

IMPERIAL

**PHYSICS-INFORMED NEURAL IMPLICIT REPRESENTATION
NETWORKS FOR MYOCARDIAL IMAGE REGISTRATION**

Author

BASTIEN BALUYOT

CID: 01874363

Supervised by

DR C. QIN

DR M. VARELA

Second Marker

Dr A. Abu Ebayyeh

A Thesis submitted in fulfillment of requirements for the degree of
Master of Engineering in Electrical and Electronic Engineering with Management

Department of Electrical and Electronic Engineering
Imperial College London
2024

Abstract

With cardiovascular diseases remaining the leading cause of death worldwide, this project focuses on a physics-informed neural network (PINN)-based approach to registering the deformation of the Left Ventricle from cine-MR images and subsequently quantifying cardiac motion and measuring cardiac strain, a valuable metric of cardiac function that is beneficial in aiding the diagnosis (by characterising) of certain cardiomyopathies. Specifically, this project investigates the use of PINNs as Implicit Neural Representations (INRs) for cardiac image registration. This project aims to address two important ill-posed problems: the difficulty of imposing biomechanical plausibility and predicting realistic deformations in medical image registration, as well as the lack of generalisability of INRs in performing pairwise optimisation. This project begins with López et al.'s WarpPINN model as a baseline and adapts the model to the popular Automated Cardiac Diagnosis Challenge (ACDC) dataset in order to compute cardiac strains for patients of varying cardiomyopathies. The image registration performance of WarpPINN with different parameters are discussed and analysed in length. By introducing learning bias through a Neo-Hookean-based composite loss, extended research in this project entails the exploration of various novel approaches for generalising physics-informed INRs for multiple patients. Discussed implementations for these approaches are based upon adapting generalisation strategies utilised in signal approximation. The justification and effectiveness of these methods are analysed and discussed with the aim of demonstrating a proof-of-concept for generalising physics-informed INRs for medical image registration.

Declaration of Originality

I hereby declare that the work presented in this thesis is my own unless otherwise stated. To the best of my knowledge the work is original and ideas developed in collaboration with others have been appropriately referenced.

I affirm that I have submitted, or will submit, an electronic copy of my final year project report to the provided EEE link.

I affirm that I have submitted, or will submit, an identical electronic copy of my final year project to the provided Blackboard module for Plagiarism checking.

I affirm that I have provided explicit references for all the material in my Final Report that is not authored by me, but is represented as my own work.

I have used ChatGPT v4o as an aid in the preparation of my report. I have used it to improve the quality of my English throughout, however all technical content and references comes from my original text.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Contents

Abstract	i
Declaration of Originality	ii
Copyright Declaration	iii
1 Introduction	1
2 Background	3
2.1 Clinical background	3
2.1.1 CVD diagnosis and challenges	3
2.1.2 Quantifying cardiac motion	3
2.2 Image registration	5
2.2.1 Existing related work	5
2.2.2 Relevant algorithms	6
2.2.3 Deep-learning approaches	9
2.3 Physics-Informed Neural Networks	13
2.3.1 WarpPINN	14
3 Methods	18
3.1 Validating WarpPINN on CMAC	19
3.1.1 Dataset	19
3.1.2 Anatomical landmark tracking	19
3.1.3 Strain estimation	20
3.1.4 Benchmark models/methodologies	21
3.1.5 Training regime	23
3.1.6 Evaluation methodology	26
3.1.7 Ablation Study: Nullifying the Physics-Informed Loss	29
3.1.8 Hyperparameter tuning	30
3.2 Adapting WarpPINN to ACDC	33
3.2.1 Dataset	33
3.2.2 Further preprocessing: Surface mesh generation	35
3.2.3 Training regime	37
3.2.4 Alternative methodologies for comparison	37

3.2.5	Evaluation method	38
3.2.6	Hyperparameter tuning	40
4	Experimental Results and Discussions	41
4.1	Validating WarpPINN on CMAC	42
4.1.1	Analysis of Predicted Images and Deformation Field	42
4.1.2	Landmark tracking	45
4.1.3	Strain analysis	46
4.1.4	Training loss curves	48
4.1.5	Discussion of Validation results	49
4.1.6	Results of ablation study	50
4.1.7	Hyperparameter tuning	54
4.2	ACDC	59
4.2.1	Selective presentation of individual patient results	59
4.2.2	Results aggregated by disease	62
4.2.3	Results across all patients	66
4.2.4	Discussion of WarpPINN-ACDC experiments	66
5	Extended Research: Generalisation of WarpPINN	68
5.1	Validating WarpPINN on CMAC	69
5.1.1	Motivation	69
5.1.2	Related work	69
5.1.3	Methods	70
5.1.4	Experimental results and discussions	72
A	ACDC Best Results	76
	Bibliography	78

1 Introduction

Cardiovascular diseases are the leading cause of death globally, emphasizing the critical need for accurate diagnostic tools and techniques in cardiology. One promising approach to enhancing diagnostic accuracy involves the use of physics-informed neural networks (PINNs) for myocardial image registration. This project focuses on developing and validating such a network, named WarpPINN, to register the deformation of the Left Ventricle (LV) from cine-MR images. By accurately quantifying cardiac motion and measuring cardiac strain, this methodology aims to improve the characterization and diagnosis of various cardiomyopathies.

The core of this project lies in addressing two significant challenges in medical image registration. First, the problem of imposing biomechanical plausibility to ensure that predicted deformations are realistic and clinically meaningful. Second, the inherent limitation of implicit neural representations (INRs) and implicit deformable image registration (IDIR) techniques, which typically require pairwise optimization. These techniques necessitate the training of a new model from scratch for each new image pair, leading to inefficiencies and extended processing times.

Building upon the WarpPINN model introduced by López et al., this project adapts the model for use with the Automated Cardiac Diagnosis Challenge (ACDC) dataset. The ACDC dataset provides a diverse range of patient data, enabling a comprehensive evaluation of WarpPINN's performance across various cardiomyopathies. Through this adaptation, the project aims to compute cardiac strains, offering insights into cardiac function and potential diagnostic improvements.

A significant part of this research involves exploring generalization strategies to extend the applicability of WarpPINN beyond pairwise optimization. Inspired by techniques in signal approximation, the project investigates novel approaches to generalize physics-informed INRs, making them capable of handling multiple patients without retraining from scratch for each new case.

The methods section details the dataset preparation, model adaptation, and the implementation of these generalization techniques. The evaluation methodology includes both qualitative and quantitative assessments, comparing WarpPINN's performance with other state-of-the-art image registration methods.

The experimental results section discusses the findings from validating WarpPINN on the CMAC dataset and its subsequent adaptation to the ACDC dataset. Key metrics such as Dice Similarity Coefficient (DSC), Hausdorff Distance (HD95), and Mean Contour Distance (MCD) are used to evaluate the model's performance in terms of image similarity and biomechanical plausibility.

In conclusion, this project not only demonstrates the feasibility of using PINNs for realistic myocardial deformation prediction but also provides a proof-of-concept for generalizing these models to handle multiple patients efficiently. Future work will focus on optimizing hyperparameters and further enhancing the generalization capabilities of physics-informed neural networks for medical image registration.

Relevant code will be uploaded and updated in this Github repository: <https://github.com/BastienB04/WarpPINN>

2

Background

Contents

2.1	Clinical background	3
2.1.1	CVD diagnosis and challenges	3
2.1.2	Quantifying cardiac motion	3
2.2	Image registration	5
2.2.1	Existing related work	5
2.2.2	Relevant algorithms	6
2.2.3	Deep-learning approaches	9
2.3	Physics-Informed Neural Networks	13
2.3.1	WarpPINN	14

2.1 Clinical background

2.1.1 CVD diagnosis and challenges

Cardiovascular diseases (CVDs) continue to be the primary cause of mortality worldwide. The World Health Organization (WHO) reports that in 2019, approximately 17.9 million lives were lost to CVDs, predominantly due to strokes and heart attacks, accounting for 32% of all deaths globally [1].

Recent studies estimate that treatment expenses for CVDs cost the European Union alone €282 billion annually [2]. Equivalently, this number is projected to exceed \$1 trillion in the US by 2030, with an estimate of >40% of the US population having heart failure or other forms of CVD [3].

The mortality numbers and prevalence, as well as the significant economic burden placed on healthcare systems, by these diseases undoubtedly make the diagnosis of CVDs an important topic of research, which is reflected with many recent studies in emerging fields, such as machine learning, branching to focus applications on this diagnosis problem [4][5].

Aside from various risk factors such as genetic predispositions, environmental influences, lifestyle factors and other medical conditions, CVD can be attributed to decreased or irregular mechanical/biophysical functionalities of the heart [6]. For example, dilated cardiomyopathy occurs when the heart's ventricles enlarge and weaken, consequently leading to the heart's diminished capacity to pump blood effectively. Therefore, the diagnosis of CVD roots with quantifying cardiac motion.

2.1.2 Quantifying cardiac motion

In this domain, many different imaging techniques have been developed to capture the intricate dynamics of cardiac function using medical imaging from distinct perspectives. Table 2.1 summarises key techniques utilised in this endeavour.

Imaging Technique	Type	Prevalence in Research	Prevalence in Hospitals	Advantages	Disadvantages
Echocardiography	Ultrasound-based	Very High	Very High	Non-invasive, widely available	Operator-dependent image quality
MRI (incl. cine SSFP, TAG, SENC, DENSE) [7][8]	Magnetic resonance	High	Moderate to High	Excellent soft tissue contrast	Expensive, contraindications for some patients
Computed Tomography (CT)	X-ray based	Moderate to High	Moderate to High	High-resolution imaging of coronary arteries	Exposure to ionising radiation
Nuclear Cardiology (SPECT, PET) [9]	Nuclear imaging	Moderate	Moderate	Functional imaging of perfusion and metabolism	Exposure to ionising radiation, requires radiotracers
Advanced Echo Techniques (TDI, IVUS, OCT) [10]	Ultrasound and intravascular	Low to Moderate	Moderate (IVUS, OCT for interventional use)	High-resolution coronary artery imaging (IVUS, OCT)	Invasive (IVUS, OCT)
Cardiac Biomarker Imaging	Molecular imaging	Moderate	Low	Targets specific molecular processes	Requires specialised equipment and radiotracers

Table 2.1: Comparison of existing imaging techniques for cardiac assessment [11]

Most notably, this project draws focus on cine SSFP (Steady-State Free Precession) MRI as a premier choice among non-invasive methods for both qualitative and quantitative evaluation of heart function. This preference is largely attributable to its superior advantages, including exceptional soft-tissue contrast, consistent reproducibility and excellent spatial and temporal resolution [12]. These features make cine SSFP MRI an indispensable tool in cardiac imaging, providing detailed insights into the heart's structure and dynamics.

Cine SSFP MRI's high soft-tissue contrast enables clear differentiation between the blood pool and myocardium, enhancing the visualisation of cardiac chambers and facilitating accurate assessment of cardiac anatomy and function [13]. Its high reproducibility ensures that measurements like ventricular volumes and mass are reliable over time, making it ideal for monitoring disease progression or response to therapy. Moreover the technique's fine spatial and temporal resolution allows for precise tracking of myocardial motion through the cardiac cycle [12]. This is crucial for effectively measuring the different metrics that have also been developed to assess the mechanical and biophysical functionality of the heart, including left ventricular ejection fraction (LVEF) [14] and cardiac strain [15], as it provides detailed data on the extent and pattern of myocardial contraction and relaxation [16].

By offering a comprehensive view of the heart's mechanical function, cine SSFP MRI supports the assessment of cardiac function metrics such as LVEF and cardiac strain. Strain measurements, which detail the myocardial deformation during the cardiac cycle, are vital for detecting subtle changes in heart function that might not be apparent through LVEF alone.

Throughout this project, the aim of attributing estimated quantification of cardiac motion (through, as will be explained later, image registration) to certain CVDs will be focussed upon and achieved through the measurement of cardiac strain.

Measuring cardiac strain provides several advantages over LVEF and other traditional metrics when it comes to assessing heart function, especially in the context of diagnosing and managing CVDs. Cardiac strain measures the deformation of the myocardium during the cardiac cycle, offering a detailed view of both global and regional myocardial mechanics. In contrast to LVEF, which offers only a global assessment based on a singular measure of global systolic function based on volume changes in the left ventricle, cardiac strain is usually reported using three regional metrics (Circumferential, Radial and Longitudinal) corresponding to specific American Heart Association (AHA) segments [17], enabling precise localisation of functional abnormalities. Furthermore, strain analysis can detect subtle changes and is more sensitive to early signs of myocardial dysfunction, particularly in conditions where LVEF remains within normal ranges, such as in heart failure with preserved ejection fraction (HFpEF) [18]. As such, cardiac strain measurements, as several studies have shown, offer significant prognostic value in predicting outcomes in ischemic heart disease, cardiomyopathies

and heart failure, making the metric an important consideration in this project.

2.2 Image registration

Image registration is the process of superimposing two or more images taken at various instances or perspectives to geometrically align them for analysis. This is done by aligning the reference and target images into the same coordinate system [19].

Involuntary motions such as the heartbeat mean that images need to be aligned to establish correspondence point by point. In the image registration task of the heart throughout the cardiac cycle, the recovered transformation would have clinical significance as the cardiac strain can subsequently be measured. Known physical constraints of the tissues involved (such as volume preservation or local rigidity) could be incorporated to constrain what would otherwise be a wide range of possible solutions. In later discussions, it becomes apparent that this still does not solve the ill-posed of such a problem and this concept of registering realistic deformations is explored. One approach is taking advantage of the quasi-incompressibility physical constraint of the myocardium through the use of physics-informed neural networks [20].

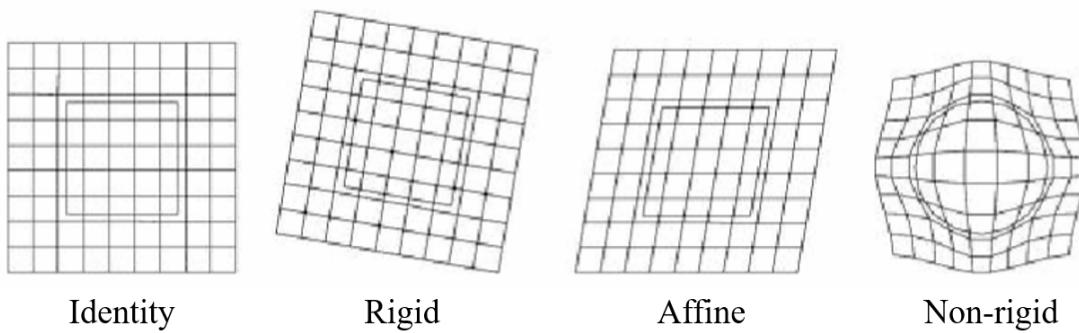


Figure 2.1: Examples of transformations on a square. [19]

Transformations in image registration tasks usually belong to four broad and fundamental categories: Identity, Rigid, Affine and Non-rigid. These are illustrated in Figure 2.1. As a baseline, identity transformation serves as the simplest form of transformation, where no change is applied to the image and every point in the image remains in its original position, meaning the transformed image is identical to the input image. Rigid transformation involves rotation and translation of the image without altering its shape or size. All points in the image move in a uniform manner, maintaining the distances and angles between points. This would be useful in cases when the object's orientation or position changes, but its structure remains constant, such as in skull radiography or bone imaging. Affine transformation extends rigid transformations to include scaling, shearing and reflection, allowing for changes in size, shape and orientation while preserving lines and planes. Points can move in a linear fashion that allows for parallel lines in the source image to remain parallel in the template image. Finally, non-rigid transformation allows localised deformations that can vary across the image. This can include model bending, stretching and more complex deformations, enabling detailed alignment of features that may undergo non-uniform changes. As such, this project is concerned with the task of registering the non-rigid transformations between a reference image (cine SSFP MR image of the heart at the start of the cardiac cycle, i.e., end-diastole) to a number of template images (across the cardiac cycle).

2.2.1 Existing related work

Non-rigid registration has been widely used for medical applications in past existing literature and has proven to have clinical significance.

For example, Hill et al. [21] used a non-rigid registration algorithm to detect and quantify intraoperative brain deformation from pre- and post-operative brain MR images in eight patients undergoing resective surgery [22]. Although the researchers had described the pattern of deformation qualitatively, the deformation was also quantified through the use of landmark tracking, where anatomical landmarks were determined and their spread over the brain across images (pre- and post-resection) were analysed and quantified. In this case, the non-rigid registration algorithm was used to validate the deformation of these known landmarks.

This method of tracking anatomical landmarks is the same concept that will be applied in a later section when discussing López et al.'s WarpPINN [20], where the landmarks' locations in the deformed images have a ground truth and hence, the error of the learnt performed image registration can be quantified.

Thirion et al. [23] and Rey et al. [24] have also shown that non-rigid registration algorithms can be used to identify multiple sclerosis (MS) lesions and characterise their change over time by, again, performing the registration between two images of the brain. This use case can be visualised with an example from Rey et al in Figure 2.2 where a non-rigid registration is carried out on the two images of the brain (taken at different times). The non-rigid registration allows for the displacement field to be computed and it is appreciated that, even though both images are not exactly the same, the alignment of the two images thanks to image registration allows the non-rigid transformation to be shown and emphasise that the lesion in the ROI (region of interest) is shrinking.

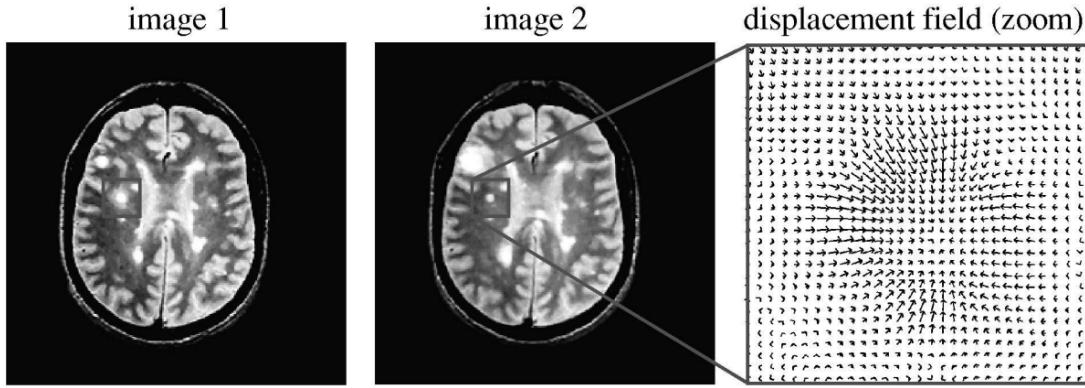


Figure 2.2: Example computation of the displacement field using a non-rigid registration algorithm in order to detect the apparent deformation of a lesion. [24]

Thompson et al. [25] extends this use case by using non-registration to not only study the growth patterns in the developing brain of children, but also uses the deformation fields to calculate local rates of tissue dilation, contraction and shear. This has clinical significance in being able to report an informative and complex quantitative map of growth patterns, which has diagnostic potential if the method is to be employed to other data.

Although the examples outlined above are not necessarily for the registration of the heart and cardiac strain analysis, these multiple studies validate the clinical and effective usability of non-rigid registration in a medical and diagnostic setting for comparing changes across time, which is applicable to this project's goal of quantifying the cardiac motion by comparing a pair (Reference and Template) of images of the heart at any given point in the cardiac cycle.

2.2.2 Relevant algorithms

Any image registration algorithm (including non-rigid registrations) can be described by three components:

1. A **transformation** which relates the target and reference images.
2. A **similarity measure** which measures the similarity between the target and reference images.
3. An **optimisation routine** which determines the optimal transformation parameters as a function of the similarity measure.

A very general form of a cost function that can be used to iteratively find a transformation between medical images is:

$$C = -C_{\text{similarity}} + C_{\text{deformation}} \quad (2.1)$$

Here, the first term, $C_{\text{similarity}}$, characterises the similarity between the reference and target images. The second term, $C_{\text{deformation}}$, acts as a regularisation/penalty function that characterises the cost associated with particular deformations and can be interpreted to represent *a priori* knowledge about the expected deformations. This is crucial in non-rigid transformation and can be used to penalise unexpected or non-plausible deformations.

Some examples of regularisation functions include the Laplacian or membrane model [26] and the biharmonic or thin-plate model [27]. The Laplacian/membrane model (Equation 2.2) minimises the first derivatives of the transformation function, \mathbf{T} , which acts to stretch the image. The integral calculates the sum (over the entire image) of the squares of the spatial gradients (derivatives) of the transformation \mathbf{T} with respect to the x , y and z coordinates. This is analogous to a membrane that resists stretching but can freely bend. The goal of the Laplacian model is to penalise large gradients, enforcing smoothness in the transformation field across the image.

The Biharmonic model (Equation 2.3), which is used by thin-plate splines and TPS registration as discussed later, minimises the second derivatives of the transformation function, analogous to bending energy. The integral sums the squares of all second-order partial derivatives (bending in all planes and directions) of the transformation \mathbf{T} . This model not only penalises stretching (as with the Laplacian model), but also bending.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\left(\frac{\partial \mathbf{T}}{\partial x} \right)^2 + \left(\frac{\partial \mathbf{T}}{\partial y} \right)^2 + \left(\frac{\partial \mathbf{T}}{\partial z} \right)^2 \right] dx dy dz \quad (2.2)$$

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\left(\frac{\partial^2 \mathbf{T}}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial y^2} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial z^2} \right)^2 \right. \\ & \quad \left. + 2 \left(\left(\frac{\partial^2 \mathbf{T}}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial x \partial z} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial y \partial z} \right)^2 \right) \right] dx dy dz \end{aligned} \quad (2.3)$$

Free-Form Deformation

One common image registration algorithm is Free-Form Deformation (FFD) [28]. FFD utilises a three-dimensional lattice or mesh of control points to define the space within which the deformation occurs. These control points are then manually or algorithmically manipulated, which deforms the lattice. The positions of the control points are iteratively adjusted to maximise/minimise the similarity metric, whilst adhering to the chosen regularisation term/method to prevent unrealistic deformations. The optimisation is usually performed using gradient-descent methods or more advanced optimisation techniques.

The deformation of the lattice/mesh through the control points, in turn, dictates the deformation of the encapsulated object. This manipulation can be as simple as moving control points to stretch or compress parts of the object, or more complex transformations to twist, bend, or otherwise deform it.

The deformation effect is applied by recalculating the positions of the points within the object based on the deformed lattice. This step involves mathematical interpolation using the chosen spline functions, which define how the movement of control points influences the deformation of the object.

B-spline registration

A common approach to FFD registration is B-spline registration [29]. B-splines (Basis splines) are piecewise polynomial functions that provide a smooth and continuous way to represent curves and surfaces. B-spline registration defines a grid of control points and optimises the positions of these points iteratively as FFD does. The deformation field can ultimately be modelled by B-splines. This means that, due to the nature of B-splines being continuous functions, the transformation is ensured to be smooth across the entire domain, maintaining the visual integrity of the image, as well as ensuring that there are no abrupt changes in the deformation, which helps in achieving a more natural and accurate alignment between images.

In the previous example of Hill et al.'s work on detecting and quantifying intraoperative brain deformation, the non-rigid registration algorithm uses a 2-stage transformation model, where the second stage captures the local motion of the brain and is modelled by a FFD based on B-splines. The non-rigid registration is achieved by maximising the normalised mutual information as a similarity measure between pre- and post-resection images [21].

Thin-Plate Splines registration

Thin-Plate Splines (TPS) registration is another common image registration algorithm which is based on the physical analogy of bending a thin metal plate [30]. TPS calculates the smoothest possible deformation that aligns points from the reference image with corresponding points in the target image by minimising a bending energy subject to landmark constraints. TPS has been shown to be a valid and effective method for cardiac image registration, in particular, Krishnaswamy et al. employed TPS alongside a diffeomorphic registration algorithm to successfully generate a true deformation field [31].

Elastic registration

The ideas and fundamental concepts of these algorithms can be extended for more advanced image registration techniques by taking account of physical constraints that are present in the image registration task. For example, elastic registration techniques were introduced by Bajcsy et al. which model the deformation of the reference image into the target image as a physical process which resembles the stretching of an elastic material such as rubber [32]. The registration is modelled with the internal force caused by the deformation of elastic material (stress) and the external force which acts on the elastic body. The deformation stops when these two forces reach an equilibrium, and the behaviour of this elastic body can be described by the Navier linear elastic PDE given in Equation 2.4.

$$\mu \nabla^2 \mathbf{u}(x, y, z) + (\lambda + \mu) \nabla(\nabla \cdot \mathbf{u}(x, y, z)) + \mathbf{f}(x, y, z) = 0 \quad (2.4)$$

where \mathbf{u} is the displacement field, \mathbf{f} is the external force acting on the elastic body and μ and λ are Lamé's elasticity constants describing the behavior of the elastic body.

Like FFD and B-Spline registration, elastic registration allows for smooth, continuous deformations but with the added unique advantage of simulating elastic properties, which can more accurately reflect the physiological movements of tissues. This can be particularly useful in incorporating biomechanical properties of tissues into the registration process.

Registration using FEM and Mechanical Models

Similarly, registration using FEM (Finite Element Methods) and mechanical models take into account the elasticity and stiffness of tissues to predict how the tissues deform under forces or constraints [33]. Edwards et al. proposed a simplified version of an FEM model to model tissue deformations in image-guided surgery where images are divided into a triangular mesh with n connected nodes (ϕ) [34]. Each node can be rigid, elastic or fluid. For example, bone is rigid, soft tissue is elastic and CSF (cerebrospinal fluid) is fluid.

Energy terms are used to constrain deformations. For example, elastic nodes can be constrained by a tension energy (2.6) or a stiffness energy term (2.5). Fluid nodes do not have any associated tension or stiffness energy. Instead, they have an associated folding energy (2.7).

$$E_{\text{stiffness}}(\phi_i, \phi_j, \phi_k) = |\phi_j - \phi_k - 2\phi_i|^2 \quad (2.5)$$

$$E_{\text{tension}}(\phi_i, \phi_j) = |\phi_j - \phi_i - \phi_{i,j}^0|^2 \quad (2.6)$$

$$E_{\text{total}}(\phi_i, \phi_j, \phi_k) = \begin{cases} \frac{A^2}{\gamma^2 A_0^2} + \frac{\gamma^2 A_0^2}{A^2}, & \text{if } \frac{A}{A_0} \leq \gamma \\ 2 & \text{otherwise} \end{cases} \quad (2.7)$$

where $\phi_{i,j}^0$ corresponds to the relaxed distance between two nodes, A_0 is the area of the undeformed triangle, A is the area of the deformed triangle and γ is a threshold for the triangular area above which the energy contribution is constant.

Finally, the registration is driven by a similarity measure which minimises the distance between corresponding landmarks. This approach can be interpreted as an advanced application of principles underlying FFD and the mathematical framework for smoothly interpolating deformations across a plane provided by TPS.

Although the Elastic registration technique is not directly applicable to the cardiac image registration task of the project (since the myocardium is not elastic), Elastic and FEM registration serve as fundamental examples of more

advanced techniques and introduce the concept of a “*physics-informed*” task/process, which will be explored in the next section.

Registration using Optical Flow

Before discussing deep learning-based approaches, it is worth discussing another image registration method, optical flow registration [19]. Optical flow methods were originally introduced in computer vision in order to recover the relative motion of an object and the viewer in between 2 successive frames of a temporal image sequence [35].

This method acts on the fundamental assumption that the image brightness (intensity) of a particular point stays constant over time (Brightness constancy assumption).

$$I(x, y, z, t) = I(x + \delta x, y + \delta y, z + \delta z, t + \delta t) \quad (2.8)$$

where $I(x, y, z, t)$ is the intensity at point (x, y, z) at time t , and $(\delta x, \delta y, \delta z)$ is the displacement of the point after a small time increment δt .

Using the brightness constancy assumption (2.8), this method aims to minimise the difference in intensities between the reference and template images (since making the intensities strictly equal may not always be possible due to noise and other factors) by iteratively adjusting δx , δy and δz .

In this method, the cost function would include the difference in intensities between the reference and target images (possibly alongside another image similarity measure).

Although useful, it has been noted that, in practice, the use of optical flow methods in cardiac MR images are not as effective. This shortcoming is attributed to the presence of blood pool areas which can introduce significant noise and artifacts [36] amongst other things.

2.2.3 Deep-learning approaches

With rapid developments in the field of machine learning, the use of deep learning models (particularly Convolutional Neural Networks (CNNs)) has gained popularity in the application of image registration [37][38][39]. CNN-based methods are able to surpass conventional iterative registration methods by learning from large amounts of training data to predict transformations.

CNN-based methods have been found to have some advantage over traditional methods in speed (in inference when trained) [40], but not as much in accuracy [41] (Table 2.2). Nevertheless, the capability to learn spatial features from images provide a powerful tool in predicting the deformation between the reference and template images.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specialised kind of neural network model designed primarily for processing structured array data such as images. These networks are modelled in part after the human visual cortex, with the ability to autonomously and progressively learn spatial hierarchies of features directly from visual inputs.

A typical CNN architecture is composed of a series of layers that each perform distinct types of operations on the input data. These layers include mainly: convolutional layers, pooling layers, and fully connected layers. In the convolutional layers, the convolution operation is applied across small regions of the input data to produce feature maps. This operation uses a set of learnable filters that slide across the input image (sliding window approach). A simple 2D convolutional layer example is illustrated in Figure 2.3. The feature maps generated encapsulate essential visual features of the input data at various levels of abstraction. As the data progresses through successive convolutional layers, the network continually adjusts the filters’ weights based on the learning process, which is aimed at optimising a predefined loss function. The feature maps are then further passed through a non-linear activation function (such as ReLU or sigmoid) which introduces non-linearity to the system, allowing the network to learn more complex patterns.

Pooling layers perform downsampling on feature maps by reducing the feature map size with a kernel-based operation. Typical pooling operations include max, mean and min pooling. Figure 2.1 demonstrates an example of max and average

pooling. Finally, fully connected (FC) or dense layers contain neurons and full connections to all activations in the previous layers, as seen in regular ANNs (Artificial Neural Networks). The role of FC layers is to output the final feature representation. Figure 2.5 illustrates the network architecture of AlexNet, a popular and relatively simple/fundamental CNN architecture.

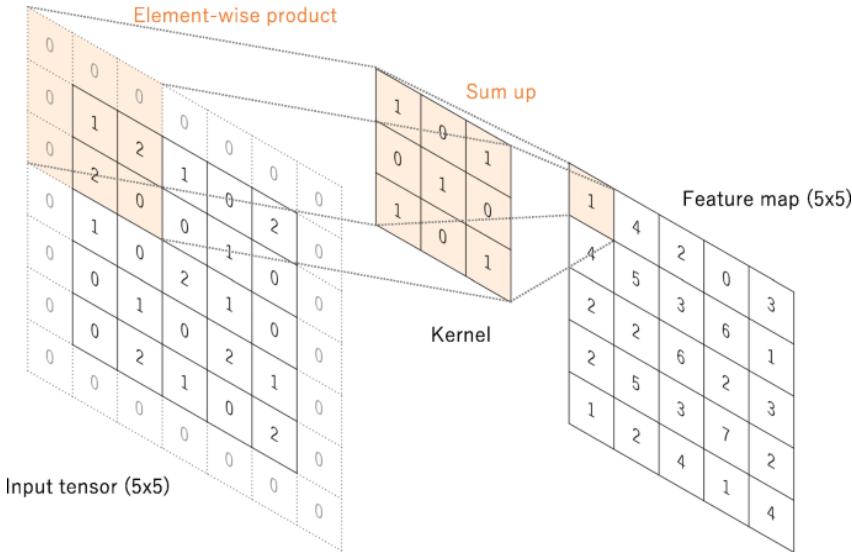


Figure 2.3: Illustration of a 2D convolutional layer sliding across an input tensor to produce a feature map [42].

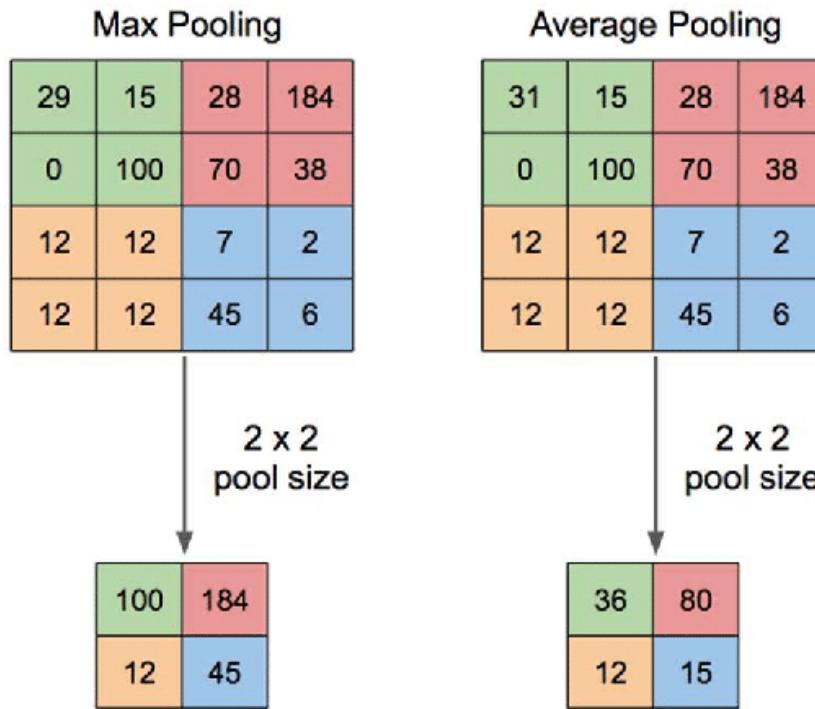


Figure 2.4: Illustration of the downsampling operation of Max and Average pooling [43].

The filter-based and kernel-based approach to convolution and non-linear systems allows the CNN to capture local spatial and temporal dependencies in the input image and learn low-level features (such as edges and textures), as well as high-level features (such as shapes and objects) within images [44][45]. This ability to learn relevant complex image features makes their use applicable in predicting a deformation vector field (DVF) for a reference and template image in image registration tasks.

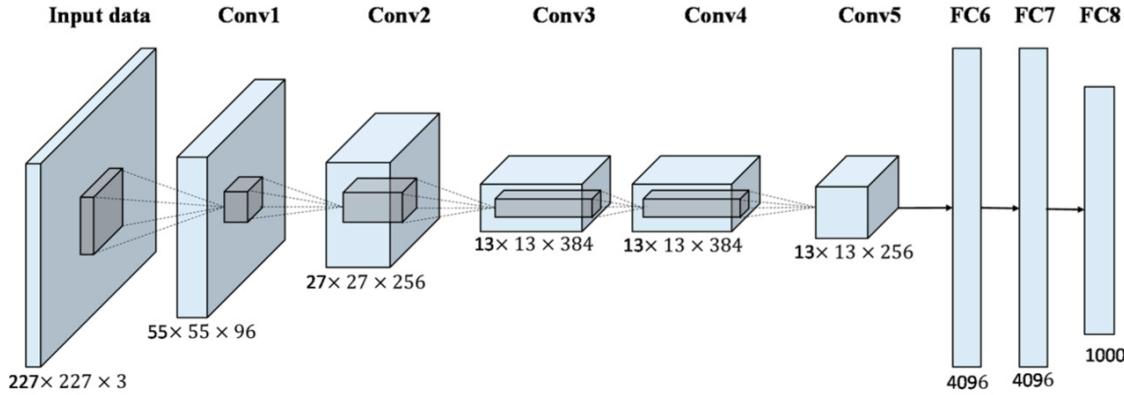
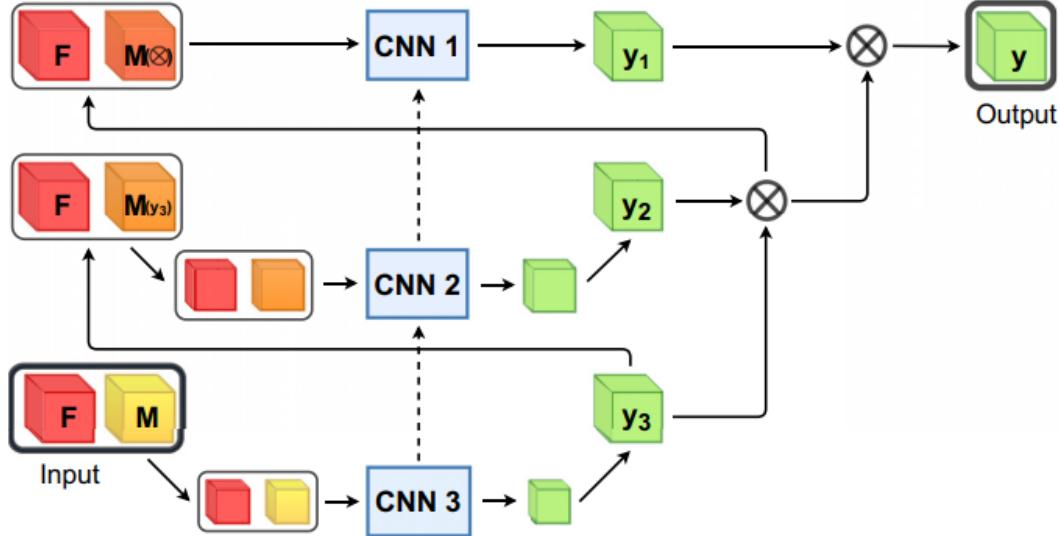


Figure 2.5: AlexNet architecture [46].

VIRNet

VIRNet (Multilevel Variational Image Registration Network) emerged as a state-of-the-art model for CNN-based approaches to image registration to lung CT scans and is a good example of the advantages of CNN-based approaches to conventional image registration methods [47].

VIRNet makes use of a Gaussian-pyramid-based multilevel framework (Figure 2.6), which is built upon another common CNN architecture, UNet [48]. This architecture allows the CNN to handle multilevel image registration tasks, which is particularly useful when dealing with a range of deformations, from large to small. The multilevel framework also allows for the computation of deformation fields at different scales, first at a broad level and then refined at finer resolutions (“coarse-to-fine fashion”).

Figure 2.6: VIRNet Architecture [47]. F indicates the fixed image, M the moving image, y the deformation field and $M(y)$ the warped image.

VIRNet combines elements of the conventional registration method of Rühaak et al. and Jacobian determinant-based regularisation [49]. This regularisation ensures the smoothness of the deformation field and penalises transformations that would result in physically implausible changes in volume. The use of this regulariser is similar to the strain energy function regularisation that is used in WarpPINN, the baseline model of this project, as discussed later [20].

The ability to progressively learn the optimal parameters for image registration from coarse to fine presents a significant advantage over traditional registration methods when handling large deformations, which struggle due to the complexity of the optimisation landscape [50][51]. The learning of hierarchical features means that DL models are able to capture such deformations more effectively, which is prevalent in the case of VIRNet’s original application on the deformations of the lungs when breathing.

Furthermore, Table 2.2 presents the results of various conventional and DL-based image registration methods and their performance on the DIRLAB 4D-CT dataset. Here, it is emphasised that, although conventional methods generally

outperform DL-based methods in achieving a lower TRE (target registration error), the runtime of DL models are much faster at inference.

Method	Mean TRE (mm)	Runtime
Conventional		
Schmidt-Richberg et al. (2011) [52]	1.38 ± 0.87	83min
Heinrich et al. (2012) [53]	1.6 ± 1.7	20min
Heinrich et al. (2013) [50]	1.43 ± 1.3	7.97min
Berendsen et al. (2014) [54]	1.36 ± 0.99	-
Rühaak et al. (2017) [49]	0.94 ± 1.06	5min
Deep Learning		
Sentker et al. (2018) [55]	2.5 ± 1.16	few seconds
de Vos et al. (2019) [38]	2.64 ± 4.32	0.63s
Eppenhof et al. (2019) [56]	2.43 ± 1.81	0.56s
Hering et al. (2019a) [57]	2.19 ± 1.62	-
Hansen and Heinrich (2020) [58]	1.97 ± 1.42	-
Jiang et al. (2020) [59]	1.66 ± 1.44	1.4s
Fu et al. (2020) [60]	1.59 ± 1.58	1min
Hansen and Heinrich (2021) [61]	1.39 ± 1.29	2s
VIRNet [47]	1.14 ± 0.76	0.75s

Table 2.2: Comparison of performances of different image registration models on the DIRLAB 4D-CT dataset [62][63][47].

Implicit Deformable Image Registration

IDIR (Implicit Deformable Image Registration) is an alternative approach to leveraging neural networks for image registration [64] and will be a focal point of the studies carried out in this thesis. Unlike most other DL-based registration methods, IDIR does not use a CNN and instead uses a multi-layer perceptron (MLP) to implicitly represent a transformation function.

The image registration and learning process is summarised in Figure 2.7. By parameterising the transformation, ϕ , as an implicit neural representation in an MLP, the network can represent a continuous and meshless function that is not restricted to any particular grid resolution. This representation is effective since any well-behaved function can be described with arbitrarily high accuracy by an appropriate neural network according to the universal approximation theorem [65].

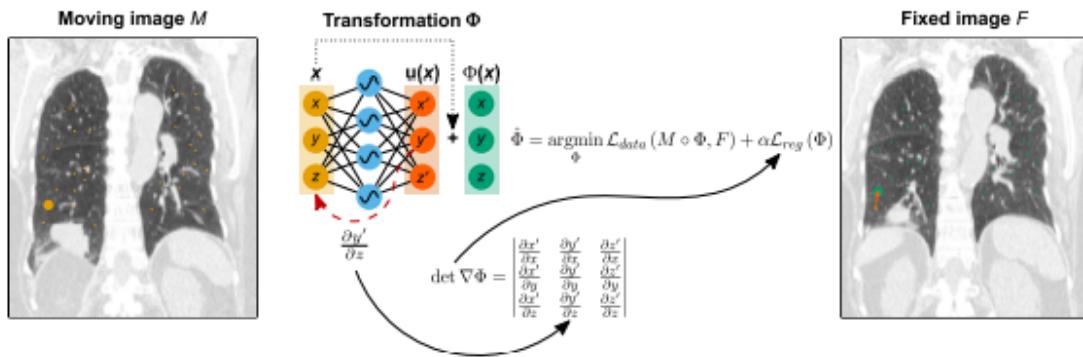


Figure 2.7: IDIR framework for image registration [64].

This problem is prevalent in conventional image registration techniques such as FFD, where the resolution of the grid of control points determines the precision of the deformation. The accuracy of representation is also a concern in CNN-based approaches where a CNN architecture would have to be adapted for different image sizes whilst the MLP can serve as a relatively lightweight model regardless of image resolution.

Furthermore, implicit neural representation through the MLP means that standard deep learning libraries (such as TensorFlow and PyTorch) can be exploited to numerically compute gradients of the transformation through automatic

differentiation as opposed to finite difference approximations [66][38], yielding more accurate gradients.

These numerous advantages, alongside the flexibility and ease of implementation offered through IDIR, have shown the method’s effectiveness in outperforming other CNN-based approaches (including VIRNet [47]) in a direct comparison on the DIR-LAB dataset. These experimental results are reported in Table 2.3. This could be due to improved precision in the continuous function that is being optimised, as well as the ability to compute accurate analytical gradients.

Finally, IDIR also does not require large sets of training data in comparison to other DL-based registration methods. Instead, a new network is initialised and optimised for each image pair. This is analogous to intentionally training on the test data in standard machine learning and is an effective method to overcome the challenge of DL-based methods over conventional image registration methods.

Later, when discussing WarpPINN as a baseline model for this project, the principles of IDIR and Neural Implicit Representation networks are used in combination with physics-informed machine learning, allowing neural networks to learn from less data by incorporating awareness of the physical constraints of a system [20]. Such constraints also allow for more effective regularisation to predict more biomechanically plausible deformations and will be discussed in the next section.

Scan	IDIR	isoPTV¹	CNN²	VIRNet	TRE before registration
4DCT 01	0.76 (0.94)	0.76 (0.90)	1.21 (0.88)	0.99 (0.47)	4.01 (2.91)
4DCT 02	0.76 (0.94)	0.77 (0.89)	1.13 (0.65)	0.98 (0.46)	4.65 (4.09)
4DCT 03	0.94 (1.02)	0.90 (1.05)	1.32 (0.82)	1.11 (0.61)	6.73 (4.21)
4DCT 04	1.32 (1.27)	1.24 (1.29)	1.84 (1.76)	1.37 (1.03)	9.42 (4.81)
4DCT 05	1.23 (1.47)	1.12 (1.44)	1.80 (1.60)	1.32 (1.36)	7.10 (5.14)
4DCT 06	1.09 (1.03)	0.85 (0.89)	2.30 (3.78)	1.15 (1.12)	11.10 (6.98)
4DCT 07	1.12 (1.00)	0.80 (1.28)	1.91 (1.65)	1.05 (0.81)	11.59 (7.87)
4DCT 08	1.21 (1.29)	1.34 (1.93)	3.47 (5.00)	1.22 (1.44)	15.16 (9.11)
4DCT 09	1.22 (0.95)	0.92 (0.94)	1.47 (0.85)	1.11 (0.66)	7.82 (3.99)
4DCT 10	1.01 (1.05)	0.82 (0.89)	1.79 (2.24)	1.05 (0.72)	7.63 (6.54)
Average	1.07	0.95	1.83	1.14	8.52

Table 2.3: Comparison of image registration performance of IDIR against a ¹then-state-of-the-art iterative algorithm using isotropic total variation regularisation [67], a ²CNN-based iterative algorithm [68] and VIRNet on the Dirlab 4D-CT dataset. The reported figures are the [Mean (Standard Deviation)] target registration error (TRE) in mm.

2.3 Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs) are a subset of neural networks that integrate established physical laws into their predictive models [69]. This means that they learn from data while aligning with embedded physical principles, aiding interpretability and explainability [70]. In the context of image registration, which poses an ill-posed problem of registering realistic deformations, the ability of PINNs to adhere to strict physical constraints not only help them learn from less data, but also allow for the prediction of more biomechanically and physically plausible transformations than previously outlined methods [20].

Biases in physics-informed machine learning

Neural networks are made “physics-informed” through the introduction of one or more of the following biases in the learning process: observational, inductive and learning bias. Observational biases are introduced directly through the dataset such that the training data reflects the underlying physical phenomena, which can be done through a variety of techniques, e.g., data augmentation. This method is conceptually the simplest mode of embedding physics into neural network training [70]. An example of this is Li et al.’s Fourier neural operator, where the integral kernel is parametrised directly in the Fourier space, allowing a neural network to be trained as a neural operator capable of modelling turbulent flows with zero-shot super-resolution [71].

Inductive biases correspond to prior assumptions of physical laws and can be incorporated through the architecture of a neural network. A common occurrence of this is through CNNs. For example, GU-Net was developed by replacing all transposed convolution and batch normalisation layers in U-Net with group equivariant counterparts in order to exploit

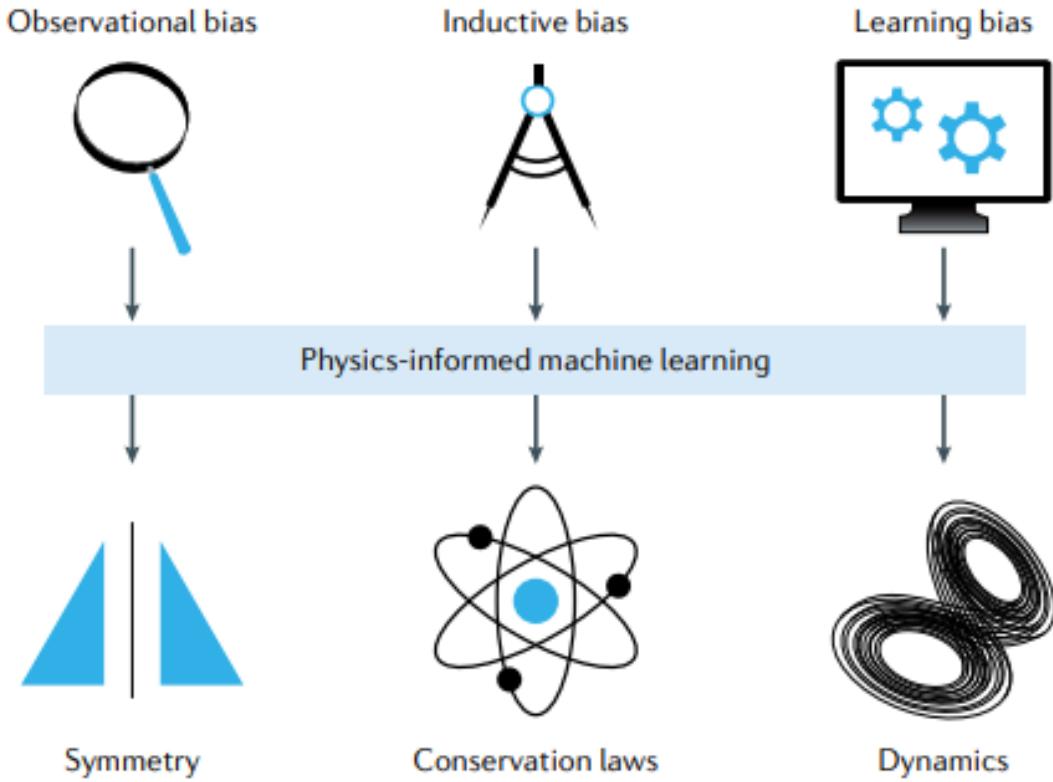


Figure 2.8: Biases that can be introduced to embed physical constraints in machine learning models, as described by Karniadakis et al. [70]

rotation and reflection symmetries inherent in histopathology images [72]. Finally, learning biases can be introduced through the loss functions used. Figure 2.9 shows an example of a “physics-uninformed” NN (a regular ANN) that has a modified loss function adhering to the viscous Burgers PDE. In the next discussion of the WarpPINN method, the quasi-incompressibility of the myocardial tissue is enforced in the process of learning from cardiac data by introducing a learning bias [20].

2.3.1 WarpPINN

This project will draw strong focus on López et al.’s WarpPINN and will serve as baseline model for performing the necessary image registration for cardiac strain analysis in this project. The WarpPINN methodology is a PINN-based approach to registering the deformations of the LV across the cardiac cycle [20]. The use of a neural network to implicitly represent the deformation field as in IDIR means that WarpPINN can learn from just one patient’s data and leverage some advantages of IDIR over other non-DL based algorithms. In addition, the incorporation of physics through the enforcement of the quasi-incompressibility of the myocardium within the loss function, means that the learnt and predicted deformations are expected to be more biomechanically plausible than traditional methods.

Methodology

The original WarpPINN architecture (Figure 2.10) is a relatively uncomplex one, consisting of 5 fully connected hidden layers (each containing 64 neurons). The input layer contains 4 neurons, corresponding to a 3-dimensional image, \mathbf{X} , as well as time, t , as an input. The output layer contains 3 neurons, corresponding to a learnt 3-dimensional non-rigid transformation, φ . In this sense, the fully connected NN represents a displacement field $\mathbf{u}(\mathbf{X}; \theta)$ where θ represents the weights/parameters of the NN and $\varphi(\mathbf{X}; \theta) = \mathbf{X} + \mathbf{u}(\mathbf{X}; \theta)$. This neural implicit representation technique for image registration is a type of IDIR [64].

In order to improve the results of image registration and yield more biomechanically plausible deformations, WarpPINN incorporates physical knowledge of cardiac contraction. In particular, the (quasi-)incompressibility [73] of the myocardial tissue is enforced through the use of a binary mask of the myocardium, indicating whether a voxel is inside or outside the LV. A learning bias is then introduced into the loss function, which takes the generalised form: $\mathcal{L}(\theta) = \|\mathbf{R} - \mathbf{T} \circ \varphi\|_p^p + \mu \mathcal{R}(\varphi)$.

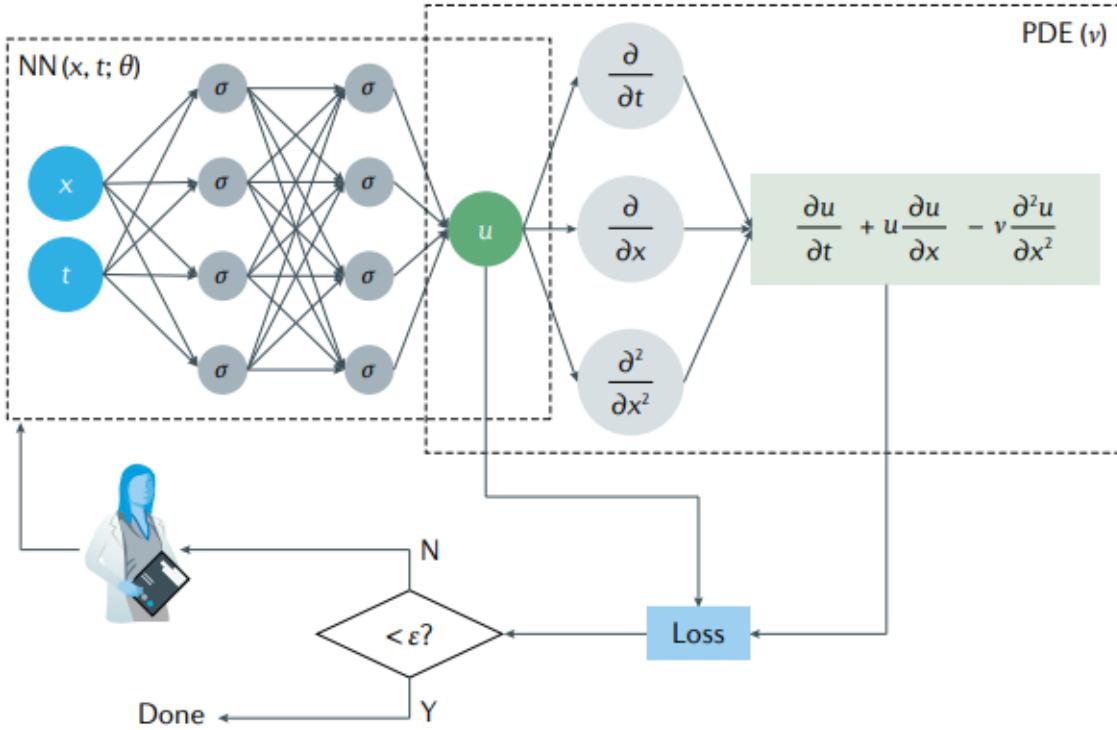


Figure 2.9: An example of a PINN framework to solve the viscous Burgers' equation. $NN(x, t; \theta)$ alone represents a "physics-uninformed" NN and the introduction of a learning bias through a modified loss function allows the PINN to solve the PDE [70].

Here, \mathbf{R} is the reference image and \mathbf{T} is the template image. Hence, the first term of the loss function corresponds to a similarity metric (between the reference image and the warped template image) to be minimised. This could be either L1 loss (MAE) or L2 loss (MSE).

On the other hand, the second term of the loss function is a regularisation term for the predicted deformation. This term is considered to be the physics-informed loss where μ is a hyperparameter and $\mathcal{R}(\varphi)$ is chosen to be the neo-Hookean hyperelastic strain energy function of the form: $W(\varphi; \lambda) = \text{Tr}(\mathbf{C}) - 3 - 2 \log(J) + \lambda(J - 1)^2$, where $\mathbf{F} = \frac{d\varphi}{d\mathbf{x}}$ is the deformation gradient, $\mathbf{C} = \mathbf{F}^T \mathbf{F}$ is the first invariant of the right Cauchy-Green deformation tensor and $J = \det(\mathbf{F})$ is the determinant of the Jacobian of the deformation gradient [74].

The introduction of the Jacobian is particularly important as it represents changes in volume in the material, where $J = 1$ indicates volume conservation (incompressible behaviour). The $\lambda(J - 1)^2$ term penalises changes in volume and is zero when incompressibility is satisfied. Thus, the incompressibility of the myocardium can be enforced by splitting the regulariser term $\mathcal{R}(\varphi)$ into 2 separate strain energy functions, for voxels inside the LV and voxels outside the LV, each with different λ parameters. Therefore, the complete loss function takes the form as in Equation 3.11.

$$\mathcal{L}(\theta) = \|\mathbf{R} - \mathbf{T} \circ \varphi\|_p^p + \mu \mathcal{R}(\varphi) \quad (2.9)$$

$$\begin{aligned} &= \frac{1}{N_d} \sum_i^{N_d} \left(R(\mathbf{X}_i^d) - \mathbf{T}(\varphi(\mathbf{X}_i^d; \theta)) \right)^p \\ &+ \mu \left(\frac{1}{N_{\text{inc}}} \sum_l^{N_{\text{inc}}} W(\varphi(\mathbf{X}_l^{\text{inc}}; \theta); \lambda_{\text{inc}}) + \frac{1}{N_{\text{bg}}} \sum_k^{N_{\text{bg}}} W(\varphi(\mathbf{X}_k^{\text{bg}}; \theta); \lambda_{\text{bg}}) \right) \end{aligned} \quad (2.10)$$

where X_i^d represents the position of the i^{th} pixel in the reference image, R of size N_d . The points $\{\mathbf{X}_l^{\text{inc}}\}_{l=1}^{N_{\text{inc}}} \subset \Omega_{\text{inc}}$ and $\{\mathbf{X}_k^{\text{bg}}\}_{k=1}^{N_{\text{bg}}} \subset \Omega_{\text{bg}}$ are collocation points, with Ω_{inc} being the nearly-incompressible region where the cardiac tissue is located and Ω_{bg} being the background of the image.

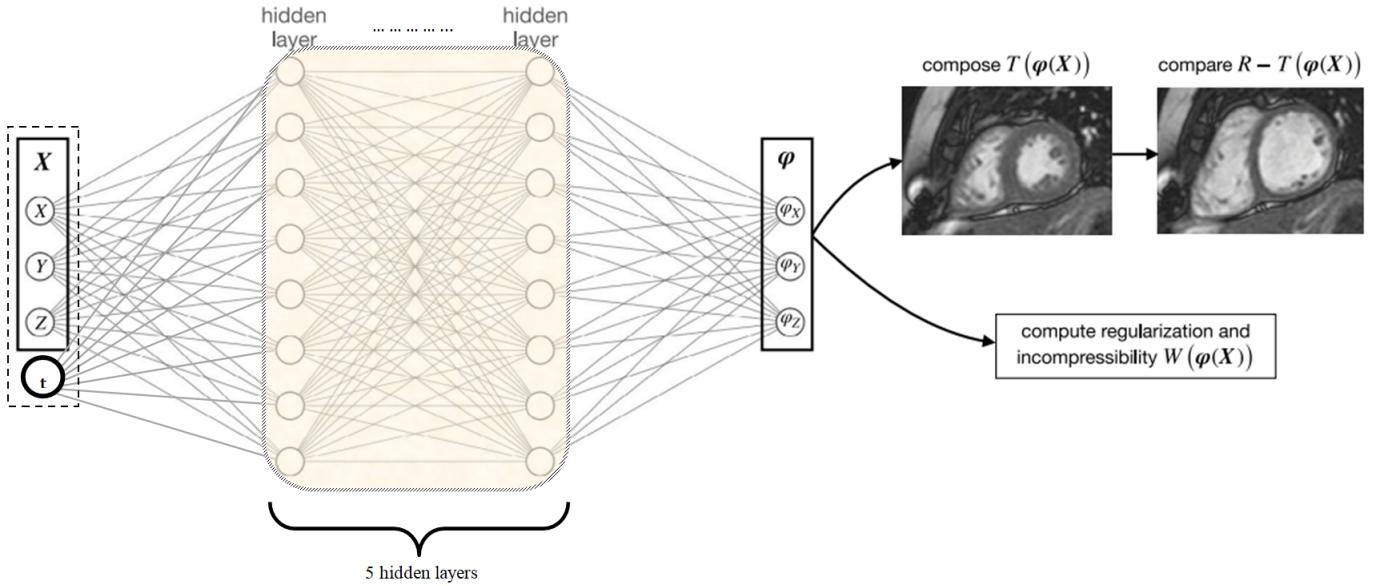


Figure 2.10: The WarpPINN architecture [20].

Open topics of exploration

As with any deep learning model, adapting WarpPINN to work with other datasets such as the ACDC (Automated Cardiac Diagnosis Challenge) [75] and UKBB (UK Biobank) [76] datasets could yield clinically valuable results and provide more insights into the image registration capability of WarpPINN. When discussing this project's goals in the next section, particular focus will be brought to adapting WarpPINN to use the ACDC dataset to leverage the diagnostic capabilities of cardiac strain analysis on patients with different types of cardiomyopathies.

Furthermore, there is value in investigating the acceleration of the training process (as WarpPINN is trained individually per patient) by employing techniques such as initial alignment and transfer learning. In this work, extended research investigates generalisation strategies for Neural Implicit Representation networks such as WarpPINN.

2.3.1.1 WarpPINN-FF

Spectral bias is a phenomenon observed during the training of neural networks, where NNs are biased towards learning less complex functions [77]. This is evident in WarpPINN where the low-frequency component of the deformation mapping is learnt first and this point is even discussed in the results obtained in López et al.'s validation of WarpPINN in section 2.3.1 where the smoothness of the predicted strain curves can be attributed through this spectral bias [20]. A consequence of this spectral bias, however, is that high-frequency features can take a relatively larger amount of training time, or the network is unable to learn sharp changes, resulting in predictions that are overly smooth and/or inaccurate. This consequence drives the motivation for the development of such models like VIRNet, as discussed previously in the background section [47].

To overcome this consequence, two common approaches is to pre-process the inputs with Fourier feature mapping [78][79][80] or to use a periodic activation function [81] (The latter is implemented in Wolterink et al.'s proposition of IDIR [64]).

López et al. chooses to implement the pre-processing approach by proposing the WarpPINN-FF model, which is a variation of the WarpPINN architecture with Fourier feature mappings. In this approach, the input of WarpPINN is modified from a low-frequency domain to a higher dimensional, higher frequency domain with Fourier feature mappings.

Let \mathbf{X} be the input images of WarpPINN. Defining the matrix $\mathbf{B} \in \mathbb{R}^{m \times d}$ where m is a positive integer, d is the input dimension and each entry $\mathbf{B}_{i,j}$ is sampled independently from a normal distribution $\mathcal{N}(0, \sigma^2)$. The Fourier feature mapping γ is thus defined as:

$$\gamma(\mathbf{X}) = \begin{bmatrix} \cos(\mathbf{BX}) \\ \sin(\mathbf{BX}) \end{bmatrix} \quad (2.11)$$

Here, the $\sigma > 0$ hyperparameter controls the frequency spectrum of the input domain and the hyperparameter m controls the dimensionality of the input space.

By using a higher dimensional vector $\gamma(\mathbf{X}) \in \mathbb{R}^{2m}$ as the input to WarpPINN instead of the original vector $\mathbf{X} \in \mathbb{R}^d$, the spectral bias can be overcome. The results of the Fourier feature mapping can be visualised with sample images as in Figure 2.11.

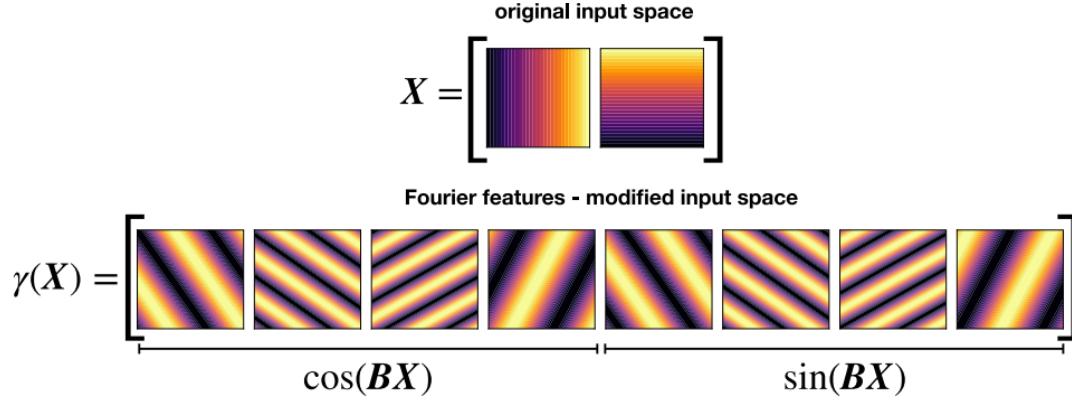


Figure 2.11: Demonstration of modifying the network input to a higher dimensional vector using Fourier feature mapping [20].

The effects of using WarpPINN-FF over WarpPINN can be visualised in Figure ??, where WarpPINN predicts a smooth Jacobian whereas WarpPINN-FF captures the sharp discontinuities on the boundary of the ring. This additional factor impacts the overall quality of predicted realistic deformations. As shown in Figure ??, WarpPINN failed to register the deformation properly for volunteer 5, thus predicting a deformation field close to the identity (Jacobian is 1 almost everywhere) whereas WarpPINN-FF was able to predict larger transformations (Figure ??) and successfully register the deformation.

These effects can be quantitatively assessed. As depicted in Figure ??, the landmark error distribution for WarpPINN-FF is lower than WarpPINN and Figure ?? shows the Fourier feature mappings' ability to predict larger strains.

It is expected that WarpPINN-FF would perform better when there is high resolution data available since the Fourier features help to fit the higher frequency component that might appear in the deformation field whereas WarpPINN would perform better when data is of lower resolution and noisy since the spectral bias will act as implicit regularisation resulting in a smoother solution. However, in all cases, WarpPINN-FF can be adapted to fit lower frequency functions either by decreasing the value of sigma in the Fourier features or by increasing the importance of the hyperelastic regularisation with the mu hyperparameter. Therefore, although all the experiments performed using WarpPINN at this stage of the project have been done with the baseline WarpPINN model, this project will incorporate the use of WarpPINN-FF in the later development and benchmarking stages.

3

Methods

Contents

3.1 Validating WarpPINN on CMAC	19
3.1.1 Dataset	19
3.1.2 Anatomical landmark tracking	19
3.1.3 Strain estimation	20
3.1.4 Benchmark models/methodologies	21
3.1.5 Training regime	23
3.1.6 Evaluation methodology	26
3.1.7 Ablation Study: Nullifying the Physics-Informed Loss	29
3.1.8 Hyperparameter tuning	30
3.2 Adapting WarpPINN to ACDC	33
3.2.1 Dataset	33
3.2.2 Further preprocessing: Surface mesh generation	35
3.2.3 Training regime	37
3.2.4 Alternative methodologies for comparison	37
3.2.5 Evaluation method	38
3.2.6 Hyperparameter tuning	40

This section outlines the different methodologies studied and discussed in this thesis.

3.1 Validating WarpPINN on CMAC

Motivation: This project utilises López et al.’s WarpPINN model as a baseline approach for cardiac image registration using neural implicit functions, emphasising the methodology. It is therefore essential to first validate the experimental setup before progressing to the adaptation of the model to the ACDC dataset, as well as further modifications/improvements to the WarpPINN approach. This validation task follows a methodology similar to López et al., evaluating image registration performance and plausibility through various metrics and methods, as discussed below.

3.1.1 Dataset

The validation experiments use the same dataset as that used in the original WarpPINN paper, that is the CMAC (Cardiac Motion Analysis Challenge) dataset from the Cardiac Atlas Project [82] [83]. This dataset corresponds to a motion tracking challenge proposed by MICCAI (Medical Image Computing and Computer Assisted Interventions) on the workshop “Statistical Atlases and Computational Models of the Heart: Imaging and Modelling Challenges” (STACOM’11) in 2011 [84]. The goal of the challenge was to accurately quantify the motion of the Left Ventricle during the cardiac cycle from different imaging modalities (Table 2.1). The dataset consists of 15 healthy volunteers imaged with cine SSFP, tagged MR and 3D ultrasound.

For each volunteer, the CMAC dataset also includes a bespoke surface mesh representing the segmentation of the LV at the end-diastolic phase, as well as 12 ground truth anatomical “landmarks” (Figure ??). There is one landmark per wall (anterior, lateral, posterior, septal) per ventricular level (basal, mid-ventricular, apical). These landmarks were generated from the 3DTAG datasets and then manually tracked by two observers. The landmarks were then post-processed and converted to SSFP coordinates with DICOM header information.

It should be noted that, during the duration of this project, the complete dataset was not available and only data for Volunteer 1 was accessible. Nevertheless, this provides an excellent situation and sufficient data for WarpPINN’s validation task.

3.1.2 Anatomical landmark tracking

The tracking of anatomical landmarks is a popular method of assessing accuracy in both 2D and 3D image registration, as mentioned previously in Section 2.2.1 with works such as Hill et al.[21].

The CMAC dataset includes the position of 12 of these anatomical landmarks for each frame in the cardiac cycle. These act as the ground truth positions for the landmarks.

In the WarpPINN approach, the ground truth landmarks accompany the ground truth surface mesh that is provided as an input (as they are in the reference frame) to be deformed. The positions of each landmark can thus be predicted for each frame and compared with the ground truth positions to obtain an error measurement. Landmark tracking error is defined as the sum of the Euclidean distances between each landmark’s ground truth position and its corresponding predicted position.

It should be noted that due to temporal misalignment between the different modalities (3DTAG and SSFP) during the generation of the landmarks’ positions, landmarks in SSFP coordinates can only be compared at the end-diastole and end-systole frame.

Therefore, in this validation task, the total landmark tracking error at end-systole can be measured and image registration performance can be measured by how close this error is to 0. Landmark tracking errors of WarpPINN can also be directly compared to that of other alternative image registration methodologies, namely UPF, INRIA and CarMEN. These methods are discussed in more detail later (Section 3.1.4).

3.1.3 Strain estimation

After training is complete and WarpPINN has predicted the deformation field φ , the computation of the myocardial strain tensor $\mathbf{E}(\mathbf{X}, t)$ is relatively straightforward (Equation 3.1) since the deformation gradients are calculated with automatic differentiation, an advantage of the WarpPINN’s architecture.

$$\mathbf{E}(\mathbf{X}, t) = \frac{1}{2}(\mathbf{C} - \mathbf{I}) = \frac{1}{2}(\mathbf{F}^T \mathbf{F} - \mathbf{I}) \quad (3.1)$$

where $\mathbf{C} = \mathbf{F}^T \mathbf{F}$ is the right Cauchy-Green deformation tensor and $\mathbf{F} = \frac{d\varphi}{d\mathbf{X}}$ is the deformation gradient tensor.

This tensor can then be projected to the heart’s coordinate system to find the radial, longitudinal and circumferential strains.

In reporting estimated myocardial strain, it is common to compute the deformation of the myocardium along three specific directions: radial, circumferential and longitudinal. These directional strains are illustrated in Figure 3.1.

To formulate the longitudinal, radial and circumferential strains, the myocardial strain tensor can be projected along each direction p :

$$\mathbf{E}_p(\mathbf{X}, t) = p^T \cdot \mathbf{E}(\mathbf{X}, t) \cdot p \quad (3.2)$$

The directional vector, p , for each of the directions is derived from the surface mesh provided in the CMAC dataset, and are defined as follows:

- **Longitudinal (l):** Aligned from the apex to the base of the Left Ventricle, along the z-axis of the coordinate system. $l = (0, 0, 1)$.
- **Radial (r):** Perpendicular to the endocardial surface at each point. To obtain the directional vector r , the normal vector e to each node of the mesh is computed, and then the vertical component is subtracted: $r = e - (e \cdot l)l$.
- **Circumferential (c):** Tangential to the myocardial surface and perpendicular to the longitudinal direction, computed as the cross product of l and r : $c = l \times r$.

Therefore, each directional strain is formulated as such:

- **Longitudinal strain:** $E_{ll}(\mathbf{X}, t) = l^T \cdot \mathbf{E}(\mathbf{X}, t) \cdot l$
- **Radial strain:** $E_{rr}(\mathbf{X}, t) = r^T \cdot \mathbf{E}(\mathbf{X}, t) \cdot r$
- **Circumferential strain:** $E_{cc}(\mathbf{X}, t) = c^T \cdot \mathbf{E}(\mathbf{X}, t) \cdot c$

3.1.4 Benchmark models/methodologies

In this validation task, the performance of WarpPINN is directly compared to that of three other different image registration methodologies:

- **Temporal Diffeomorphic Free Form Deformation (TDFFD) or “UPF” [85]:** This method models a $3D+t$ velocity field using B-Spline spatiotemporal kernels. The displacement field is computed via Euler integration of the velocity field. Image similarity is achieved through the minimisation of the Mean Squared Error (MSE) between the intensities of the reference frame and other frames. To enforce quasi-incompressibility, a regulariser that favours a zero-divergence velocity field is added. This method allows for accurate tracking of deformations over time by capturing the temporal evolution of the velocity field. TDFFD was developed at the Universitat Pompeu Fabra and hence, for convenience, this method will often be referred to as “UPF” in following experiments.
- **iLogDemons (“INRIA”)** [86]: iLogDemons is an advanced modification of the widely used logDemons algorithm, specifically tailored for cardiac motion tracking. This method incorporates elasticity and incompressibility constraints through a diffeomorphic transformation parameterised by a stationary velocity field. This ensures that the transformation is smooth and invertible, which is critical for accurately modelling biological tissues, such as the myocardium. iLogDemons enforces full incompressibility by including a divergence-free constraint. iLogDemons was proposed by the Inria-Asclepios project and hence, for convenience, this method will often be referred to as “INRIA” in following experiments.
- **DeepStrain or CarMEN** [87]: CarMEN (Cardiac Motion Estimation Network) is a CNN-based approach that is used to estimate the deformation field between reference and template images. The network processes two input

images and outputs the corresponding deformation field. The template image is deformed using trilinear interpolation, and an L1-norm loss function measures the similarity between the deformed template and the reference image. Additionally, a diffusion regulariser based on the Jacobian of the deformation field is applied to ensure smooth and realistic deformations, ensuring realistic and plausible deformations.

The use of these three methods, a mix of deep learning-based (CarMEN) and non-deep learning-based (UPF and INRIA) methods as comparisons will provide useful benchmark models for the validation of WarpPINN, when performing the image registration task on the same CMAC dataset. As will be investigated thoroughly in later sections and observe throughout this work, the use of the Jacobian determinant of the deformation field as a regulariser to ensure volume preservation is an important and relevant method of ensuring biomechanical plausibility. This technique is employed in both the WarpPINN approach and CarMEN, whereas UPF and INRIA utilise constraints on the velocity field (which captures spatial and temporal deformations).

3.1.5 Training regime

In this validation task, WarpPINN is first trained with an identical regime as that originally proposed by López et al.

Architecture: A Multi-Layer Perceptron (MLP) with 5 hidden layers. This architecture is illustrated in Figure 2.10.

- **Input layer:** The input layer consists of 4 neurons. These correspond to the 3D spatial coordinates (x, y, z) of the Template image to be deformed, at a given time t ($3D + t$).
- **Hidden layers:** Each of the 5 hidden layers consists of 64 neurons per layer. A hyperbolic tangent (\tanh) is used between the layers to introduce non-linearity.
- **Output layer:** The output layer consists of 3 neurons. These correspond to the network's output of a 3-dimensional vector representing the components of the deformation field (u_1, u_2, u_3) .

Weight/Bias Initialisation: Weights are initialised with Xavier initialisation and bias vectors are initialised as 0.

Pretraining: In addition to the network's weights and biases' initialisation, a pretraining step is done in order to initialise the network's parameters to approximate the identity mapping. The objective of this pretraining step is to aid the network in starting close to a meaningful solution, avoiding potential problems that could arise from the random initialisation of the neural network, such as poor local minima or predicted displacement fields that produce negative volumes.

Pretraining is done by training the network to predict deformations that are close to zero, by minimising the MSE/L2 loss function over the entire domain, as in Eq. 3.3. This pretraining is done until the MSE or pretraining loss attains a value below a predefined tolerance. In this validation task, this tolerance is chosen to be 10^{-6} .

$$\mathcal{L}_{\text{pretrain}}(\theta) = \frac{1}{N_d} \sum_i^{N_d} \|\mathbf{u}(\mathbf{X}_i^d)\|_2^2 \quad (3.3)$$

Loss Function: During training, the composite loss function shown in Eq. 3.5, which was previously discussed in Section 2.3.1 is minimised. In these validation experiments, the image similarity loss component of $\mathcal{L}(\theta)$ is the L1 loss

or Mean Absolute Error (MAE) between the Reference and Warped Template images. The Warped Template image is composed through the use of trilinear interpolation of the predicted deformation field (the output vector of the MLP) on the Template image.

$$\mathcal{L}(\theta) = \|\mathbf{R} - \mathbf{T} \circ \varphi\|_p^p + \mu \mathcal{R}(\varphi) \quad (3.4)$$

$$\begin{aligned} &= \frac{1}{N_d} \sum_i^{N_d} \left(R(\mathbf{X}_i^d) - \mathbf{T}(\varphi(\mathbf{X}_i^d; \theta)) \right)^p \\ &+ \mu \left(\frac{1}{N_{\text{inc}}} \sum_l^{N_{\text{inc}}} W(\varphi(\mathbf{X}_l^{\text{inc}}; \theta); \lambda_{\text{inc}}) + \frac{1}{N_{\text{bg}}} \sum_k^{N_{\text{bg}}} W(\varphi(\mathbf{X}_k^{\text{bg}}; \theta); \lambda_{\text{bg}}) \right) \end{aligned} \quad (3.5)$$

where X_i^d represents the position of the i^{th} pixel in the reference image R of size N_d .

The points $\{\mathbf{X}_l^{\text{inc}}\}_{l=1}^{N_{\text{inc}}} \subset \Omega_{\text{inc}}$ and $\{\mathbf{X}_k^{\text{bg}}\}_{k=1}^{N_{\text{bg}}} \subset \Omega_{\text{bg}}$ are collocation points, with Ω_{inc} being the nearly-incompressible region where the cardiac tissue is located and Ω_{bg} being the background of the image.

The incompressible (inc) and background (bg) regions of the images correspond to regions segmented by a boolean mask derived from the ground truth surface mesh provided in the CMAC dataset at end-diastole (which is the reference frame). The incompressible and background regions of the image are subject to separate physics-informed losses (which are modelled with the neo-Hookean hyperelastic strain energy function), which are parameterised by two different λ parameters, λ_{inc} and λ_{bg} .

These λ parameters describe the degree that the strain energy function penalises changes in volume ($W(\varphi; \lambda) = \text{Tr}(\mathbf{C}) - 3 - 2 \log(J) + \lambda(J - 1)^2$). In these validation experiments, λ_{inc} is chosen to be 10^5 whilst λ_{bg} is chosen to be 1. It is noted that it is possible to ignore the physics-informed loss on the background region altogether ($\lambda_{\text{bg}} = 0$), and doing such is found to have little effect on the image registration performance of WarpPINN. However, this parameter is set to a small constant as López et al. have originally done so. Later sections will explore the behaviour of these various training hyperparameters.

Finally, the μ hyperparameter controls the ratio and weight of the physics-informed loss to image similarity loss in the composite loss function. These validation experiments consider cases where $\mu = 1 \times 10^{-5}$ and $\mu = 5 \times 10^{-6}$. Later studies in this section explores cases where $\mu = 0$ (Section 3.1.7), as well as the impact of varying μ for hyperparameter tuning (Section 3.1.8).

Batching and Mini-batching Strategy: For each volunteer in the CMAC dataset, a sequence of 3D images (across the cardiac cycle) of the LV is provided. These represent the template images to be deformed by WarpPINN with the

objective of the warped template image being similar to the reference image. The reference image is chosen to be the first image in the data sequence.

For each template image T_j in the sequence of f frames, the loss function at every iteration is evaluated between T_j and R for a randomly chosen integer $j \in [1, f]$, i.e., the template image to be deformed at each iteration is randomly chosen and the corresponding time t_j is fed as an input to the network (The MLP's input layer has 4 neurons, corresponding to $3D+t$). The justification for this batching strategy is that the composition of T_j with trilinear interpolation is computationally expensive to perform for all j at each iteration.

The neo-Hookean loss function is also evaluated in mini-batches at each iteration for each region, Ω_{inc} and Ω_{bg} . This mini-batching strategy considers the space-time collocation points $\{(\mathbf{X}_i^{\text{inc}}, t_i)\}_i \subset \Omega_{\text{inc}} \times [0, 1]$ and $\{(\mathbf{X}_k^{\text{bg}}, t_k)\}_k \subset \Omega_{\text{bg}} \times [0, 1]$. The times $\{t_i\}_{i=1}^{N_{\text{inc}}}$ and $\{t_k\}_{k=1}^{N_{\text{bg}}}$ are randomly taken in $[0, 1]$ and resampled at every training iteration. In these validation experiments, a batch size of 1000 collocation points within the myocardium and 1000 collocation points in the background region is chosen.

Optimiser and Iterations: The WarpPINN model is trained for 300k iterations with an Adam optimiser, with a learning rate of 10^{-3} .

Data Preprocessing: In order to reduce the size of the data, as well as to accelerate the training process, the images are cropped around the region of interest (ROI), i.e., the LV.

Fourier Feature Mappings: As discussed in Section 2.11, the use of Fourier feature mappings (FFMs) as a pre-processing step to the input images serves to alleviate the spectral bias that is commonly observed during the training of neural networks, thus allowing WarpPINN to predict larger and more local deformations.

In these validation experiments, in addition to cases where $\mu = 1 \times 10^{-5}$ and $\mu = 5 \times 10^{-6}$, the use of FFMs are also considered. In such cases, the parameters are chosen to be $m = 8$ and $\sigma = 1$.

This mapping is represented by

$$\gamma(\mathbf{X}) = \begin{bmatrix} \cos(B\mathbf{X}) \\ \sin(B\mathbf{X}) \end{bmatrix} \quad (3.6)$$

such that $B \in \mathbb{R}^{m=8 \times d=3}$ and each entry $B_{i,j}$ in B is sampled independently from a normal distribution $\mathcal{N}(0, \sigma^2 = 1)$.

$\gamma(\mathbf{X}) \in \mathbb{R}^{2m}$ is used as an input to the network instead of the original vector $\mathbf{X} \in \mathbb{R}^d$. Therefore, the dimension of the input layer is changed to $2m + 1 = 17$ and a wider network is used. The number and dimensions/size of the hidden and output layers remain the same.

Throughout this work, the WarpPINN model that uses FFMs is referred to as WarpPINN-FF.

3.1.6 Evaluation methodology

Image registration performance can be evaluated with two main criteria: image similarity (between the warped and reference images) and biomechanical plausibility (volume preservation). These criteria are reflected in components of the loss function used for WarpPINN (Eq. 3.5).

3.1.6.1 Predicted images

The learnt deformation field φ , which has Cartesian components u_1 , u_2 , and u_3 (the outputs of the WarpPINN MLP) is applied to the Template images to produce the Warped Template image. In this image registration task, this warped template image is expected to be similar to the reference image. The reference image, R is the image at the end-diastolic frame and, while WarpPINN performs image registration to every T_j in the cardiac cycle (sequence of j frames), it is sensible and common practice in cardiac image registration to compare the end-diastolic and end-systolic frames. For Volunteer 1 of the CMAC dataset, the ES frame is the 12th image in the sequence.

Qualitative comparisons and visualisations of warped images can provide a good first indication of the biomechanical plausibility and can serve as an inspective measure of how realistic the predicted deformations are, whilst quantitative results of image similarity can measure how well the image registration error has been minimised.

Representative slices: Furthermore, it is common practice in cardiac image registration and in related work to compare 3D images using representative 2D z-axis slices. These are the basal, mid-ventricular and apical slices, corresponding to the ventricular levels of the Left Ventricle. The selection of these slices are illustrated in Figure 3.1 and defined as follows:

- Basal slice: 25% of the LV length.
- Mid-ventricular slice: 50% of the LV length.
- Apical slice: 75% of the LV length.

Performing the evaluation (both quantitative and qualitative) on these three representative slices provides a sufficient and good understanding of the motion and deformation at different planes of the cardiac volume, as well as having the

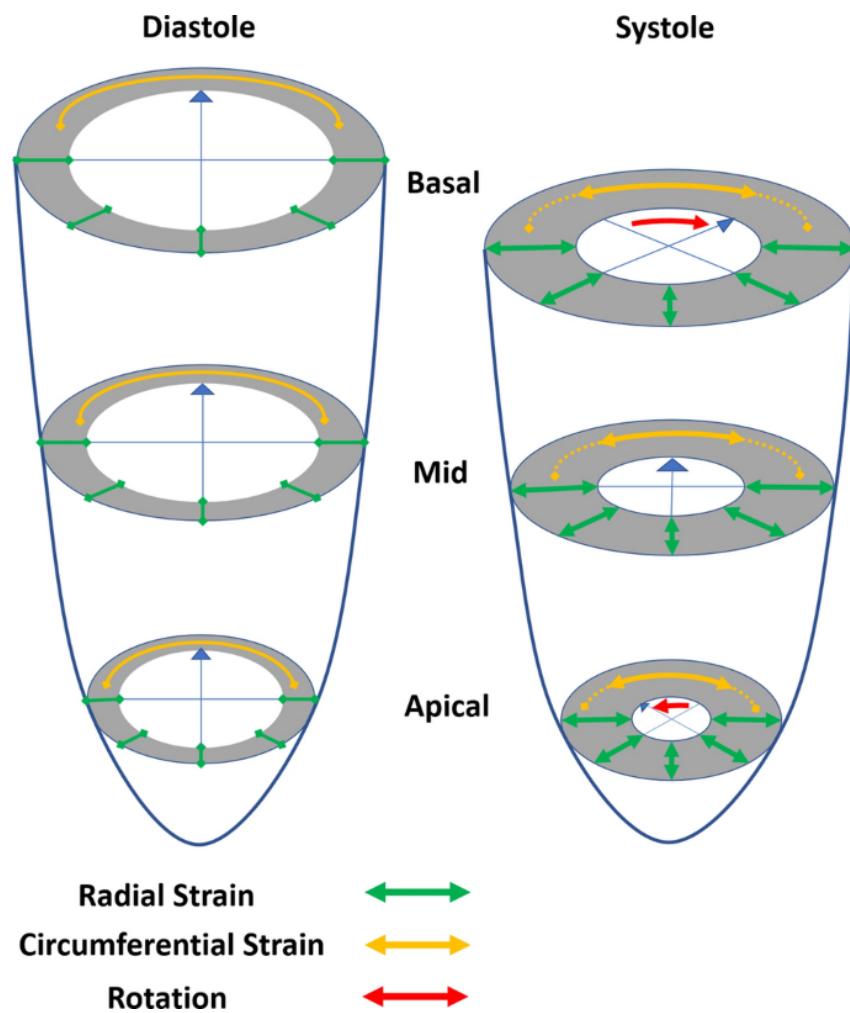


Figure 3.1: Illustration of the three representative short-axis slices of the Left Ventricle and directional strains. This image is reused from [88].

advantage of being more robust to outliers, which can be prevalent when directly comparing two 3D volumes, especially images with a much lower resolution in the z-axis than x and y.

Evaluation metrics (Image similarity): In order to measure the image similarity between images, several metrics are calculated between the representative slices of the warped template (ES) and reference (ED) images, namely the Normalised Root Mean Square Error (NRMSE), Structural Similarity Index Measure (SSIM) and Normalised Cross Correlation (NCC). These metrics are defined as shown in Eqs. 3.7–3.9.

$$\text{NRMSE}(I_1, I_2) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (I_1(i) - I_2(i))^2}}{I_{\max} - I_{\min}} \quad (3.7)$$

$$\text{NCC}(I_1, I_2) = \frac{\sum_{i=1}^N (I_1(i) - \bar{I}_1)(I_2(i) - \bar{I}_2)}{\sqrt{\sum_{i=1}^N (I_1(i) - \bar{I}_1)^2 \sum_{i=1}^N (I_2(i) - \bar{I}_2)^2}} \quad (3.8)$$

$$\text{SSIM}(I_1, I_2) = \frac{(2\mu_{I_1}\mu_{I_2} + C_1)(2\sigma_{I_1I_2} + C_2)}{(\mu_{I_1}^2 + \mu_{I_2}^2 + C_1)(\sigma_{I_1}^2 + \sigma_{I_2}^2 + C_2)} \quad (3.9)$$

It is noted that there exists more popular and commonly used image similarity metrics in medical image registration, such as the Dice Score Coefficient and Mean Contour Distance. However, since the CMAC dataset does not include ground truth segmentations for the myocardium or LV, these metrics are less appropriate. These segmentation mask-based metrics are utilised later on in experiments with the ACDC dataset.

3.1.6.2 Predicted meshes

WarpPINN learns a deformation field φ by warping a template image to match the reference image during training. Once WarpPINN is trained, a deformation field φ_j is learnt for every time index j (frame in the cardiac cycle / sequence). At inference, this can be used to predict the deformation of the provided ground truth surface mesh of the LV at ED, to predict the location of the mesh points, as well as landmarks, at ES ($j = 12$).

A predicted 3D surface mesh of the LV at ES provides two important opportunities for evaluation measure in the WarpPINN approach: the tracking of anatomical landmarks, as discussed previously, and a visualisation of the 3D deformation.

3D Deformation field and Jacobian: Visualising the 3D deformation field and predicted ES mesh allows for qualitative visualisation of how realistic the learnt deformation is. Furthermore, the determinant of the Jacobian (of the deformation

gradient), $\det(\mathbf{J})$, is an important and widely used measure of volume preservation in medical image registration. A Jacobian closer to 1 indicates a greater degree of volume preservation and thus, physical plausibility during registration.

The Mean Absolute Deviation of the Jacobian to 1, denoted by $|\mathbf{J}| - 1$ is used as another evaluation metric, in order to measure volume preservation and quantify the degree of biomechanical plausibility of learnt deformations.

3.1.6.3 Estimated strains

The computation of myocardial strain, as discussed previously, serves as another measure of validation for WarpPINN. Although there is no ground truth for cardiac strains, the estimated measurements of WarpPINN can be compared with that of other alternative methodologies, namely UPF and INRIA from the benchmark models.

Regional Strain Analysis: In addition to reporting the myocardial strain along the set of three directions (radial, longitudinal, circumferential), as previously discussed, related work on cardiac strain analysis have also reported myocardial strain in two separate regions of the left ventricle: the septum and the free wall. These regions correspond to the American Heart Association (AHA) 17-segment model, which is a standardised framework used in cardiac imaging to systematically describe the anatomy of the LV. This segmentation framework is visualised in Figure 3.2

Regional strain analysis for the Septum and the Free Wall, as opposed to Global strain analysis, has been recommended for comprehensive cardiac assessment. In this particular study, the use of regional strains could yield more robust estimates of strain, given anatomical and functional differences between the septum and the free wall, as well as different mechanical behaviours. For example, the septum experiences complex strain patterns due to its role in both left and right ventricular function, whilst the free wall typically shows distinct strain patterns, often with higher strain values in the lateral segments compared to the septum.

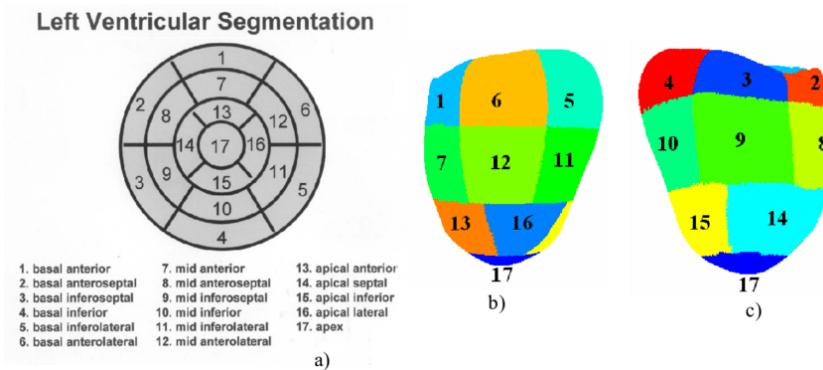


Figure 3.2: Visualisation of the AHA-17 segment model. The image is repurposed from [3]

3.1.7 Ablation Study: Nullifying the Physics-Informed Loss

Motivation: As discussed in the Background chapter (2), many different image registration approaches exist, including more accurate conventional methods and faster DL-based models. However, in the context of registering the LV's motion and deformation, and cardiac/medical image registration in general, the problem is ill-posed as transformations could be unrealistic, even with the use of regularisers. Therefore, López et al.'s PINN-based approach to impose a learning bias and incompressibility constraint through the use of the neo-Hookean hyperelastic strain energy function is well justified.

The aim of this experiment is to quantify the impact of the physics-informed component of WarpPINN's loss function by nullifying the neo-Hookean loss and observing the effects on image registration. Often, medical image registration algorithms must carefully balance between the two criteria of image similarity and biomechanical plausibility (in the context of cardiac image registration, myocardial volume preservation) and a trade-off is faced. A quantified understanding of this physics-based loss component, controlled by the μ hyperparameter, could well inform future hyperparameter tuning approaches for other patients and datasets throughout this project.

3.1.7.1 Approach

In this study, the WarpPINN model is trained under the same methodology and regime as in the Validation experiments, which will act as a baseline. The key difference is the change made to the loss function, such that the physics-informed or neo-Hookean component is no longer used. This is effectively setting the μ hyperparameter to 0:

$$\mathcal{L}(\theta) = \|\mathbf{R} - \mathbf{T} \circ \varphi\|_p^p + \mu \mathcal{R}(\varphi)^0 \quad (3.10)$$

$$= \frac{1}{N_d} \sum_i^{N_d} \left(R(\mathbf{X}_i^d) - \mathbf{T}(\varphi(\mathbf{X}_i^d; \theta)) \right)^p$$

$$+ \mu \left(\frac{1}{N_{\text{inc}}} \sum_l^{N_{\text{inc}}} W(\varphi(\mathbf{X}_l^{\text{inc}}; \theta); \lambda_{\text{inc}}) + \frac{1}{N_{\text{bg}}} \sum_k^{N_{\text{bg}}} W(\varphi(\mathbf{X}_k^{\text{bg}}; \theta); \lambda_{\text{bg}}) \right)^0 \quad (3.11)$$

Results for this experiment will follow the same evaluation and presentation as the validation experiments, with a particular focus on the trend in changes in image similarity performance and volume preservation as μ is changed from 1×10^{-5} to 5×10^{-6} and to 0.

3.1.8 Hyperparameter tuning

Motivation: While the ablation study on nullifying the physics-informed loss aims to quantify the impact of the μ hyperparameter on performance, there are still several hyperparameters in the WarpPINN method. Tuning hyperparameters for several WarpPINN use cases (such as each patient in the ACDC dataset) will become an important task in later sections of this work. Therefore, this section aims to study the relationship between different hyperparameters and parameter groups, and their impact on image registration performance.

3.1.8.1 Approaches

This study is separated into two subtasks, for quantifying the impact of two groups of hyperparameters:

- **Training parameters:** The μ parameter, learning rate and batch size. These are parameters that are commonly tuned in neural networks to ensure that models are optimised and converge to a suitable loss.
- **NeoHookean loss parameters:** μ , λ_{inc} and λ_{bg} . These parameterise the physics-informed regularisation component of the loss function. While the ablation study aims to focus on the impact of the Mu hyperparameter, studying the impact of λ_{inc} and λ_{bg} , which parameterise the degree that incompressibility is enforced in the myocardium and background regions, respectively, is important in order to understand the performance of WarpPINN when all three parameters are tuneable.

Since the training of a WarpPINN model from scratch is significantly time-consuming, hyperparameter tuning tasks are carried out with a reduced number of 100k iterations. It is observed in later documentation of results that, at 100k iterations, the loss of a WarpPINN model with the default hyperparameters reaches a reduction of $40 \sim 50\%$ of its convergence loss.

This reduced training iterations is, however, not used for tuning the Training parameters, since analysis of the behaviour of the loss curves presents meaningful results. Therefore, the Training parameters are tuned with a **One-Factor-at-a-Time (OFAT)** approach, where each parameter will be varied independently whilst the other two parameters are fixed. Therefore, WarpPINN is tested with each parameter independently changed to pre-determined variations, as such:

- μ : $[5 \times 10^{-6}, 1 \times 10^{-5}, 5 \times 10^{-4}, 5 \times 10^{-3}]$
- Learning rate: $[1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}]$
- Batch size: $[1000, 2000, 5000]$

On the other hand, the NeoHookean loss parameters are tuned with a grid search approach, where every point on the 3x3 grid of discrete values denotes a combination of hyperparameter values. This grid of points is pre-defined with $\mu =$

$[5 \times 10^{-6}, 1 \times 10^{-5}, 5 \times 10^{-4}]$, $\lambda_{\text{inc}} = [10^4, 10^5, 10^6]$ and $\lambda_{\text{bg}} = [0, 1, 10]$. These ranges are chosen based on the current default hyperparameters (3.1.5).

Furthermore, results of WarpPINN being trained with a $\lambda_{\text{bg}} = 0$ effectively serve to accompany the preceding ablation study, as this represents a nullification of the physics-informed loss on only the background region. This has significance since one consideration of the WarpPINN method is that the modelling of the background region (with artefacts such as blood and tissues) with the strain energy function is not well justified.

In each search method, image registration performance of WarpPINN on Volunteer 1 of the CMAC dataset is evaluated with measures of image similarity (SSIM) and volume preservation ($||J| - 1|$).

It is noted that these approaches do not constitute an extensive search of a large hyperparameter search space. While this is partly driven by the training times of WarpPINN being significantly longer with CMAC data in comparison to ACDC data (over twice as long), this study serves as an introduction to the behaviour of the NeoHookean loss parameters. Hence, the reduced training time. This study also aims to affirm that the ranges of the learning rate and batch size parameters used in the Validation study are sufficient to use in later experiments. A more extensive hyperparameter tuning approach is done in later experiments with the ACDC dataset.

3.2 Adapting WarpPINN to ACDC

Motivation: The Automated Cardiac Diagnosis Challenge (ACDC) dataset is a more comprehensive and larger dataset than CMAC (which WarpPINN was originally developed to use) containing information about patients from different pathological groups. Being able to perform and evaluate image registration on different types of cardiomyopathies provide a more detailed and robust evaluation of WarpPINN’s performance, as well as having the advantage of being able to compare results to more alternative methodologies employed in previous related work. This section is therefore concerned with adapting the WarpPINN methodology to be able to train and test on the ACDC dataset. These results will also serve as a basis for further extended work discussed in later sections.

3.2.1 Dataset

The ACDC dataset consists of cine MRI sequences for 150 patients, with sequences spanning the entire cardiac cycle from end-diastole to end-systole. Each time frame in the sequence of images corresponds to a series of short-axis slices that capture the LV from the base to the apex, similar to the CMAC dataset. The in-plane resolution of these slices are quoted to be in the range of 1.37 to 1.68mm. Across the 150 patients, the length of each MRI sequence range from 12 to 35 time frames [75].

The total ACDC cohort is further separated into a training (100 patients) and testing (50 patients) set. A segmentation mask of the myocardium is also provided at the end-diastolic and end-systolic frames, these were manually labelled by experts. Furthermore, the cohort is evenly distributed amongst 5 subgroups (4 pathological and 1 healthy set of patients). These groups are defined as follows:

- **NOR:** “Normal” or healthy subjects.
- **DCM:** Patients with **dilated cardiomyopathy (DCM)**, characterised by a diastolic left ventricular volume of $> 100 \text{ mL/m}^2$ and an ejection fraction of the LV of $< 40\%$.
- **HCM:** Patients with **hypertrophic cardiomyopathy (HCM)**, characterised by a left ventricular cardiac mass of $> 110 \text{ g/m}^2$, several myocardial segments with a thickness $> 15 \text{ mm}$ in diastole and a normal ejection fraction.
- **MINF:** Patients with previous **myocardial infarction (MINF)**, characterised by an ejection fraction of the LV of $< 40\%$ and several myocardial segments with abnormal contraction.
- **ARV:** Patients with **abnormal right ventricle (ARV)**, characterised by the volume of the RV cavity $> 110 \text{ mL/m}^2$ or an ejection fraction of the RV $< 40\%$.

The table below shows the distribution of patients across the different subsets:

Disease	Training Set	Testing Set
DCM	20	10
HCM	20	10
MINF	20	10
ARV	20	10
NOR	20	10

Table 3.1: Distribution of patients across training and testing sets for different pathology groups in the entire ACDC dataset.

It is noted that, unlike the CMAC dataset, there are no anatomical landmarks to facilitate landmark tracking as an evaluation measure in the ACDC dataset. However, a significant advantage is the inclusion of ground truth segmentations of the LV, which enable the use of more appropriate medical image registration metrics of image similarity, which are discussed further below.

3.2.1.1 Patient selection

Since the WarpPINN method is a type of Implicit Deformable Image Registration (IDIR), acting as a pairwise optimisation of images, it is not required to train the neural network on a training and validation set, as neural networks are traditionally done. However, in order to validly compare this work’s results with alternative methodologies in previous related work, a sufficient number of 10 patients from each disease (excluding ARV) is chosen to evaluate WarpPINN on. It is also noted in reporting that one patient (Patient 47 from the MINF group) was excluded from this evaluation due to errors in the surface mesh generation stage (3.2.2).

Disease	Number of patients trained on	Patient numbers
DCM	10	1-10
HCM	10	21-30
MINF	9	41-50 (excluding Patient 47)
NOR	10	61-70

Table 3.2: The chosen patients from each pathology group of the ACDC dataset that the WarpPINN method is evaluated with.

3.2.1.2 Preprocessing

As with the CMAC dataset, images from the ACDC dataset are cropped to reduce the number of pixels and accelerate the training process of WarpPINN. These cropped images enclose the ROI of the myocardium. To maintain consistency, this cropping process was automated by finding a unified 3D bounding box (with the minimum dimensions) which encapsulates the ROI of each short-axis slice, with a 5-pixel buffer.

3.2.2 Further preprocessing: Surface mesh generation

In addition to the absence of anatomical landmarks, a notable exclusion of the ACDC dataset is the lack of a ground truth surface mesh of the LV at the end-diastolic phase, as in CMAC. The surface mesh facilitates the learning of WarpPINN by scaling the magnitude of predicted deformations but, more importantly, is used to provide a coordinate system to project the computed myocardial strain tensor to estimate direction (Radial, Circumferential and Longitudinal) strains. Thus, this subsection discusses approaches employed to generate surface meshes from the provided ground truth segmentations of the LV in the ACDC dataset.

3.2.2.1 Basic mesh generation

The first discussion is the basic mesh generation method used to generate surface meshes as an input to WarpPINN. This method begins with the 3D binary mask of the myocardium, which is provided as ground truth in the ACDC data as a stack of short-axis segmentation masks.

This 3D binary mask is scaled using the known resolution of images and an initial surface mesh is created using the marching cubes algorithm, specifically, the Lewiner method is used to avoid degenerate triangles. This step is implemented using Python’s `scikit-image` library. An isosurface value close to 0 (10^{-12}) is used and a descending gradient direction is chosen (normal of the surface point in the direction of decreasing intensity), in order to accurately capture the geometry information provided by the binary masks in this initial step.

The initial surface mesh is cleaned to remove any irregularities, artefacts or noise. This step includes removing points that are not used by any cell in the mesh, as well as merging points that are close together (a tolerance of 10^{-8} is chosen for this step). The mesh is then triangulated to facilitate further refinement processes.

Adaptive subdivision is performed to refine the mesh, followed by volume-preserving Taubin smoothing, and then a small amount of Laplacian smoothing is applied to remove any remaining outliers. Finally, the mesh is cleaned again and re-meshed using the ACVD (Adaptive Clustering and Voronoi Diagram) method [89] to ensure a uniform distribution of mesh points and equilateral triangle faces.

The steps in this basic mesh generation method are relatively fast to perform and the process is automated to generate a surface mesh for each patient. This basic mesh is used as an input to WarpPINN since the smaller number of points in the surface mesh (in comparison to the advanced mesh discussed next) facilitate a significantly faster training time, since there are less points to sample in each iteration in WarpPINN’s mini-batching strategy.

It is noted that, during training and inference, the method of surface mesh generation (basic vs. advanced) was found to have little to no impact on overall performance. However, in the computation of directional strains after training, the use of a smoother and more biophysically plausible surface mesh coordinates is desired.

3.2.2.2 Advanced mesh generation

In contrast to the basic mesh generation technique, a more advanced approach is also used. This method begins with an existing LV mesh to act as a template. This initial template is chosen based on a previous study by Joyce et al.[90] and based upon capturing the morphological variation observed in 40 3D cardiac images.

Using the series of short-axis slices, and their binary segmentation masks of the myocardium, equivalent slices of the template mesh is taken. The mesh-fitting task then becomes an image registration task between the template mesh (short-axis slices of the mesh) to the reference input/ground truth slices. To perform this image registration task, the mesh slices are parameterised by three parameters: the mode weights of the shape model (α), the global position/orientation of the mesh (γ), which is described with 6 elements corresponding to 6 DOF in the registration, and finally, a per-slice offset (β).

These parameters are optimised in a gradient descent process that minimises the mean dice score (3.12) between the template mesh and reference masks, at each slice. A further regularisation term is also used which penalises large alpha and large beta values. The penalisation of large α values is justified from an observation that mode weights follow a standard Gaussian distribution.

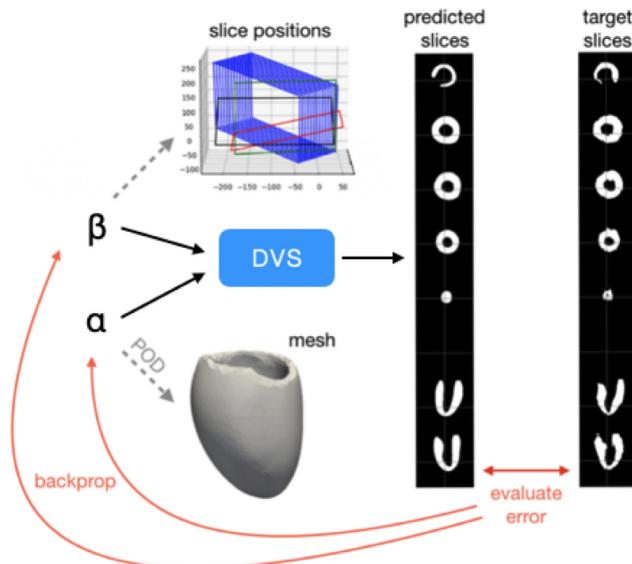


Figure 3.3: An illustration of the iterative mesh fitting process from Joyce et al. [90]

The gradient descent process is facilitated by Joyce et al.'s method of approximate differentiable mesh voxelisation, in which the mesh is voxelised to a volume and thin-plate-spline based warp is performed during the fitting step. The voxelisation step is done outside of the gradient descent loop allows for the deformation to be performed differentiably.

This mesh-fitting process allows for the generation of a personalised surface mesh for each patient, starting from the same template mesh. By using a pre-defined shape model, the final predicted/generated mesh is constrained to only biophysically plausible variations of the template and the complexity of the optimisation process is found to be significantly reduced, as well as increasing the robustness to noise from the input short-axis slices. However, the disadvantages of using a shape model are also noted, notably the restrictions in capturing specific patient (e.g., patients within the same disease

group) geometric abnormalities.

This advanced mesh generation method provides a superior surface mesh model in comparison to the basic method. It is noted that this advanced mesh generation process is time-consuming, specifically the differentiable voxelisation process. This constraint is mitigated to an extent by performing the deformation during the mesh-fitting process only every 50 gradient descent steps. Nevertheless, the duration of mesh generation is not the reason for not using the advanced meshes for the training of WarpPINN, but rather the significantly higher number of points (which is required to perform the approximate differentiable voxelisation and mesh-fitting), resulting in significantly longer training times.

The advanced meshes are instead used to evaluate and project the estimated myocardial strain tensor to find directional strains after WarpPINN has finished training as the advanced meshes are more biophysically plausible and smoother, facilitating a lower standard deviation in computed directional strains.

3.2.3 Training regime

WarpPINN is trained with the same regime as that discussed in the use of the CMAC dataset (Section 3.1.5). The μ and λ hyperparameters of the composite loss function (Eq. 3.5) are more thoroughly tuned and studied in each individual patient and pathology group.

3.2.4 Alternative methodologies for comparison

With ACDC being a popular dataset for cardiac image registration, there are many alternative methodologies from related previous work with easily reproducible corresponding results for WarpPINN. The results of WarpPINN are compared to several benchmark methodologies. As in the benchmark models for the evaluation of WarpPINN on CMAC data (3.1.4), these benchmark methods consist of a mix of both conventional and DL-based image registration approaches.

Classical benchmark methods include:

- **Free-Form Deformation with Volumetric Preservation (FFD-VP):** Similarly to TDFFD, this approach parameterises the deformation field using B-Splines and optimises the displacement field using an image similarity measure in the cost function whilst enforcing volumetric preservation through a regulariser term. Unlike TDFFD, FFD-VP models deformations in static images without an explicit temporal component (TDFFD models a $3D + t$ velocity field).
- **Diffeomorphic Demons (dDemons):** Similarly to INRIA / iLogDemons, this approach is another modification of the Demons algorithm, parameterising the transformation by a stationary velocity field, which is integrated to compute the displacement field. A key difference is the focus on ensuring diffeomorphism rather than elasticity and incompressibility constraints, as INRIA does.

DL-based benchmark methods include:

- **Motion-Net:** Motion-Net is a CNN-based approach for simultaneously learning motion estimation and image segmentation. Biomechanical plausibility is enforced through the use of an explicit regulariser in the form of an approximation of Huber loss to penalise displacement gradients[91].
- **Biomechanics-Informed Neural Network (BINN):** This method leverages regularisation for an image registration network in the form of a separate Variational Autoencoder (VAE), as opposed to an explicit regularisation term in the loss function. The VAE is trained to learn the probability distribution of simulated biomechanically plausible deformations. Simulations of biomechanically plausible deformations are generated by solving PDEs for a neo-Hookean constitutive model [92].
- **Biomechanics-Informed Generative Model (BIGM-VAE):** This method extends BINN’s methodology of simulating biomechanically plausible deformations. Image registration is done through a pre-trained temporal VAE as a generative transformation model, leveraging the simulations as a data-driven prior. A manifold of plausible deformations is generated with the VAE’s latent space, which is traversed to optimise image registration [93].

Finally, the estimated cardiac strains are also compared to those predicted by DeepStrain / CarMEN, as discussed further below.

3.2.5 Evaluation method

As discussed previously in the evaluation of WarpPINN on the CMAC data, image registration performance can be evaluated with the criteria of image similarity and biomechanical plausibility (volume preservation of the myocardium). These can be inspected with qualitative results and measured with quantitative metrics, as discussed below, as well as compared to results produced from alternative methodologies as previously discussed.

3.2.5.1 Visual inspection of predicted images and deformation field

As with the evaluation in (Section 3.1.6), the predicted reference (ED) image and the predicted deformation field is visually inspected and compared to the ground truth ED and ES images to gain an initial informative assessment of both image similarity between the predicted and ground truth images at the ED phase (the predicted image is constructed by warping the ground truth ES image using the predicted deformation field output through trilinear interpolation), as well as biomechanical plausibility. In both qualitative and quantitative evaluations, 3D images are compared using the representative short-axis slices (Basal, Mid-Ventricular and Apical).

3.2.5.2 Evaluation metrics

To quantitatively assess image registration performance, several commonly used metrics in medical image registration are measured. As noted above, an advantage of the ACDC dataset over CMAC data is the existence of ground truth

segmentation masks of the myocardium at the ED and ES frames. This allows the use of more appropriate image similarity metrics that measure similarities/dissimilarities between binary images (masks), which are more robust to artefacts than measures such as the MSE/SSIM.

The following image similarity metrics are measured in the evaluation of warped and reference images:

- **Dice-Similarity Coefficient (DSC):** The Dice score measures the overlap between two binary masks (e.g., the predicted and ground truth segmentation masks).

$$\text{DSC}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (3.12)$$

where:

- A and B are the sets of pixels in the predicted and ground truth masks.
- $|A \cap B|$ is the number of overlapping pixels between the two masks.
- $|A|$ and $|B|$ are the number of pixels in each mask.
- **Mean Contour Distance (MCD):** The Mean Contour Distance measures the average distance between the surfaces of the predicted and ground truth masks.

$$\text{MCD}(A, B) = \frac{1}{|A| + |B|} \left(\sum_{a \in A} \inf_{b \in B} d(a, b) + \sum_{b \in B} \inf_{a \in A} d(b, a) \right) \quad (3.13)$$

where:

- $d(a, b)$ is the Euclidean distance between points a and b .
- **Hausdorff Distance (HD):** The Hausdorff distance measures the distance between two subsets of a metric space. In the context of image registration, HD is used to measure the maximum distance between the surfaces of the predicted and ground truth masks. The **95% Hausdorff Distance (HD95)** is also used, which is a robust version of HD that considers the 95th percentile of distances between the surfaces, reducing sensitivity to outliers.

It is also noted that the calculation of Hausdorff distance has been documented to be computationally expensive. In order to facilitate the use of the metric, calculation of HD is implemented using Python's `monai` library.

$$\text{HD}(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(b, a) \right\} \quad (3.14)$$

where:

- $d(a, b)$ is the Euclidean distance between points a and b .
- sup and inf denote the supremum and infimum, respectively.

$$\text{HD95}(A, B) = \text{percentile}_{95}(\{d(a, B) \mid a \in A\} \cup \{d(b, A) \mid b \in B\}) \quad (3.15)$$

where:

- $d(a, B)$ is the distance from point a in set A to the closest point in set B .

The MAD of the Jacobian determinant and 1, denoted as $\|J| - 1\|$, is also measured to quantitatively assesses the level of volume preservation over the LV during the predicted deformation.

3.2.5.3 Estimated strains

The estimated mean cardiac strains for each disease group is also computed across the cardiac cycle and compared to alternative methodologies, namely DeepStrain / CarMEN, as similarly done in the CMAC validation task (3.1.6.3).

3.2.5.4 Aggregation of results

A notable advantage of utilising the ACDC dataset is the aggregation of results by disease group, which many previous related works have done for image registration, classification, strain estimation and classification (e.g., cardiomyopathy diagnosis) tasks. This serves as another dimension of interpreting results, as well as another point of validation when comparing with other alternative methodologies. Thus, in the Results section, experimental results of image similarity, volume preservation and strain estimation are also aggregated by disease.

3.2.6 Hyperparameter tuning

An extensive hyperparameter search is performed for the WarpPINN model for each patient in the training data. This search tunes the μ and λ_{inc} parameters in the loss function (3.5). The results of this extensive search is discussed further in the Results section, employing a Bayesian optimisation-like approach, as opposed to OFAT or grid search.

4

Experimental Results and Discussions

Contents

4.1	Validating WarpPINN on CMAC	42
4.1.1	Analysis of Predicted Images and Deformation Field	42
4.1.2	Landmark tracking	45
4.1.3	Strain analysis	46
4.1.4	Training loss curves	48
4.1.5	Discussion of Validation results	49
4.1.6	Results of ablation study	50
4.1.7	Hyperparameter tuning	54
4.2	ACDC	59
4.2.1	Selective presentation of individual patient results	59
4.2.2	Results aggregated by disease	62
4.2.3	Results across all patients	66
4.2.4	Discussion of WarpPINN-ACDC experiments	66

This chapter outlines the experiments, results and subsequent discussions of methods employed as discussed in the previous section.

4.1 Validating WarpPINN on CMAC

This section documents and presents results and subsequent discussions for the Validation experiments of WarpPINN on Volunteer 1 of the CMAC dataset, as outlined in Section 5.1 .

4.1.1 Analysis of Predicted Images and Deformation Field

The warped Template image predicted by WarpPINN (with $\mu = 1 \times 10^{-5}$) on Volunteer 1 of the CMAC dataset is presented in Figure 4.1, along with the ground truth Reference and Template images, which are the ground truth images at the end-diastolic and end-systolic phases, respectively. 3D images are presented with the three representative slices.

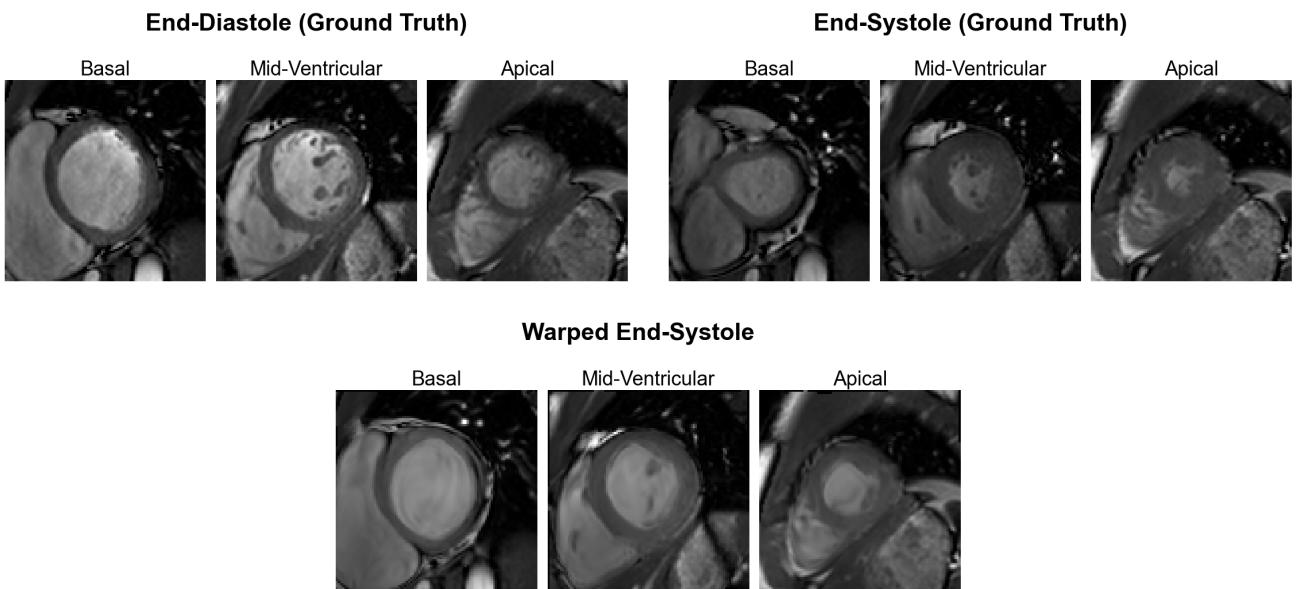


Figure 4.1: Ground truth images for the LV of Volunteer 1 of the CMAC dataset at the end-diastolic (Reference) and end-systolic (Template) phases, as well as the warped Template image as predicted by the WarpPINN ($\mu = 1 \times 10^{-5}$) method.

At an initial visual inspection, it is observed that the Template image has been warped to be similar to the Reference image. Table 4.1 presents quantitative results for image similarity measures between these images for all tested models. The $||J| - 1|$ metric is also reported for the entire LV surface at the ES phase. In this table, *Initial* refers to the similarity between the ground truth ED and ES images, this provides a baseline to compare the image similarity measures after the deformation is applied. WarpPINN refers to the vanilla WarpPINN model without FFM and WarpPINN-FF refers to the model with FFM.

At inference, the ground truth surface mesh of the LV segmentation at ED is deformed using trilinear interpolation by the trained WarpPINN to obtain a prediction of the mesh at ED. The 3D deformation field for this test is visualised in Figure 4.2 for the WarpPINN model with $\mu = 1 \times 10^{-5}$. The predicted ES meshes for all tested models are visualised in Figure 4.3, which also shows the calculated determinant of the Jacobian at each cell of the surface mesh.

Method	Basal			Mid-Ventricular			Apical			$ J -1 $
	NRMSE	SSIM	NCC	NRMSE	SSIM	NCC	NRMSE	SSIM	NCC	
<i>Initial</i>	0.541	0.278	0.619	0.459	0.373	0.706	0.267	0.649	0.882	
WarpPINN										
$\mu = 1 \times 10^{-5}$	0.274	0.591	0.902	0.270	0.615	0.898	0.191	0.779	0.939	0.018(0.014)
$\mu = 5 \times 10^{-6}$	0.273	0.592	0.902	0.264	0.612	0.904	0.204	0.760	0.931	0.029(0.029)
WarpPINN-FF										
$\mu = 1 \times 10^{-5}, \sigma = 1$	0.255	0.626	0.915	0.253	0.656	0.911	0.193	0.779	0.938	0.017(0.015)
$\mu = 5 \times 10^{-6}, \sigma = 1$	0.202	0.701	0.948	0.213	0.710	0.938	0.219	0.791	0.919	0.027(0.025)

Table 4.1: Image registration results of different tested WarpPINN models on Volunteer 1 of the CMAC dataset. $||J|-1|$ is reported as mean(std deviation).

Here, the wireframe represents the original ground truth surface mesh at ED. Since there is no ground truth for the mesh configuration at ES, the equivalent results of testing on INRIA is also presented as for comparison.

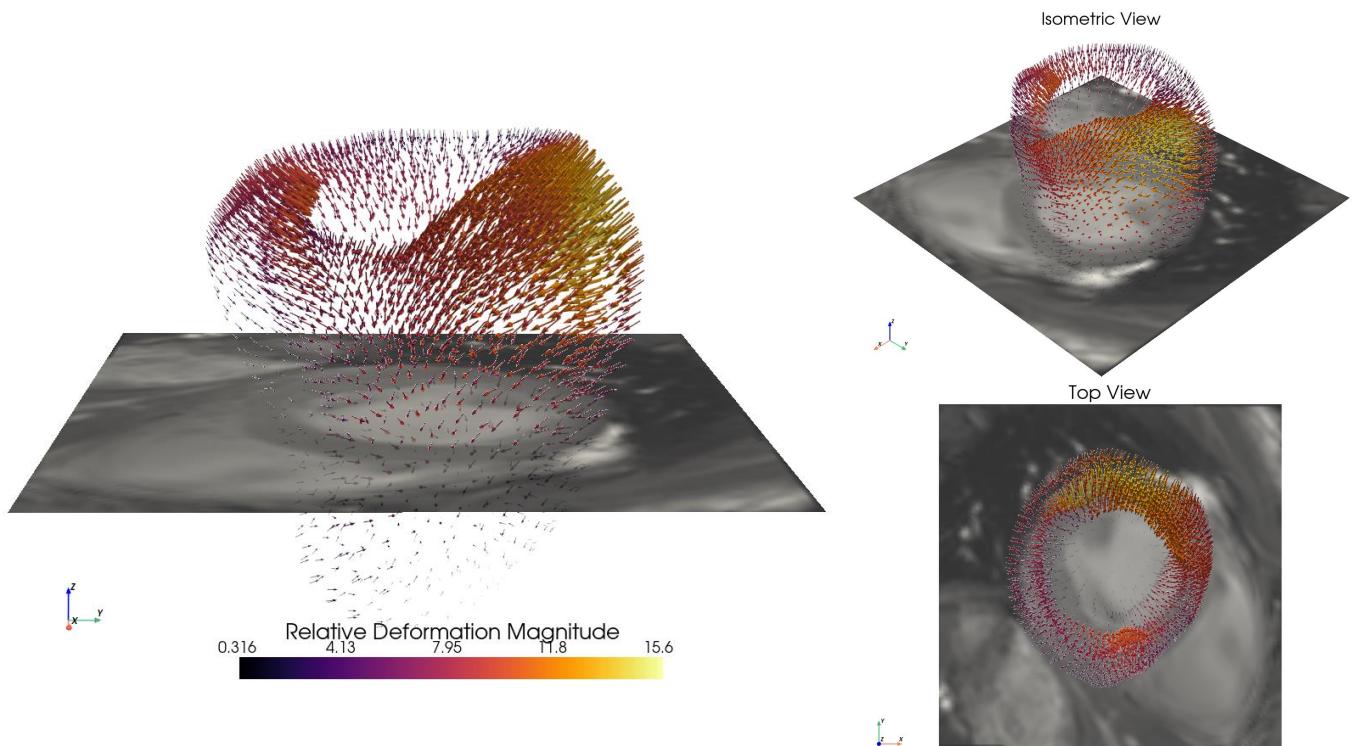


Figure 4.2: Visualisation of the predicted deformation field by the WarpPINN method when deforming the input surface mesh (ED) at inference to estimate the point-by-point coordinates of the predicted ES surface mesh, represented by the shown glyphs.

From the quantitative results in Table 4.1, it is observed that all models result in an increase in SSIM and NCC for all representative slices in comparison to the *Initial* values, indicating that the Template image has been deformed to successfully be more similar to the Reference image. It is, however, noted that this improvement is not observed when considering NRMSE.

When comparing the differences in image similarity for the WarpPINN and WarpPINN-FF models when μ is changed, there are small differences in results that are too statistically small to draw conclusions from. It is noted, however, that a decrease in μ results in an increase in $||J| - 1|$ for both WarpPINN and WarpPINN-FF. This is expected as the μ

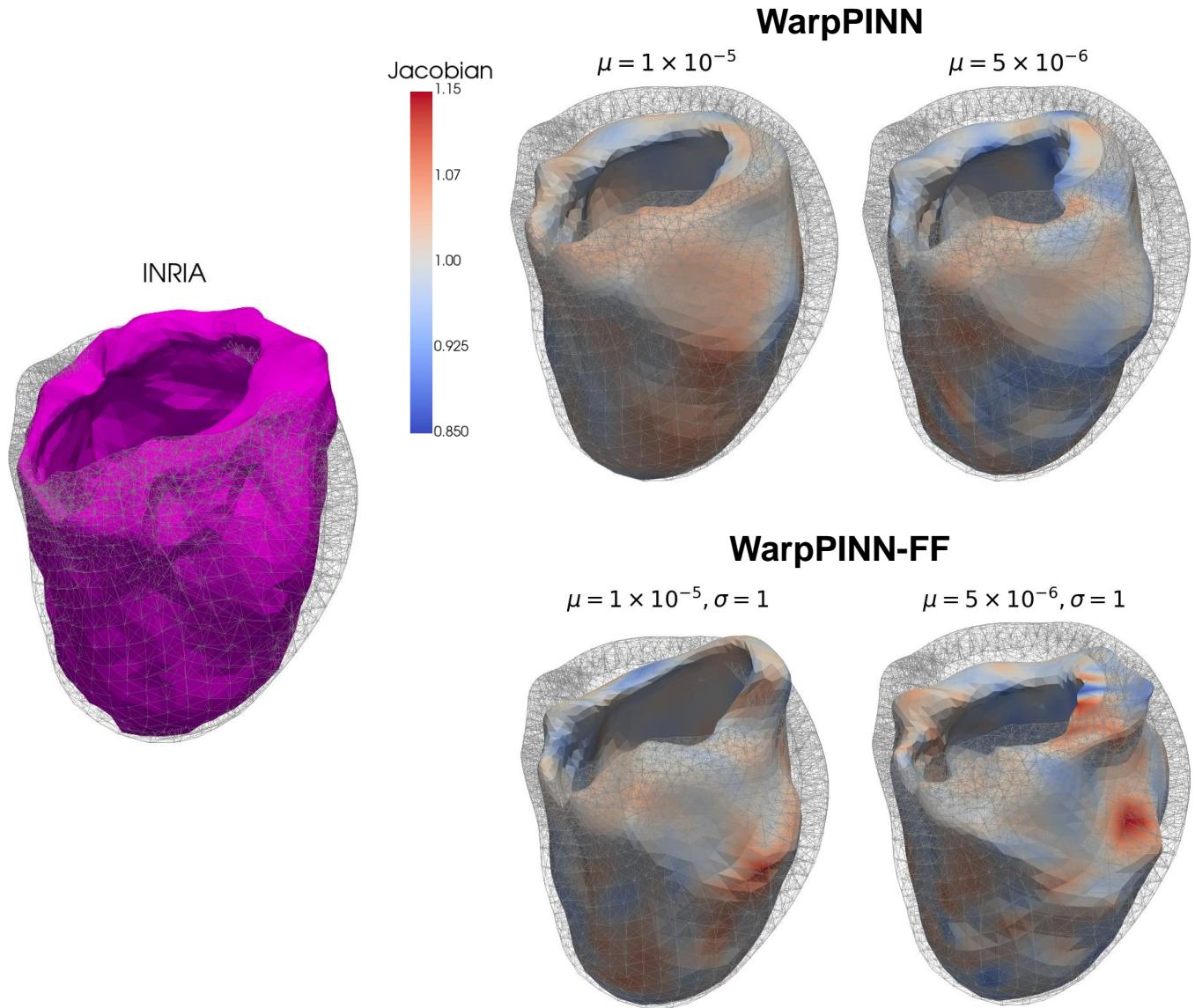


Figure 4.3: Visualisation of the predicted surface mesh at the ES phase, with the wireframe representing the ground truth surface mesh of Volunteer 1 provided in the CMAC dataset. The Jacobian determinant value at every point of the mesh is also visualised.

parameter determines the weight of the NeoHookean loss in the composite loss function. A lower μ value means less physics-informed constraints of incompressibility during training, leading to less volume preservation. This relationship is visually confirmed with the results in Figure 4.3.

A lower mean $||J| - 1|$ is also achieved with WarpPINN-FF, in comparison to WarpPINN, for both Mu parameter values. This is expected as the use of FFM's allow for less spectral bias and larger deformations. For the same value of μ (i.e., same level of incompressibility constraint), the higher dimensional input space allow WarpPINN-FF to have more capability to predict the plausible deformations. This ability to predict larger and more local deformations can be observed in Figure 4.3, where WarpPINN-FF predicts less smooth or more discontinuous deformations, but on average has a Jacobian closer to 1 than WarpPINN.

The uncertainty observed across the three image similarity metrics used (NRMSE, SSIM and NCC) suggest that these measures may not be as suitable for this task. These metrics, unlike those used in later experiments with the ACDC dataset (DSC, MCD, HD) are vulnerable and less robust to artefacts and varying intensities found in medical images. This is further emphasised with Table 4.1 showing *Initial* values for NRMSE being higher than any of the tested models' predicted deformations, implying that the Reference image is more similar to the Template image than the Warped Template image. From Figure 4.1, one can observe that this is indeed not the case.

The justification for the use of NRMSE, SSIM and NCC as evaluation metrics for image similarity in this Validation task, as opposed to DSC, MCD, etc., is due to the absence of a ground truth segmentation mask in the CMAC dataset. Hence, the use of anatomical landmark tracking could be more valuable to assess the WarpPINN models' image registration performance.

4.1.2 Landmark tracking

Ground truth positions for twelve anatomical landmarks are included in the ED and ES timeframes for Volunteer 1 of the CMAC data. Similar to deforming the ground truth ED surface mesh at inference to predict the ES mesh, the learnt deformation field is applied to these landmarks at the ED timeframe to predict their positions at the ES timeframe, and then compared to the ground truth positions. This landmark tracking task is visualised and presented in Figure 4.4, where the red and green points indicate the positions of the ground truth and predicted landmarks, respectively, at the ES timeframe. The predicted ES mesh (grey) is included for visualisation purposes of the landmark coordinates. For each landmark, the black line measures the Euclidean distance between the ground truth location and the corresponding predicted position. Landmark tracking error is thus defined as the Euclidean distance for each landmark.

Figure 4.4 shows the landmark tracking errors for each tested WarpPINN model, as well as a box plot of landmark tracking errors for each model and alternative methodology, namely UPF, INRIA and CarMEN, for comparison. It is observed that all tested WarpPINN models achieve lower landmark tracking errors than each benchmark method, indicating greater image registration performance.

It is expected that a decrease in μ results in an increase in image similarity performance. From the box plots, it can be

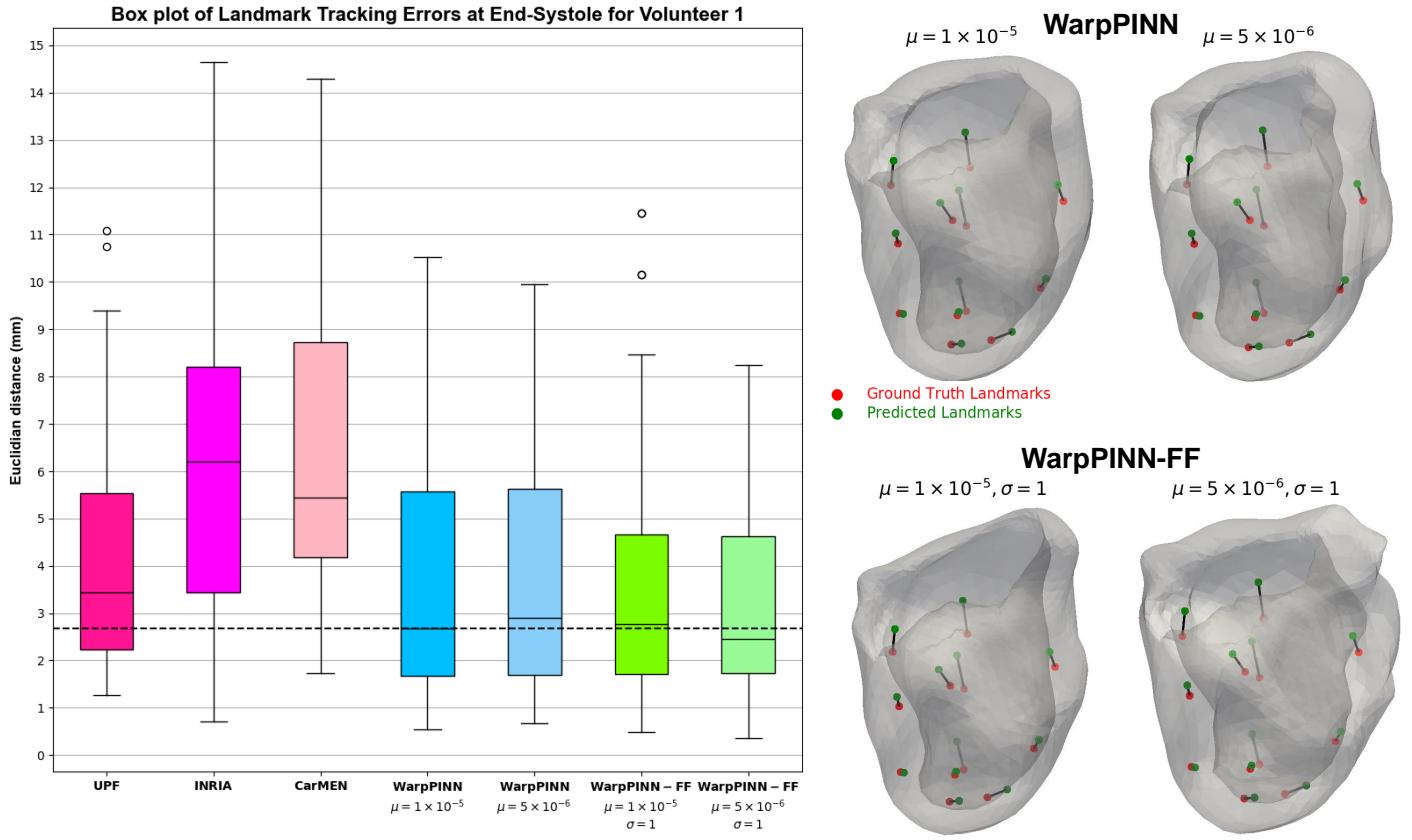


Figure 4.4: Visualisation of the the motion tracking of 12 anatomical landmarks provided in the CMAC dataset, with a box plot of landmark tracking errors for each tested WarpPINN model and benchmark methodology.

observed that a decrease in μ results in a lower landmark tracking error for WarpPINN-FF, but not the vanilla WarpPINN model. It should be noted, however, that a decrease in μ for the vanilla WarpPINN model resulted in a lower range of errors.

Furthermore, in the comparison of WarpPINN and WarpPINN-FF, WarpPINN-FF achieves a lower inter-quartile range of landmark tracking errors, but results for the median values are mixed and the differences are too small to interpret as they are all relatively close to each other. It is at this point that it should be noted that in the generation of landmark positions for the CMAC dataset, the registration to cine MRI from tagged MRI was reported to have an inter-observer variability of 0.84mm between two observers. Therefore, there is no total agreement for the landmark errors. Nevertheless, a lower median of all tested WarpPINN models, in comparison to all three alternative methodologies, as well as comparable ranges/IQRs, serve as another point of validation for the WarpPINN method.

Finally, from the 3D visualisation in Figure 4.4, it is noted that longitudinal deformation is difficult to predict for WarpPINN. This discussion is extended further in the analysis of computed directional strains.

4.1.3 Strain analysis

Figure 4.5 presents the estimated directional strains by WarpPINN ($\mu = 1 \times 10^{-5}$) and WarpPINN-FF ($\mu = 1 \times 10^{-5}, \sigma = 1$) at every time frame for Volunteer 1. These strain curves are further segmented for the Septum and the Free Wall.

Equivalent strain curves predicted by INRIA and UPF are also included for comparison.

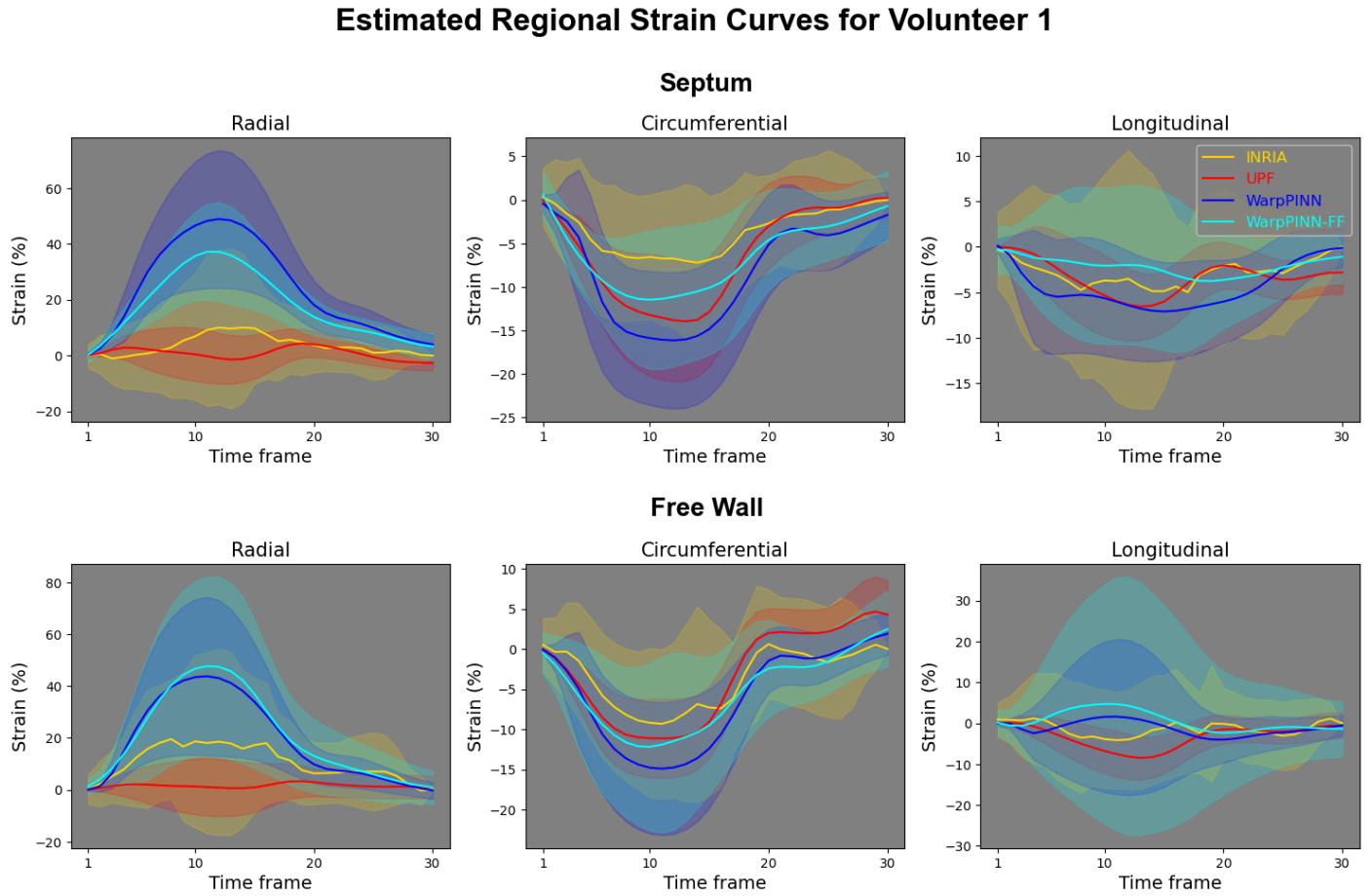


Figure 4.5: Estimated regional and directional strain curves by the WarpPINN (vanilla and WarpPINN-FF) and benchmark methodologies considered.

It is observed that WarpPINN (both vanilla and FF) predicts smoother strain curves in comparison to INRIA and UPF. This can be attributed to the spectral bias in PINNs [77]. Whereas this spectral bias is a factor that was aimed to be mitigated through the use of FFM to increase the dimensions of the input images, no such encoding is used in the time variable (the input of the WarpPINN architecture is $3D + t$). As a consequence, low-frequency behaviour are dominant, resulting in smoother strain curves (akin to WarpPINN without FFM being less able to predict local deformation in the 3D-space variables).

Furthermore, the estimated global (at all segments) strains at ES are presented in Figure 4.6 as a violin plot. It is observed from both the strain curves and violin plots that the estimated strain values of WarpPINN are in agreement with that of the chosen benchmark methodologies, with the exception of radial strain. WarpPINN and WarpPINN-FF predict higher values for the radial strain, in comparison to UPF and INRIA. Since there is no ground truth for cardiac strain measurements, the comparison of estimated strains to UPF and INRIA only serve as a point of validation. In reality, the high peaks of the radial strain curve and the higher median radial strain in the violin plots for both WarpPINN models are actually closer to the peak value quoted in clinical studies, around 45%, suggesting more biomechanically plausible deformations.

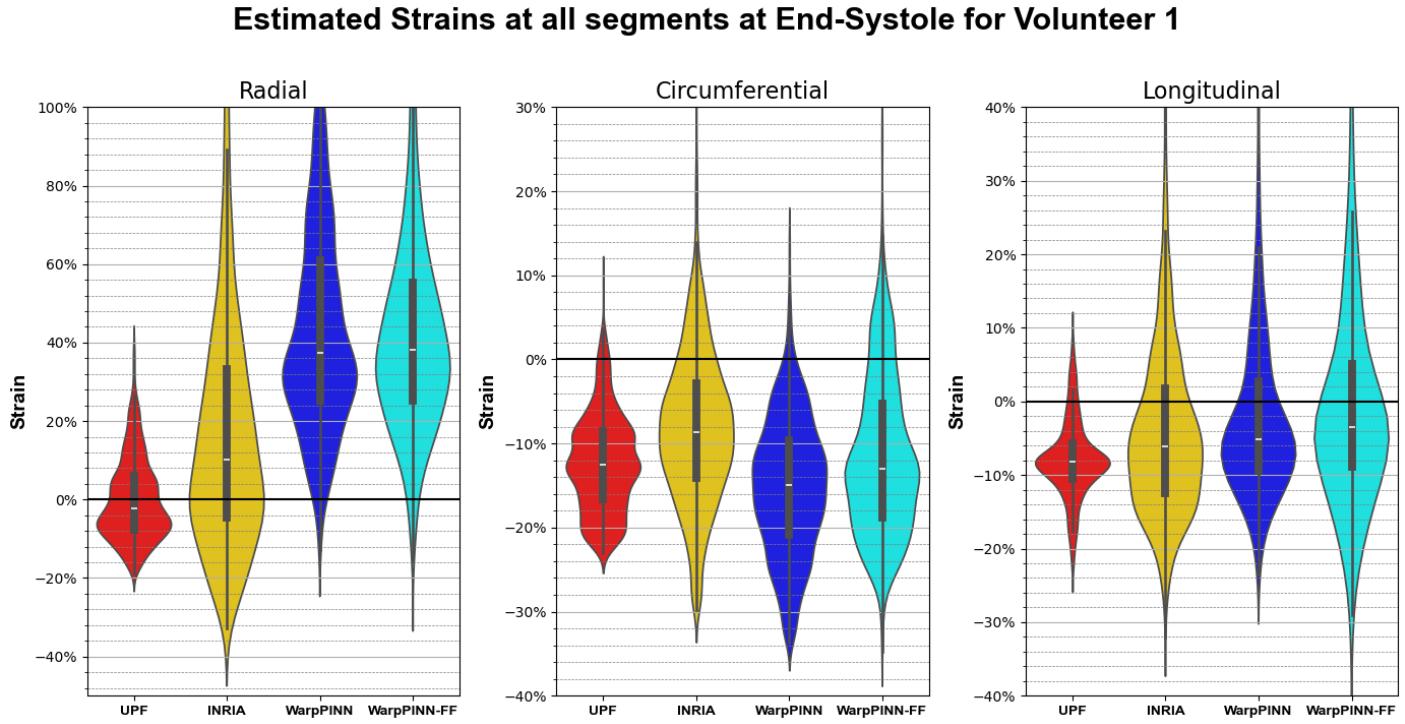


Figure 4.6: Violin Plot of global directional strains as predicted by the WarpPINN (vanilla and WarpPINN-FF) and benchmark methodologies considered.

When comparing the predicted strains of the vanilla and FF WarpPINN models, it is expected that WarpPINN-FF is better able to predict larger strain values. While this is not observed in the strain curves (which plot the mean strain value), this is observed in the median strain values shown in the violin plots.

Finally, the low longitudinal strain values predicted by all methodologies reinforces the previously discussed idea of the challenge of predicting large longitudinal strains. One possible explanation could be the nature of the MRI data, which is a stack of short-axis slices, meaning that the resolution along the z-axis (where longitudinal strain, which is in the direction of the LV base to apex, would mostly be acting) is significantly lower in comparison to the x- and y-axes.

4.1.4 Training loss curves

The training loss curve of WarpPINN (vanilla, $\mu = 1 \times 10^{-5}$) is presented in Figure 4.7, where it is observed that the WarpPINN method is minimising both components (L1 and NeoHookean) of the loss function as expected. However, it is noted that the loss curves present with significant variability and oscillations, a behaviour that is not typically observed in loss curves of neural networks, in general. This is explained by the batching strategy used during the training of WarpPINN, where Target images are fed one at a time for each training iteration. This results in the observed variability since some Target images (for example, frames closer to ED) have more similarity than others (such as frames closer to ES). The Reference image remains the same (ED) throughout training.

It is also observed that the NeoHookean loss first steadily increases as it is first calculated, starting from 0, with the mini-batching strategy used, before beginning to decrease at around 100k iterations and showing convergence behaviour.

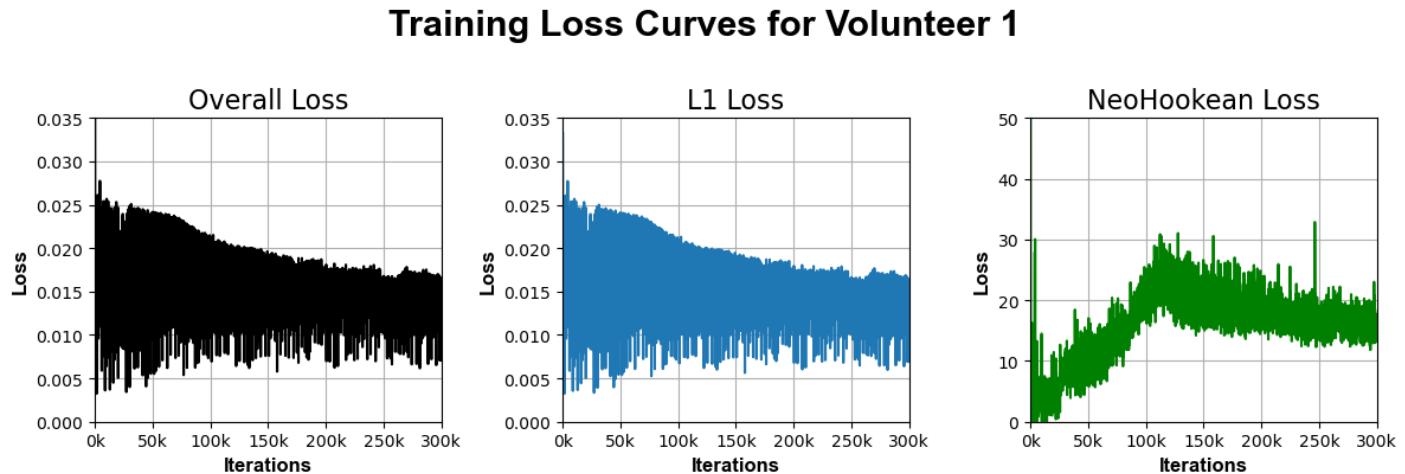


Figure 4.7: Training loss curves for the considered WarpPINN model.

4.1.5 Discussion of Validation results

In this Validation task thus far, the results for the image registration performance of WarpPINN have been analysed both qualitatively and quantitatively, with the criteria of image similarity and biomechanical plausibility (volume preservation). Qualitative results visualise the learnt deformation field applied to the Template (ES) image to compose a Warped image similar to the Reference (ED). Visualisation of the predicted ES surface mesh by deforming the ground truth ED surface mesh was also presented, where qualitative inspections of the deformation were made to assess the realism of such deformations. Quantitative results also showed that the Jacobian determinant is calculated to be close to 1, on average, implying that the quasi-incompressibility of the myocardium is enforced to an extent. Jacobian values were also confirmed to be positive in the entire domain, implying the learnt deformation mapping to be a diffeomorphism.

The assessment of image similarity in this Validation task brought to attention the ineffectiveness of the employed evaluation metrics for quantitative analysis of results. In particular, the use of NRMSE may be deemed unsuitable for comparing MR images with varying intensities, due to the low robustness to such artefacts. The need for evaluation metrics that measure similarities between binary masks is emphasised.

The image registration performance of WarpPINN was also compared to three other discussed alternative methodologies on the same Volunteer, showing agreement in anatomical landmark tracking results and estimated cardiac strains. In particular, all tested WarpPINN models achieved lower landmark tracking errors and more realistic radial strain predictions than all three benchmark methods.

The effect of using Fourier Feature mappings with the WarpPINN method was also studied in this Validation task, confirming hypotheses with WarpPINN-FF predicting larger, more local deformations and higher median strain values, than its vanilla counterpart, while achieving a greater level of volume preservation.

Finally, this Validation task also introduced the effect of varying the μ hyperparameter, which controls the degree of regularisation (NeoHookean) during training. The results from this Validation task affirm that the μ hyperparameter, tested at 1×10^{-5} and 5×10^{-6} , control the balance between image similarity and volume preservation. The next subsections

will present and discuss the results of more comprehensive tuning of hyperparameters.

4.1.6 Results of ablation study

This subsection presents and documents the results of the ablation study carried out for nullifying the physics loss in the WarpPINN method. This nullification was achieved by using a μ value of 0. Results in this section draw focus on comparing three tested WarpPINN models, with a variation in the value of the μ hyperparameter, $\mu = [1 \times 10^{-5}, 5 \times 10^{-6}, 0]$

Figure 4.8 presents the Reference (ground truth ED) and Template (ground truth ES) images, alongside the Warped Template images of each tested model for each representative slice of Volunteer 1. Table 4.2 also reports measurements of image similarity between each of these images, as well as the $||J| - 1|$. The predicted 3D deformation field (from ED to ES) is also visualised in Figure 4.9.

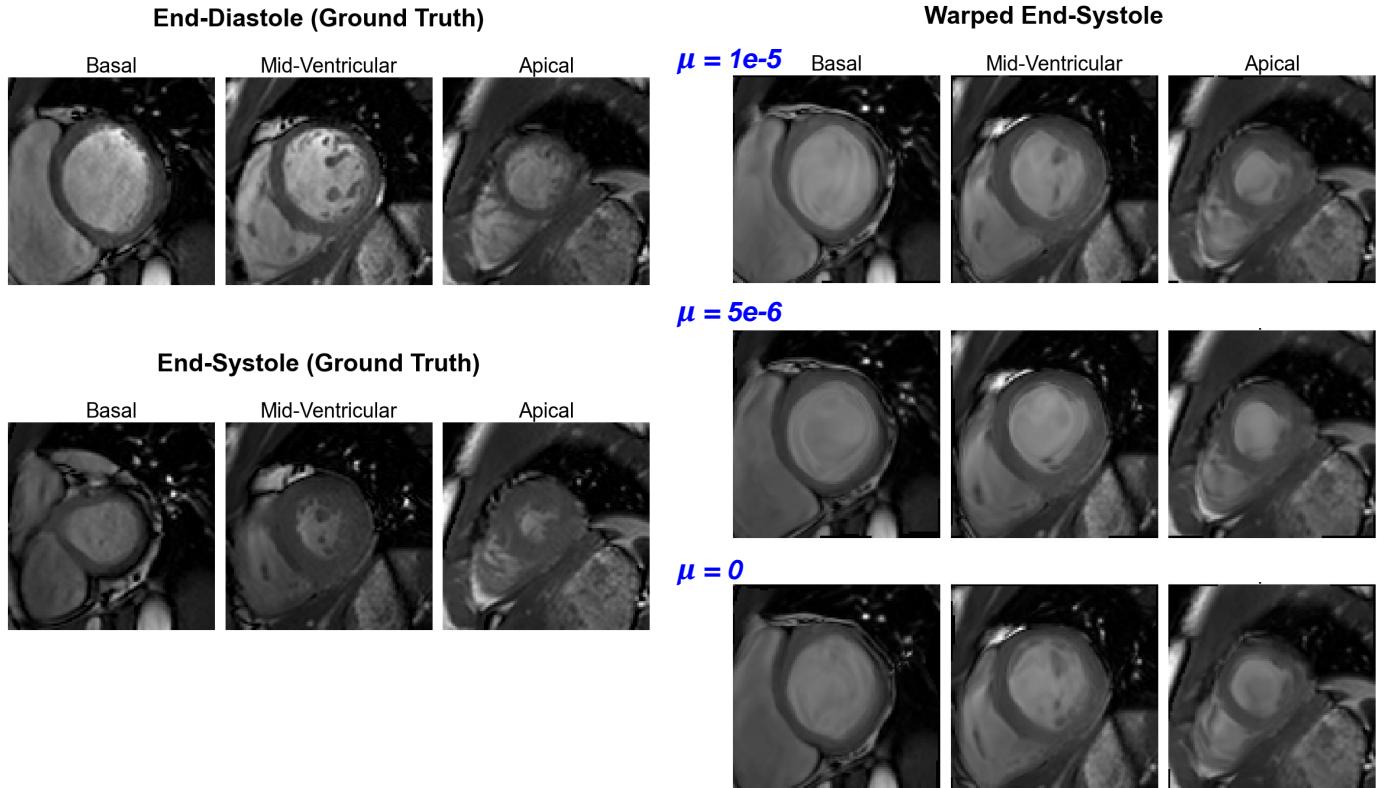


Figure 4.8: Reference, Template and Warped Template images at the representative slices for WarpPINN models of varying μ .

Method	Basal			Mid-Ventricular			Apical			$ J -1 $
	NRMSE	SSIM	NCC	NRMSE	SSIM	NCC	NRMSE	SSIM	NCC	
Initial	0.541	0.278	0.619	0.459	0.373	0.706	0.267	0.649	0.882	
WarpPINN										
$\mu = 1 \times 10^{-5}$	0.274	0.591	0.902	0.270	0.615	0.898	0.191	0.779	0.939	0.018(0.014)
$\mu = 5 \times 10^{-6}$	0.273	0.592	0.902	0.264	0.612	0.904	0.204	0.760	0.931	0.029(0.029)
$\mu = 0$	0.270	0.577	0.905	0.267	0.608	0.900	0.214	0.755	0.923	0.357(0.277)

Table 4.2: Quantitative results of image registration for WarpPINN with varying values of μ .

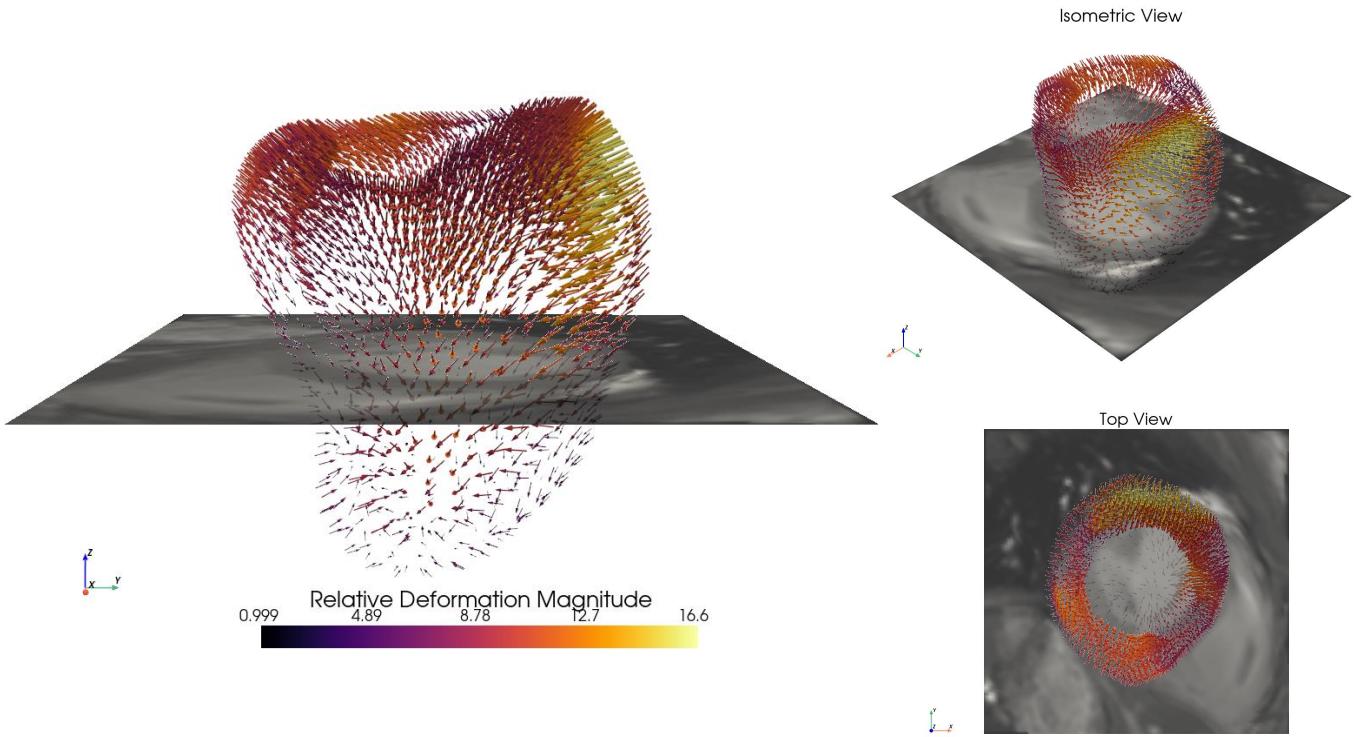


Figure 4.9: Visualisation of the predicted deformation field applied to the ED surface mesh at inference to predict the ES surface mesh for WarpPINN ($\mu = 0$)

From visually inspecting the predicted deformation field and the predicted images, it is difficult to observe neither the differences in image similarity nor plausibility when varying μ . Table ?? also shows comparable image similarity measures across all three models, though, as previously discussed in this section, these evaluation metrics have been found to be ineffective and unsuitable for this task. It is observed, however, that there is a clear trend of $|J| - 1$ increasing as μ is decreased. This reinforces the hypotheses and results found in the Validation task, where the objective of μ in the loss function is to control the degree of (NeoHookean) regularisation, thus a lower value results in a heavier minimisation of the L1 loss relative to the NeoHookean, consequently resulting in unrealistic deformations. This is also visualised in the computed Jacobian at each cell of the predicted ES mesh in Figure 4.10.

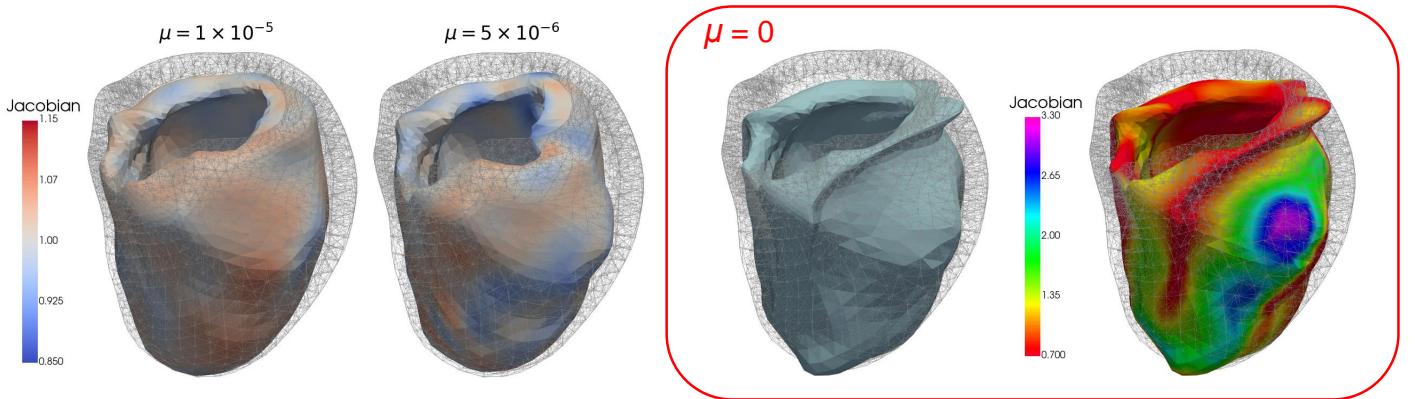


Figure 4.10: Visualisation of the computed Jacobian across the surface mesh at the ES phase for WarpPINN with varying values of μ .

The landmark tracking performances of the 3 WarpPINN models is also presented in Figure 4.11. It is difficult to

confidently infer from the medians of the box plot whether the decrease in μ correlates to better image registration performance as all the medians are relatively close together. As discussed in the Validation task, the inter-observer variability of 0.84mm during the generation of the landmark locations mean that there is no total agreement between landmarks. However, a trend is observed where a decrease in μ correlated to a decrease in the inter-quartile range of landmark tracking errors.

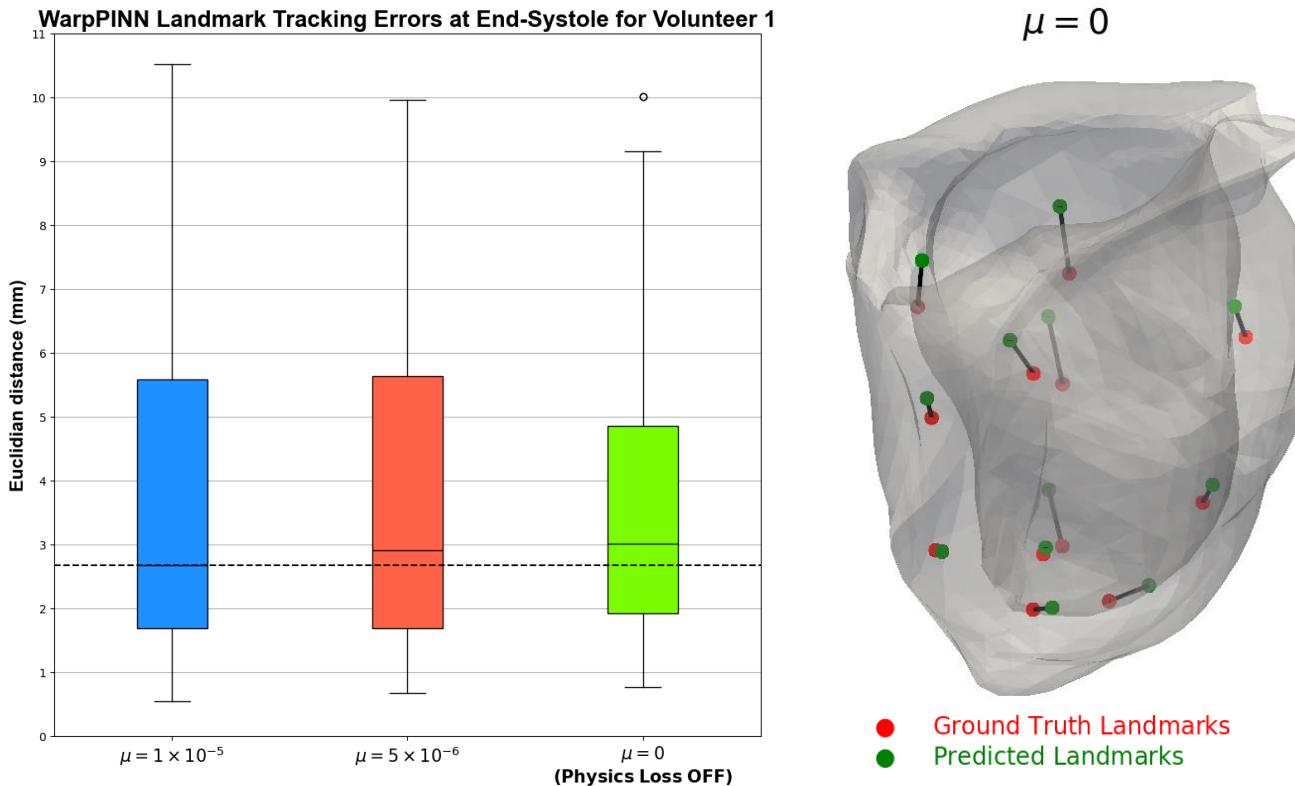


Figure 4.11: Visualisation of anatomical landmark tracking for WarpPINN with $\mu = 0$.

Figures 4.12 and 4.13 report the regional strain curves and violin plots of global strains predicted at the ES phase, respectively. It is also difficult to infer if there is a consistent trend between the μ value and both the peak and median magnitude of strain values. However, both figures display a trend of increasing variability (the standard deviation in the strain curves and the range in the violin plots) as μ is decreased.

Finally, the loss curve of WarpPINN trained with $\mu = 0$ is shown in Figure 4.14, where it is observed that the loss (which is only the L1 loss in this case) is minimised significantly faster in comparison to the previously presented model where $\mu = 1 \times 10^{-5}$ (Figure 4.7), resulting in an earlier convergence to a similar value of 0.015.

The results of this ablation study of nullifying the physics-informed loss have shown that the size of μ (i.e., how much the NeoHookean loss impacts training) has a significant effect in predicting volume-preserving/biomechanically plausible deformations, with higher μ values corresponding to the Jacobian values closer to 0.

Conversely, the formulation of the composite loss function means that a decrease in μ , i.e. a larger weight of L1 minimisation in the loss function, is expected to lead to an increase in image similarity in registration. Although some trends in decreasing variability in landmark tracking error are observed, the lack of useful image similarity metrics presents

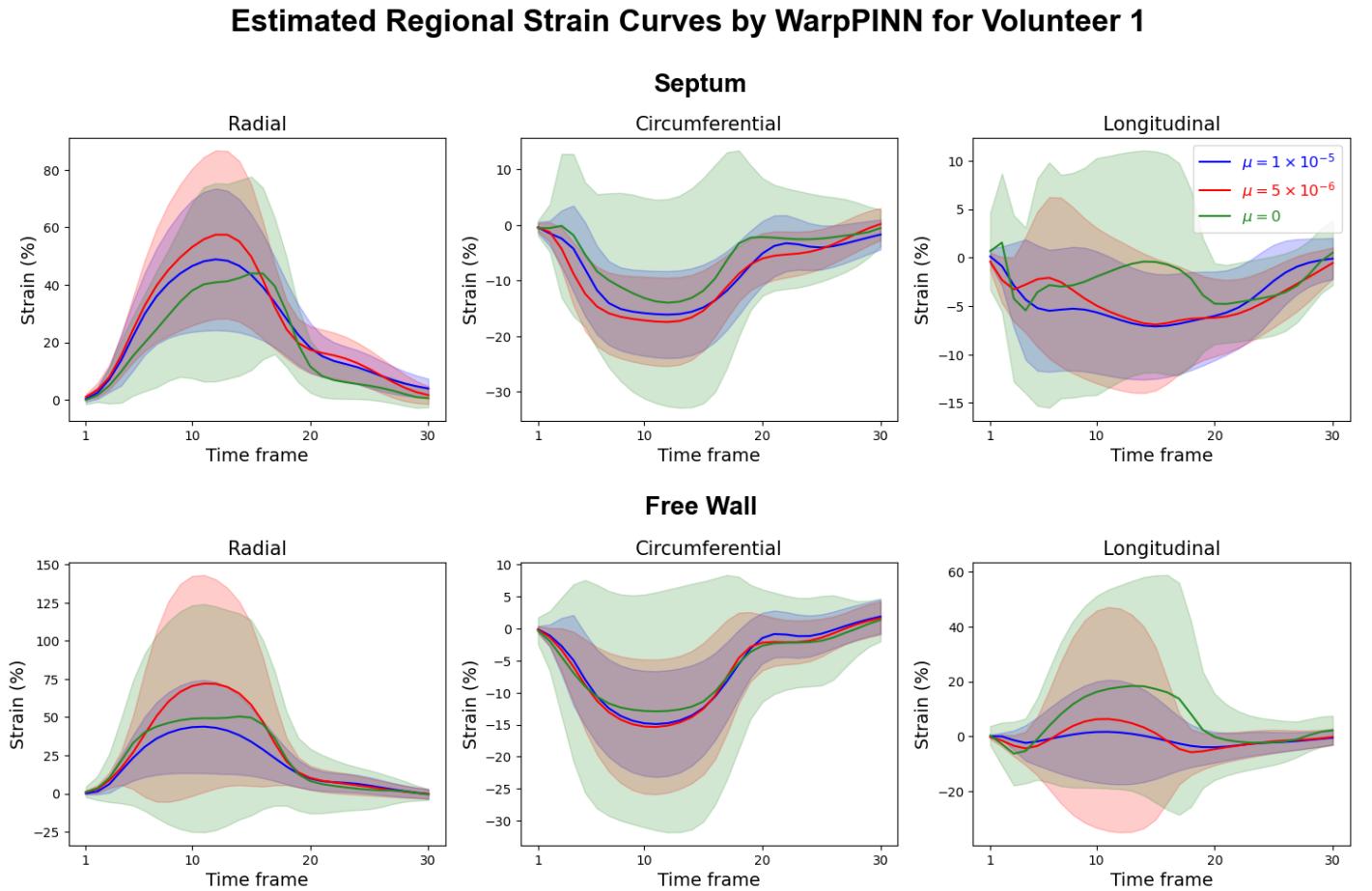


Figure 4.12: Estimated regional strain curves for WarpPINN with $\mu = 0$.

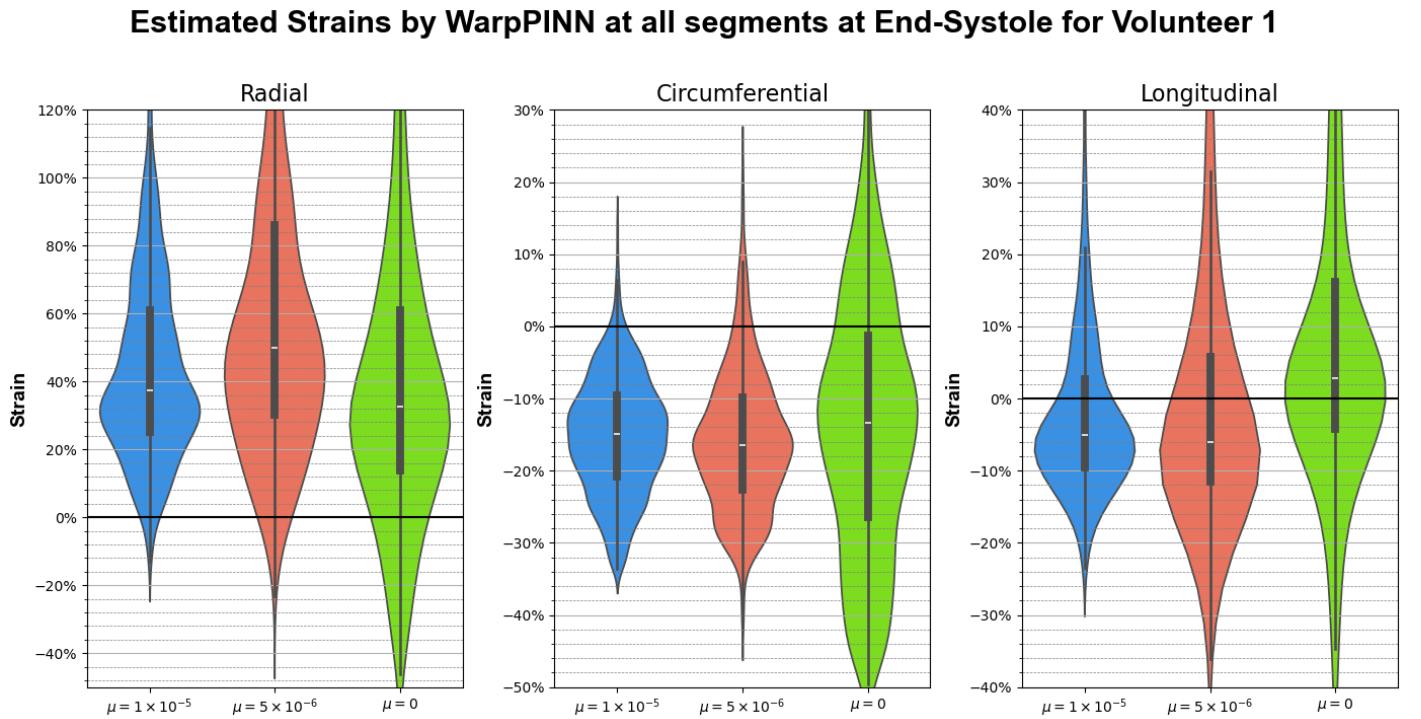


Figure 4.13: Violin plots for predicted global strains at the ES phase for WarpPINN with varying values of μ .

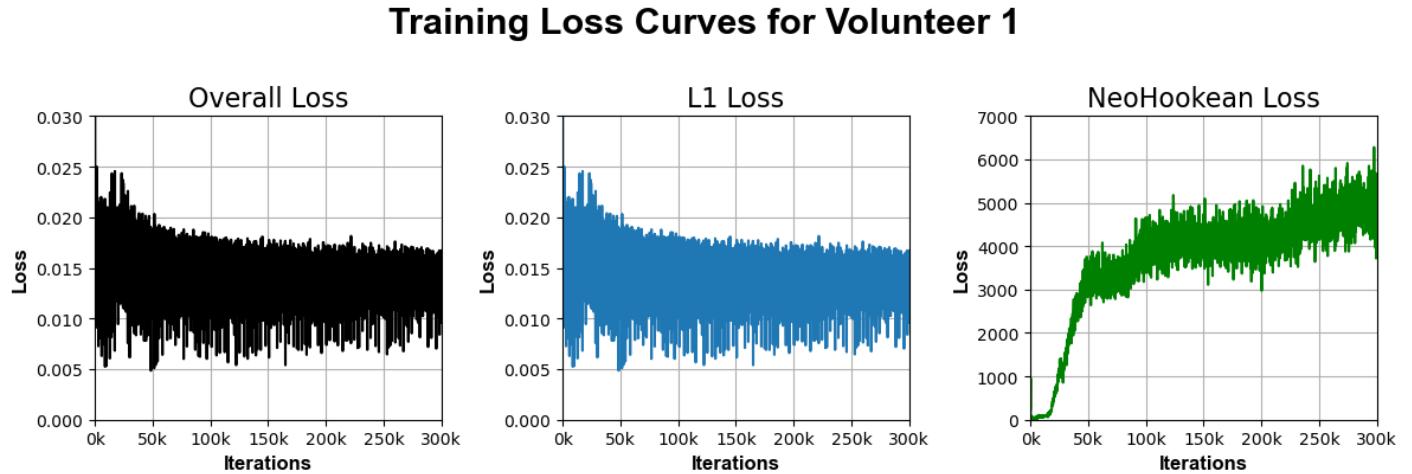


Figure 4.14: Training loss curves for WarpPINN with $\mu = 0$.

challenges to the task of quantifying the impact of the Neo-Hookean loss on image similarity performance. Therefore, this topic is studied and analysed further in the following subsections of results of hyperparameter tuning experiments.

4.1.7 Hyperparameter tuning

This subsection presents and documents the results of the hyperparameter tuning experiments performed on the WarpPINN method for Volunteer 1 of the CMAC dataset. Five different parameters were grouped into 2 groups, “Training parameters” and “NeoHookean loss parameters”, and 2 tuning approaches were carried out.

4.1.7.1 OFAT search: Training paramters

The Mu parameter, learning rate and batch size were tuned using a One-Factor-at-a-Time (OFAT) approach, where each parameter is varied independently whilst the other two parameters are fixed. The image registration performance of each model was evaluated with the criteria of image similarity and volume preservation of the LV. Image similarity performance for each model was measured by calculating the Mean SSIM (MSSIM) between the warped Template (ES) and Reference (ED) images over all three representative slices. Volume preservation was measured with the $||J| - 1|$ metric over the LV at the ES phase. Figure 4.15 shows the results of each WarpPINN model tested for this hyperparameter tuning method.

It is first noted that a trend in tuning μ is observed: an increase in μ correlates to a decrease in image similarity performance (MSSIM) and a decrease in $||J| - 1|$. This result affirms previous subsections’ discussions in this chapter of the impact of μ on volume preservation, as well as image similarity.

Secondly, it is observed that there is an optimum learning rate for the set of parameters and iterations tested, with the highest MSSIM being achieved with a learning rate of 10^{-3} . As learning rate is progressively increased to 10^{-3} , image registration performance is improved (with MSSIM increasing and $||J| - 1|$ decreasing). However, higher learning rates caused performance to deteriorate. This behaviour is expected and is typical in training deep learning models. Low learning rates mean that the model makes small but progressive changes in each iteration. When learning rate is increased,

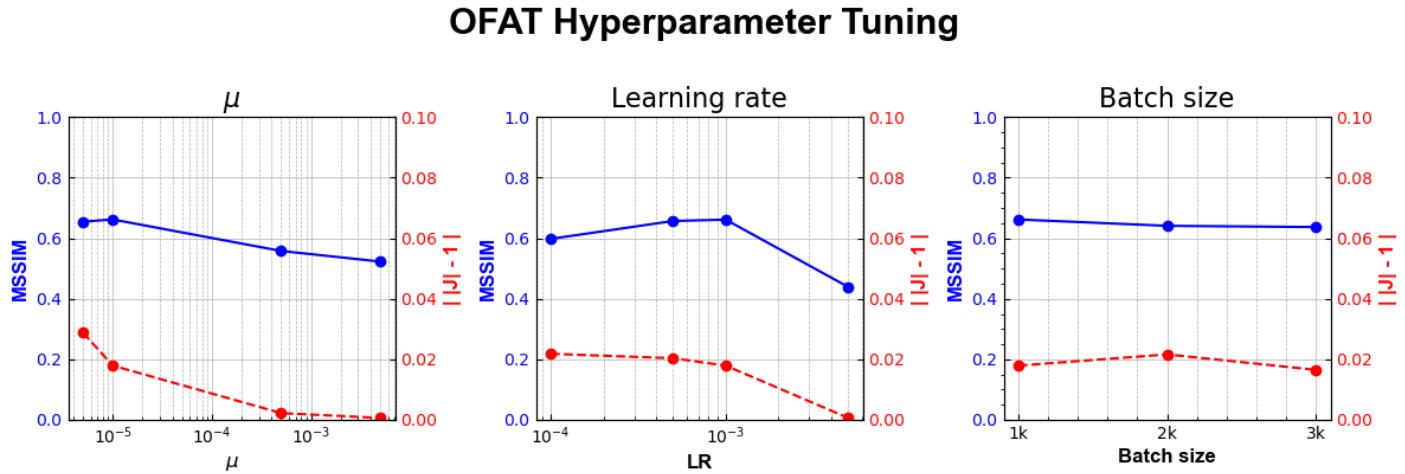


Figure 4.15: OFAT Hyperparameter tuning results

the model is able to make more significant changes to weights, resulting in increased performance given the same number of iterations. This can be observed when analysing the training loss curves for WarpPINN models of different learning rates in Figure 4.16, where a learning rate of 10^{-4} exhibit a slow, gradual decrease in loss, in comparison to higher learning rates of 5×10^{-4} and 10^{-3} . Lower learning rates also have a risk of being stuck in local minima, which further inhibit performance.

Conversely, a learning rate that is too high means that the model makes changes at each iteration that are too drastic to result in any learning to take place. This is observed in the training loss curve when the learning rate is 5×10^{-3} .

It is also observed that variations in the batch size within the tested range, which determines the number of collocation points sampled in the myocardium and background regions, has no significant effect on the performance of WarpPINN.

This series of hyperparameter tuning tests affirm that the default learning rate of 10^{-3} and batch size of 1000 are suitable to be used. Thus, further experiments can, instead, optimise the NeoHookean loss parameters, which have more significant impact on the image registration performance.

4.1.7.2 Grid search: NeoHookean loss paramters

The μ , λ_{inc} and λ_{bg} parameters were tuned together using a grid search approach on a pre-defined grid of 27 sample points (27 possible combinations). The results of this grid search are reported in Table 4.3, with each WarpPINN model again being evaluated using the MSSIM and $|J| - 1$ metrics. These results are also visualised in the response surface plots in Figure 4.17, where a manifold interpolated from the datapoints is generated to represent the optimisation landscape.

The relationship between μ , image similarity performance and volume preservation is once again observed, with lower μ values resulting in higher MSSIM and higher $|J| - 1$ values, as expected. It is also observed that λ_{inc} exhibits a similar behaviour. λ_{inc} parameterises the strain energy function used in the NeoHookean loss function for the incompressible region and determines the degree that incompressibility is enforced. It is, therefore, expected that higher λ_{inc} values result in lower $|J| - 1$ values. Lower λ_{inc} values resulting in lower MSSIM values indicates that there is a trade-off between

μ	λ_{inc}	λ_{bg}	MSSIM	$ \mathbf{J} - \mathbf{1} $
5×10^{-6}	10^4	0	0.628	0.079
		1	0.604	0.102
		10	0.632	0.084
	10^5	0	0.530	0.028
		1	0.503	0.034
		10	0.484	0.024
	10^6	0	0.530	0.007
		1	0.516	0.008
		10	0.494	0.005
1×10^{-5}	10^4	0	0.606	0.065
		1	0.577	0.070
		10	0.602	0.076
	10^5	0	0.547	0.021
		1	0.537	0.020
		10	0.525	0.021
	10^6	0	0.505	0.003
		1	0.485	0.004
		10	0.490	0.003
5×10^{-4}	10^4	0	0.516	0.009
		1	0.524	0.010
		10	0.512	0.010
	10^5	0	0.470	0.001
		1	0.507	0.002
		10	0.497	0.002
	10^6	0	0.494	0.000
		1	0.493	0.000
		10	0.469	0.001

Table 4.3: Grid search results for different hyperparameter values of μ , λ_{inc} , and λ_{bg} , evaluated using MSSIM and $||\mathbf{J} - \mathbf{1}||$ metrics.

image similarity accuracy and biomechanical plausibility for any WarpPINN model.

Finally, it is observed that the variations made to λ_{bg} have little significance in impacting the performance of the WarpPINN models. There is limited justification for using the NeoHookean model on the background region of the heart, which includes other tissues, blood and artefacts. Therefore, λ_{bg} should be set to zero, nullifying the physics-informed loss for the background region. The use of a small λ_{bg} in the default WarpPINN parameters was based upon previous research by López et al. producing mixed results when $\lambda_{\text{bg}} = 0$, correlating to lower landmark errors for WarpPINN, but higher for WarpPINN-FF.

It is noted that this grid search of 27 points is limited and the surface response plot is only an approximation. Nevertheless, the results of this study successfully informs the trade-off between image similarity and predicting realistic deformations, depicting the ill-posed problem in medical image registration.

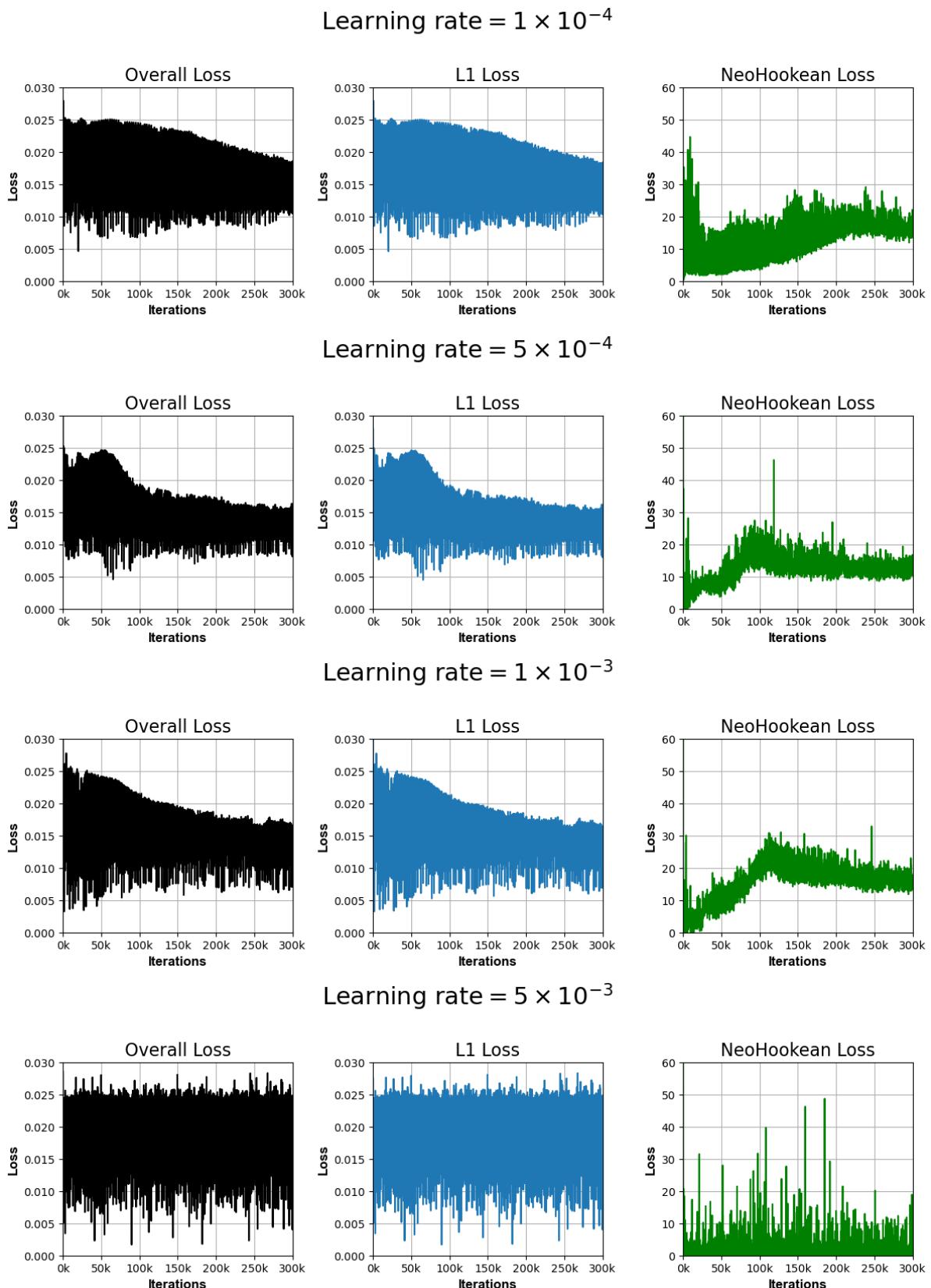


Figure 4.16: Training loss curves for WarpPINN with varying learning rates.

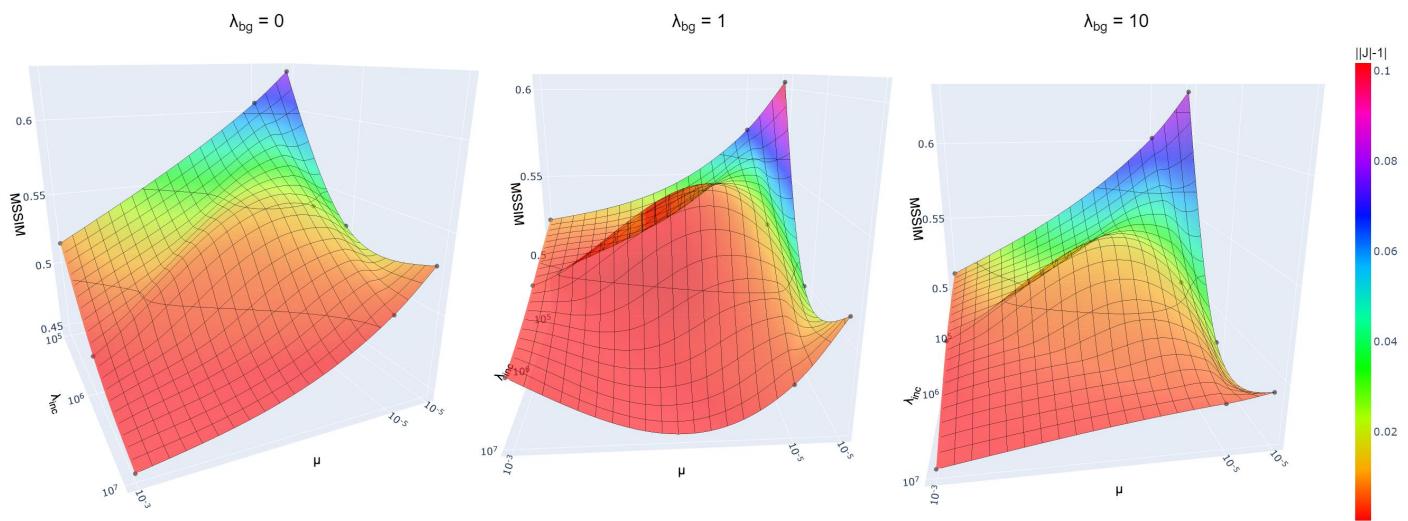


Figure 4.17: Estimated response surface plots for WarpPINN in the tuning of NeoHookean loss parameters

4.2 ACDC

This section documents and presents results and subsequent discussions for the adaptation of the WarpPINN method for image registration on patients from the ACDC dataset, following the methodology outlined in Section 3.2.

4.2.1 Selective presentation of individual patient results

Firstly, to assess and qualitatively inspect the image registration process and performance of the WarpPINN method on ACDC data, the predicted/warped template images, segmentation masks and deformation fields for patients 10 (DCM), 30 (HCM), 50 (MINF) and 70 (NOR) are visualised in Figures 4.18- 4.21.

These visualisations provide a qualitative understanding and validation of the image registration performance on each patient by observing the warped template image being similar to the Reference (ED) image, as well as observing the overlapping area between the corresponding binary masks. The 3D deformation field visualised in each slice also affirms that the predicted deformation is as expected, showing that the ES image is warped in the direction that the LV expands to match the ED (R) image.

Quantitative results for the image registration performance for each of the four considered patients is also presented in Table 4.4. These figures affirm the qualitative observations by quantitatively showing the warped template image for each patient having higher similarity with the reference image than the *Initial* configuration (with the exception of the Apical slice for Patient 10), for all employed image similarity metrics. It is also noted that the mean Jacobian determinant being close to 1 (measured by $||J| - 1|$ being close to 0) also affirms that the quasi-incompressibility constraint of the myocardium is satisfied.

	Basal				Mid-Ventriele				Apical			
	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $
Patient 10												
<i>Initial</i>	0.637	30.000	2.303		0.736	13.504	1.334		0.855	10.000	0.346	
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.901	<u>9.375</u>	0.250	0.046 (0.044)	0.802	10.000	0.561	0.044 (0.043)	0.801	<u>10.704</u>	<u>0.673</u>	0.023 (0.020)
Patient 30												
<i>Initial</i>	0.637	50.000	3.580		0.670	25.654	2.450		0.692	30.139	2.177	
$\mu = 5 \times 10^{-2}, \lambda = 10^8$	<u>0.869</u>	13.385	<u>0.408</u>	0.146 (0.116)	0.751	21.680	1.276	0.109 (0.135)	0.748	20.010	1.168	0.073 (0.079)
Patient 50												
<i>Initial</i>	0.498	26.454	3.318		0.696	12.305	1.592		0.837	9.914	0.409	
$\mu = 5 \times 10^{-2}, \lambda = 10^9$	0.839	7.178	0.423	<u>0.140</u> (0.130)	<u>0.770</u>	<u>13.939</u>	<u>0.823</u>	<u>0.101</u> (0.100)	<u>0.788</u>	10.000	0.653	<u>0.058</u> (0.060)
Patient 70												
<i>Initial</i>	0.482	30.000	3.855		0.512	16.890	2.839		0.430	20.000	3.481	
$\mu = 5 \times 10^{-2}, \lambda = 10^9$	0.786	10.121	0.703	0.194 (0.191)	0.762	18.986	1.060	0.204 (0.208)	0.726	14.041	0.880	0.185 (0.198)

Table 4.4: Quantitative results for the image registration performance of WarpPINN on Patients 10, 30, 50 and 70.

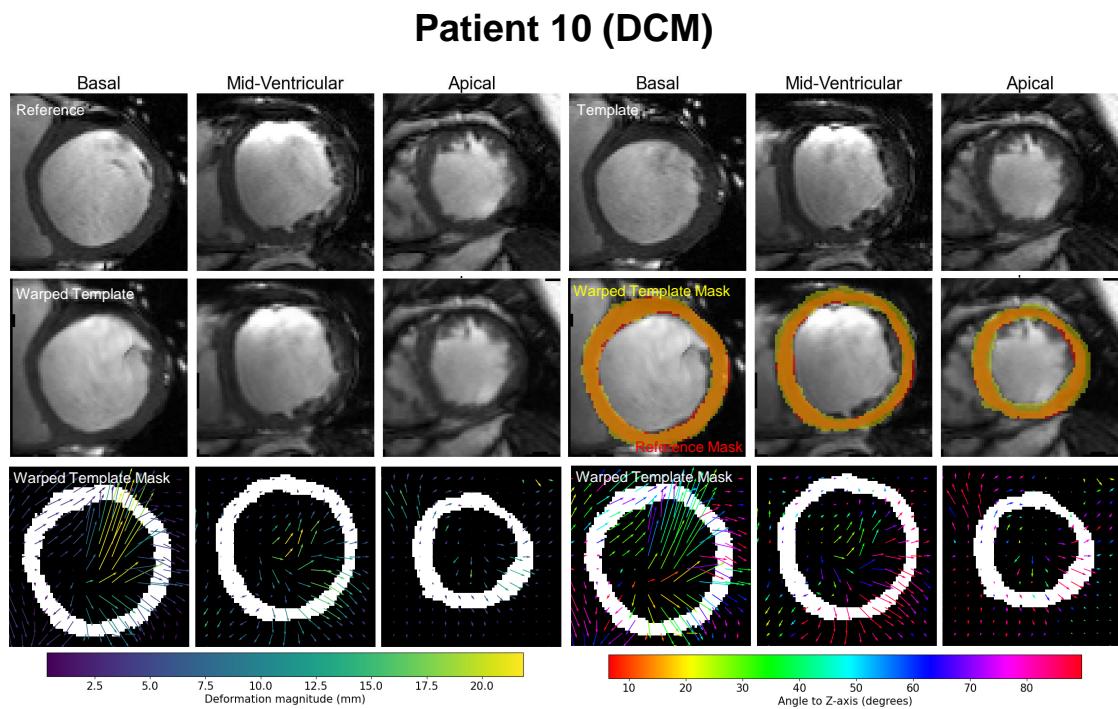


Figure 4.18: Visualisation of Reference, Template and Warped Template images and segmentation masks for Patient 10.

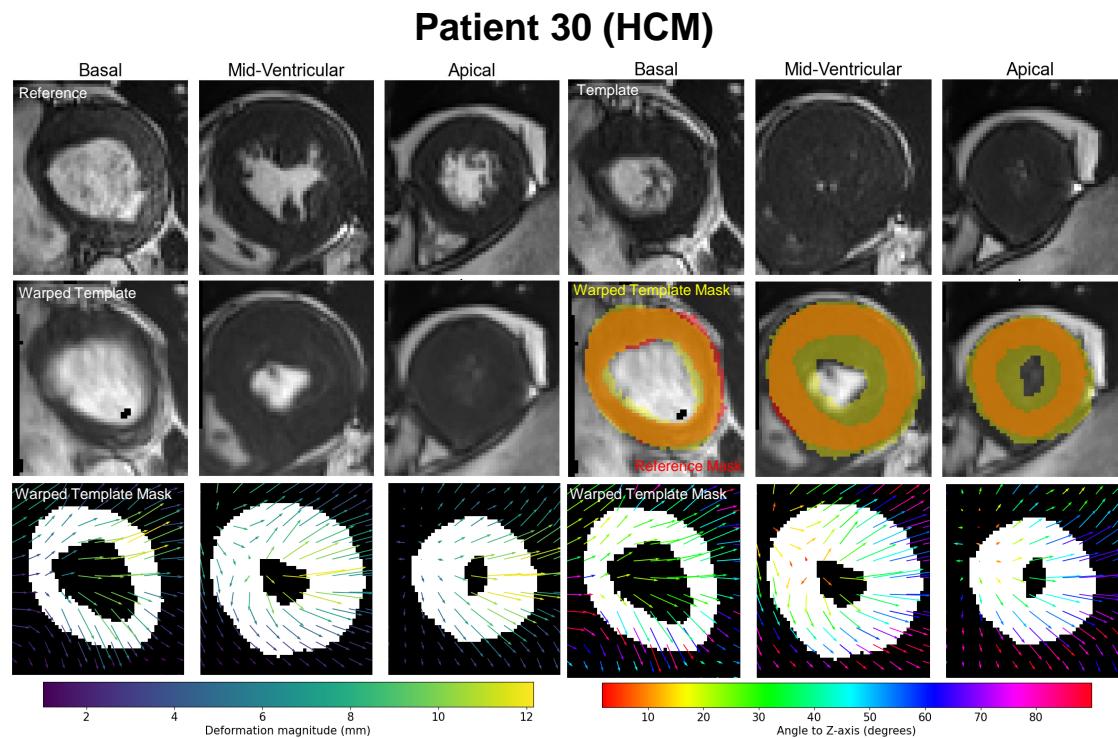


Figure 4.19: Visualisation of Reference, Template and Warped Template images and segmentation masks for Patient 30.

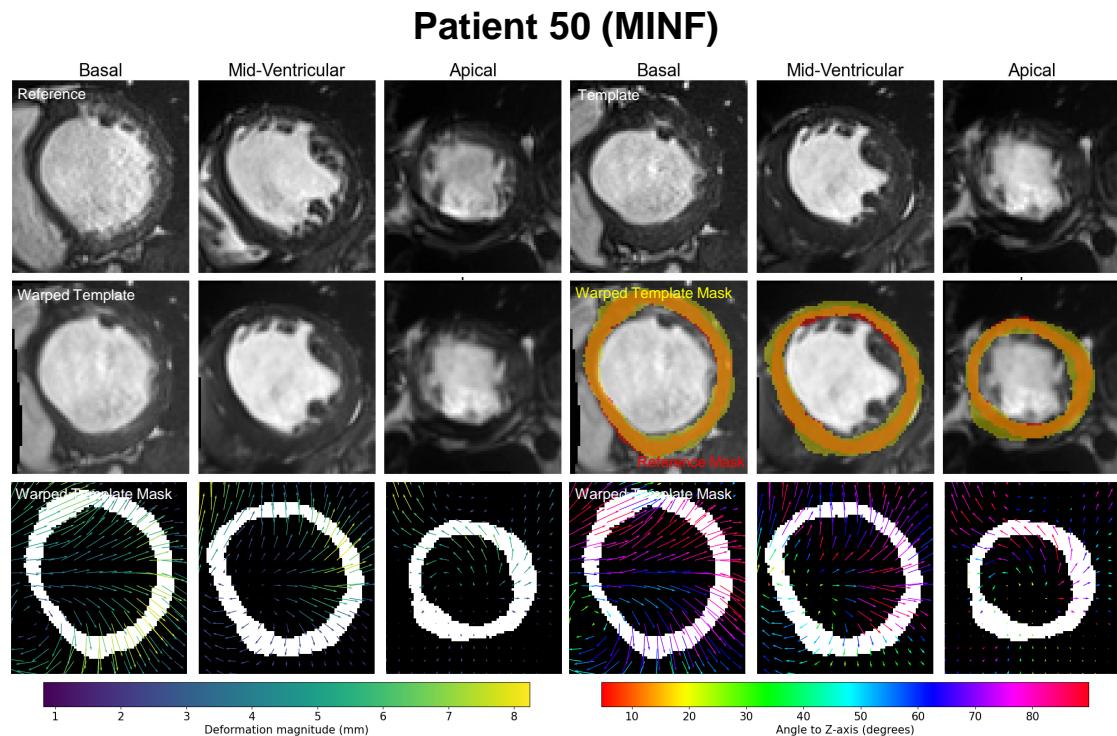


Figure 4.20: Visualisation of Reference, Template and Warped Template images and segmentation masks for Patient 50.

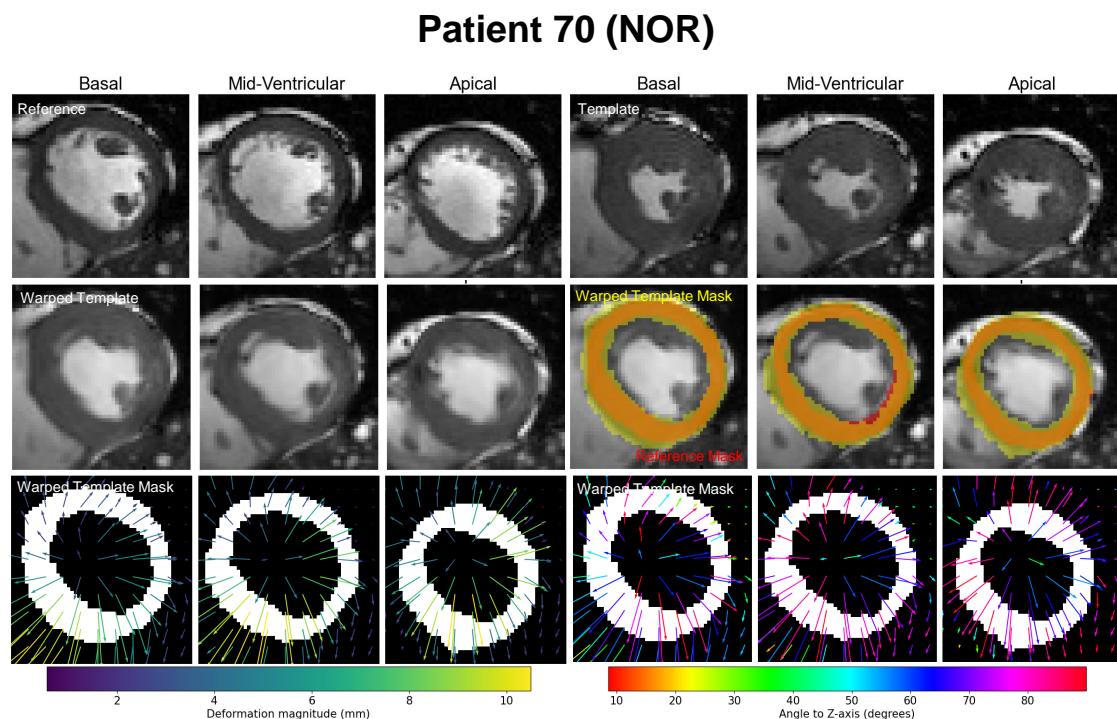


Figure 4.21: Visualisation of Reference, Template and Warped Template images and segmentation masks for Patient 70.

4.2.2 Results aggregated by disease

4.2.2.1 Hyperparameter tuning

In performing the WarpPINN pairwise optimisation for each of the 39 chosen patients from the ACDC dataset (Section 3.2.1.1), an extensive hyperparameter tuning search was carried out to optimise image registration performance by varying the μ and λ parameters. In this section, λ refers specifically to the λ_{inc} hyperparameter of the WarpPINN loss function (3.5).

This hyperparameter search employed a Bayesian optimisation-like approach where the default μ and λ values of 1×10^{-5} and 10^5 , respectively, were chosen as a starting point for each patient. The μ and λ parameters were then tuned in order to achieve a desired volume preservation, determined as a $||J| - 1|$ in the range of 0.1-0.15, which previous related studies have quoted to be a sufficient benchmark value for realistic deformations.

The remainder of the hyperparameter search then constituted maximising the image similarity performance, as measured with the three considered evaluation metrics (DSC, MCD and HD95). The direction that the μ and λ parameter values were varied in was based upon the quantified understanding of their impact on the image similarity-volume preservation trade-off, as studied in Sections 3.1.7 and 3.1.8.

A total of 305 hyperparameter tuning tests were completed before results were satisfactory (Table ??). Since WarpPINN is an IDIR pairwise optimisation method, each test required training a new model from scratch for each patient. With the same training regime outlined in the Validation experiments (3.1.5), total training times ranged from 2.5 ~ 7 hours, although it was noted that the WarpPINN loss for training on ACDC patients converged significantly earlier than in the Validation experiments with CMAC data, and that the number of iterations and consequently, training time, could have been reduced significantly, but kept at 300k to maintain consistency and fairness across all test results.

Disease group	DCM	HCM	MINF	NOR	All patients
Number of tests done	61	93	81	70	305

Figures 4.22- 4.25 show the approximate surface response plots for each test done, aggregated by disease group, to visualise the optimisation landscape when tuning μ and λ . These visualisations affirm previous discussions (3.1.8) of the relationship of the μ and λ parameters with image similarity and volume preservation performances. The optimal parameters for the determined “best” (4.2.2.2) performance correspond to μ values with order of magnitude of -2 and λ values with order of magnitude of 8~9. It is also observed that the HCM group exhibits the least smooth manifold, which is in line with the understanding of mechanical function of the LV of patients with HCM (Section 3.2.1), where there is a thicker myocardium and thus, less motion to be registered.

4.2.2.2 Selection of best results

The approach taken to determining the “best” result (optimal hyperparameter values) for each patient involved choosing the model that yielded the best image similarity performances while maintaining the desired volume preservation, as

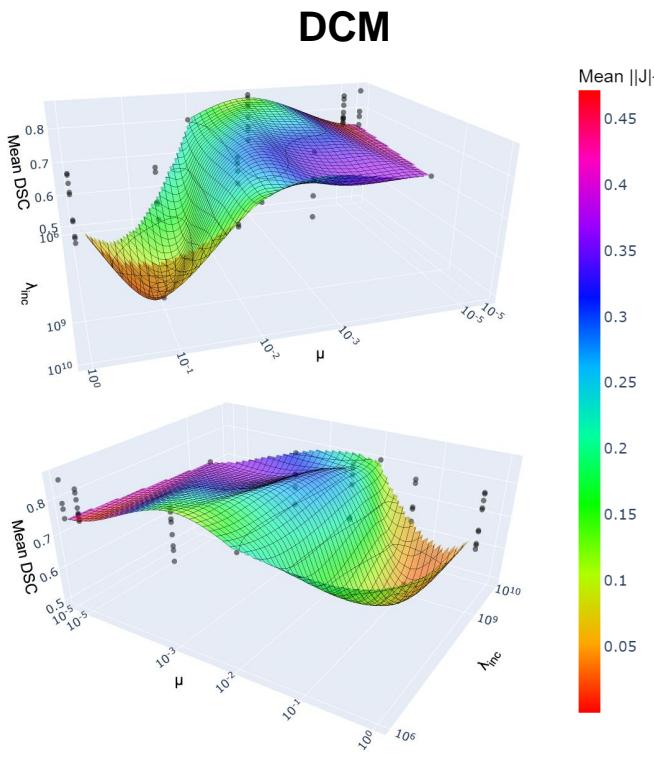


Figure 4.22: Surface response plot approximated from all hyperparameter tuning tests done with the DCM disease group.

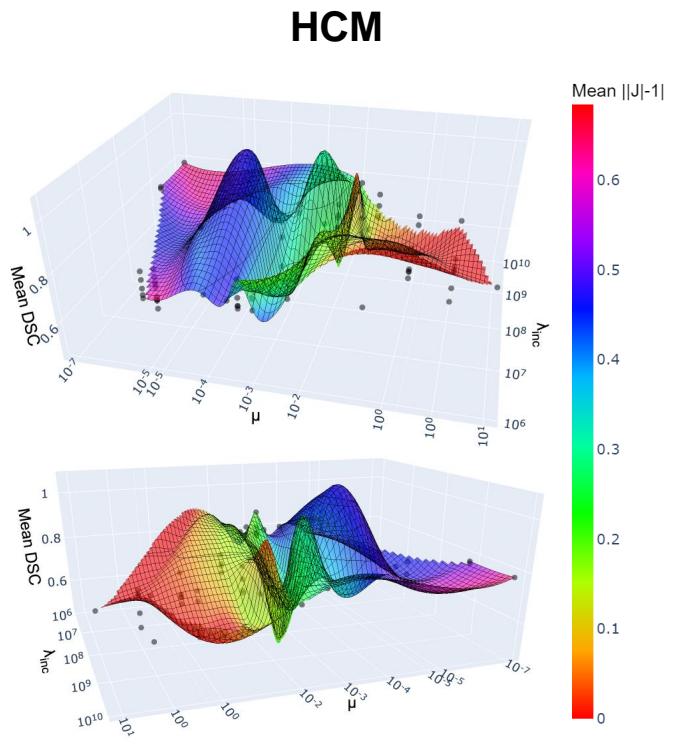


Figure 4.23: Surface response plot approximated from all hyperparameter tuning tests done with the HCM disease group.

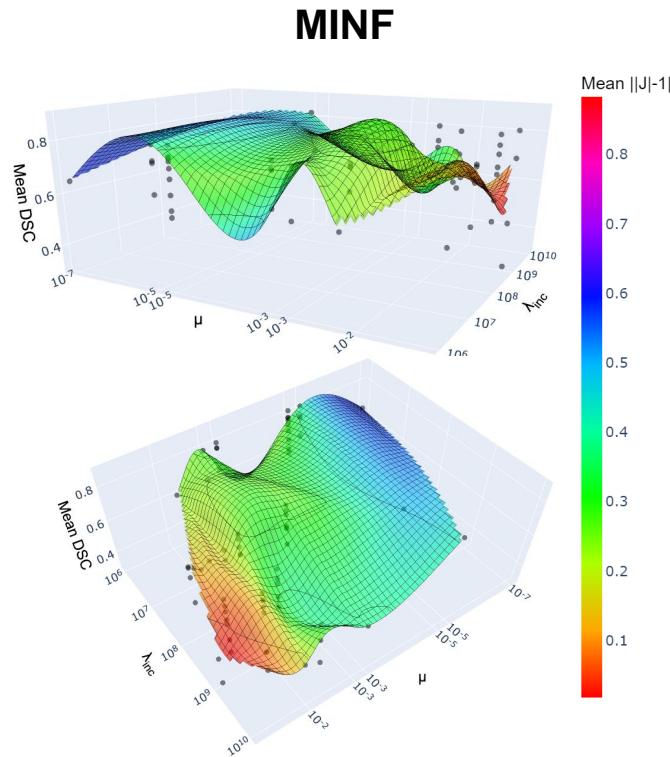


Figure 4.24: Surface response plot approximated from all hyperparameter tuning tests done with the MINF disease group.

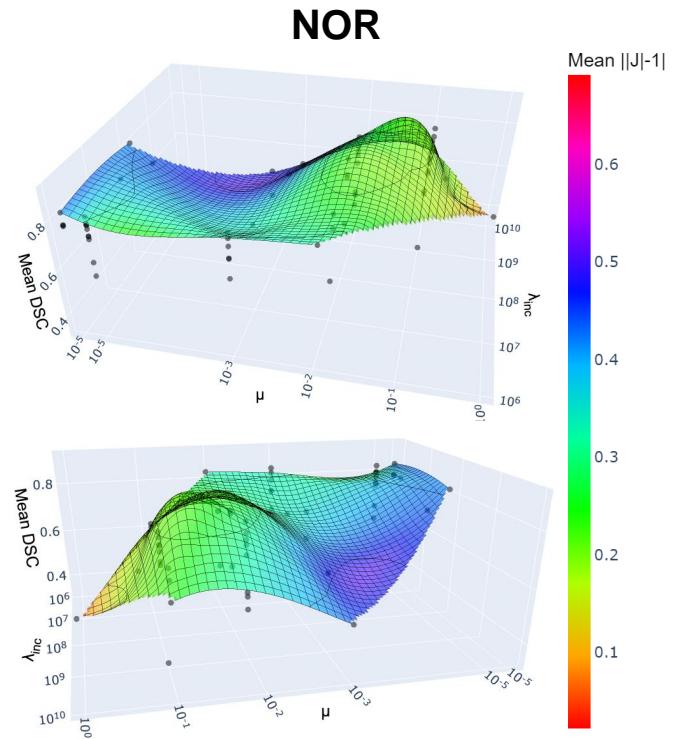


Figure 4.25: Surface response plot approximated from all hyperparameter tuning tests done with the NOR disease group.

outlined. Appendix A has been included to demonstrate this decision, with tables to show 4 results for each patient: the performance when $\mu = 0$ (nullified physics-informed loss), the model with the highest image similarity performance, the model with the lowest $|J| - 1$, as well as the chosen best model. It is noted that these multiple criteria often overlap.

4.2.2.3 Evaluation of best results

The quantified performances of the best chosen models for each patient in each disease group is presented in the box plots of Figure 4.26. Here, it is observed that, across all 3 image similarity metrics, patients in the DCM group achieved the best median image similarity performances. As initially drawn from the example of 4 patients (4.2.1), this could be explained with DCM patients having the highest similarity in the *initial* configuration of ground truth images at the ED and ES frames.

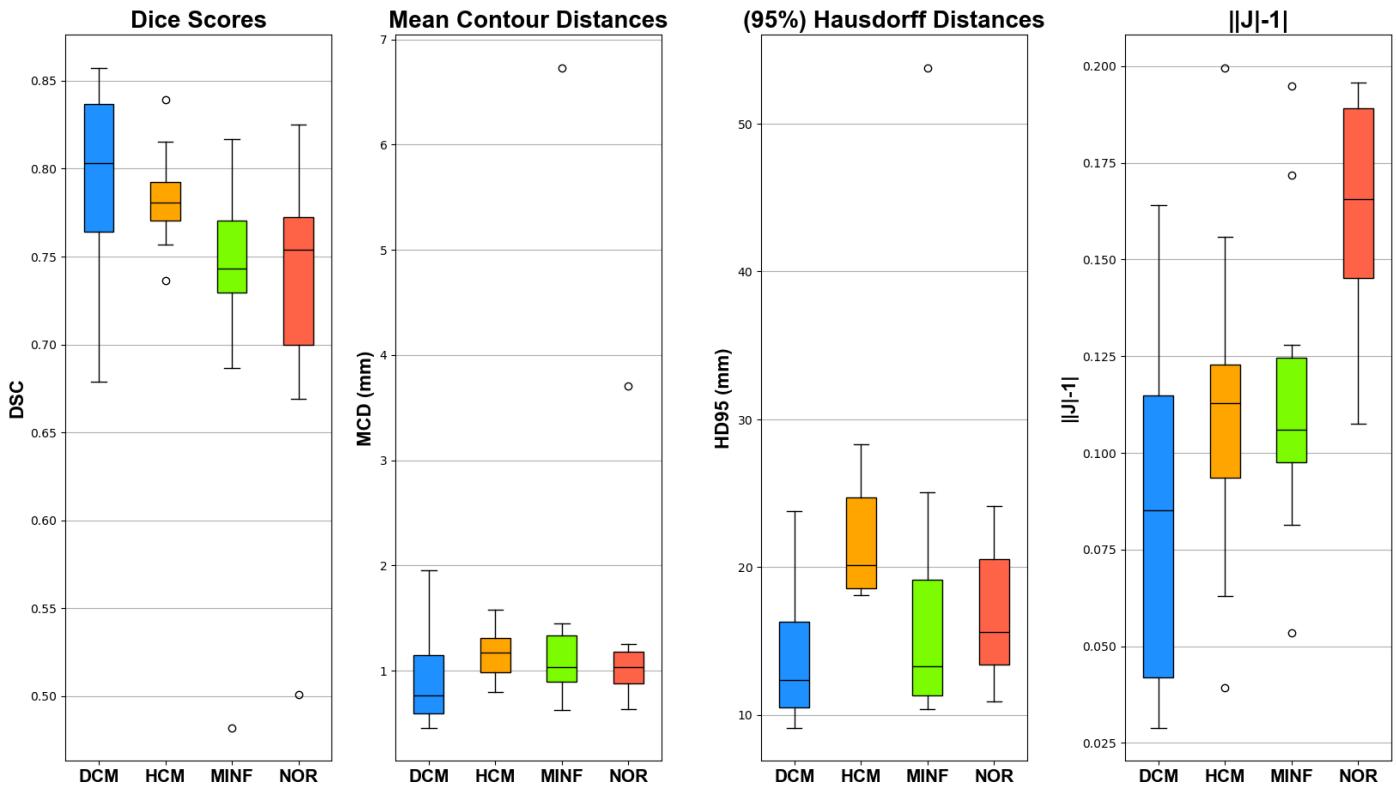


Figure 4.26: Box plots of measures of evaluation metrics for all considered patients from the ACDC cohort for WarpPINN with optimal hyperparameters.

Table 4.5 compares these results with the BIGM-VAE approach and it is observed that WarpPINN achieves competitive image registration performances in comparison to this benchmark model. It is appreciated that these comparison of results show that the MCD is lower for each patient group for WarpPINN in comparison to the BIGM-VAE approach (introduced in Section 3.2.4).

The trend of DCM and MINF patient groups yielding the best image similarity performances is observed in both WarpPINN and BIGM-VAE, which serve as a measure of validation for the adaptation of WarpPINN to the ACDC dataset. It is also observed from the box plots that the median $|J| - 1$ values were lowest for DCM, followed by MINF, HCM and then NOR. A similar trend is observed in the results of BIGM-VAE, with the difference of the NOR patient

group performing slightly better than MINF in volume preservation.

Disease	Method	DSC	HD95	MCD	$\ J\ -1\ $
DCM	WarpPINN	0.793(0.018)	14.383(1.928)	0.912(0.145)	0.085(0.012)
	BIGM-VAE	0.832(0.021)		1.260(0.221)	0.084(0.021)
HCM	WarpPINN	0.783(0.045)	21.899(4.877)	1.173(0.259)	0.113(0.015)
	BIGM-VAE	0.764(0.047)		3.398(0.934)	0.149(0.025)
MINF	WarpPINN	0.725(0.029)	19.228(1.323)	1.655(0.329)	0.116(0.024)
	BIGM-VAE	0.801(0.038)		1.549(0.196)	0.116(0.017)
NOR	WarpPINN	0.727(0.017)	16.915(3.518)	1.256(0.076)	0.164(0.015)
	BIGM-VAE	0.704(0.076)		3.123(0.939)	0.140(0.037)

Table 4.5: Comparison of WarpPINN and BIGM-VAE image registration performance on the ACDC dataset, aggregated by disease group.

4.2.2.4 Global strain analysis

The estimated strains of WarpPINN on the ACDC data is also compared with equivalent estimations from the previously used benchmark, the DeepStrain (CarMEN) method. Figure 4.27 shows these global strain curves, where it is observed that the order of peak mean magnitudes of radial and circumferential strains follow the same order (from lowest to highest, DCM, MINF, HCM, NOR) for both the WarpPINN and DeepStrain methods. It is also observed, as with the equivalent strain curve comparisons of WarpPINN and DeepStrain in the Validation experiments on CMAC (Section 4.1.3), that WarpPINN is able to predict larger strain values. It is also, once again, observed that the image registration methods considered thus far face difficulty in predicting large longitudinal strains, which is discussed in Section 4.1.3.

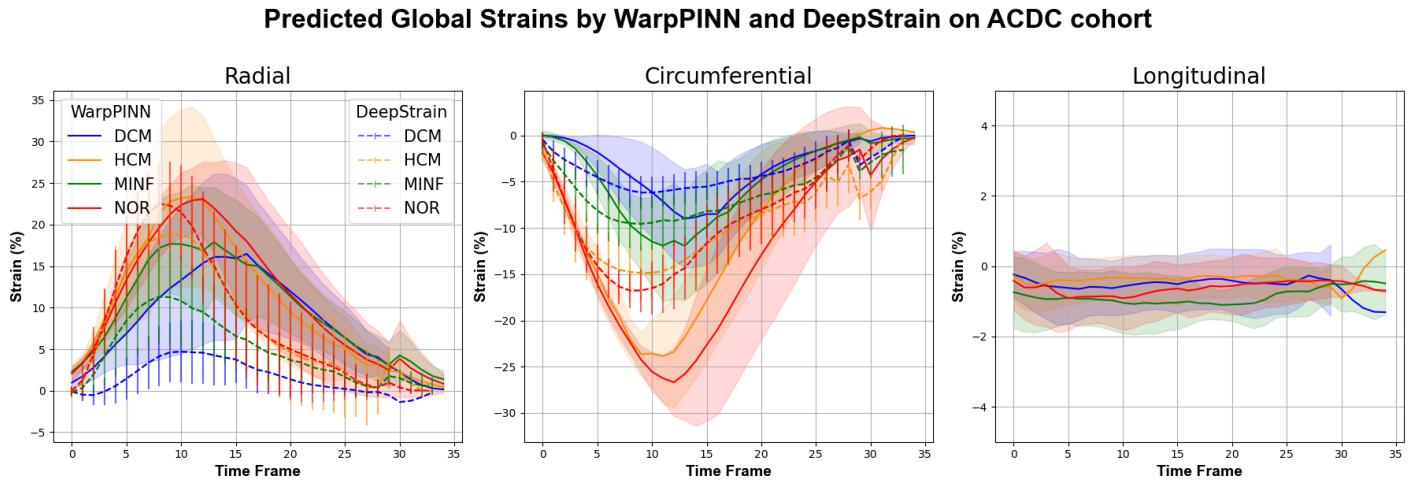


Figure 4.27: Estimated global strain curves by WarpPINN and DeepStrain on the ACDC dataset, aggregated by disease group.

	Basal			Mid-Ventrie			Apical		
	DSC	MCD	$ J -1 $	DSC	MCD	$ J -1 $	DSC	MCD	$ J -1 $
Method									
FFD-VP	0.731 (0.12)	2.925 (1.328)	0.135 (0.052)	0.756 (0.088)	2.442 (1.352)	0.147 (0.072)	0.684 (0.143)	2.784 (1.648)	0.148 (0.075)
dDemons	0.778 (0.102)	2.295 (1.174)	<u>0.132</u> <u>(0.031)</u>	<u>0.788</u> <u>(0.075)</u>	1.993 (1.112)	0.141 (0.048)	0.707 (0.128)	2.437 (1.453)	<u>0.132</u> <u>(0.043)</u>
Motion-Net	0.751 (0.123)	2.814 (1.236)	0.167 (0.049)	0.745 (0.105)	2.799 (1.007)	0.167 (0.056)	0.656 (0.142)	2.853 (1.301)	0.171 (0.06)
BINN	<u>0.789</u> <u>(0.091)</u>	2.229 (0.86)	0.161 (0.063)	0.783 (0.097)	2.21 (0.918)	0.148 (0.056)	0.707 (0.147)	2.450 (1.253)	0.158 (0.081)
BIGM-VAE	0.829 (0.067)	<u>1.66</u> <u>(0.671)</u>	0.138 (0.060)	0.818 (0.055)	1.646 (0.971)	0.134 (0.067)	<u>0.731</u> <u>(0.137)</u>	<u>2.101</u> <u>(1.312)</u>	0.134 (0.091)
WarpPINN	0.781 (0.095)	1.106 (1.072)	0.144 (0.062)	0.759 (0.082)	1.236 (0.932)	0.126 (0.0503)	0.733 (0.085)	1.374 (1.439)	0.089 (0.042)

Table 4.6: Comparison of image registration performance of all considered methodologies on the ACDC dataset.

4.2.3 Results across all patients

Finally, the image registration performance of WarpPINN across all patients is presented and compared to the alternative methodologies in Table 4.6. These results are evaluated with the representative slices and it is observed that WarpPINN achieves comparable results in the image similarity measured by DSC. It is also appreciated that, across all representative slice levels, WarpPINN achieves the lowest MCD and $||J| - 1|$ in comparison to all compared methodologies.

4.2.4 Discussion of WarpPINN-ACDC experiments

In this section, the task of adapting the WarpPINN method to perform image registration on patients from the ACDC dataset is successfully implemented. In particular, the extensive hyperparameter tuning results of 305 total tests demonstrate the capability of WarpPINN to achieve comparable Dice scores and superior Mean Contour Distances, while achieving the greatest level of volume preservation, in comparison to the set of considered alternative methodologies (Table 4.6). The superiority in MCD values, as opposed to DSC, could be a result of the WarpPINN method predicting 3D deformation fields from $3D + t$ inputs (as reflected in its architecture).

Through different levels of aggregation (individual patients, pathology group, and the entire dataset), the WarpPINN method’s implementation on the ACDC data is validated through comparisons to the alternative methodologies. It is also appreciated that WarpPINN achieves the best (lowest) $||J| - 1|$ in comparison to these benchmark techniques, indicating that the unique choice of regularisation term (NeoHookean hyperelastic strain energy) is well-justified in modelling and enforcing the quasi-incompressibility constraint of the myocardium, ultimately resulting in the predictions of realistic and biomechanically plausible deformations.

In this section, the vanilla WarpPINN model was successfully and extensively tuned to find the optimal hyperparameters for image registration performance. Due to this time-consuming task, however, the WarpPINN-FF model was not considered and, as such, presents a basis for future work. As seen in the results of the hyperparameter tuning tests, the range of optimal parameters (as well as other factors, such as training time-to-converge) is very different for the ACDC

and CMAC datasets (ACDC: Figures 4.22- 4.25 , CMAC: Figure 4.17), naturally due to several varying factors across each patient’s images, such as data resolution, dimensionality (including the time domain, i.e., the number of frames representing the cardiac cycle), quality, etc. It is, therefore, noted that the task of hyperparameter tuning is crucial at each step when employing the WarpPINN method for future use.

The hyperparameter tuning approach employed in these experiments employed a Bayesian optimisation-like methodology, which has proven to be successful in achieving sufficient performance, but also time consuming, given the pairwise initialisation-optimisation nature of IDIR networks. The consequences of an uninformed relationship between hyperparameters and image registration performance are observed in the next section of results for the generalisation study on the ACDC dataset, where the pairwise optimisation limitation is also addressed.

5

Extended Research: Generalisation of WarpPINN

Contents

5.1	Validating WarpPINN on CMAC	69
5.1.1	Motivation	69
5.1.2	Related work	69
5.1.3	Methods	70
5.1.4	Experimental results and discussions	72

This section outlines the different methodologies studied and discussed in this thesis.

5.1 Validating WarpPINN on CMAC

5.1.1 Motivation

As discussed previously, and as emphasised throughout this work, the use of Neural Implicit Representations for image registration (Implicit Deformable Image Registration) faces challenges and is limited by its pairwise optimisation approach. Such NIRs require a new MLP to be initialised and trained from scratch for every new image pair. Whilst the WarpPINN approach does train on multiple Template images by using the same Reference image of the LV at ED phase and the batching strategy outlined in Section 3.1.5, a new WarpPINN model is required to be initialised for every patient, which results in long training times for each patient. These neural networks are also trained to overfit on only two images, as in the nature of IDIRs acting as a parameterisation of an optimisation algorithm minimising an image similarity loss, hence not being justified to be used for inference on other images or unseen data.

Therefore, this section is concerned with approaches to generalising the WarpPINN method to perform image registration on unseen patients' data. The experiments and evaluation methods outlined in this section aim to show a proof of concept of generalisation, rather than achieving state-of-the-art or competitive results with other image registration methodologies, such as those discussed in Section 4.2.3. Achieving such proofs of concepts have meaningful implications as comparative results of the WarpPINN method on the ACDC dataset suggests that the advantages of physics-informed neural networks (specifically, the technique of introducing a learning bias modelled after the NeoHookean) over traditional and other DL-based methods can result in predicted deformations with competitive accuracy and superior volume preservation, i.e., more realistic deformations.

5.1.2 Related work

As mentioned briefly in the justification of the use of Fourier Feature mappings (FFMs), another approach to alleviate the natural spectral bias of neural networks is through the use of periodic activation functions. Sitzmann et al. introduced the use of such periodic activation functions to capture high-frequency details in the data through modulation, as opposed to traditional activation functions such as the ReLU, sigmoid or hyperbolic tangent (which is used in WarpPINN) functions [94]. Sitzmann et al. introduces SIRENs (Sinusoidal Representation Networks) and demonstrate the effectiveness of modulating the frequency of periodic activation functions in tasks requiring generalisation across different instances, in particular, by dynamically adjusting to the level of detail required by data.

Related works such as Zimmer et al. have extended the adaptability advantage of SIRENs to Neural Implicit Representations in order to achieve generalisation through the use of latent feature encodings. In particular, Zimmer et al. have demonstrated that different strategies utilising latent encodings produced by an autoencoder can allow an IDIR method to be generalised in registering images of lung CT data. Neural implicit representations are inherently signal-specific and thus, such generalisation strategies were based upon approaches used in signal approximation, including conditioning and

signal modulation [95].

Implementations of a generalised WarpPINN in this section are inspired by methods tested by Zimmer et al., namely:

- **Conditioned SIRENs:** where the latent feature encoding vectors are appended as inputs to the SIREN MLP.
- **Modulated SIRENs:** Zimmer et al. explored the use of Amplitude Modulation (AM) and Quadrature Amplitude Modulation (QAM) of the sinusoidal activation functions, where the modulation is determined by the latent feature encodings.

These conditioned/modulated SIRENs were evaluated through different tests. In particular, a generalisation experiment was carried out by training the modified SIREN on a small dataset of 10 image pairs, and subsequently testing its image registration performance (inference) on the same ten pairs, with results compared to the pairwise MLP trained and tested on each pair individually as a benchmark comparison. As with this section’s motivation, the objective of these tests by Zimmer et al. was to show proof of concept rather than state-of-the-art results.

5.1.3 Methods

5.1.3.1 Implementation

Given the existing implementation of FFMs in WarpPINN-FF, the work of this thesis explores the use of Fourier Feature encoding, as opposed to latent space encoding, as parameters to the SIREN. Therefore, WarpPINN-FF was modified to first leverage sinusoidal activation functions, as well as the weights being initialised with SIREN initialisation (the technique originally proposed by Sitzmann et al.) as opposed to the current hyperbolic tangent activation functions and Xavier initialisation of weights used in the original implementation of WarpPINN.

Weights from (scaled) uniform distribution:

For a weight matrix W_ℓ in layer ℓ :

$$W_\ell \sim \mathcal{U} \left(-\sqrt{\frac{6}{n_\ell}} \frac{1}{\omega}, \sqrt{\frac{6}{n_\ell}} \frac{1}{\omega} \right)$$

where n_ℓ is the number of input neurons for layer ℓ .

For the first layer:

$$W_1 \sim \mathcal{U} \left(-\frac{1}{\omega_0}, \frac{1}{\omega_0} \right)$$

where ω_0 is a hyperparameter that controls the frequency of the sinusoidal activations.

Similar to Zimmer et al.’s use of AM and QAM, the use of various signal modulation techniques were implemented to be used on the sinusoidal activation functions of the modified WarpPINN-FF network. These include:

AM

$$\text{AM: } \Psi(x) = \alpha \sin(x)$$

PSK

$$\text{PSK: } \Psi(x) = \sin(x + \phi)$$

QPSK

$$\text{QPSK: } \Psi(x) = \sin(x + \phi) + \cos(x + \phi)$$

Where $\Psi(x)$ represents the activation functions at every layer of the modified WarpPINN-FF model.

The alpha, beta and phi parameters that describe the above three modulation techniques are derived from the Fourier Feature mappings (which are used as inputs in WarpPINN-FF). As an exploratory research task, this work investigates the use of three basic example computation methods for these parameters:

- Statistical: Parameters are calculated based on the mean and variance of the FFMs.
- Adaptive Frequency Response: Parameters are calculated based upon calculating the total Energy of the signal and applying the sigmoid and tanh functions.
- Phase synchronisation: Parameters are calculated based upon calculating the coupling strength and magnitude of the mean phase.

The justification for the use of adaptive frequency response and phase synchronisation are based upon signal processing techniques that could present a stronger correlation to the available Fourier Feature mappings and thus an improved performance in parameterising alpha, beta and phi, as opposed to the basic statistical approach.

The equivalent of the conditioned SIREN model used by Zimmer et al. is the Fourier Feature mapping equivalent for WarpPINN and is referred to as CSIREN in this section.

5.1.3.2 Evaluation methodology

The evaluation of the SIREN models for generalisation will follow Zimmer et al.'s experimental design and method, where the modified WarpPINN-FF models are trained on a subset of the ACDC dataset and tested on the same subset in order to evaluate generalised image registration performance, as well as robustness to the size of the training set. In particular, the first 8 patients of the NOR (healthy) group of the ACDC dataset are chosen in these experiments.

The performance of these generalised models are benchmarked against the equivalent WarpPINN-vanilla model (with the same hyperparameters) trained “pairwise” on the corresponding patient, as from the experiments in Section 4.2.2.1.

5.1.4 Experimental results and discussions

5.1.4.1 Pairwise image registration performance of WarpPINN-FF with sinusoidal activations

The image registration performance, without generalisation, of the newly implemented sinusoidal WarpPINN models are first tested on Patient 61 with the objective of validating their use as IDIR methods. As discussed in previous sections, the task of validating techniques serves as an important first step in extensions to the discussed approaches.

Table 5.1 presents the quantitative results of each SIREN in predicting the deformation of the LV image of Patient 61 at the ES phase (Template) to the ED phase (Reference), alongside the performance of the pairwise WarpPINN-vanilla for comparison. It is observed that all SIREN models, with the exception of the SIREN trained with statistical AM, achieve a higher similarity score with the warped Template images than the Initial configuration, indicating a successful image registration in the image similarity criteria. It is simultaneously noted that the $||J|-1|$ of these SIRENs are higher than that of the vanilla model, indicating less volume preservation and less realistic deformations. However, it is now known that the hyperparameters being used ($\mu = 1 \times 10^{-5}$ and $\lambda = 10^5$) are not in the optimal range found and documented in Section 4.2.2.1, with the benchmark vanilla WarpPINN model also having a relatively high mean $||J|-1|$ of 0.491 across all three representative slices.

	Basal				Mid-Ventrie				Apical			
	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $
INITIAL	0.448	31.218	4.493	0.643	0.443	40.250	4.832	0.582	0.483	20.000	2.662	0.539
C SIREN	0.797	12.305	0.791	0.643	0.882	8.203	0.293	0.807	10.000	0.588	0.000	0.539
AM (statistical)	0.478	30.869	4.113	0.000	0.488	40.122	4.510	0.000	0.547	17.152	2.281	0.000
AM (adaptive freq)	0.809	10.093	0.619	0.722	0.845	10.000	0.524	0.662	0.779	10.088	0.789	0.492
PSK (statistical)	0.824	11.145	0.621	0.896	0.893	10.000	0.363	0.685	0.806	10.000	0.521	0.775
PSK (adaptive freq)	0.816	10.093	0.588	0.672	0.875	5.469	0.313	0.575	0.829	10.000	0.488	0.546
PSK (phase sync)	0.854	10.367	0.455	1.548	0.878	10.000	0.462	1.921	0.805	10.000	0.554	1.599
QPSK (statistical)	0.846	10.189	0.447	0.941	0.876	6.836	0.334	0.770	0.783	10.000	0.728	0.846
QPSK (adaptive freq)	0.819	10.808	0.717	0.671	0.849	10.000	0.506	0.684	0.789	10.000	0.627	0.643
WarpPINN-vanilla	0.728	14.150	1.076	0.575	0.864	19.072	0.562	0.469	0.782	10.019	0.853	0.427

Table 5.1: Quantitative pairwise image registration results of considered models on Patient 61.

5.1.4.2 Image registration performance on unseen data

The trained SIREN models (on Patient 61’s data) were used to predict the deformation field and test on Patient 62 in order to evaluate the dynamic inference performance of these modified SIRENs on unseen data. These results are presented in Table 5.2. It is first observed that all tested methods predicted a deformation mapping of ES to ED that resulted in an increase in image similarity, indicating a successful registration in comparison to the Initial configuration. Although this image similarity is lower than that compared to the baseline of WarpPINN-vanilla, as well as having a higher $||J|-1|$ indicating less biomechanical plausibility, it is noted that the objective of these tests are not to show state of the art results, but instead to show proof of concept for generalisation, as Zimmer et al. have done. The previous point of sub-optimal hyperparameter being used in experiments, as discussed in the previous experiment, is also recalled.

From assessing these results, it is observed that both image registration performance, both in the criteria of image similarity and volume preservation, are best in the adaptive frequency response method of computing modulation param-

	Basal				Mid-Ventricule				Apical			
	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $
INITIAL	0.487	30.769	3.836	0.502	19.725	3.267	0.454	20.000	3.530	20.655	2.220	0.613
C SIREN	0.650	13.842	1.820	0.678	0.608	23.737	2.334	0.578	0.635	21.076	1.766	0.488
AM (adaptive freq)	0.666	13.672	1.798	0.721	0.675	20.000	1.754	0.644	0.675	20.000	1.628	1.066
PSK (statistical)	0.657	10.938	1.611	1.049	0.623	21.569	2.115	0.733	0.667	20.000	1.632	0.502
PSK (adaptive freq)	0.646	15.176	1.792	0.616	0.644	20.000	1.922	0.606	0.654	13.229	1.632	0.502
PSK (phase sync)	0.621	19.803	2.364	1.685	0.604	30.000	2.402	2.161	0.692	13.740	1.515	1.749
QPSK (statistical)	0.636	14.575	1.897	1.072	0.639	14.701	1.881	0.744	0.683	13.057	1.753	0.995
QPSK (adaptive freq)	0.658	16.269	1.774	0.748	0.668	20.000	1.818	0.646	0.633	14.561	1.795	0.582
WarpPINN-vanilla	0.828	11.541	0.702	0.373	0.785	20.000	0.934	0.319	0.776	40.000	1.583	0.240

Table 5.2: Quantitative image registration results of considered models on Patient 62.

eters, for all modulation methods amongst SIREN models, also including CSIREN which performed well relative to the statistical and phase synchronisation-based SIRENs. As mentioned in the justification of the three methods of computing modulation parameters, this relative success could be attributed to the adaptive and frequency-based nature of the computation method, as opposed to the statistical and phase synchronisation methods, although this hypothesis would require again a comprehensive study of hyperparameters α , β and ϕ to understand their impact on performance.

The evaluation of the adaptive frequency-based SIRENs and CSIREN was extended to test the performance of inference on all patients in the NOR group. The results of this experiment is presented in Tables 5.3 and 5.4, where the mean DSC across representative slices is taken as the image similarity measure. These results are also illustrated in Figure 5.1.

	Patient 61		Patient 62		Patient 63		Patient 64		Patient 65	
	DSC	$ J -1 $								
INITIAL	0.458	0.481	0.340	0.569	0.217					
C SIREN	0.829	0.588	0.631	0.623	0.688	0.516	0.743	0.553	0.743	0.620
AM (adaptive freq)	0.811	0.626	0.672	0.618	0.697	0.590	0.739	0.609	0.644	0.566
PSK (adaptive freq)	0.840	0.598	0.648	0.574	0.667	0.604	0.732	0.571	0.669	0.569
QPSK (adaptive freq)	0.819	0.666	0.653	0.659	0.696	0.648	0.738	0.626	0.717	0.603
WarpPINN-vanilla	0.791	0.491	0.796	0.311	0.735	0.549	0.824	0.329	0.641	0.492

Table 5.3: Inference on NOR patients

	Patient 66		Patient 67		Patient 68		Patient 69		Patient 70	
	DSC	$ J -1 $								
INITIAL	0.373	0.410	0.174	0.693	0.651					
C SIREN	0.739	0.569	0.640	0.616	0.673	0.515	0.693	0.530	0.651	0.578
AM (adaptive freq)	0.743	0.620	0.640	0.616	0.658	0.657	0.624	0.659	0.652	0.660
PSK (adaptive freq)	0.759	0.606	0.593	0.591	0.676	0.563	0.654	0.575	0.685	0.590
QPSK (adaptive freq)	0.752	0.674	0.603	0.622	0.642	0.614	0.656	0.611	0.684	0.620
WarpPINN-vanilla	0.746	0.485	0.791	0.453	0.583	0.692	0.776	0.463	0.748	0.439

Table 5.4: Inference on NOR patients

Again, it is observed that each tested SIREN model have achieved a higher DSC than the Initial configuration, suggesting success in image similarity registration to an extent. It is still noted that the Jacobians are still high to be considered biomechanically plausible deformations, but these (DSC, $||J|-1|$) results fall within the range observed during the extensive hyperparameter tuning tests of WarpPINN-vanilla on ACDC for untuned models, suggesting potential to maintain image registration performance in similarity whilst improving volume preservation if a similar hyperparameter tuning is done, or alternatively using informed optimal ranges of μ and λ , as quoted in 4.2.2.1. This potential is justified especially with the observation of result for Patient 68, where the SIREN models outperformed the baseline WarpPINN-vanilla model in both image similarity and volume preservation.

5.1.4.3 Robustness test for sequential learning

Following Zimmer et al.'s evaluation approach for generalisation robustness, the considered SIREN models were trained sequentially from Patient 61 to Patient 68. For each model trained with N patients, the model's image registration perfor-

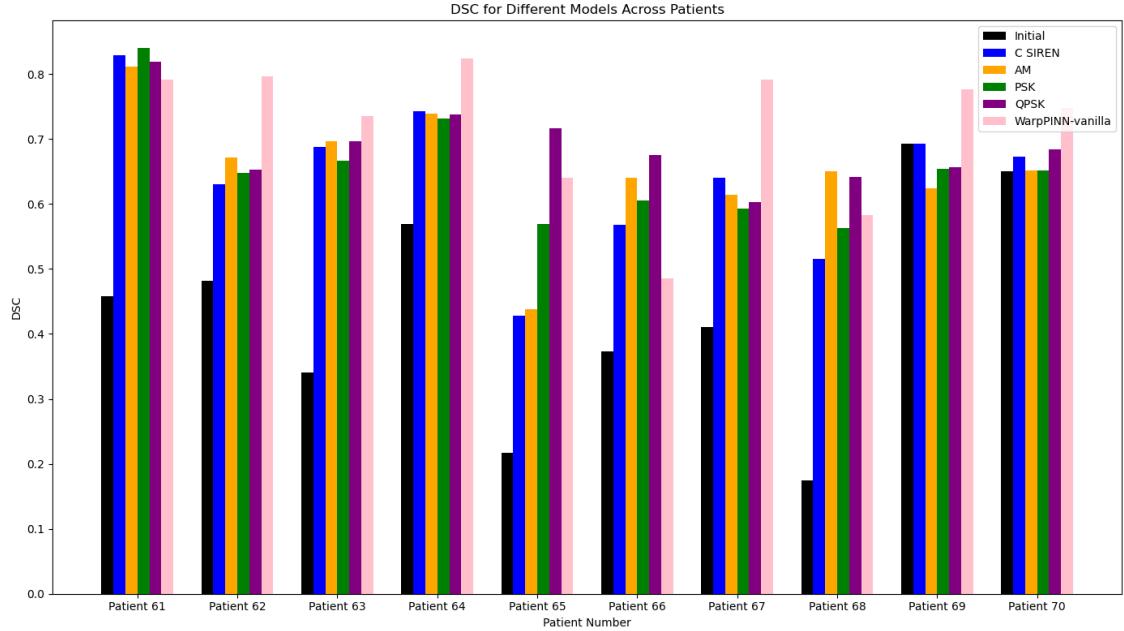


Figure 5.1: Inference results on every NOR patient.

mance at inference when testing on the same N number of patients is measured and the mean DSC across representative slices of each tested patient is taken as the evaluation measure of image registration performance. The results, for each tested SIREN model and at each $N \in [1, 8]$, are compared to the equivalent corresponding mean DSC of the pairwise vanilla WarpPINN models. The results of this robustness test are presented in Figure 5.2. In addition to testing the robustness of generalisation performance of the SIREN models to the size of the training set N , this experiment can also be considered as a test of the forgetting feature of neural networks when training on multiple sets of data, especially since the implementation of multi-patient learning was done through implementing sequential learning. It can be visually observed from Figure X that the image registration performance at $N = 1 - 4$ remain relatively constant, but deteriorates at $N \geq 5$. This result is interpreted as the SIRENs being able to generalise well (with the current hyperparameter settings and multi-patient learning approach) to a threshold of 4 patients, after which a larger training set causes the model to forget learnt generalised or personalised features. This observation of deterioration is confirmed with performed Wilcoxon and Man U Whitney tests confirming the statistical similarity of results of all models at $N = 1$ to 4, as well as a statistical difference of results of SIRENs when $N = 1$ to 4 and $N = 5$ to 8.

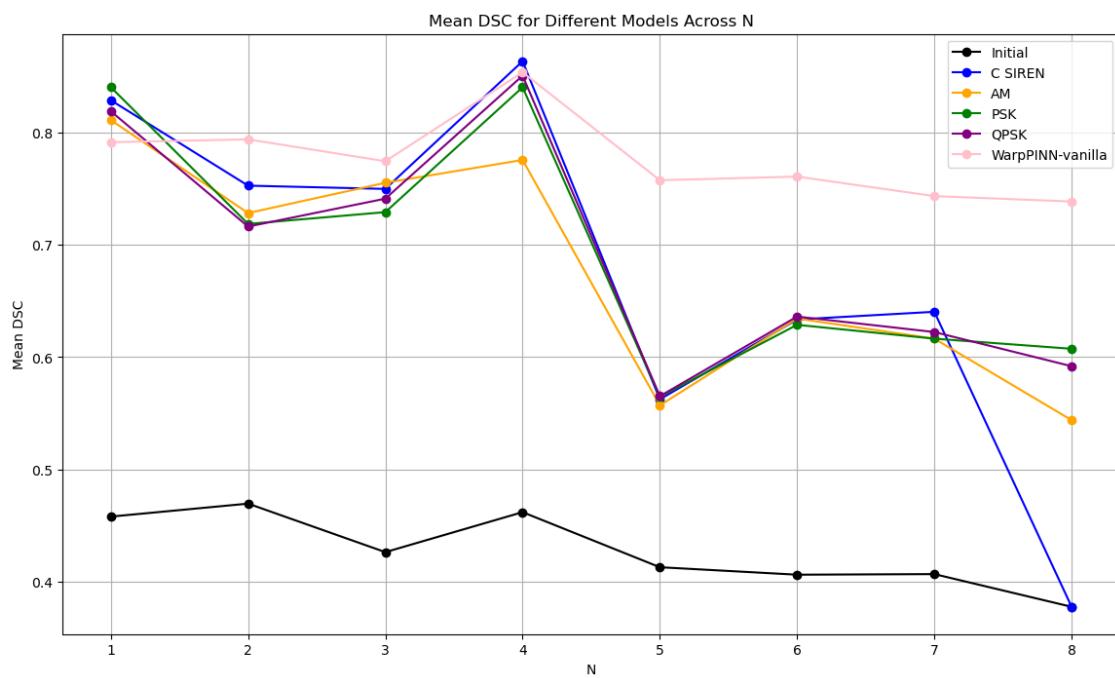


Figure 5.2: Robustness test

A

ACDC Best Results

	Basal				Mid-Ventriele				Apical			
	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $	DSC	HD95	MCD	$ J -1 $
Patient 1												
<i>Initial</i>												
$\mu = 0$	0.713	12.500	1.422	0.363	0.787	20.000	0.866	0.312	0.781	10.000	0.726	0.222
$\mu = 5 \times 10^{-3}, \lambda = 10^8$	0.750	20.000	1.232	0.244	0.804	20.000	0.961	0.293	0.785	17.528	0.973	0.174
$\mu = 5 \times 10^{-2}, \lambda = 10^8$	0.418	60.510	7.466	0.053	0.591	41.018	3.978	0.048	0.534	30.273	3.745	0.030
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.635	20.542	2.231	0.054	0.796	14.626	0.732	0.080	0.758	15.614	0.889	0.057
Patient 2												
<i>Initial</i>												
$\mu = 0$	0.829	10.000	0.572	0.521	0.735	15.039	1.311	0.528	0.800	8.203	0.556	0.189
$\mu = 5 \times 10^{-3}, \lambda = 10^8$	0.857	10.000	0.433	0.315	0.823	10.938	0.639	0.216	0.803	10.184	0.625	0.096
$\mu = 5 \times 10^{-2}, \lambda = 10^8$	0.734	15.244	1.216	0.236	0.752	10.938	0.843	0.164	0.778	17.092	0.884	0.092
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.656	30.000	2.160	0.050	0.711	15.039	1.273	0.047	0.705	19.806	1.157	0.020
Patient 3												
<i>Initial</i>												
$\mu = 0$	0.900	4.688	0.231	0.109	0.832	11.792	0.647	0.117	0.829	10.406	0.530	0.075
$\mu = 5 \times 10^{-6}, \lambda = 10^5$	0.907	4.688	0.213	0.134	0.834	12.690	0.632	0.109	0.830	10.000	0.531	0.077
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.827	10.500	0.610	0.029	0.807	11.365	0.698	0.025	0.814	11.082	0.660	0.021
Patient 4												
<i>Initial</i>												
$\mu = 0$	0.867	8.203	0.395	0.247	0.892	9.367	0.322	0.183	0.755	12.113	1.053	0.118
$\mu = 1 \times 10^{-5}, \lambda = 10^5$	0.865	5.469	0.341	0.255	0.892	5.469	0.266	0.209	0.776	10.938	0.983	0.133
$\mu = 5 \times 10^{-4}, \lambda = 10^8$	0.873	6.357	0.371	0.175	0.881	10.000	0.392	0.122	0.759	12.113	1.090	0.102
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.625	30.186	2.723	0.000	0.782	17.163	1.088	0.000	0.792	10.742	0.730	0.000
Patient 5												
<i>Initial</i>												
$\mu = 0$	0.781	14.063	0.849	0.481	0.787	12.376	1.102	0.356	0.795	18.281	1.002	0.245
$\mu = 5 \times 10^{-3}, \lambda = 10^8$	0.621	30.784	2.804	0.054	0.732	11.473	1.190	0.024	0.684	29.171	1.882	0.009
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.494	41.516	4.541	0.002	0.769	15.469	1.014	0.000	0.773	28.237	1.322	0.000
Patient 6												
<i>Initial</i>												
$\mu = 0$	0.742	12.305	1.361	0.374	0.765	20.363	1.709	0.422	0.789	10.153	0.897	0.376
$\mu = 5 \times 10^{-3}, \lambda = 10^8$	0.765	20.000	1.304	0.150	0.811	20.000	1.079	0.105	0.816	30.000	1.224	0.091
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.630	14.063	2.014	0.025	0.741	12.225	1.178	0.032	0.714	20.000	1.383	0.024
Patient 7												
<i>Initial</i>												
$\mu = 0$	0.795	10.938	0.956	0.205	0.826	10.000	0.613	0.170	0.739	10.000	0.936	0.091
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.805	5.625	0.583	0.030	0.831	10.174	0.674	0.048	0.738	20.000	1.112	0.027
Patient 8												
<i>Initial</i>												
$\mu = 0$	0.826	10.938	0.707	0.197	0.824	11.269	1.075	0.170	0.799	10.000	0.631	0.108
$\mu = 5 \times 10^{-3}, \lambda = 10^8$	0.824	10.210	0.652	0.166	0.826	20.000	0.943	0.088	0.777	7.813	0.624	0.084
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.782	20.000	0.879	0.045	0.795	20.000	1.033	0.042	0.811	10.000	0.574	0.026
Patient 9												
<i>Initial</i>												
$\mu = 0$	0.790	20.186	1.260	0.553	0.831	10.938	0.618	0.405	0.824	13.487	0.724	0.235
$\mu = 5 \times 10^{-4}, \lambda = 10^5$	0.856	10.938	0.568	0.463	0.864	20.000	0.644	0.287	0.879	8.887	0.344	0.158
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.835	12.934	0.749	0.060	0.856	10.000	0.478	0.059	0.821	13.330	0.547	0.046
Patient 10												
<i>Initial</i>												
$\mu = 0$	0.886	4.688	0.263	0.312	0.788	10.000	0.730	0.259	0.763	11.082	0.875	0.145
$\mu = 5 \times 10^{-1}, \lambda = 10^8$	0.901	9.375	0.250	0.046	0.802	10.000	0.561	0.044	0.801	10.704	0.673	0.023

Table A.1: Selection of the optimal hyperparameters for WarpPINN evaluating on the DCM cohort of the considered ACDC dataset. Green indicates chosen, bold indicates best DSC and underline indicates best $||J|-1|$.

Bibliography

- [1] WHO, *Cardiovascular diseases (CVDs)*, [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), Accessed: 08-Feb-2024.
- [2] R. Luengo-Fernandez, M. Walli-Attaei, A. Gray, *et al.*, “Economic burden of cardiovascular diseases in the European Union: a population-based cost study,” *European Heart Journal*, vol. 44, no. 45, pp. 4752–4767, Aug. 2023, ISSN: 0195-668X. DOI: 10.1093/eurheartj/ehad583. [Online]. Available: <https://doi.org/10.1093/eurheartj/ehad583>.
- [3] J. A. Hill, R. Ardehali, K. T. Clarke, *et al.*, “Fundamental cardiovascular research: Returns on societal investment: A scientific statement from the american heart association,” *Circulation research*, vol. 121, no. 3, e2–e8, 2017.
- [4] S. Vyshnya, R. Epperson, F. Giuste, W. Shi, A. Hornback, and M. D. Wang, “Optimized clinical feature analysis for improved cardiovascular disease risk screening,” *IEEE Open Journal of Engineering in Medicine and Biology*, 2024.
- [5] M. Alsanea and A. K. Dutta, “Detecting cardiovascular diseases from radiographic images using deep learning techniques,” *Expert Systems*, e13542, 2024.
- [6] M. O'Rourke, “Mechanical principles in arterial disease,” *Hypertension*, vol. 26, no. 1, pp. 2–9, 1995.
- [7] R. Wayne Alexander, R. C. Schlant, V. Fuster, *et al.*, *Hurst's the heart*, 2000.
- [8] E.-S. H. M. Ibrahim, *Techniques for cardiac tissue characterization using magnetic resonance imaging*. The Johns Hopkins University, 2008.
- [9] J. A. Case and T. M. Bateman, “Taking the perfect nuclear image: Quality control, acquisition, and processing techniques for cardiac spect, pet, and hybrid imaging,” *Journal of Nuclear Cardiology*, vol. 20, pp. 891–907, 2013.
- [10] B. B. Das, S. Deshpande, and T. Hussain, “Multimodality imaging to detect rejection, and cardiac allograft vasculopathy in pediatric heart transplant recipients—an illustrative review,” *Transplantology*, vol. 3, no. 3, pp. 241–256, 2022.
- [11] H. Kasban, M. El-Bendary, and D. Salama, “A comparative study of medical imaging techniques,” *International Journal of Information Science and Intelligent System*, vol. 4, no. 2, pp. 37–58, 2015.
- [12] A. Bustoquet, J. Oshinski, and O. Škrinjar, “Myocardial deformation recovery from cine mri using a nearly incompressible biventricular model,” *Medical image analysis*, vol. 12, no. 1, pp. 69–85, 2008.
- [13] A. A. Malayeri, W. C. Johnson, R. Macedo, J. Bathon, J. A. Lima, and D. A. Bluemke, “Cardiac cine mri: Quantification of the relationship between fast gradient echo and steady-state free precession for determination of myocardial mass and volumes,” *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 28, no. 1, pp. 60–66, 2008.
- [14] G. C. Fonarow and J. J. Hsu, *Left ventricular ejection fraction: What is “normal”?* 2016.
- [15] A. Scatellia, A. Baritussio, and C. Bucciarelli-Ducci, “Strain imaging using cardiac magnetic resonance,” *Heart failure reviews*, vol. 22, pp. 465–476, 2017.
- [16] P. Bucius, J. Erley, R. Tanacli, *et al.*, “Comparison of feature tracking, fast-senc, and myocardial tagging for global and segmental left ventricular strain,” *ESC heart failure*, vol. 7, no. 2, pp. 523–532, 2020.
- [17] B. S. Selvadurai, V. O. Puntmann, D. A. Bluemke, *et al.*, “Definition of left ventricular segments for cardiac magnetic resonance imaging,” 2017.
- [18] M. A. Pfeffer, A. M. Shah, and B. A. Borlaug, “Heart failure with preserved ejection fraction in perspective,” *Circulation research*, vol. 124, no. 11, pp. 1598–1617, 2019.
- [19] J. V. Hajnal and D. L. Hill, *Medical image registration*. CRC press, 2001.
- [20] P. A. López, H. Mella, S. Uribe, D. E. Hurtado, and F. S. Costabal, “Warppinn: Cine-mr image registration with physics-informed neural networks,” *Medical Image Analysis*, vol. 89, p. 102925, 2023.
- [21] D. L. Hill, C. R. Maurer, A. J. Martin, *et al.*, “Assessment of intraoperative brain deformation using interventional mr imaging,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI'99: Second International Conference, Cambridge, UK, September 19–22, 1999. Proceedings 2*, Springer, 1999, pp. 910–919.
- [22] O. Devinsky, W. Barr, B. Vickrey, *et al.*, “Changes in depression and anxiety after resective surgery for epilepsy,” *Neurology*, vol. 65, no. 11, pp. 1744–1749, 2005.
- [23] J.-P. Thirion and G. Calmon, “Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences,” *IEEE transactions on medical imaging*, vol. 18, no. 5, pp. 429–441, 1999.

- [24] D. Rey, G. Subsol, H. Delingette, and N. Ayache, "Automatic detection and segmentation of evolving processes in 3d medical images: Application to multiple sclerosis," *Medical image analysis*, vol. 6, no. 2, pp. 163–179, 2002.
- [25] P. M. Thompson, J. N. Giedd, R. P. Woods, D. MacDonald, A. C. Evans, and A. W. Toga, "Growth patterns in the developing brain detected by using continuum mechanical tensor maps," *Nature*, vol. 404, no. 6774, pp. 190–193, 2000.
- [26] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Transactions on pattern analysis and Machine Intelligence*, no. 4, pp. 413–424, 1986.
- [27] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
- [28] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 1986, pp. 151–160.
- [29] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast mr images," *IEEE transactions on medical imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [30] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Transactions on medical imaging*, vol. 20, no. 6, pp. 526–534, 2001.
- [31] 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.
- [32] R. Bajcsy and S. Kovačić, "Multiresolution elastic matching," *Computer vision, graphics, and image processing*, vol. 46, no. 1, pp. 1–21, 1989.
- [33] J. A. Schnabel, C. Tanner, A. D. Castellano-Smith, et al., "Validation of nonrigid image registration using finite-element methods: Application to breast mr images," *IEEE transactions on medical imaging*, vol. 22, no. 2, pp. 238–247, 2003.
- [34] P. J. Edwards, D. L. Hill, J. A. Little, and D. J. Hawkes, "A three-component deformation model for image-guided surgery," *Medical Image Analysis*, vol. 2, no. 4, pp. 355–367, 1998.
- [35] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM computing surveys (CSUR)*, vol. 27, no. 3, pp. 433–466, 1995.
- [36] T. Küstner, J. Pan, H. Qi, et al., "Lapnet: Non-rigid registration derived in k-space for magnetic resonance imaging," *IEEE transactions on medical imaging*, vol. 40, no. 12, pp. 3686–3697, 2021.
- [37] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: A review," *Physics in Medicine & Biology*, vol. 65, no. 20, 20TR01, 2020.
- [38] 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.
- [39] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part I*, Springer, 2018, pp. 729–738.
- [40] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.
- [41] L. Hansen and M. P. Heinrich, "Revisiting iterative highly efficient optimisation schemes in medical image registration," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, Springer, 2021, pp. 203–212.
- [42] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into imaging*, vol. 9, pp. 611–629, 2018.
- [43] M. Yani, S. S. M. Budhi Irawan, and C. Setiningsih, "Application of transfer learning using convolutional neural network method for early detection of terry's nail," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1201, 2019, p. 012052.
- [44] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.
- [45] L. Zhang, G. Zhu, P. Shen, J. Song, S. Afaq Shah, and M. Bennamoun, "Learning spatiotemporal features using 3dcnn and convolutional lstm for gesture recognition," in *Proceedings of the IEEE international conference on computer vision workshops*, 2017, pp. 3120–3128.

- [46] X. Han, Y. Zhong, L. Cao, and L. Zhang, “Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification,” *Remote Sensing*, vol. 9, no. 8, p. 848, 2017.
- [47] A. Hering, S. Häger, J. Moltz, N. Lessmann, S. Heldmann, and B. van Ginneken, “Cnn-based lung ct registration with multiple anatomical constraints,” *Medical Image Analysis*, vol. 72, p. 102139, 2021.
- [48] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, Springer, 2015, pp. 234–241.
- [49] J. Rühaak, T. Polzin, S. Heldmann, *et al.*, “Estimation of large motion in lung ct by integrating regularized keypoint correspondences into dense deformable registration,” *IEEE transactions on medical imaging*, vol. 36, no. 8, pp. 1746–1757, 2017.
- [50] M. P. Heinrich, M. Jenkinson, M. Brady, and J. A. Schnabel, “Mrf-based deformable registration and ventilation estimation of lung ct,” *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1239–1248, 2013.
- [51] T. Polzin, J. Rühaak, R. Werner, *et al.*, “Combining automatic landmark detection and variational methods for lung ct registration,” in *Fifth international workshop on pulmonary image analysis*, 2013, pp. 85–96.
- [52] A. Schmidt-Richberg, R. Werner, J. Ehrhardt, J.-C. Wolf, and H. Handels, “Landmark-driven parameter optimization for non-linear image registration,” in *Medical Imaging 2011: Image Processing*, SPIE, vol. 7962, 2011, pp. 263–270.
- [53] M. P. Heinrich, M. Jenkinson, S. M. Brady, and J. A. Schnabel, “Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1–5, 2012, Proceedings, Part III 15*, Springer, 2012, pp. 115–122.
- [54] F. F. Berendsen, A. N. Kotte, M. A. Viergever, and J. P. Pluim, “Registration of organs with sliding interfaces and changing topologies,” in *Medical Imaging 2014: Image Processing*, SPIE, vol. 9034, 2014, pp. 95–101.
- [55] T. Sentker, F. Madesta, and R. Werner, “Gdl-fire: Deep learning-based fast 4d ct image registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 765–773.
- [56] K. A. Eppenhof, M. W. Lafarge, M. Veta, and J. P. Pluim, “Progressively trained convolutional neural networks for deformable image registration,” *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1594–1604, 2019.
- [57] A. Hering, B. van Ginneken, and S. Heldmann, “Mlvirnet: Multilevel variational image registration network,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, Springer, 2019, pp. 257–265.
- [58] L. Hansen and M. P. Heinrich, “Tackling the problem of large deformations in deep learning based medical image registration using displacement embeddings,” *arXiv preprint arXiv:2005.13338*, 2020.
- [59] Z. Jiang, F.-F. Yin, Y. Ge, and L. Ren, “A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration,” *Physics in Medicine & Biology*, vol. 65, no. 1, p. 015011, 2020.
- [60] Y. Fu, Y. Lei, T. Wang, *et al.*, “Lungregnet: An unsupervised deformable image registration method for 4d-ct lung,” *Medical physics*, vol. 47, no. 4, pp. 1763–1774, 2020.
- [61] L. Hansen and M. P. Heinrich, “Graphregnet: Deep graph regularisation networks on sparse keypoints for dense registration of 3d lung cts,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2246–2257, 2021.
- [62] R. Castillo, E. Castillo, R. Guerra, *et al.*, “A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets,” *Physics in Medicine & Biology*, vol. 54, no. 7, p. 1849, 2009.
- [63] R. Castillo, E. Castillo, D. Fuentes, *et al.*, “A reference dataset for deformable image registration spatial accuracy evaluation using the copdgene study archive,” *Physics in Medicine & Biology*, vol. 58, no. 9, p. 2861, 2013.
- [64] J. M. Wolterink, J. C. Zwienenberg, and C. Brune, “Implicit neural representations for deformable image registration,” in *International Conference on Medical Imaging with Deep Learning*, PMLR, 2022, pp. 1349–1359.
- [65] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [66] 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*, vol. 38, no. 8, pp. 1788–1800, 2019.
- [67] 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, 2016.
- [68] T. Fechter and D. Baltas, “One-shot learning for deformable medical image registration and periodic motion tracking,” *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2506–2517, 2020.

- [69] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.
- [70] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.
- [71] Z. Li, N. Kovachki, K. Azizzadenesheli, *et al.*, “Fourier neural operator for parametric partial differential equations,” *arXiv preprint arXiv:2010.08895*, 2020.
- [72] J. Winkens, J. Limmans, B. S. Veeling, T. S. Cohen, and M. Welling, “Improved semantic segmentation for histopathology using rotation equivariant convolutional networks,” 2018.
- [73] J. M. Guccione and A. D. McCulloch, “Mechanics of active contraction in cardiac muscle: Part i—constitutive relations for fiber stress that describe deactivation,” 1993.
- [74] R. S. Rivlin and D. Saunders, “Large elastic deformations of isotropic materials vii. experiments on the deformation of rubber,” *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 243, no. 865, pp. 251–288, 1951.
- [75] O. Bernard, A. Lalande, C. Zotti, *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.
- [76] S. E. Petersen, N. Aung, M. M. Sanghvi, *et al.*, “Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (cmr) in caucasians from the uk biobank population cohort,” *Journal of cardiovascular magnetic resonance*, vol. 19, no. 1, p. 18, 2016.
- [77] N. Rahaman, A. Baratin, D. Arpit, *et al.*, “On the spectral bias of neural networks,” in *International Conference on Machine Learning*, PMLR, 2019, pp. 5301–5310.
- [78] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [79] M. Tancik, P. Srinivasan, B. Mildenhall, *et al.*, “Fourier features let networks learn high frequency functions in low dimensional domains,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.
- [80] S. Wang, H. Wang, and P. Perdikaris, “On the eigenvector bias of fourier feature networks: From regression to solving multi-scale pdes with physics-informed neural networks,” *Computer Methods in Applied Mechanics and Engineering*, vol. 384, p. 113938, 2021.
- [81] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.
- [82] C. Tobon-Gomez, M. De Craene, K. Mcleod, *et al.*, “Benchmarking framework for myocardial tracking and deformation algorithms: An open access database,” *Medical image analysis*, vol. 17, no. 6, pp. 632–648, 2013.
- [83] C. G. Fonseca, M. Backhaus, D. A. Bluemke, *et al.*, “The cardiac atlas project—an imaging database for computational modeling and statistical atlases of the heart,” *Bioinformatics*, vol. 27, no. 16, pp. 2288–2295, 2011.
- [84] O. C. E. Konukoglu, M. P. K. Rhode, and M. S. A. Young, “Statistical atlases and computational models of the heart,”
- [85] M. De Craene, C. Tobon-Gomez, C. Butakoff, *et al.*, “Temporal diffeomorphic free form deformation (tdffd) applied to motion and deformation quantification of tagged mri sequences,” in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges: Second International Workshop, STACOM 2011, Held in Conjunction with MICCAI 2011, Toronto, ON, Canada, September 22, 2011, Revised Selected Papers 2*, Springer, 2012, pp. 68–77.
- [86] 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, pp. 92–111, 2011.
- [87] M. A. Morales, M. Van den Boomen, C. Nguyen, *et al.*, “Deepstrain: A deep learning workflow for the automated characterization of cardiac mechanics,” *Frontiers in Cardiovascular Medicine*, vol. 8, p. 730316, 2021.
- [88] P. Hofrichter, A. Hagendorff, U. Laufs, *et al.*, “Analysis of left ventricular rotational deformation by 2d speckle tracking echocardiography: A feasibility study in athletes,” *The International Journal of Cardiovascular Imaging*, vol. 37, pp. 2369–2386, 2021.
- [89] S. Valette, J. M. Chassery, and R. Prost, “Generic remeshing of 3d triangular meshes with metric-dependent discrete voronoi diagrams,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 2, pp. 369–381, 2008.
- [90] T. Joyce, S. Buoso, C. T. Stoeck, and S. Kozerke, “Rapid inference of personalised left-ventricular meshes by deformation-based differentiable mesh voxelization,” *Medical image analysis*, vol. 79, p. 102445, 2022.

-
- [91] C. Qin, W. Bai, J. Schlemper, *et al.*, “Joint learning of motion estimation and segmentation for cardiac mr image sequences,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11*, Springer, 2018, pp. 472–480.
 - [92] C. Qin, S. Wang, C. Chen, H. Qiu, W. Bai, and D. Rueckert, “Biomechanics-informed neural networks for myocardial motion tracking in mri,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23*, Springer, 2020, pp. 296–306.
 - [93] C. Qin, S. Wang, C. Chen, W. Bai, and D. Rueckert, “Generative myocardial motion tracking via latent space exploration with biomechanics-informed prior,” *Medical Image Analysis*, vol. 83, p. 102682, 2023.
 - [94] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.
 - [95] V. A. Zimmer, K. Hammernik, V. Sideri-Lampretsa, *et al.*, “Towards generalised neural implicit representations for image registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2023, pp. 45–55.