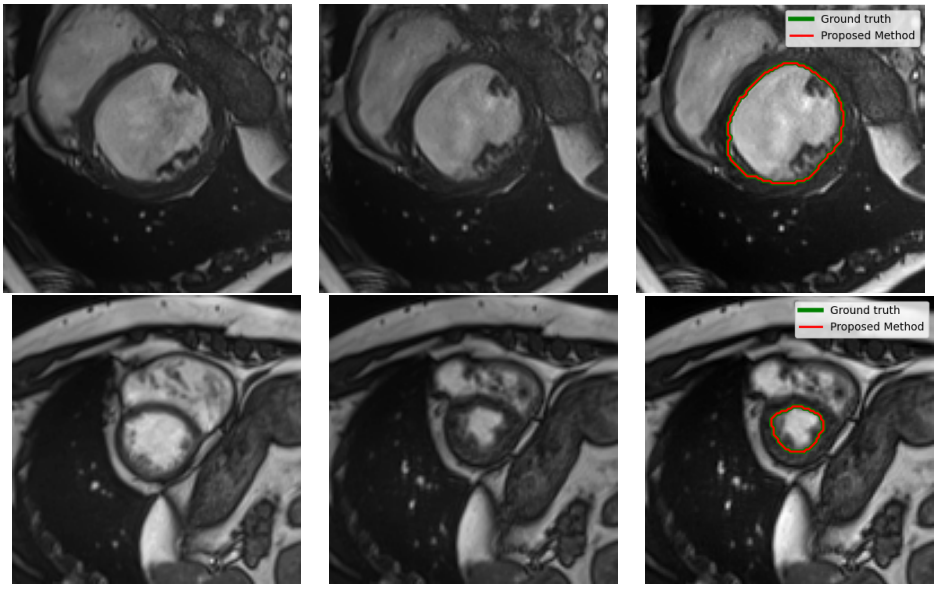


Figure 7.2: Samples of registered images on the ACDC with the correspondence segmentation. End-diastolic (ED) frame is I_1 and end-systolic (ES) frame is I_2. The true segmentation and predicted segmentation are shown by the green line and red line respectively.

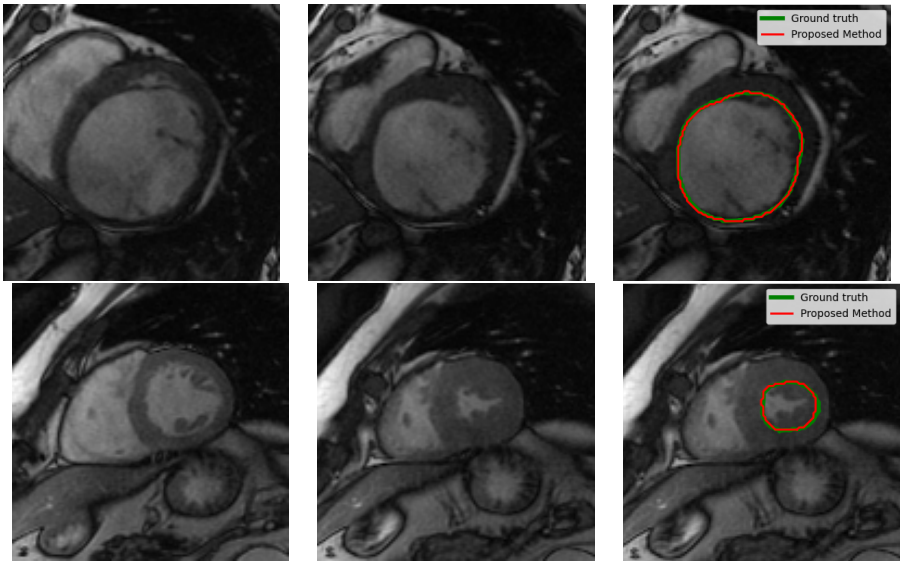


Figure 7.3: Samples of registered images on the SCD with with the correspondence segmentation. End-diastolic (ED) frame is I_1 and end-systolic (ES) frame is I_2. The true segmentation and predicted segmentation are shown by the green line and red line respectively.

Table 7.1: Quantitative evaluation of the results for cardiac MRI registration on the 2D ACDC data set. The following metrics are reported for each method: The Dice score $Dice$ (mean± standard deviation), and Hausdorff distance HD, and reliability $R(0.75)$. Smaller values of HD and larger values of $Dice$ indicate more accurate results. The higher probability values of $R(0.75)$ show that more patients have the dice score higher or equal to %0.75. Values that are highlighted in bold indicate the metric that gave the best performance compared to the other algorithms.

Method	Dice	HD	$R(0.75)$
Undeformed	0.71 ± 0.145	10.1	–
Demon[93]	0.76 ± 0.096	8.3	0.36
SyN[27]	0.80 ± 0.091	8.1	0.66
LPM[7]	0.79 ± 0.095	7.6	0.46
MM[30]	0.83 ± 0.153	5.64	0.81
Elastix[69]	0.84 ± 0.138	4.51	0.82
UDCR6	0.88 ± 0.11	3.85	0.89
Proposed Method	**0.98 ± 0.03**	**0.35**	**0.99**

Table 7.2: Quantitative evaluation of the results for cardiac MRI registration on the 2D SCD data set. The following metrics are reported for each method: The Dice score $Dice$ (mean± standard deviation), Hausdorff distance HD, and reliability $R(0.75)$. Smaller values of HD and larger values of $Dice$ indicate more accurate results. The higher probability values of $R(0.75)$ show that more patients have the dice score higher or equal to %0.75. Values that are highlighted in bold indicate the metric that gave the best performance compared to the other algorithms.

Method	Dice	HD	$R(0.75)$
Undeformed	0.62 ± 0.15	16.02	–
Demons[71]	0.68 ± 0.18	12.46	0.36
SyN[27]	0.81 ± 0.16	8.9	0.70
LPM[7]	0.78 ± 0.08	7.6	0.63
MM[30]	0.72 ± 0.12	12.53	0.59
Elastix[69]	0.79 ± 0.08	11.12	0.62
UDCR6	0.88 ± 0.09	5.25	0.90
Proposed Method	**0.92 ± 0.10**	**3.18**	**0.96**

Since the ACDC data set is originally a 3D data set where a set of 2D axial slices are stacked to form a 3D volume, we computed 2D metrics on each slice separately and aggregated the results over all slices to obtain the final values, reported in Table 7.1.

The presented method has shown significantly better performance among other compared methods in all aspects. The Dice score and HD of ACDC data set have been improved by 10% and 3% respectively. For LA data set, the proposed method yielded an improvement of 10%, 11%, and 11% for Dice score and 5%, 4%, and 4% for HD for 2-chamber, 3-chamber, and 4-chamber, respectively. The accuracy of the results by the proposed method over the SCD data set has also been improved by 4% for Dice score and 2% for HD.

3D Segmentation Results

Table. 7.4 provides the summary of the mean and standard deviations of Dice score, HD and reliability $R(0.75)$ metrics on 3D ACDC. Similar to the 2D version the end-systolic frame and end-diastolic frame are used as the moving and fixed images respectively. The time to segment an unseen 3D volume is 0.07 ± 0.001 seconds on a GPU. As it can be seen in Table 7.4, the proposed method improved the segmentation results by 8% in Dice score and 2% in HD.

Analysis of cardiac data requires the calculation of clinical metrics which are crucial for diagnosis. One such metric is the volume of the LV, which when calculated at the end-diastolic and the end-systolic frames can be used to calculate the ejection fraction, an indication of the ability of the heart to pump blood. The Bland-Altman plot can be used to describe the agreement between the ground truth volume and the proposed method volume, as seen in Figure 7.5. The blue dots represent each patient volume, and the red line indicates the bias, or how much underestimation or overestimation of the volume occurs. The two yellow lines indicate the limits of agreement, which are two standard deviations from the bias. In this case the bias

was calculated to be -0.539 mL, indicating a slight overestimation of the volume by the proposed method compared to the ground truth. It can be seen that almost all patients fall within the limits of agreement, demonstrating the high agreement of the proposed volume to the ground truth.

The correlation can also be calculated between the ground truth and the proposed volume, as demonstrated in Fig 7.6. Each patient is represented by a blue dot and the reference y=x line is shown by a yellow dotted line. The resulting correlation is 0.999 indicating a high agreement between the two sets of volumes.

7.3.4 Implementation and Parameters Analysis

The proposed method is implemented in Python programming language using Pytorch module. The network includes three downsampling layers with 32,64,64 convolutional filters and a stride of two, and 5 upsampling convolutional layers with 64,64,32,32,16 filters. The neural net framework is evaluated on an NVIDIA GeForce GTX 1080 Ti GPU. Different activation functions are used to apply constraints on μ and γ and keep their range in $(0, 1)$ and $(-\lambda, +\lambda)$ respectively. Where λ can be any value in range of $(1, \infty)$, we set $\lambda = 10$ in our experiment. The two main hyper-parameters, lower bound $\tau_{lb} \in (0, 1)$ and upper bound $\tau_{ub} \in (1, \infty)$ of the transformation Jacobian determinant $|J_\theta|$ which are set by the user can control the diffeomorphism and the amount of movement which directly affects the evaluation metrics. By increasing the values of τ_{lb} and τ_{ub}, each node in a grid (each pixel) can have a larger displacement; however, after a certain point, the results do not change significantly. We vary the precision τ_{lb}, τ_{ub} and set them to 0.2, 8.0 respectively. The chosen values resulted in the best Dice score and HD distance.

Table 7.3: Quantitative evaluation of the results for cardiac MRI registration on the 2D LA data set. The following metrics are reported for each method: The Dice score *Dice* (mean± standard deviation), Hausdorff distance HD, and reliability $R(0.75)$. Values in bold indicate the best performance.

(a) 2ch

Methods	Dice	HD	$R(0.75)$
Undeformed	0.79 ± 0.07	7.37	–
Demons[71]	0.84 ± 0.08	7.41	0.89
SyN[27]	0.87 ± 0.06	6.38	0.95
MM[30]	0.84 ± 0.06	6.58	0.92
Elastix[69]	0.82 ± 0.11	7.28	0.74
UDCR6	0.88 ± 0.04	6.54	0.95
Proposed Method	**0.98 ± 0.01**	**1.28**	**1.0**

(b) 3ch

Methods	Dice	HD	$R(0.75)$
Undeformed	0.78 ± 0.08	7.70	–
Demons[71]	0.85 ± 0.06	7.33	0.94
SyN[27]	0.86 ± 0.13	7.52	0.93
MM[30]	0.83 ± 0.06	6.48	0.88
Elastix[69]	0.86 ± 0.10	6.82	0.9
UDCR6	0.87 ± 0.05	6.3	0.94
Proposed Method	**0.98 ± 0.02**	**1.44**	**1.0**

(c) 4ch

Methods	Dice	HD	$R(0.75)$
Undeformed	0.78 ± 0.09	8.66	–
Demons[71]	0.82 ± 0.10	7.84	0.77
SyN[27]	0.84 ± 0.11	7.51	0.86
MM[30]	0.83 ± 0.08	6.77	0.87
Elastix[69]	0.82 ± 0.10	7.56	0.64
UDCR6	0.87 ± 0.05	6.1	0.99
Proposed Method	**0.98 ± 0.01**	**1.5**	**1.0**

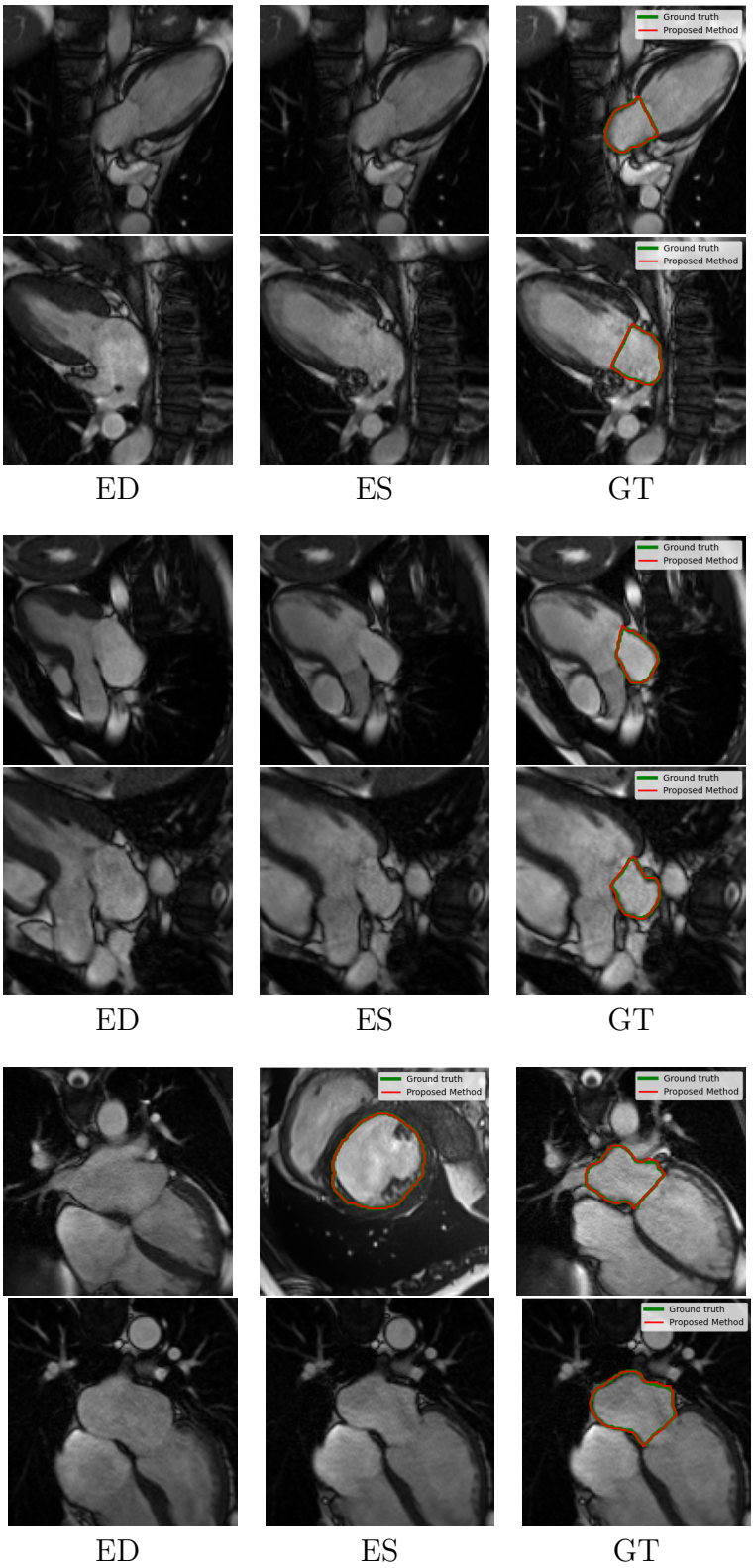


Figure 7.4: Samples of registered images on the left atrium data set with the correspondence segmentation. The first two rows are 2ch, the next rows are 3ch and the last tow rows are 4ch views. End-diastolic (ED) frame is I_1 and end-systolic (ES) frame is I_2. The true segmentation and predicted segmentation are shown by the green line and red line respectively.

Table 7.4: Quantitative evaluation of the results for cardiac MRI registration on the 3D ACDC data set. The following metrics are reported for each method: The Dice score *Dice* (mean± standard deviation), Hausdorff distance HD, and reliability $R(0.75)$. Smaller values of HD and larger values of *Dice* indicate more accurate results. The higher probability values of $R(0.75)$ show that more patients have the dice score higher or equal to %0.75. Bold values indicate the best performance.

Method	Dice	HD	$R(0.75)$
Undeformed	0.71 ± 0.145	10.1	–
Demon[93]	0.80 ± 0.17	8.3	0.28
SyN[27]	0.801 ± 0.091	8.1	0.51
LPM[7]	0.812 ± 0.085	7.3	0.52
LapIRN [95]	0.74 ± 0.162	7.4	0
MM[94]	0.75 ± 0.156	7.03	0
Elastix[69]	0.83 ± 0.161	5.75	0.60
UDCR [101]6	0.84 ± 0.06	5.3	0.78
Proposed Method	**0.92 ± 0.04**	**3.3**	**1.0**

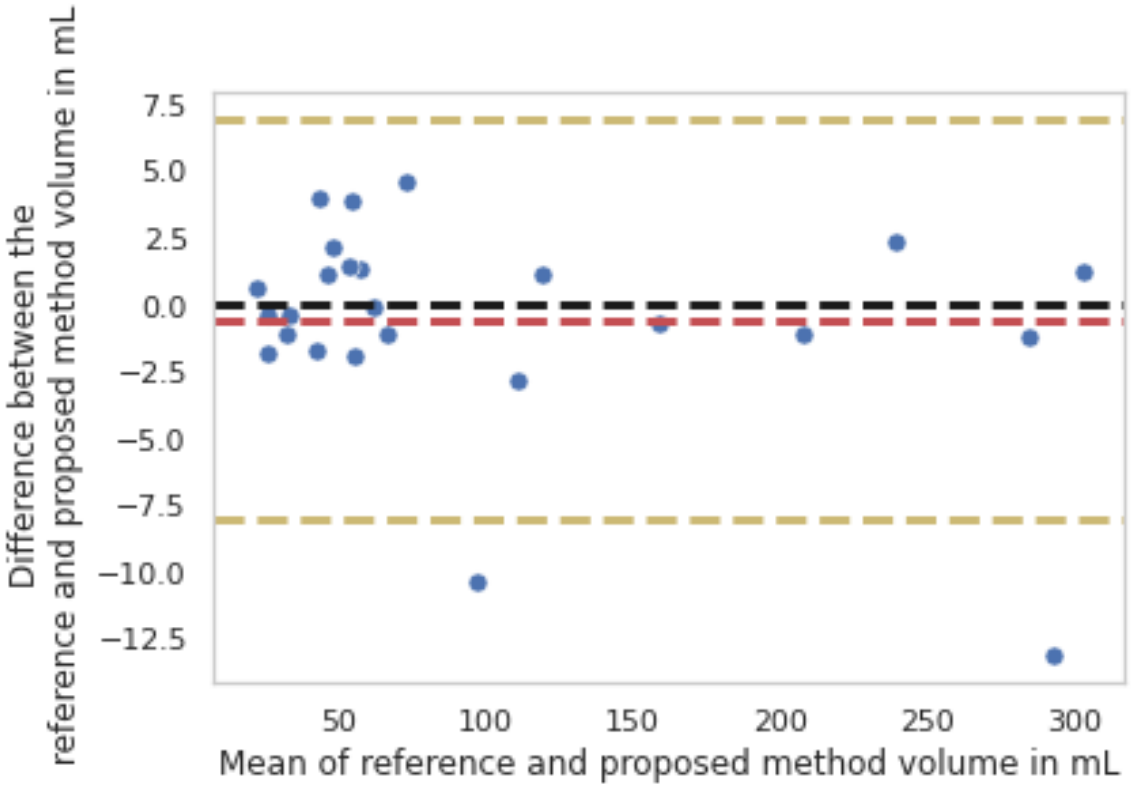


Figure 7.5: The figure on the left displays the Bland-Altman plot for the 3D segmentation results on the ACDC dataset. The blue dots represent each of the patients analyzed, and the bias is represented by the red dotted line and the limits of agreement are represented by the yellow dotted lines. The bias is -0.539 mL which indicates a slight overestimation of the volume of the chamber.

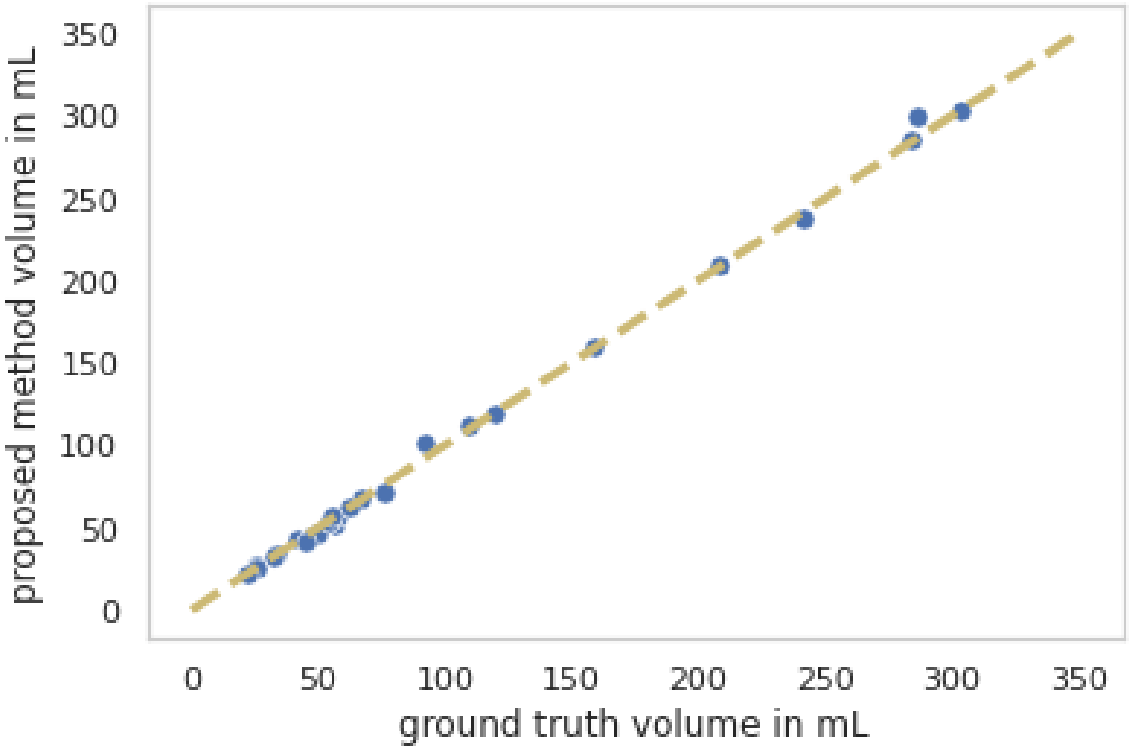


Figure 7.6: This figure displays correlation between the ground truth and the proposed method volume, where each blue dot represents a patient and the yellow dotted line is the reference y=x line. The resulting correlation is 0.999 indicating high agreement between the ground truth and the proposed method volumes.

7.4 Conclusion

We propose an end-to-end supervised framework for 2D and 3D cardiac MRI segmentation. To preserve the topology and find the points correspondence, we used a diffeomorphic deformable image registration algorithm. Which formulate a transformation in terms of radial and rotational components and make it suitable for processing cardiac imaging data.

The proposed algorithm was evaluated on cardiac cine-MRI images of two publicly available data sets as well as a set of left atrial images obtained from the Mazankowski Alberta Heart Institute against ground truth segmentation by expert clinicians. The reported results demonstrated significant improvement of proposed method in comparison to the existing learning and non-learning based methods.

Acknowledgment

The authors wish to thank Alberta Innovates for the AICE Concepts funding that supported this research work.

Chapter 8

Conclusion

8.1 Summary

Cardiac deformable registration is a fundamental step in quantitatively assessing cardiac function and generating diagnostic measurements. The interpretations of medical data are mostly being done by medical experts that are quite limited due to their subjectivity, the complexity of the image, extensive variations exist across different interpreters, and fatigue. Therefore, there is a strong need for automated algorithms for medical image processing and my goal is to develop deep learning approaches to meet the demands of this emerging area of the healthcare technology sector. The lack of reliable and automatic tools significantly limits the usage of registration which is heavily underutilized in the course of clinical treatment. Several medical media are required to provide accurate and complete information about a patient. For example, each CT, MRI, and ultrasound images provide different information in which the combination of them will be useful for the treatment. At the same time, medical image registration techniques serve as the fundamental basis for procedures such as image-guided radiation therapy, image-guided radiation surgery, and image-guided minimally invasive treatments.

The main goal of this thesis was to develop automated diffeomorphic methods to address the shortcomings listed above. Particularly improving the accuracy and robustness of image registration. In chapter 3 and chapter 4, we proposed two training-free

deformable image registration frameworks that eliminated the need for a dedicated training set while exploiting the capabilities of neural networks. Both are suitable for group-wise cardiac MRI registration. In chapter 5, a training-free multi-resolution framework for diffeomorphic deformable image registration was proposed. The proposed recursive algorithm also eliminates the need for dedicated training data. Since the presented framework was formulated in the diffeomorphic setting, it is suitable for pairwise registration. All the proposed methods in chapters 3, 4, 5 achieved the highest accuracy in comparison to learning and non-learning based methods. As these methods are iterative, they are suitable for the problems with the lack of suitable training data. However, they are not qualified for real-time applications. To address the above issues, in chapter 6, we proposed an unsupervised diffeomorphic cardiac image registration using the moving mesh parameterization. Since the presented method is learning-based, the time to register an unknown pair of images, during the deployment phase is less than a second. Which makes it appropriate for real-time applications. In addition, parametrizing the deformation field using radial and rotational components made this method more suitable for cardiac registration. As shown in chapter 7 the proposed method in chapter 6 can be used for supervised segmentation tasks as well.

8.2 Contributions

The list of original accomplishments described in this thesis can be summarized as the following:

- A training-free generative neural network for group-wise cardiac deformable image registration [72]
- A training-free generative multi-resolution framework for group-wise cardiac deformable registration

- A training-free recursive multi-resolution framework for diffeomorphic deformable image registration for cardiac images [102]
- An unsupervised 2D-to-2D and 3D-to-3D diffeomorphic learning-based algorithm with moving mesh correspondence for cardiac image registration [101]
- A supervised 2D and 3D diffeomorphic learning-based segmentation framework using deformable image registration

In conclusion,

8.3 Future Considerations

This section lists some of the limitations of the proposed work, and ways to remedy them and improve on the algorithms developed.

- Multimodality registration

  Multimodality image registration refers to those applications where the moving and targeted fixed images are acquired using different imaging modalities and sometimes even in different dimensions. Multimodality image registration and fusion have a key role in routine diagnosis, staging, restaging, and the assessment of response to treatment, surgery, and radiotherapy planning of malignant disease. The complementarity between anatomic (CT and MR imaging) and molecular (SPECT and PET) imaging modalities is well established and the role of fusion imaging is widely recognized as a central piece of the general tree of clinical decision making. Moreover, dual-modality imaging technologies including SPECT/CT, PET/CT, and, in the future, PET/MR imaging, now represent the leading component of contemporary health care institutions.

  In addition to the challenge caused by tissue deformation presented in single-modality image registration, gauging image similarity between multimodality

images can be difficult, even when they are well aligned. So that, multimodality image registration can be considered in the future.

- Different modalities:

  Cardiac MRI can provide 3D cross-section images of the heart and produce images with high resolution as well as high tissue contrast which is useful for different assessments such as heart chambers, heart valves, etc. Also, MRI is used in diagnosing a variety of cardiovascular disorders. Unfortunately, MRI suffers from several drawbacks. Compared with cardiac CT, these examinations are much more expensive and high cost and the fact that it is not portable[103]. Echocardiography or ultrasound imaging has therefore been used by clinicians to address the above issues. All different proposed methods in this thesis are tested on MRI images. In the future, it would be beneficial to perform these methods on different modalities such as ultrasound.

- Improvements to the 3D-to-3D registration algorithm

  The proposed a 2D-to-2D/3D-to-3D diffeomorphic registration algorithm that was applied to cardiac MR images. The main idea of using moving mesh generation is to parameterize the deformation field with radial and rotational components which are more suitable for cardiac images. Since heart spatially the left ventricle twists in systole storing potential energy and untwists (recoils) in diastole releasing the energy. Twist aids left ventricular ejection and untwist aids relaxation and ventricular filling. Therefore, rotation and torsion are important in cardiac mechanics. It would also be beneficial to validate the method on different tissues such as the brain, lung, etc.

Thesis Publication List

This section lists the academic contributions made during the course of this doctoral thesis, including journal papers and conference papers.

Thesis-Related Publications

Journals

1. Sheikhjafari Ameneh, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray. "A training-free recursive multiresolution framework for diffeomorphic deformable image registration." *Applied Intelligence (2022): 1-10.*

2. Sheikhjafari Ameneh, Michelle Noga, Ahmed Ahmed, Kumaradevan Punithakumar, and Nilanjan Ray. "GMCNet: A generative multi-resolution framework for cardiac registration." *Submitted to IEEE Open Access (2022) - Under Review.*

3. Sheikhjafari Ameneh, Krishnaswamy Deepa, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray, "Unsupervised diffeomorphic cardiac image registration using parameterization of the deformation field." *Medical Image Analysis (2022) - Under Review.*

Conferences

1. Sheikhjafari Ameneh, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray. "Unsupervised deformable image registration with fully connected generative neural network." *Medical Imaging with Deep Learning 2018.*

2. Sheikhjafari Ameneh, Krishnaswamy Deepa, Michelle Noga, Kumaradevan Punithakumar, and Nilanjan Ray. "Deep Learning Based Parametrization of Diffeomorphic Image Registration for the Application of Cardiac Image Segmentation." *Submitted to International Conference on Bioinformatics and Biomedicine BIBM (2022) - Under Review.*

Bibliography

[1] L. G. Brown, "A survey of image registration techniques," *ACM computing surveys (CSUR)*, vol. 24, no. 4, pp. 325–376, 1992.

[2] W. R. Crum, T. Hartkens, and D. Hill, "Non-rigid image registration: Theory and practice," *The British journal of radiology*, vol. 77, no. suppl_2, S140–S153, 2004.

[3] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in medicine & biology*, vol. 46, no. 3, R1, 2001.

[4] C. A. Pelizzari, G. Chen, D. R. Spelbring, R. R. Weichselbaum, and C.-T. Chen, "Accurate three-dimensional registration of ct, pet, and/or mr images of the brain.," *Journal of computer assisted tomography*, vol. 13, no. 1, pp. 20–26, 1989.

[5] D. L. Hill, D. J. Hawkes, J. Crossman, M. Gleeson, T. Cox, E. Bracey, A. Strong, and P Graves, "Registration of mr and ct images for skull base surgery using point-like anatomical features," *The British journal of radiology*, vol. 64, no. 767, pp. 1030–1035, 1991.

[6] D. J. Hawkes, D Barratt, J. M. Blackall, C Chan, P. J. Edwards, K Rhode, G. P. Penney, J. McClelland, and D. L. Hill, "Tissue deformation and shape models in image-guided interventions: A discussion paper," *Medical Image Analysis*, vol. 9, no. 2, pp. 163–175, 2005.

[7] J. Krebs, H. e Delingette, B. Mailhé, N. Ayache, and T. Mansi, "Learning a probabilistic model for diffeomorphic registration," *IEEE transactions on medical imaging*, 2019.

[8] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," *arXiv preprint arXiv:1903.02026*, 2019.

[9] M. Holden, "A review of geometric transformations for nonrigid body registration," *IEEE transactions on medical imaging*, vol. 27, no. 1, pp. 111–128, 2007.

[10] H. R. Boveiri, R. Khayami, R. Javidan, and A. R. MehdiZadeh, "Medical image registration using deep neural networks: A comprehensive review," *arXiv preprint arXiv:2002.03401*, 2020.

[11] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," *Machine Vision and Applications*, vol. 31, no. 1, p. 8, 2020.

[12] T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in neural information processing systems*, 2018, pp. 6571–6583.

[13] H. J. Johnson and G. E. Christensen, "Consistent landmark and intensity-based image registration," *IEEE transactions on medical imaging*, vol. 21, no. 5, pp. 450–461, 2002.

[14] S. Klein, M. Staring, and J. P. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines," *IEEE transactions on image processing*, vol. 16, no. 12, pp. 2879–2890, 2007.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[16] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.

[17] A. V. Dalca, A. Bobu, N. S. Rost, and P. Golland, "Patch-based discrete registration of clinical brain images," in *International Workshop on Patch-based Techniques in Medical Imaging*, Springer, 2016, pp. 60–67.

[18] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through mrfs and efficient linear programming," *Medical image analysis*, vol. 12, no. 6, pp. 731–741, 2008.

[19] A. Sheikhjafari, H. Talebi, and M. Zareinejad, "3d visual stabilization for robotic-assisted beating heart surgery using a thin-plate spline deformable model," in *2015 3rd RSI International Conference on Robotics and Mechatronics (ICROM)*, IEEE, 2015, pp. 743–748.

[20] A Sheikhjafari, H. A. Talebi, and M. Zareinejad, "Robust and efficient 3d motion tracking in robotic assisted beating heart surgery," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, 2015, pp. 1828–1833.

[21] J.-P. Thirion, "Image matching as a diffusion process: An analogy with maxwell's demons," *Medical image analysis*, vol. 2, no. 3, pp. 243–260, 1998.

[22] P. Cachier and X. Pennec, "3d non-rigid registration by gradient descent on a gaussian-windowed similarity measure using convolutions," in *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. MMBIA-2000 (Cat. No. PR00737)*, IEEE, 2000, pp. 182–189.

[23] C. Le Guyader and L. A. Vese, "A combined segmentation and registration framework with a nonlinear elasticity smoother," *Computer Vision and Image Understanding*, vol. 115, no. 12, pp. 1689–1709, 2011.

[24] M. Burger, J. Modersitzki, and L. Ruthotto, "A hyperelastic regularization energy for image registration," *SIAM Journal on Scientific Computing*, vol. 35, no. 1, B132–B148, 2013.

[25] T. Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache, "Ilogdemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues," *International journal of computer vision*, vol. 92, no. 1, pp. 92–111, 2011.

[26] M. Zhang and P. T. Fletcher, "Finite-dimensional lie algebras for fast diffeomorphic image registration," in *International Conference on Information Processing in Medical Imaging*, Springer, 2015, pp. 249–260.

[27] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[28] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Symmetric log-domain diffeomorphic registration: A demons-based approach," in *International conference on medical image computing and computer-assisted intervention*, Springer, 2008, pp. 754–761.

[29] K. Punithakumar, M. Noga, I. B. Ayed, and P. Boulanger, "Right ventricular segmentation in cardiac mri with moving mesh correspondences," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 15–25, 2015.

[30] K. Punithakumar, P. Boulanger, and M. Noga, "A GPU-accelerated deformable image registration algorithm with applications to right ventricular segmentation," *IEEE Access*, vol. 5, pp. 20 374–20 382, 2017.

[31] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Non-parametric diffeomorphic image registration with the demons algorithm," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2007, pp. 319–326.

[32] M. Lorenzi, N. Ayache, G. B. Frisoni, X. Pennec, A. D. N. I. (ADNI, *et al.*, "Lcc-demons: A robust and accurate symmetric diffeomorphic registration algorithm," *NeuroImage*, vol. 81, pp. 470–483, 2013.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[34] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, "Svf-net: Learning deformable image registration using shape matching," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 266–274.

[35] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen, "Deformable image registration based on similarity-steered cnn regression," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 300–308.

[36] M. C. Lee, O. Oktay, A. Schuh, M. Schaap, and B. Glocker, "Image-and-spatial transformer networks for structure-guided image registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 337–345.

[37] Y. Hu, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, T. Vercauteren, J. A. Noble, and D. C. Barratt, "Adversarial deformation regularization for training image registration neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 774–782.

[38] Y. Hu, M. Modat, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Label-driven weakly-supervised learning for multimodal deformable image registration," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018, pp. 1070–1074.

[39] P. Yan, S. Xu, A. R. Rastinehad, and B. J. Wood, "Adversarial image registration with application for mr and trus image fusion," in *International Workshop on Machine Learning in Medical Imaging*, Springer, 2018, pp. 197–204.

[40] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks," in *Advances in neural information processing systems*, 2015, pp. 2017–2025.

[41] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 204–212.

[42] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical image analysis*, vol. 52, pp. 128–143, 2019.

[43] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9252–9260.

[44] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, 2019.

[45] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 729–738.

[46] J. Krebs, T. Mansi, B. Mailhé, N. Ayache, and H. Delingette, "Unsupervised probabilistic deformation modeling for robust diffeomorphic registration," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 101–109.

[47] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018, pp. 1449–1453.

[48] J. Fan, X. Cao, Z. Xue, P.-T. Yap, and D. Shen, "Adversarial similarity network for evaluating image alignment in deep learning based registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 739–746.

[49] D. Mahapatra, Z. Ge, S. Sedai, and R. Chakravorty, "Joint registration and segmentation of xray images using generative adversarial networks," in *International Workshop on Machine Learning in Medical Imaging*, Springer, 2018, pp. 73–80.

[50] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[51] H. Lester and S. R. Arridge, "A survey of hierarchical non-linear medical image registration," *Pattern recognition*, vol. 32, no. 1, pp. 129–149, 1999.

[52] A. Nan, M. Tennant, U. Rubin, and N. Ray, "Drmime: Differentiable mutual information and matrix exponential for multi-resolution image registration," *arXiv preprint arXiv:2001.09865*, 2020.

[53] A. A. Goshtasby, *Image registration: Principles, tools and methods*. Springer Science & Business Media, 2012.

[54] R. Wang. (2013). "Gaussian-laplacian pyramid image coding," [Online]. Available: http://fourier.eng.hmc.edu/e161/lectures/canny/node3.html.

[55] C. Twining, T. Cootes, S. Marsland, V. Petrovic, R. Schestowitz, and C. Taylor, "Information-theoretic unification of groupwise non-rigid registration and model building," *Proceedings of Medical Image Understanding and Analysis (MIUA)*, vol. 2, pp. 226–230, 2006.

[56] S. Marsland, C. J. Twining, and C. J. Taylor, "A minimum description length objective function for groupwise non-rigid image registration," *Image and Vision Computing*, vol. 26, no. 3, pp. 333–346, 2008.

[57] Cardio-Research. (). "Basic anatomy of the human heart," [Online]. Available: http://www.cardio-research.com/basic-anatomy-of-the-human-heart. (accessed: 02.06.2022).

[58] H. Sokooti, B. de Vos, F. Berendsen, B. P. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 232–239.

[59] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration–a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.

[60] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, "Robust non-rigid registration through agent-based action learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 344–352.

[61] I. Yoo, D. G. Hildebrand, W. F. Tobin, W.-C. A. Lee, and W.-K. Jeong, "ssEMnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 249–257.

[62] H. Li and Y. Fan, "Non-rigid image registration using fully convolutional networks with deep self-supervision," *arXiv preprint arXiv:1709.00799*, 2017.

[63] A. Myronenko, "Non-rigid image registration regularization, algorithms and applications.," 2010.

[64] V. Vishnevskiy, T. Gass, G. Szekely, C. Tanner, and O. Goksel, "Isotropic total variation regularization of displacements in parametric image registration," *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 385–395, 2017.

[65] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[66] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester, *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[67] N. Ray, "Computation of fluid and particle motion from a time-sequenced image pair: A global outlier identification approach," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2925–2936, 2011.

[68] P Radau, Y Lu, K Connelly, G Paul, A Dick, and G Wright, "Evaluation framework for algorithms segmenting short axis cardiac mri," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 49, 2009.

[69] K. Marstal, F. Berendsen, M. Staring, and S. Klein, "Simpleelastix: A user-friendly, multi-lingual library for medical image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 134–142.

[70] C. Zachiu, N. Papadakis, M. Ries, C. Moonen, and B. D. de Senneville, "An improved optical flow tracking technique for real-time mr-guided beam therapies in moving organs," *Physics in Medicine & Biology*, vol. 60, no. 23, p. 9003, 2015.

[71] M. M. McCormick, X. Liu, L. Ibanez, J. Jomier, and C. Marion, "ITK: enabling reproducible research and open science," *Frontiers in neuroinformatics*, vol. 8, p. 13, 2014.

[72] A. Sheikhjafari, M. Noga, K. Punithakumar, and N. Ray, "Unsupervised deformable image registration with fully connected generative neural network," in *Medical Imaging with Deep Learning*, 2018.

[73] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[74] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the **hausdorff** distance," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[75] J. Ashburner, J. L. Andersson, and K. J. Friston, "High-dimensional image registration using symmetric priors," *NeuroImage*, vol. 9, no. 6, pp. 619–628, 1999.

[76] R. Beare, B. Lowekamp, and Z. Yaniv, "Image segmentation, registration and characterization in r with simpleitk," *Journal of statistical software*, vol. 86, 2018.

[77] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "Simpleitk image-analysis notebooks: A collaborative environment for education and reproducible research," *Journal of digital imaging*, vol. 31, no. 3, pp. 290–303, 2018.

[78] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of simpleitk," *Frontiers in neuroinformatics*, vol. 7, p. 45, 2013.

[79] D. Krishnaswamy, M. Noga, and K. Punithakumar, "Validation of a diffeomorphic registration algorithm using true deformation computed from thin plate spline interpolation," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2020, pp. 1351–1354.

[80] A. H. A. W. G. on Myocardial Segmentation, R. for Cardiac Imaging: M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, *et al.*, "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association," *Circulation*, vol. 105, no. 4, pp. 539–542, 2002.

[81] K. Punithakumar, I. B. Ayed, A. Islam, A. Goela, I. G. Ross, J. Chong, and S. Li, "Regional heart motion abnormality detection: An information theoretic approach," *Medical image analysis*, vol. 17, no. 3, pp. 311–324, 2013.

[82] X. Zhang, M. Noga, D. G. Martin, and K. Punithakumar, "Fully automated left atrium segmentation from anatomical cine long-axis mri sequences using deep convolutional neural network with unscented kalman filter," *Medical Image Analysis*, vol. 68, p. 101 916, 2020.

[83] M. Regehr, A. Volk, M. Noga, and K. Punithakumar, "Machine learning and graph based approach to automatic right atrial segmentation from magnetic resonance imaging," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA: IEEE, Apr. 2020, pp. 826–829.

[84] N. J. Higham, "The scaling and squaring method for the matrix exponential revisited," *SIAM review*, vol. 51, no. 4, pp. 747–764, 2009.

[85] T. C. Mok and A. Chung, "Fast symmetric diffeomorphic image registration with convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4644–4653.

[86] M. I. Belghazi, A. Baratin, S. Rajeswar, S. Ozair, Y. Bengio, A. Courville, and R. D. Hjelm, "Mine: Mutual information neural estimation," *arXiv preprint arXiv:1801.04062*, 2018.

[87] I. B. Ayed, K. Punithakumar, S. Li, A. Islam, and J. Chong, "Left ventricle segmentation via graph cut distribution matching," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2009, pp. 901–909.

[88] E. Haber and J. Modersitzki, "Numerical methods for volume preserving image registration," *Inverse problems*, vol. 20, no. 5, p. 1621, 2004.

[89] H.-m. Chen, A. Goela, G. J. Garvin, and S. Li, "A parameterization of deformation fields for diffeomorphic image registration and its application to myocardial delineation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2010, pp. 340–348.

[90] D. Cheng, *Field and wave electromagnetics*. Pearson Education India, 1989.

[91] J. Liu, *New development of the deformation method*. The University of Texas at Arlington, 2006.

[92] X. Zhou, "On uniqueness theorem of a vector function," *Progress in Electromagnetics Research*, vol. 65, pp. 93–102, 2006.

[93] T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, "Engineering and algorithm design for an image processing api: A technical report on itk-the insight toolkit," in *Medicine Meets Virtual Reality 02/10*, IOS press, 2002, pp. 586–592.

[94] D. Krishnaswamy, "A diffeomorphic 3d-to-3d registration algorithm for the segmentation of the left ventricle in ultrasound sequences," Ph.D. dissertation, University of Alberta, 2021.

[95] T. C. Mok and A. Chung, "Large deformation diffeomorphic image registration with laplacian pyramid networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 211–221.

[96] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *International journal of computer vision*, vol. 61, no. 2, pp. 139–157, 2005.

[97] N. M. Noble, D. L. Hill, M. Breeuwer, J. A. Schnabel, D. J. Hawkes, F. A. Gerritsen, and R. Razavi, "Myocardial delineation via registration in a polar coordinate system," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2002, pp. 651–658.

[98] O Bernard, A Lalande, C Zotti, F Cervenansky, X Yang, P. Heng, I Cetin, K Lekadir, O Camara, M. Ballester, and G Sanroma, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–25, 2018.

[99] A. Klein, S. S. Ghosh, B. Avants, B. T. Yeo, B. Fischl, B. Ardekani, J. C. Gee, J. J. Mann, and R. V. Parsey, "Evaluation of volume-based and surface-based brain image registration methods," *Neuroimage*, vol. 51, no. 1, pp. 214–220, 2010.

[100] D. P. Shamonin, E. E. Bron, B. P. Lelieveldt, M. Smits, S. Klein, M. Staring, and A. D. N. Initiative, "Fast parallel image registration on cpu and gpu for diagnostic classification of alzheimer's disease," *Frontiers in neuroinformatics*, vol. 7, p. 50, 2014.

[101] A. Sheikhjafari, D. Krishnaswamy, M. Noga, N. Ray, and K. Punithakumar, "Unsupervised diffeomorphic cardiac image registration using parameterization of the deformation field," in *arXiv preprint arXiv:2208.13275*, 2022.

[102] A. Sheikhjafari, M. Noga, K. Punithakumar, and N. Ray, "A training-free recursive multiresolution framework for diffeomorphic deformable image registration," in *Applied Intelligence*, Springer, 2022, pp. 1–10.

[103] A. Khalil, S.-C. Ng, Y. M. Liew, and K. W. Lai, "An overview on image registration techniques for cardiac diagnosis and treatment," vol. 10, 2018.