Journal of Machine Learning for Biomedical Imaging 2024:031 vol. 2, pp. 2293–2325
Submitted 03/2024
Published 11/2024

Echocardiography Analysis with Deep Learning using Priors: Multi-centric Evaluation of Generalisation

Yingyu Yang https://orcid.org/0009-0009-5700-054X yingyu.yang@inria.fr
Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France

Marie Rocher rocher.m@chu-nice.fr
CHU de Nice - Hôpital Pasteur, Nice, France

Pamela Moceri https://orcid.org/0000-0002-0741-1609 moceri.p@chu-nice.fr
CHU de Nice - Hôpital Pasteur, Nice, France

Maxime Sermesant https://orcid.org/0000-0002-6256-8350 maxime.sermesant@inria.fr
Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France

Abstract

Echocardiography is a widely used medical modality to evaluate cardiac function. Automatic solutions for 2D echocardiography analysis could greatly influence cardiac disease diagnosis and prognosis. Among its diverse analysis tasks, segmentation and motion estimation are critical. Segmentation provides insights into cardiac structures, while motion estimation depicts myocardial deformation during cardiac cycles via image registration. These two tasks help extract useful clinical indexes, including ejection fraction and heart wall strain. Deep learning-based methods that aim to tackle each of the above tasks have demonstrated efficient performance. However, the predominant focus on a single research dataset raises concerns about the generalisation of these approaches to other datasets and their applicability to downstream clinical tasks, such as diagnosis and prognosis. In this paper, we propose a comprehensive echocardiography analysis pipeline, consisting of a segmentation module, a motion tracking module, and a classifier. The segmentation module incorporates a shape prior, leveraging a contour-aware loss function for the endocardium and epicardium. With appropriate data augmentation strategies, this module consistently demonstrates robust segmentation and ejection fraction estimation performance across diverse datasets from distinct medical centres. The motion tracking module integrates a motion prior for the myocardium through a differentiable poly-affine transformation, ensuring smooth and anatomy-aware deformation estimation. This module exhibits improved registration performance and ensures greater consistency in the deformation field across datasets. Finally, the classifier leverages features extracted from the preceding segmentation and motion tracking modules for downstream classification, providing interpretable inputs. The segmentation and motion tracking models undergo individual training on two publicly available datasets. Subsequently, the classifier undergoes further training on a third pathology dataset using the estimated features from the previously trained models. To evaluate real-world performance, we directly apply the three trained models to a clinical dataset from a local hospital. Our successful application of these proposed methods in detecting patients with myocardial infarction (an F1 score of 0.84 and an accuracy of 0.77), a life-threatening cardiac disease, highlights their potential clinical utility. Our codes are available at `https://gitlab.inria.fr/epione/echoanalysis`.

Keywords: Image segmentation, Image registration, Generalisation of deep learning methods, priors, Echocardiography analysis

https://doi.org/10.59275/j.melba.2024-58fd

1. Introduction

According to the 2023 annual report of the World Heart Organisation, cardiovascular diseases (CVD) accounted for nearly one third of all global deaths in 2021 (Federation, 2023). Compared to high-income countries, more than 80% of CVD-related deaths occurred in low and middle-income countries. The inequitable distribution of CVD diagnosis and treatment resources still remains unresolved. Addressing the prevention and management of CVD patients requires effective diagnostic methods, with point-of-care ultrasound proving to enhance the diagnostic landscape (King et al., 2016). However, the intricacies of echocardiography image analysis demand extensive medical training and experience from cardiologists to provide reliable diagnoses. Unfortunately, the unbalanced distribution of cardiovascular medical professionals also hinders the CVD diagnostics and management (Narang et al., 2016). In the face of such challenge, automatic cardiac function analysis emerges as a potential solution. This approach not only improves the management workflow for cardiologists but also enables trainees/nurses to help, particularly in emergencies and under-served regions. Recent advances in deep learning have significantly enhanced the efficiency of automatic cardiac function analysis across both segmentation tasks (Jafari et al., 2019; Arafati et al., 2020; Leclerc et al., 2019b) and registration tasks (Ahn et al., 2020; Ta et al., 2020; Østvik et al., 2021). Despite these advancements, it is recognised that deep learning models often face over-fitting issues; thus, generalisable solutions are desired. Moreover, most of the existing literature predominantly concentrates on enhancing individual tasks within a single dataset. Generalisation of these approaches, along with studying its impact on downstream task (e.g. patient classification), is crucial in the domain of 2D echocardiography analysis.

In this paper, we propose a comprehensive deep-learning based echocardiography analysis pipeline, comprised of a segmentation module, a motion tracking module and a classifier (Figure 1(a)). For the first two modules, we consider incorporating prior knowledge with the intention to enhance their robustness against unseen data. Various means exist to integrate such prior knowledge, encompassing: data augmentation; output parameterisation; and, loss functions. To cultivate more universally applicable AI solutions for echocardiography-based cardiac function analysis, we introduce the following propositions in this paper:

- Segmentation module: a shape prior is achieved through a contour-based loss function as well as a stack of echo-specific data augmentation strategies
- Motion tracking module: a motion prior is incorporated via a poly-affine transformation framework for weakly-supervised motion estimation in echocardiography

These two modules are trained individually on two public datasets respectively and we evaluate their performance across public datasets from different centres. Subsequently, the classifier is cross-validated on another pathology dataset using features extracted by the former trained models as input to aid in myocardial infarction patient detection (Figure 1(b)).

Furthermore, we present findings from a preliminary pilot study conducted at a local hospital. This real-world evaluation allows us to monitor the proposed tool performance in practical scenarios and aligns with our commitment to rigorous assessment.

This paper builds upon our preliminary work, which introduced a segmentation method at the FIMH 2021 conference (Yang and Sermesant, 2021), and a motion estimation method

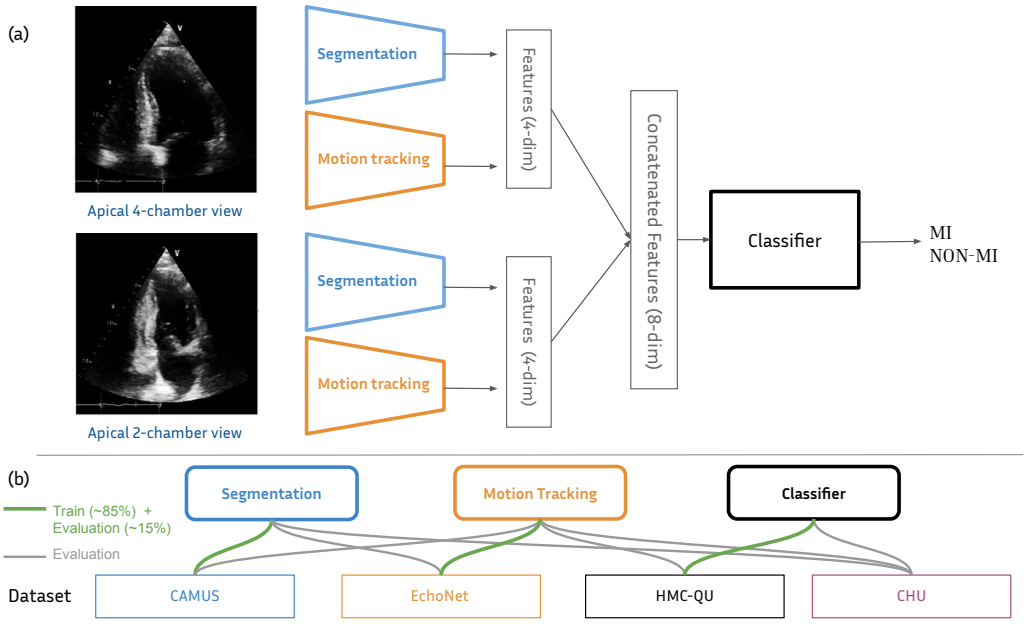

Figure 1: (a) Proposed multi-view echocardiography analysis pipeline. (b) Segmentation model and motion tracking model are first trained using CAMUS and EchoNet-Dynamic dataset respectively (green line) and are used to extract clinical features for classifier training, which employs the HMC-QU dataset for training. Each step of the pipeline are also evaluated on unseen datasets (gray line) for generalisation assessment. Detailed dataset information can be found in Table 1 and Section 4.1).

presented at the FIMH 2023 conference (Yang and Sermesant, 2023). Building upon the initial findings from (Yang and Sermesant, 2021, 2023), our current research introduces the following novel contributions:

- We enhance the generalisation evaluation process by encompassing more diverse datasets and comparing our results with a wider range of state-of-the-art methods for segmentation and motion tracking respectively. In addition, for motion tracking task, we further propose an incompressibility regularisation term for more accurate and realistic cardiac motion estimation.

- We introduce a classification task based on the output of former-proposed segmentation and motion tracking methods for downstream task evaluation.

- We design a comprehensive echocardiography analysis pipeline, showcasing its clinical potential through the detection of myocardial infarction in 2D echocardiography data on public research data as well as on real-world clinical data.

2. Related Works

2.1 Segmentation in Echocardiography

Deep learning methods have significantly advanced in the medical image segmentation domain, including automatic segmentation in 2D echocardiography. The U-Net architecture and its variants (Leclerc et al., 2019b; Ling et al., 2022) have demonstrated exceptional performance. In order to tackle the incorrect anatomy problem in cardiac segmentation, researchers have proposed regularisation techniques and refinement models (Oktay et al., 2017; Leclerc et al., 2019a). Further, to incorporate temporal information into the segmentation process, Wei et al. (Wei et al., 2020) proposed two co-learning strategies of parallel segmentation and motion estimation from 2D echocardiography. Painchaud et al. (Painchaud et al., 2022) introduced a post-processing auto-encoder that corrects the temporal inconsistency of sequence segmentation outputs.

One crucial application of cardiac segmentation is the estimation of left ventricular ejection fraction (LVEF), a fundamental cardiac measurement in assessing cardiac systolic function. LVEF plays an important role in accessing the severity of systolic dysfunction and in management of various cardiovascular diseases (Folland et al., 1979). The quantification of LVEF involves the employment of the bi-plane or single-plane Simpson's method (Folland et al., 1979) for volume assessment, thus is sensitive to image quality and segmentation process. Automatic identification of end-diastole (ED) and end-systole (ES) frames (Smistad et al., 2020; Leclerc et al., 2019b) are also necessary. An alternative approach for LVEF estimation involves the analysis of entire echocardiography sequences. This includes methods employing recurrent neural networks or transformer architectures to predict LVEF (Kazemi Esfeh et al., 2020; Reynaud et al., 2021).

For cardiac segmentation, our objective centres on creating a straightforward yet anatomically robust model. We employ a U-Net architecture, enhancing it by integrating a shape prior within the loss function. As for the LVEF estimation, we rely on the segmentation mask, prioritising its interpretability in our approach.

2.2 Motion estimation in Echocardiography

Traditional motion estimation methods such as block matching (Touil et al., 2010; Boukerroui et al., 2003), optical-flow (Ahn, 2013) and non-rigid registration (Vercauteren et al., 2008; Chakraborty et al., 2016) usually demand heavy computations. Recent works using deep learning models for echocardiography motion estimation have achieved good trade-off between run time and accuracy (Ahn et al., 2020; Ta et al., 2020; Østvik et al., 2021). Unsupervised motion estimation relies solely on image pairs (or annotated masks) for training (Ahn et al., 2020; Ta et al., 2020; Balakrishnan et al., 2019; Wang et al., 2022). In contrast, supervised motion estimation employs dense displacement fields from ground truth data in order to penalise predicted deformation fields (Østvik et al., 2021). However, obtaining accurate ground-truth myocardial displacement is challenging, often necessitating the use of synthetic echocardiography images with known motion (Alessandrini et al., 2018; Evain et al., 2022). It remains unclear how synthetic data aligns with *in vivo* data due to potential domain shifts (Deng et al., 2022).

From deformation field, we can obtain another important index of cardiac function: the global longitudinal strain (GLS). GLS reveals cardiac contractility and is a robust prognostic parameter for LV strain. It has been reported to be more sensitive in detecting early cardiac diseases than LVEF (Cikes and Solomon, 2016). Computation of GLS can stem from cardiac landmarks (Smistad et al., 2020; Østvik et al., 2021) or dense deformation fields (Morales et al., 2021). The strain tensor sensitivity to the smoothness of the deformation field necessitates regularisation in deep learning models. Common strategies include Gaussian smoothing layers (Krebs et al., 2020) or rigidity penalties (Staring et al., 2007; De Vos et al., 2019). However, these techniques often apply within limited pixel neighbourhoods in the image domain and might not consistently adhere to anatomical constraints. In our approach, we address this issue through a poly-affine motion framework, which serves as a well-established motion approximation for myocardial movement (McLeod et al., 2015). Thus, by integrating a motion prior, we achieve the capability to track myocardial motion with a small number of parameters while respecting the anatomy of the heart.

2.3 Myocardial infarction detection in echocardiography

Myocardial infarction (MI), also known as heart attack, is the most significant contributor to sudden cardiac death. Myocardial infarction happens when there is lack of oxygen in the heart muscle due to the blockage in heart arteries. 2D echocardiography is one of the clinical tools used to detect MI by evaluating motion of the myocardium. Deep learning and machine learning based classification methods have demonstrated good performance in myocardial infarction (MI) detection using 2D echocardiography. Convolutional neural networks are commonly employed to extract deep features from 2D echocardiography images for classification purposes (Kusunose et al., 2020; Huang et al., 2020; Omar et al., 2018). Additionally, handcrafted features, including frequency-related attributes, have been extracted from 2D echocardiography images to aid in MI detection (Liu et al., 2023; Raghavendra et al., 2018). However, much of the existing research relies on training models with local large datasets, posing limitations when it comes to external validation by other researchers. A notable exception is the work of Degerli et al. (2024), who introduced a publicly available dataset for MI detection and assessed its performance using segment displacement features. Given the relatively small size of this dataset (130 patients), we opt to extract interpretable features from echocardiography using segmentation and motion estimation trained models. Subsequently, we employ machine learning techniques for MI detection.

3. Methods

In this section, we begin by the introduction of the shape prior (Sec.3.1) and motion prior (Sec.3.2) for the segmentation and motion estimation models, respectively. Subsequently, we delve into how these models can help build classifiers for multi-view myocardial infarction detection (Sec.3.3).

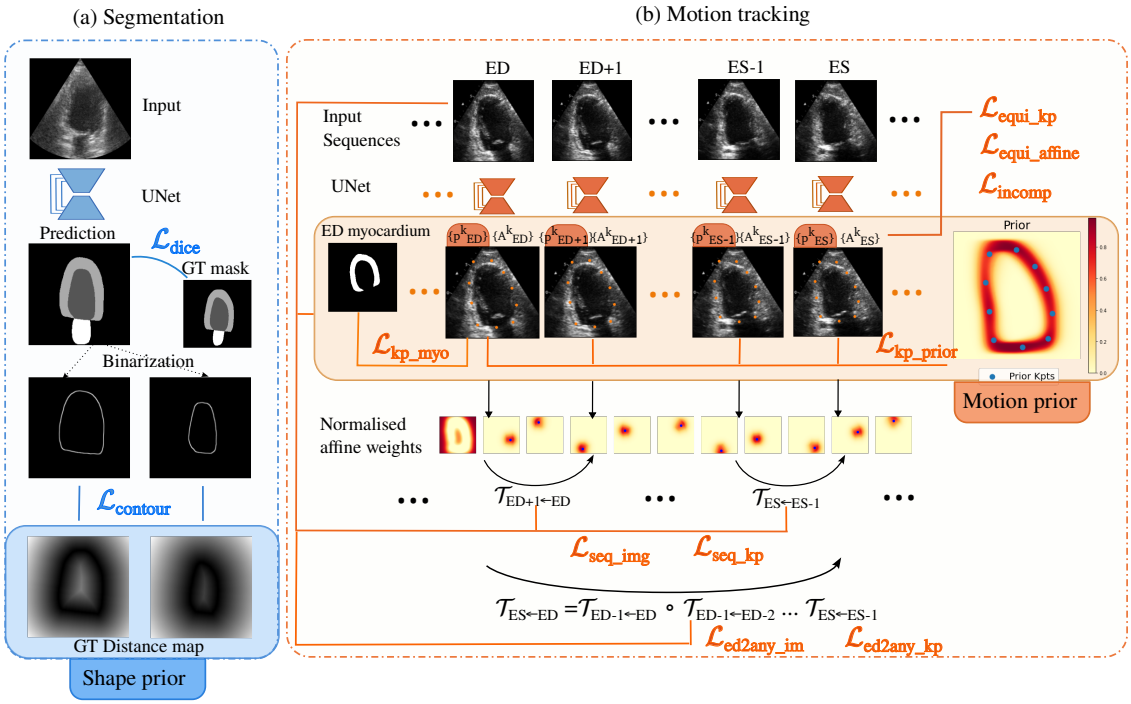

Figure 2: Shape and motion priors for echocardiography analysis. (a) The shape prior refers to the application of contour loss which maximises the matching of ground truth distance map with the contour of predicted segmentation mask (detailed in Section 3.1). (b) The motion prior refers to the precomputed key-point prior as well as the regularisation terms related to the myocardium key points (detailed in Section 3.2.1).

3.1 Left-ventricle segmentation

3.1.1 Shape prior

In order to detect the left-ventricle, we use U-Net (Ronneberger et al., 2015) as backbone network, where the shape prior is introduced as part of the loss function, noted as the contour loss. The contour loss was first proposed to segment the 3D left-atrium from CT images (Jia et al., 2019) and preliminary results on echocardiography were presented in (Yang and Sermesant, 2021). It was observed that this loss produces smoother shapes with less outliers. The idea of contour loss is to penalise the distance from predicted border pixels to ground-truth border pixels.

$$\mathcal{L}_{\text{contour}} = \sum (D(M) \circ contour(B(P))) \tag{1}$$

where $\circ$ performs element-wise multiplication. $D(M)$ represents the distance map computed from ground-truth mask, which demonstrates the shortest distance between all pixels and the mask contour. P refers to the Softmax output of U-Net. $B(P)$ stand for a differentiable

thresholded Sigmoid function for binarisation $B(P) = \frac{1}{1+\exp^{-\gamma(P-T)}}$, where $\gamma = 20$ and $T = 0.5$.

The contour of the binarised mask is obtained by applying a 2D Sobel filter

$$contour(P) = |G_x * P| + |G_y * P| \tag{2}$$

where $*$ denotes a 2D convolution and G_x, G_y are 2D Sobel kernels in x-,y- dimension:

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \tag{3}$$

In echocardiography analysis, it is important to achieve good endocardial and epicardial contours, as they play an important role in evaluating the volume and myocardial motion. We apply the contour loss for the contour of endocardium (Endo) and epicardium (Epi) respectively (illustration in Figure 2). The final loss function for optimisation is defined as:

$$\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{dice}} + \beta * \mathcal{L}_{\text{contour}} \tag{4}$$

where $\mathcal{L}_{\text{dice}}$ refers to the multi-class Dice loss.

3.1.2 Data augmentation

In order to improve the generalisation of image processing methods, we apply an ensemble of augmentation strategies to the original training dataset. Specifically the implemented augmentation consists of the following related changes, designed for echocardiography (adapted from (Zhang et al., 2020)):

- **Rotation**: Rotation is usually caused by different probe orientation and we set the rotation range to [-15,15] degrees.
- **Crop and Scaling**: Scaling is due to the heart shape variability and myocardium motion. We set the scaling factor to [0.7,1.3] and then crop (or pad) the resized image to be the same size.
- **Brightness adjustment**: Image intensity characteristics often result from device vendors and image system settings. For brightness adjustment, we shift the histogram of image intensities by factor of [0.9,1.1].
- **Contrast adjustment**: Contrast adjustment is a popular tool to improve the visualisation of cardiac structures. We apply a Sigmoid correction $I_{new} = \frac{1}{1+\exp C*(S-I)}$ on the input image with a random range of cutoff S in [0.4,0.6] and a constant multiplier C from [4,10].
- **Sharpness adjustment**: To modify the sharpness of the image, we apply an unsharp mask $I_{new} = I + A * (I - I_\sigma)$, where A ranges from [1,2] and the σ of the Gaussian filter from [0.25,1.5].
- **Blurriness adjustment**: For image blurring, we apply a Gaussian filter with σ in (0.25,1.5).

- **Speckle noise**: Due to the acoustic interference phenomenon, speckle is the main kind of noise present in echocardiography. Speckle is usually modelled as a Rayleigh multiplicative noise. Since the training data is already log-compressed, we add speckle noise following an additive fashion $I_{new} = I + \sqrt{I} * G_\sigma$, where G is a multiplicative Gaussian noise with σ in range [0.01,0.1].

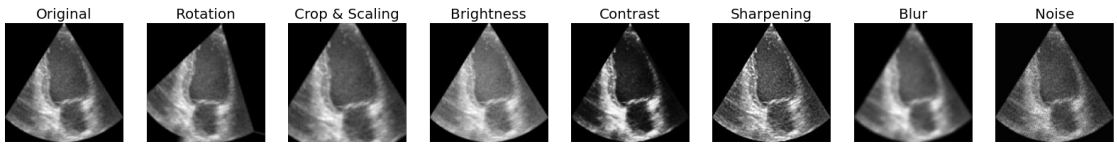

Figure 3: Data augmentation examples.

3.2 Polyaffine transformation for motion estimation

We estimate the motion of the myocardium in 2D echocardiography through a non-rigid image registration task. For a source image I_S and a target image I_T, our aim is to estimate a dense motion field $\mathcal{T}_{S\leftarrow T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that

$$I_S(\mathcal{T}_{S\leftarrow T}(\cdot)) = I_T(\cdot). \tag{5}$$

Instead of estimating the dense motion field directly from neural networks, we parameterise the motion field with a poly-affine transformation. We also introduce a motion prior in the loss function to guide the network to focus on the myocardium region. Our work was inspired by the poly-affine motion fusion framework proposed by Arsigny et al. (2009) and the motion estimation module from Siarohin et al. (2019).

We adapt the network architecture presented in (Siarohin et al., 2019) to obtain the key points position and their corresponding local affine transformation. From predefined local effective region and weights, we obtain the final dense motion field by poly-affine motion fusion.

3.2.1 Key points and affine transformations estimation

We assume that there exists an abstract reference frame R as introduced in (Siarohin et al., 2019). Given an input image $X, X \in \{S, T\}$, the encoder-decoder network predict the estimated key points $p_X^k, k = 1, 2, ..., K$, as well as the corresponding linear mappings $A_{X\leftarrow R}^k \in \mathbb{R}^{2\times 2}, k = 1, 2, ..., K$. The local transformation from target image I_T to source image I_S is modelled by applying an affine transformation passing through the reference frame R:

$$\mathcal{T}_{S\leftarrow T}^k(z) = \underbrace{\bar{A}_{S\leftarrow T}^k \cdot z}_{\text{Linear mapping}} + \underbrace{(p_S^k - \bar{A}_{S\leftarrow T}^k \cdot p_T^k)}_{\text{Translation}}, \tag{6}$$

where $z \in \mathbb{R}^2$ represents the coordinate in the target image, and $\bar{A}_{S\leftarrow T}^k = A_{S\leftarrow R}^k (A_{T\leftarrow R}^k)^{-1}$.

Motion prior In order to capture motion around the myocardium, we introduce a myocardium key-point prior. The prior is computed using myocardium masks from the training

dataset, no matter whether they come from ground-truth segmentation masks or from generated pseudo-masks (Yang and Sermesant, 2023). We calculate a mean mask from available myocardium masks and cluster all the pixels of the mean mask into K groups using a K-Means algorithm implemented by Scikit-learn (Pedregosa et al., 2011). The centre of each cluster is regarded as one myocardium key-point prior (illustrated in Figure 2).

The myocardium key-point prior is integrated in two loss functions to enforce the position of key points close to the myocardium region. The first loss $\mathcal{L}_{\text{kp\_prior}}$ minimises the L_2 norm between the predicted key-point position and the prior key-point position. This ensures that the order of predicted key points would be consistent across different patients. The second loss $\mathcal{L}_{\text{kp\_myo}}$ constrains key points to lie within a known myocardium region (either within ground truth mask or pseudo mask) by penalising the mapping between the heat-map of each key point and the myocardium mask.

$$\mathcal{L}_{\text{kp\_myo}} = -\sum_{k=1}^{K} H(p^k, \sigma_H^2) * (\text{Mask}_{\text{myo}} - 0.5), \tag{7}$$

where $H(p^k, \sigma_H^2)$ is designed as the isotropic Gaussian heat-map centred at p^k, with variance σ_H^2 and Mask_{myo} the binary mask of myocardium available. This loss help the network to predict patient-specific key point positions.

In contrast from the network in (Siarohin et al., 2019) which learned the position of keypoints in a self-supervised way, the integration of myocardium prior in our work encourages the network to learn key points close to the myocardium. Such approach would be much easier for clinicians to interpret.

Incompressibility penalisation In order to constrain the compressibility of the myocardium within physiological ranges, we also apply an incompressibility penalisation to the deformation field. Different from former work that applied divergence-free regularisation to the dense motion field (Mansi et al., 2011; Ahn et al., 2020), here we are able to efficiently constrain the Jacobian determinant of each local affine transformation to be close to 1:

$$\mathcal{L}_{\text{incomp}} = \sum_{k=1}^{K} (|\det(A^k_{X \leftarrow R})| - 1)^2 \tag{8}$$

3.2.2 Polyaffine motion fusion

The final dense motion field is the sum of weighted local transformations. Specifically, we define a 2D isotropic Gaussian spatial weight $W_k(p_T^k, \sigma^2) : z \rightarrow \exp^{-\frac{(z-p_T^k)^2}{2\sigma^2}}$ centred at each key point p_T^k with variance σ^2. The weight W_0 for zero-motion region is the residual of other local weights:

$$W_0 = \textbf{RELU}(1 - \sum_{k=1}^{K} W_k(p_T^k, \sigma^2)). \tag{9}$$

All weights are normalised $\bar{W}_k = \frac{W_k}{\sum_{k=0}^{K} W_k}$, where $k = 0, 1, 2...K.$, and are used for dense motion field computation:

$$\mathcal{T}_{S \leftarrow T}(z) = \bar{W}_0 z + \sum_{k=1}^{K} \bar{W}_k(p_T^k) \mathcal{T}^k_{S \leftarrow T}(z). \tag{10}$$

The original motion module in (Siarohin et al., 2019) estimated the normalised spatial weights by deploying a second encoder-decoder network in a self-supervised way. However, it could not guarantee the predicted weight close to the corresponding key point. Our design of 2D Isotropic Gaussian weights is consistent with the predicted key point position, thereby providing more stable and explainable parameterisation for myocardial motion.

Registration losses In order to learn the motion field from images in an unsupervised way, we optimise the following terms that are related to image similarity and key-point similarity. First, for each pair of consecutive frames in the input sequence, we constrain the following similarity terms:

$$\mathcal{L}_{\text{seq\_im}} = \sum_{seq} NCC(I_T, I_S(\mathcal{T}_{I_S \leftarrow I_T})), \tag{11}$$

$$\mathcal{L}_{\text{seq\_kp}} = \sum_{seq} \frac{1}{K} \sum_{k=1}^{K} |H(p_T^k, \sigma_H^2) - H(p_S^k, \sigma_H^2)(\mathcal{T}_{I_S \leftarrow I_T})|, \tag{12}$$

where NCC represents the normalised cross correlation as image similarity metric.

Second, in order to force the model to learn large motion (such as the cardiac motion between end-diastole and end-systole), we apply a second pair of registration losses between the image at end-diastole and other frames including the end-systole one, denoted as $\mathcal{L}_{\text{ED2any\_im}}$ and $\mathcal{L}_{\text{ED2any\_kp}}$ for image similarity and key-point similarity, respectively.

3.2.3 Loss functions

We train the motion estimation model in an end-to-end manner using a combination of loss functions: keypoint myocardium losses; registration losses; incompressibility loss; and, the following keypoint equivalence losses.

Equivalence losses As proposed in (Siarohin et al., 2019), we also include equivalence losses $\mathcal{L}_{\text{equi\_kp}}$ and $\mathcal{L}_{\text{equi\_affine}}$. Equivalence losses constrain the detected key points and the local affine transformations to be consistent when applying a known transformation (either an affine transform or a thin plane transform) to the input image. Given an image X and a known spatial transformation $\mathcal{T}_{X \leftarrow Y}$, we obtain a new image Y transformed from X. Key points and local affine transformation of images X, Y predicted by the motion estimation network should preserve the following relationship:

$$\begin{aligned} p_X^k &= \mathcal{T}_{X \leftarrow Y}(p_Y^k), k = 1, 2, ..., K \\ A_{X \leftarrow R}^k &= A_{X \leftarrow Y}^k A_{Y \leftarrow R}^k, k = 1, 2, ..., K \end{aligned} \tag{13}$$

, where $A_{X \leftarrow Y}^k$ is the local gradient around p_Y^k calculated from $\mathcal{T}_{X \leftarrow Y}$.

The final loss function involved for the polyaffine motion model, is as follows:

$$\begin{aligned} \mathcal{L}_{\text{total}} =& \lambda_1 \mathcal{L}_{\text{seq\_im}} + \lambda_2 \mathcal{L}_{\text{seq\_kp}} \\ &+ \lambda_3 \mathcal{L}_{\text{equi\_kp}} + \lambda_4 \mathcal{L}_{\text{equi\_affine}} \\ &+ \lambda_5 \mathcal{L}_{\text{kp\_prior}} + \lambda_6 \mathcal{L}_{\text{kp\_myo}} \\ &+ \lambda_7 \mathcal{L}_{\text{ED2any\_im}} + \lambda_8 \mathcal{L}_{\text{ED2any\_kp}} \\ &+ \lambda_9 \mathcal{L}_{\text{incomp}} \end{aligned} \tag{14}$$

3.3 Multi-view myocardial infarction detection

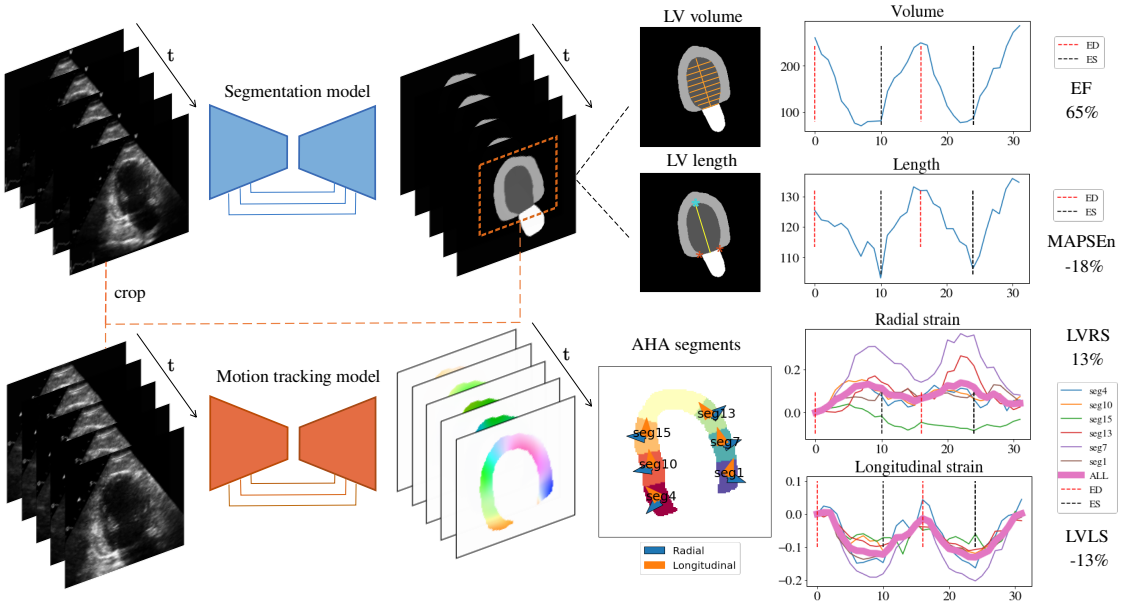

Figure 4: Pipeline for interpretable cardiac features extraction. For segmentation task, a sequence of echocardiography images are processed by the segmentation model and we compute the temporal left ventricle volume change and left ventricle length change. Ejection fraction (EF) and MAPSEn are extracted after identifying the end-diastole and end-systole frames. For motion tracking task, we obtain the deformation motion field from the input sequence (which was cropped according to the left ventricle segmentation) and identify the 7-AHA segments. Averaged strain along radial direction and longitudinal direction can be calculated.

In our study, we choose to formulate an 8-dimension feature vector for multi-view myocardial infarction classification. Specifically, the 8-dimension feature consists of a concatenation of two 4-dimension vectors extracted from apical 2-chamber view and apical 4-chamber view, respectively. Each 4-dim feature is comprised of features that globally measures cardiac function, including left ventricle ejection fraction (LVEF), left ventricle longitudinal strain (LVLS) at end-systole, left ventricle radial strain (LVRS) at end-systole and normalised mitral annular plane systolic excursion (MAPSEn) estimated from each view. LVEF and nMAPSE are two measurements for systolic function evaluation, which play important role in cardiac disease detection. LVLS and LVRS reflect the global contractility of left ventricle and may reveal the abnormality of cardiac wall, which is usually present in myocardial infarction patients. The clinical GLS is estimated from 4-chamber, 3-chamber and 2-chamber views jointly. Here we only have 2 views so we use LVLS and LVRS to refer to the averaged strain from each corresponding view. The pipeline is depicted in Figure 4. Once the 8-dimension feature vector is formed, a classifier is trained to predict the classification result: MI or non-MI (Figure 1(a)). Two popular machine learning

methods from the Scikit-Learn package (Pedregosa et al., 2011) are used for classification: random forest (RF) and support vector machine (SVM), respectively.

3.3.1 Ejection fraction

We apply the modified Simpson's rule (Folland et al., 1979) to the estimated left ventricle segmentation output to obtain an approximated volume curve along cardiac sequences (refer to subfigure entitled 'LV volume' in Figure 4). The end-diastole and end-systole frames are identified as the frames when the volume curve is at its largest and smallest point, respectively. Ejection fraction is then calculated using the end-diastole volume (EDV) and end-systole volume (ESV): $\text{EF} = 1 - \frac{ESV}{EDV}$.

3.3.2 Mitral annular plane systolic excursion (MAPSE)

MAPSE measures the displacement of mitral annular plane at systolic phase, a helpful cardiac index for LV systolic function evaluation. MAPSEn normalises MAPSE for LV length (Støylen et al., 2018).

$$\text{MAPSEn} = \frac{\text{MAPSE}}{L_{ED}} - 1 = \frac{L_{ES} - L_{ED}}{L_{ED}} \tag{15}$$

where L_t refers to the LV length at time t. We compute the LV length by connecting the apex and mid-basal points from segmentation outputs (refer to subfigure entitled 'LV length' in Figure 4).

3.3.3 Myocardial strain

We take the myocardium segmentation output obtained for end-diastole as the region of interest (ROI), and further obtain the pixel displacement within this ROI from motion estimation output across all cardiac sequences. For all myocardial pixels at end-diastole, we compute the 2D Lagrangian finite strain tensor $\mathbf{E}$ using:

$$\mathbf{E} = \frac{1}{2}(\mathbf{F}^T\mathbf{F} - \mathbf{I}), \tag{16}$$

where $\mathbf{I}$ represents the identity matrix and $\mathbf{F}$ refers to the deformation gradient tensor. The radial direction is computed as the tangent field of the correspondence trajectories between endocardial contour and epicardial contour (Yezzi and Prince, 2003). The longitudinal direction is taken perpendicular to the radial direction. We compute segment strain using AHA segment protocol (refer to sub-figure entitled 'AHA segments' in Figure 4) and the global strain is averaged across segments except for apex region. Strains at end-systolic phase are extracted as features for myocardial infarction detection.

4. Experiments

4.1 Datasets

Three public datasets (CAMUS, EchoNet-Dynamic, HMCQU) and one private dataset (CHU-Nice) of 2D echocardiography were used in our study.

Table 1: Datasets involved in this study

Dataset	Training Task	Pathology	View		
			A2C	A4C	2+4
CAMUS	Segmentation	–	500	500	500
EchoNet-Dynamic	Motion estimation	–	–	10030	–
HMC-QU	Classification	MI/Non-MI	130	162	130
CHU-Nice	–	MI/Non-MI	66	76	64

Table 2: Datasets with pathology diagnostics

Dataset	MI	Non-MI	Total
HMC-QU	88	42	130
CHU-Nice	42	22	64

CAMUS dataset (Leclerc et al., 2019b) contains apical 4-chamber (A4C) view and apical 2-chamber (A2C) view of 500 patients. The whole dataset was acquired using GE Vivid E95 ultrasound scanners. The ratio of good, medium and poor image quality is 43%, 43% and 14%. The mean frame rate of the dataset is 55 frame per second (fps). The following annotations are also publicly available: segmentation (of the left ventricle, the myocardium and the left atrium), volume information and ejection fraction of the left ventricle. The dataset contains an official train-test split of 450 patients vs. 50 patients.

EchoNet-Dynamic (Ouyang et al., 2020) is a large open dataset of 10030 A4C echocardiography videos. All videos were acquired by experienced sonographers using machines from different manufacturers, including iE33, Sonos, Acuson SC2000, Epiq 5G and Epiq 7C. The original videos were cropped and masked to remove irrelevant information and down-sampled to 112×112 pixel videos (Ouyang et al., 2019). The mean frame rate of the full dataset is 51 fps. Left ventricle traces at ED and ES as well as LVEF are provided for all videos. The official split of the train data vs. validation data vs. test data is 7465/1288/1277.

HMC-QU (Degerli et al., 2024) is the first public dataset for myocardial infarction detection, and contains: 162 A4C video and 130 A2C videos, where A2C and A4C views of 130 patients are both available. Videos were acquired using Phillips or GE Vivid machines. The mean frame rate is 25 fps. Moreover, for 109 of the A4C videos, the dataset supplies segmentation masks for the myocardium, generated through a semi-automatic approach (Degerli et al., 2021).

CHU-Nice dataset contains 2D echocardiography sequences of 76 patients collected retrospectively from Nice University hospital. This study protocol was in compliance with the declaration of Helsinki and was approved by the local research Ethics committee. These examinations were interpreted by experienced cardiologists, with only one cardiologist per examination. The ultrasound machines used for acquisition were either from Philips or GE Vivid E95. The mean frame rate of the dataset is 52 fps. The ratio of good, medium and poor image quality is 30%, 47% and 23%. It has to be noted that 20% of the images from

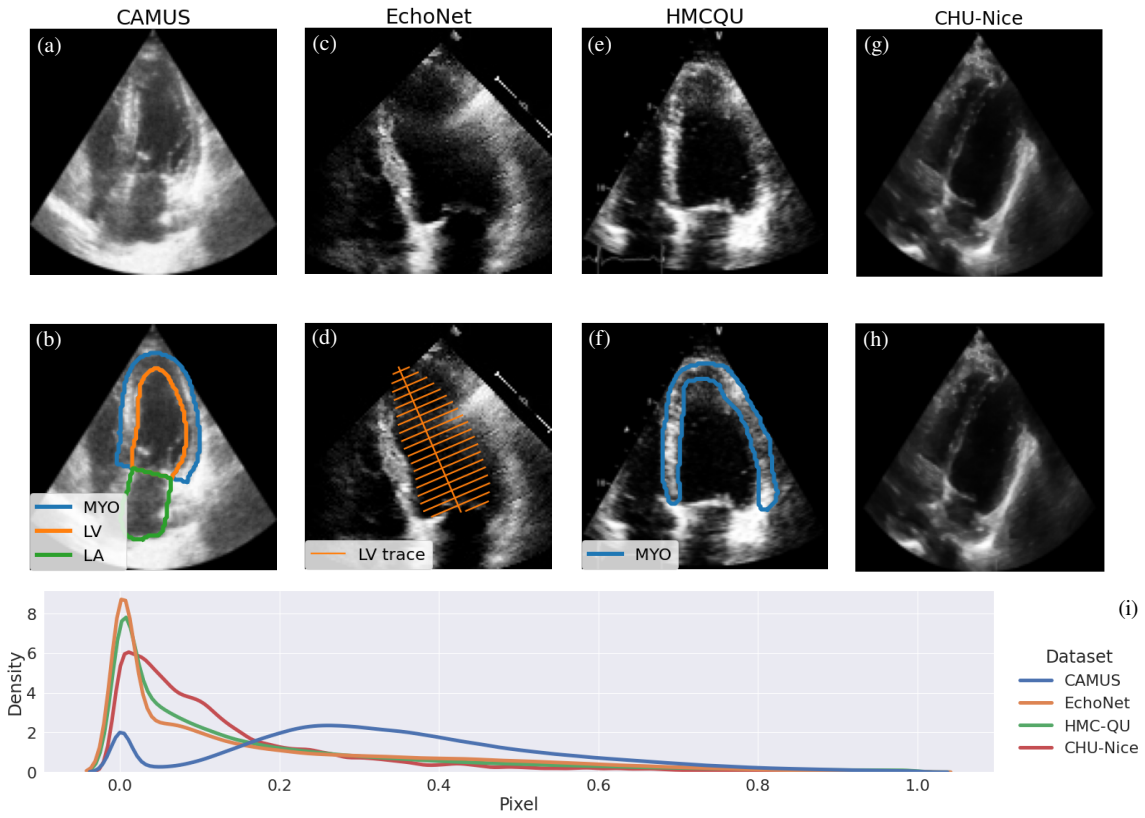

Figure 5: (a, c, e, g) Examples of echocardiography from each dataset involved in this study. (b, d, f, h) Anatomy-related annotations for each corresponding image sample. (i) The pixel value distribution using 100 random samples drawn from each dataset. Pixels outside the cone view area are excluded.

CHU-Nice were acquired within the Cardiology Emergency and Intensive Care Unit by the cardiologist on call at the time of the emergency imaging. This enables to evaluate the generalisation ability of the method on various levels of image quality. This dataset serves as an evaluation set for the real-world patient monitoring.

A detailed summary of all datasets is provided in Table 1 and Table 2. Image examples and their corresponding anatomy-related annotations are shown in Figure 5. The distribution of pixel values among the four datasets exhibits distinct patterns (Figure 5 (i)).

4.2 Experiments and implementation

4.2.1 Segmentation

For the segmentation task, we used the training split of CAMUS dataset (450 patients), as it contained the most detailed segmentation annotation of the left heart. Evaluation was performed on the test split of CAMUS dataset (50 patients: 50 A2C + 50 A4C), on the test split of EchoNet-Dynamic dataset (1277 A4C sequences) and on the CHU-Nice dataset (76

A4C sequences + 66 A2C sequences). We evaluated anatomical metrics including Dice score, Hausdorff distance (HD) and mean surface distance (MSD) for estimated segmentation, and clinical metrics such as mean absolute error (MAE) and correlation coefficient (Corr.) for predicted LVEF. In addition, the outliers of segmentation prediction for CAMUS dataset were established from the inter-variability tests with the upper limit values for HD and MSD, and lower limit values for simplicity and convexity (Leclerc et al., 2019a). Geometrical and anatomical outlier computation was only conducted on CAMUS dataset.

Implementation: The segmentation model was implemented in Pytorch (Paszke et al., 2019) with a training batch of 8. The input images were resized to 256×256 and augmented following the data augmentation strategies presented in Section 3.1 during training phase. An Adam optimizer was applied with a learning rate of 1e-4 and $\beta = 1e-6$ in (4). We ran 10-fold cross-validation on the 450 patients (A2C+A4C) of CAMUS train data with a maximum epoch of 100. The best model was identified as the one with the best performance on validation data. The model contains 1.14M trainable parameters. It takes less than 20 hours for 100 epochs' training on a NVIDIA GTX 1080Ti GPU. Inference running time in average is 0.067s per frame on a Dell laptop (Intel© Core™ i7-8650U CPU @ 1.90GHz × 4).

4.2.2 motion estimation

For the motion estimation task, we used the training split of EchoNet-Dynamic dataset (7465 A4C samples), as its echocardiography video contained at least 1-2 cardiac cycles. Evaluation was performed on the test split of EchoNet-Dynamic dataset (1277 A4C sequences), on the train split of CAMUS dataset (450 patients) and on the HMC-QU dataset (109 A4C sequences with provided myocardium masks). Since no ground truth data was available, we evaluated the registration performance of frames that contained segmentation annotation, such as ED-ES pairs.

Implementation: Since there was no myocardium annotation in the training data, we generated pseudo myocardium masks from the training data as described in (Yang and Sermesant, 2023). The poly-affine motion model (PAM) was trained using an Adam optimiser with learning rate of 1e-4 and a batch size of 1. The maximum number of frames per sample is set to 50. The training will be . The input image was resized to 128×128. The optimal hyper-parameters were identified by conducting experiments on a random subset of 1000 examples from the EchoNet-Dynamic training dataset: $\lambda_1 = \lambda_2 = 100, \lambda_3 = \lambda_4 = 50, \lambda_5 = 20, \lambda_6 = 0.1, \lambda_7 = \lambda_8 = 100, \lambda_9 = 1$ in (14), $\sigma^2 = 0.05$ in (9), and $\sigma_H^2 = 0.005$ in (7) and (12). In particular, the hyper-parameters were selected through a incremental search. We first fixed the parameter of $\lambda_1 = \lambda_2 = 100$, and tuned the other weights according to the following order and principle: $\lambda_3 = \lambda_4 \in \{1, 50, 100\}$; $\lambda_5 \in \{1, 10, 20\}$; $\lambda_6 \in \{0.1, 1, 10\}$; $\lambda_7 = \lambda_8 \in \{1, 50, 100\}$; and $\lambda_9 \in \{0.1, 1, 10\}$.

We compared our PAM model with one state-of-art conditional variational autoencoder (CVAE), which was re-implemented according to (Krebs et al., 2020). In order to leverage segmentation information, an extra Dice loss for registering LV masks between ED and ES was added to the final loss function. We also integrated a divergence-free term into the final loss function for incompressibility regularisation. All models were trained with 100 epochs at maximum and early stopped when there was no decrease in validation loss for more than

10 epochs. The PAM model contains 0.25M trainable parameters and it takes around 3-4 days on a NVIDIA GTX 1080Ti GPU. The inference running time on average is 0.9s for 50 frames (one sequence) on a Dell laptop (Intel© Core™ i7-8650U CPU @ 1.90GHz × 4).

Ablation study: The backbone FOMM model (Siarohin et al., 2019) only optimises the first four terms of (14). *+Prior* refers to the integration of prior-related losses (5th and 6th loss terms in (14)). *+Polyaffine* denotes replacing the original encoder-decoder module for weight estimation by designed poly-affine weights. *+ED* refers to the incorporation of registration losses between ED frame and others (7th and 8th terms in (14)). PAM(all) represents the model that optimises the total loss function of (14).

4.2.3 Myocardial infarction detection

In order to further evaluate the generalisation potential of the proposed methods, we extracted cardiac features from the segmentation and motion estimation outputs to detect myocardial infarction patients on HMC-QU dataset and CHU-Nice dataset. In practice, we performed 5-fold cross-validation on HMC-QU dataset using multi-view features (130 patients) and evaluated the classifier directly without retraining or fine tuning on CHU-Nice dataset (64 patients).A resampling strategy was conducted before classification training for the unbalanced HMC-QU dataset.

The best performing method on HMC-QU dataset used normalised segment displacement as features (Degerli et al., 2024).

$$f_k = \frac{\max D_k(t)}{\min I_{(k,p)}(t)} \tag{17}$$

where $D_k(t)$ referred to the mean displacement of segment k at time t, and $I_{(k,p)}(t)$ represented the averaged Manhattan distance between segment k and its opposite segment p. Authors in (Degerli et al., 2024) used active polynomials to extract displacement information. In our case, we obtained the normalised segment displacement using the predicted segmentation and motion estimation results, which resulted in a 6-dimension vector for each apical view. We denoted the method using the proposed global features as 'global', and the method using normalised displacement as 'local'.

4.3 Results

4.3.1 Segmentation

CAMUS On the unseen test split of CAMUS dataset, our proposed method achieved comparable performance (ranked No.2) with the best performing method either on cardiac structure segmentation or segmentation-based ejection fraction estimation (Table 3). In terms of correct anatomy, comparing with the previous work of R-Unet (Leclerc et al., 2019a), who added a refinement network after the U-Net work to improve the shape regularity of predicted left ventricle structures, our U-Net model with a contour prior (U-Net Contour loss) can directly generate less geometrical and anatomical outliers on the same dataset (Table 4).

EchoNet-Dynamic On the test split of EchoNet-Dynamic, our model segmented the left ventricle (endocardial region) with a Dice Coefficient of 0.895, only 2.7% lower than the

Table 3: Segmentation evaluation on CAMUS test split data. (Values in **bold** represent the best performance, while values with underline represent the second best.)

ED	**Endo**			**Epi**		
	Dice	HD	MSD	Dice	HD	MSD
U-Net(Leclerc et al., 2019b)	0.939	5.3	1.6	0.954	6.0	1.7
ACNNs(Oktay et al., 2017)	0.936	5.6	1.7	0.953	5.9	1.9
CLAS(Wei et al., 2020)	0.947	4.6	**1.4**	0.961	4.8	**1.5**
nnUNet(Ling et al., 2022)	**0.952**	**4.3**	**1.4**	**0.963**	**4.6**	**1.5**
Ours	0.949	4.8	**1.4**	0.961	5.0	1.6
ES	**Endo**			**Epi**		
	Dice	HD	MSD	Dice	HD	MSD
U-Net(Leclerc et al., 2019b)	0.916	5.5	1.6	0.945	6.1	1.9
ACNNs(Oktay et al., 2017)	0.913	5.6	1.7	0.945	5.9	2.0
CLAS(Wei et al., 2020)	0.929	4.6	1.4	0.955	4.9	1.6
nnUNet(Ling et al., 2022)	**0.935**	**4.2**	**1.3**	**0.959**	**4.4**	**1.5**
Ours	0.930	4.7	1.4	0.955	5.0	1.6
Volume	**EDV**		**ESV**		**EF**	
	Corr.	MAE	Corr.	MAE	Corr.	MAE
U-Net(Leclerc et al., 2019b)	0.926	11.2	0.960	7.5	0.845	5.0
ACNNs(Oktay et al., 2017)	0.928	9.7	0.954	6.9	0.807	5.5
CLAS(Wei et al., 2020)	0.958	7.7	0.979	4.4	**0.926**	**4.0**
nnUNet(Ling et al., 2022)	**0.977**	**5.9**	**0.987**	**4.0**	0.857	4.7
Ours	0.965	6.4	0.983	4.6	0.912	4.4

Table 4: CAMUS Segmentation shape outliers evaluated on 500 patients (2,000 images in total) with geometrical and anatomical outliers.

Method	**Geometrical**	**Anatomical**
R-Unet (Leclerc et al., 2019a)	16%	1.2%
U-Net (Leclerc et al., 2019b)	18%	3.75%
Ours	**15.4%**	**1.15%**

model trained on EchoNet-Dynamic dataset (Dice 0.92) (Ouyang et al., 2020). In terms of LVEF estimation, our model demonstrated only 1% decrease in MAE (MAE: 8.32%, Corr. 0.7) when compared with the benchmark model trained on the EchoNet-Dynamic training split (MAE: 7.35%) (Ouyang et al., 2020). Compared with the LVEF estimation of nnUNet (MAE: 14.3%, Corr. 0.27), our model achieves better generalisation performance.

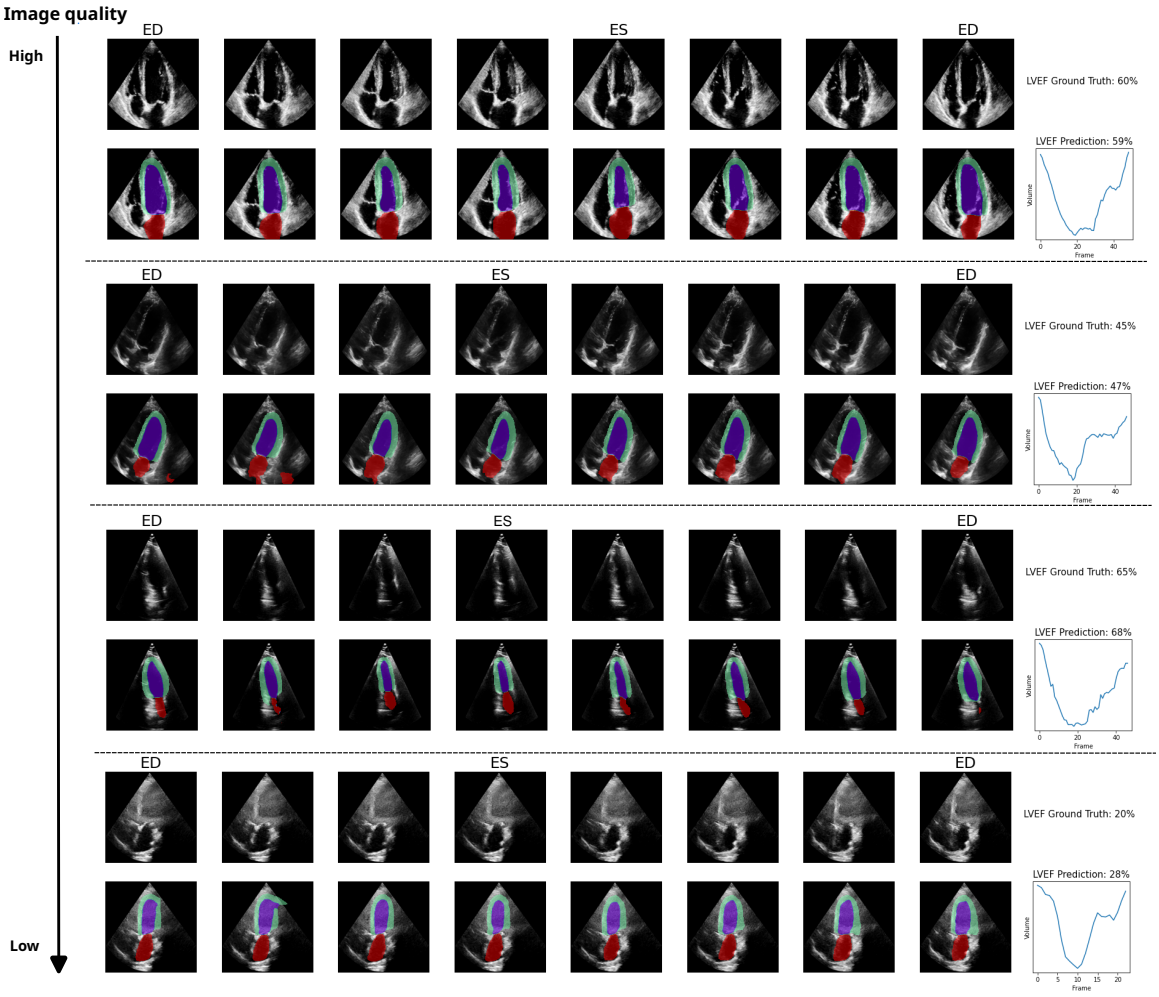

Figure 6: Examples of segmentation and LVEF estimation from CHU-Nice datasets. Each example comprises two rows: the first row displays the original image sequence, while the second row presents the predicted segmentation masks (Red: the left atrium, Cyan: the myocardium, and Purple: the left ventricle blood pool). Additionally, volume changes throughout one cardiac cycle are displayed alongside the corresponding predicted LVEF at the end of each example row.

CHU-Nice Since no segmentation ground truth was available for CHU-Nice dataset, we compared the LVEF estimated from segmentation model with the LVEF value provided by expert cardiologists. We achieved a MAE of 8.3% and correlation coefficient of 0.74. Examples of segmentation output and LVEF prediction are visualised in Figure 6 with various image qualities. The trained nnUNet model demonstrated inferior generalisable performance (corr. 0.50 and MAE 11.7%), which in turn reflects the superior performance of the proposed method.

4.3.2 Motion estimation

Table 5: Registration evaluation on EchoNet-Dynamic test split.

Method	**Endo**		
	Dice	HD (*pixels*)	MSD (*pixels*)
CVAE (Krebs et al., 2020)	0.91	7.97	2.30
FOMM (Siarohin et al., 2019)	0.75	22.99	6.07
FOMM+*Polyaffine*	0.82	11.50	4.61
FOMM+*Prior*	0.77	18.36	5.46
FOMM+*Prior*+*Polyaffine*	0.91	7.53	2.36
FOMM+*Prior*+*Polyaffine*+*ED*	**0.92**	<u>7.34</u>	<u>2.23</u>
PAM(all)	**0.92**	**7.17**	**2.14**

EchoNet-Dynamic As illustrated in Table 5, our proposition PAM(all) model out-performed all the other methods, including the state-of-art method CVAE and the backbone model FOMM. The ablation study pointed out that the introduction of the motion prior (*+Prior)* and poly-affine motion framework (*+Polyaffine*) improved largely the performance of registration compared with the original FOMM. Addition of the long-term similarity (*+ED*) and the incompressibility regularisation further contributed to the improvement of the registration accuracy of the PAM model.

Table 6: Registration evaluation on CAMUS training set (450 patients with 2CH and 4CH samples). The MCLAS method was trained on the CAMUS training set using 10-fold cross-validation, serves as the upper bound of the performance. Other methods (trained on EchoNet-Dynamic dataset) were evaluated directly on the same CAMUS training split.

Method	**Epi**		**Endo**	
	Dice	MSD *(mm)*	Dice	MSD *(mm)*
MCLAS (Wei et al., 2023)	0.931	2.0	0.903	1.9
CVAE (Krebs et al., 2020)	0.917	3.0	0.859	3.3
FOMM (Siarohin et al., 2019)	0.883	4.3	0.778	5.2
FOMM+*Polyaffine*	0.881	4.3	0.796	4.8
FOMM+*Prior*	0.893	3.8	0.804	4.3
FOMM+*Prior*+*Polyaffine*	<u>0.932</u>	<u>2.5</u>	<u>0.885</u>	<u>2.6</u>
FOMM+*Prior*+*Polyaffine*+*ED*	**0.933**	**2.4**	**0.894**	**2.4**
PAM(all)	0.927	2.6	0.883	2.6

CAMUS Table 6 summarises the generalisation performance of our proposed PAM model on CAMUS dataset without any retraining or fine-tuning. Notably, the proposed PAM model out-performed the backbone FOMM and the state-of-art CVAE method. In addition, it achieved the closest registration performance with the best tracking model (MCLAS) (Wei et al., 2023) which was trained on CAMUS training set using 10-fold cross-validation, while our PAM model has never seen any data from CAMUS. Although the introduction of the incompressibility term degraded the performance of PAM, it still outperformed other competitive methods.

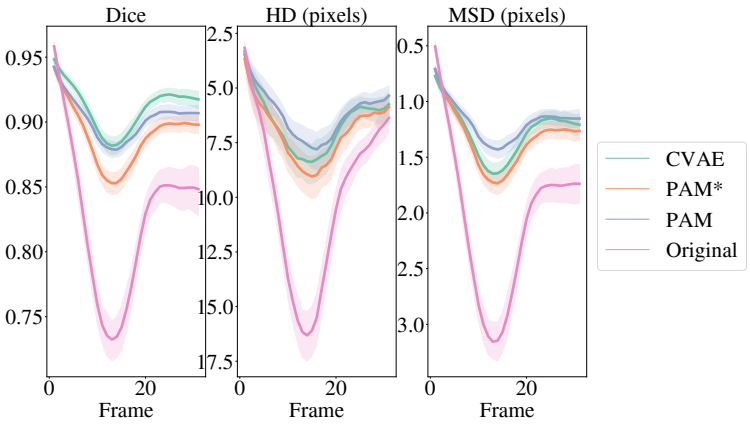

Figure 7: Registration results on HMC-QU dataset (109 A4C samples) using frame-wise myocardium masks. Original: Comparison between the ground truth masks along one cardiac cycle and that of end-diastole. Curves of all samples are interpolated to the same length. PAM*: without incompressibility term.

HMC-QU Figure 7 summarises the performance of the proposed PAM model, the PAM model without incompressibility term and the CVAE model on HMC-QU dataset by evaluating the registration accuracy of the myocardium mask.

The introduction of the incompressibility penalisation improved the tracking performance of PAM model across three metrics. In particular, the PAM model demonstrates smaller values than the CVAE in terms of Hausdorff distance and mean surface distance around end-systole frames, where largest deformation has occurred. An visualisation example from HMC-QU dataset (Figure 8) further demonstrates the tracking consistency. Our prediction shows that the patient has a small radial strain in SEG14 region (decreased regional radial strain and longitudinal strain highlighted by a red rectangle), which is in accordance with the fact that the patient is diagnosed with MI and in particular the SEG14 is affected.

Motion regularity The extensive evaluation of the proposed PAM model across different datasets demonstrated its robustness and its better performance compared with other competing methods in motion tracking in echocardiography. The introduction of the incompressiblity term improved the performance of the original PAM* model (Yang and

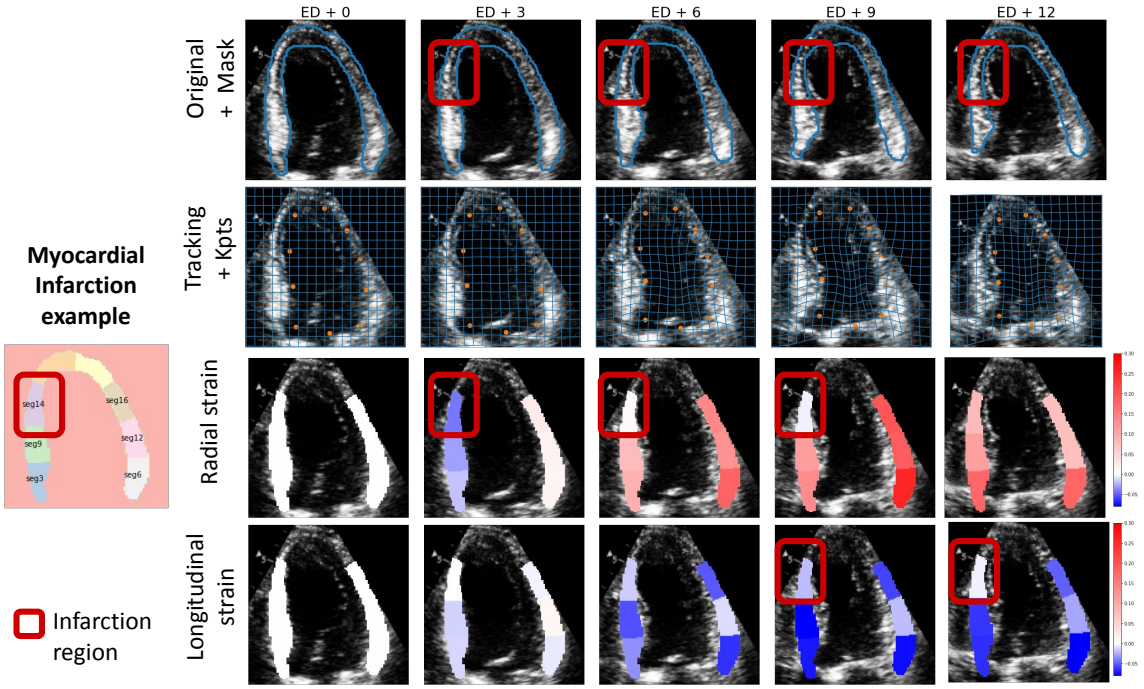

Figure 8: Visual results of a myocardial infarction sequence from HMC-QU dataset where SEG14 is annotated as infarcted region. Red rectangles highlight the abnormal local strain values.

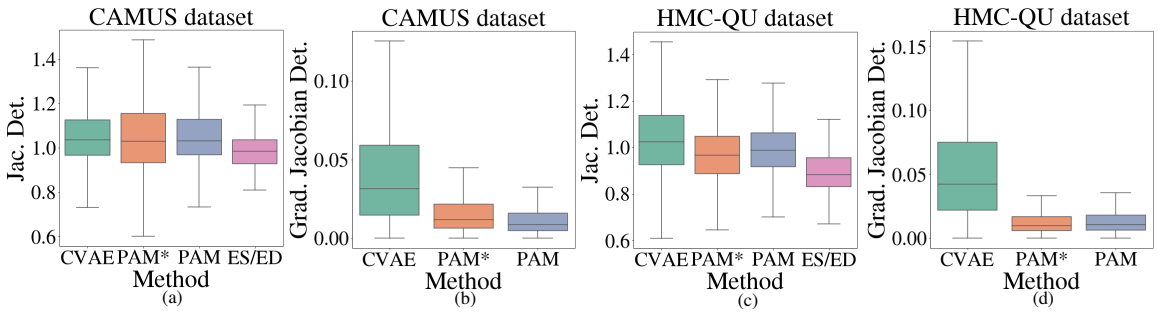

Figure 9: Evaluation results of Jacobian Determinant and its gradient in the myocardium region on the CAMUS dataset (a-b) and on the HMC-QU dataset (c-d). ES/ED represents the area ratio of the myocardium between end-systole and end-diastole using ground truth masks. PAM*: without incompressibility term.

Sermesant, 2023) on two evaluation datasets and improved the regularity of predicted motion fields. This was evidenced by the Jacobian Determinant which was closer to 1 and its smaller gradient (Figure 9). Further visualisation of Jacobian Determinant under CVAE

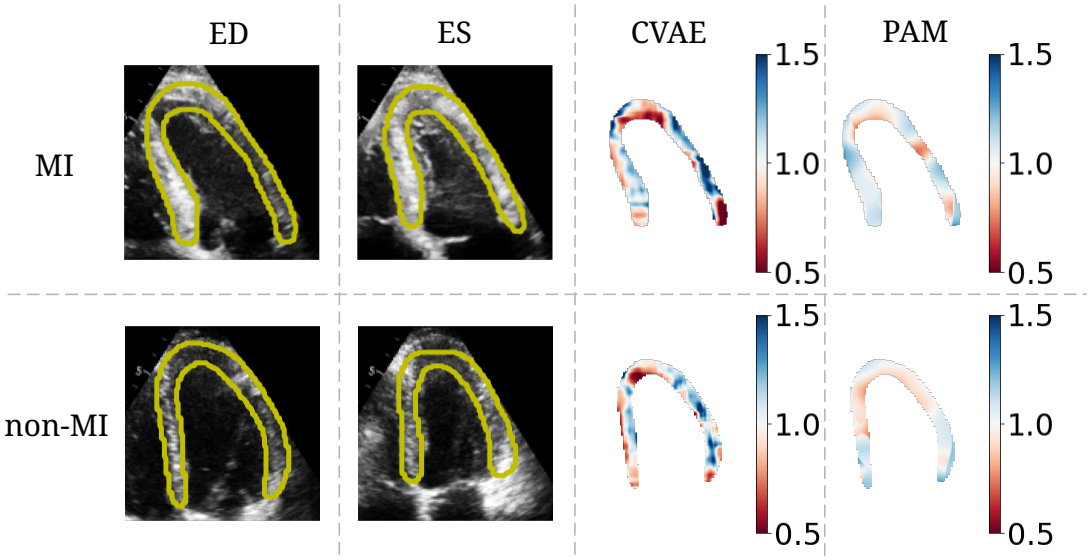

Figure 10: Examples of Jacobian determinant map in the myocardium region from HMC-QU dataset using CVAE and PAM methods.

and PAM methods (Figure 10) supported the former conclusion, witnessing the superiority of using poly-affine motion model for cardiac motion tracking.

4.3.3 Myocardial infarction detection

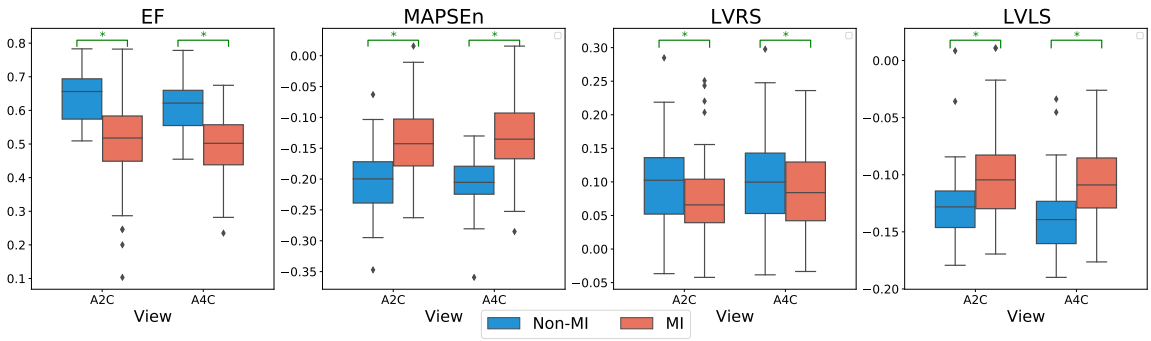

Figure 11: Distribution of global features between MI patients and non-MI patients on HMC-QU dataset. Wilcoxon rank-sum test is performed between non-MI and others. (Green $*$: $p\text{-}value < 0.05$)

HMC-QU First, a significant group difference was observed between MI patients and non-MI patients in the four global parameters on the HMC-QU dataset, as illustrated in Figure 11, except for the LVRS feature for the apical 4-chamber view.

Table 7: Myocardial infarction detection results. *Local features: normalised segment displacement (Section 4.2.3); Global features: LVEF, nMAPSE, LVLS, LVRS (Section 3.3).*

Method	Feature	Sen.	Spe.	Prec.	F1.	Accuracy
HMC-QU (5 fold cross-validation)						
RF	(Degerli et al., 2024)	0.875	0.619	0.828	**0.951**	0.792
SVM	(Degerli et al., 2024)	**0.910**	0.429	0.769	0.877	0.754
RF	Local	0.773	0.714	0.850	0.810	0.754
SVM	Local	0.761	0.786	0.881	0.817	0.769
RF	Global	0.841	0.786	0.891	0.865	**0.823**
SVM	Global	0.784	**0.833**	**0.908**	0.841	0.800
CHU (Evaluation)						
RF	Local	**1.000**	0.091	0.677	0.808	0.688
SVM	Local	0.952	0.273	0.714	0.816	0.719
RF	Global	0.952	0.409	0.754	**0.842**	**0.766**
SVM	Global	0.881	**0.545**	**0.787**	0.831	**0.766**

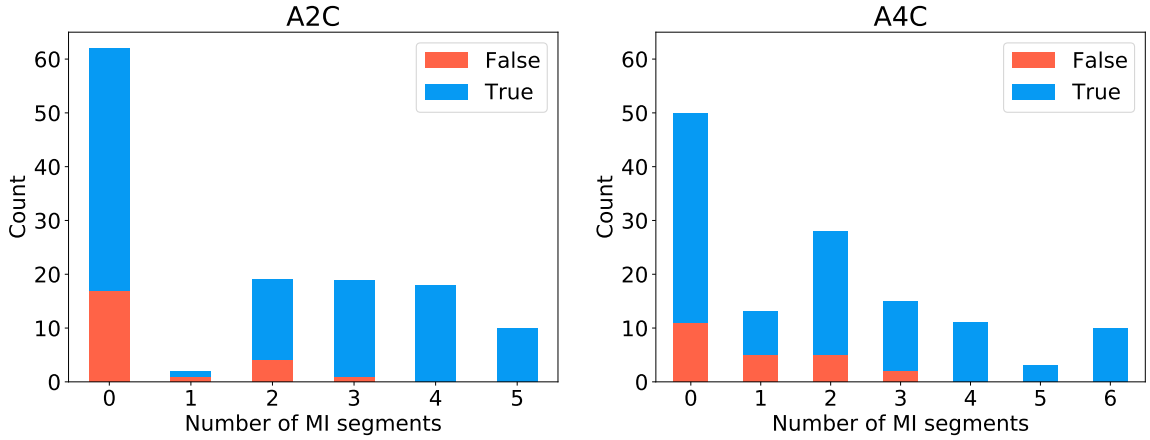

Figure 12: 5-fold cross validation result of MI detection using global features and SVM classifier. False: wrong classification, True: correct classification.

Second, the results presented in Table 7 demonstrate that our method, utilising global features, outperformed the use of local features, including the state-of-the-art result reported in (Degerli et al., 2024). This showcases the effectiveness and generalisability of the proposed pipeline for myocardial infarction detection. Notably, local features extracted using a trained segmentation model and motion tracking model exhibited similar classification accuracy compared to those obtained through the time-consuming active polynomial method.

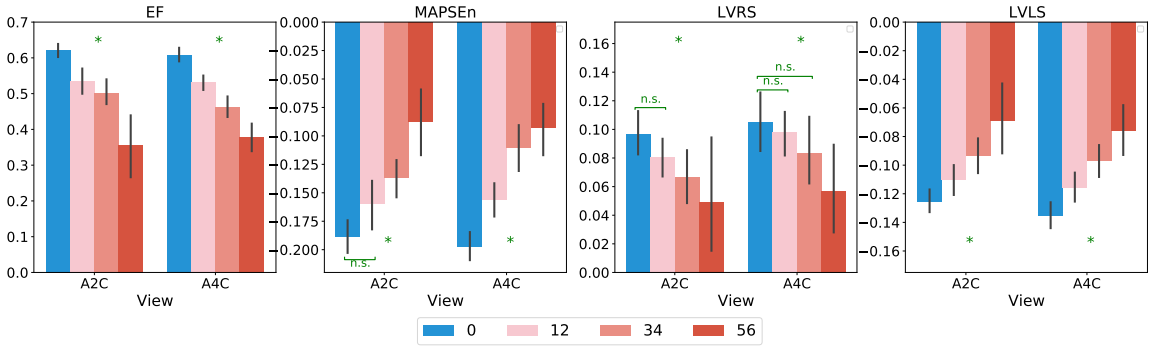

Figure 13: Distribution of global features between patients with different number of infarct segments from HMC-QU dataset. 0: non-MI, 12: MI patients with 1-2 infarct segments, 34: MI with 3-4 segments, 56: MI with 5-6 segments. Wilcoxon rank-sum test is performed between non-MI and others. (Green $*$: $p\text{-}value < 0.05$)

However, the use of global features for MI detection had its limitations. Further analysis, as illustrated in Figure 12, revealed that our proposed method using global features faced challenges in differentiating healthy samples from MI patients with fewer infarct segments. This observation aligns with the linear relationship observed between the number of infarct segments and global features, as shown in Figure 13. It is noteworthy that there is no significant difference in MAPSEn and LVRS values between non-MI patients and MI patients with 1-2 infarct segments. However, the difference between non-MI patients and MI patients with more than 2 segments is statistically significant in all four measurements.

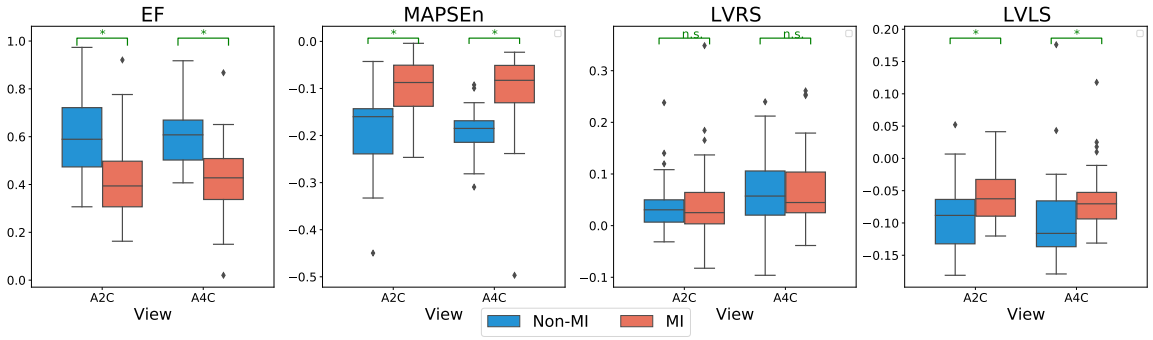

Figure 14: Distribution of global features between MI patients and non-MI patients on CHU-Nice dataset. Wilcoxon rank-sum test is performed between non-MI and others. (Green $*$: $p\text{-}value < 0.05$)

CHU-Nice As shown in the lower part of Table 7, classifiers that utilised global features demonstrated better classification performance on the CHU-Nice dataset. This is evident from comparable sensitivity, higher specificity, higher precision, higher F1-score, and higher

accuracy. Local features tend to be more sensitive to noisy output from segmentation and motion tracking models due to the domain shift between the test dataset (CHU-Nice) and the training dataset (HMC-QU). In contrast, global features, often derived from averaged measurements, exhibit greater stability compared to local features, resulting in superior performance.

Additionally, a significant group difference between MI patients and non-MI patients on the CHU-Nice data was observed (Figure 14). The evaluation results on the CHU-Nice dataset strongly reinforce the robustness and good generalisation of our proposed pipeline for echocardiography analysis.

5. Discussion

In this work, we introduced a generalisable echocardiography analysis pipeline, including one segmentation model, one motion tracking model and a classifier that takes inputs from the predictions of the two former models to detect myocardium infarction patients. We have extensively evaluated each part of the pipeline thoroughly across various test datasets.

Segmentation The proposed segmentation model not only demonstrated segmentation performance on the test split from the training dataset very close to the state-of-the-art method but also exhibited generalisation capabilities on two other datasets without retraining or fine-tuning. In addition, regarding the left ventricular ejection fraction (LVEF) estimation from segmentation masks, it is noteworthy that the three datasets involved in the segmentation evaluation phase had different annotation schemes, impacting the estimation performance of LVEF. For instance, the CAMUS dataset provided detailed annotations of the left ventricle (LV) for end-diastole (ED) and end-systole (ES) (see Figure 5 (b)), while EchoNet-Dynamic only provided endocardial traces used for volume estimation (see Figure 5 (d)). Additionally, the ground truth LVEF of the CHU-Nice dataset was measured by expert cardiologists either visually or using the Simpson's method. Despite the diverse annotation approaches across datasets, our model consistently predicted LVEF results on external evaluation datasets (MAE 8%, Corr. 0.7), closely aligning with the inter-observer variability range (MAE: 10%, Corr. 0.8) reported in a previous study (Leclerc et al., 2019b).

Motion tracking The motivation behind the Polyaffine Affine Motion (PAM) tracking model lies in addressing the absence of regularity in deformation fields found in current deep learning-based methods, particularly in the myocardium region. The proposed PAM method models the motion of the entire myocardium through poly-affine fusion, which ensures local smoothness in a larger neighbourhood. This is quantitatively reflected in the smaller Jacobian determinant gradient observed in the myocardium region, as well as through visualisation examples from test datasets. Additionally, extensive evaluation across public echocardiography datasets demonstrated the comparable or even superior tracking performance of the PAM model when compared with other competing methods, such as the Conditional Variational Autoencoder (CVAE) model and the First-Order Motion Model (FOMM). This supports the conclusion that the proposed PAM model is more suitable for cardiac motion tracking in echocardiography.

Classification Initially, all features of the HMC-QU dataset and CHU-Nice dataset were extracted from segmentation and motion estimation outputs, derived from models trained

on other datasets. We observed significant differences in specific measurements between healthy and Myocardial Infarction (MI) patients on these two previously unseen datasets, underscoring the effectiveness and robustness of our segmentation and motion estimation methods in distinguishing between the two groups. The application of the trained classifier on a real-world unseen dataset (CHU-Nice dataset) further supports the robustness of utilising global features for MI detection. However, we acknowledge that using global features for MI detection faces challenges in accurately identifying patients with only one or two segments containing infarcted myocardium. Moreover, local features such as the displacement or strain of myocardium segments necessitate precise segmentation of the myocardium, rendering them sensitive to domain shift and incorrect segmentation masks, consequently making them less stable for generalisable classification.

Limitations We acknowledge that the segmentation model was applied on separate frames of echocardiography sequences, while temporal information of cardiac sequences was neglected by current model. Enforcing temporal consistency is a good way to reduce segmentation outliers and to engage unlabelled data for training (Wei et al., 2020; Painchaud et al., 2022). In addition, for the motion tracking model, we imposed a strong regional weight for each key point without considering the varying thickness of the LV wall across patients and across regions. Thus, it would be interesting to further combine temporal motion tracking model with temporal segmentation model, letting the predicted mask of the myocardium guide the region of influence of each local key point. Furthermore, motion tracking of LV regularises vice versa the segmentation of cardiac structures for unlabelled temporal frames, accordingly. Finally, it is crucial to estimate the uncertainty of the predicted segmentation and motion transformation for clinical applications. Methods proposed in (Grzech et al., 2022; Judge et al., 2023) are of great interest for the further improvement of our echocardiography analysis pipeline.

6. Conclusion

In this paper, we proposed a novel and generalisable pipeline for 2D echocardiography analysis using deep learning. Specifically, we introduced a segmentation model incorporating a shape prior to enhance regular anatomy and improve EF estimation. Furthermore, a motion estimation model was presented, leveraging a motion prior for superior tracking accuracy and increased regularity of the motion field. Lastly, a classification module was proposed using clinically interpretable features for robust detection of myocardial infarction patients. Our experiments and evaluations clearly demonstrated the robustness and strong generalisation ability of our novel method compared to other competing methods at different stages. We utilised various metrics and validated our approach on different public datasets and real-world clinical data, providing benchmark results for generalisable echocardiography analysis.

Acknowledgments

This work has been supported by the French government through the National Research Agency (ANR) Investments in the Future with 3IA Côte d'Azur (ANR-19-P3IA-0002). The

authors are grateful to the OPAL infrastructure from Université Côte d'Azur for providing resources and support.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare that we do not have conflicts of interest.

References

Chi Young Ahn. Robust myocardial motion tracking for echocardiography: Variational framework integrating local-to-global deformation. *Computational and Mathematical Methods in Medicine*, 2013, 2013. ISSN 1748670X. .

Shawn Ahn, Kevinminh Ta, Allen Lu, John Stendahl, Albert Sinusas, and James Duncan. Unsupervised motion tracking of left ventricle in echocardiography. *Proceedings of SPIE–the International Society for Optical Engineering*, 11319, 02 2020.

Martino Alessandrini, Bidisha Chakraborty, Brecht Heyde, Olivier Bernard, Mathieu De craene, Maxime Sermesant, and Jan D'hooge. Realistic vendor-specific synthetic ultrasound data for quality assurance of 2-d speckle tracking echocardiography: Simulation pipeline and open access database. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65:411–422, 2018.

Arghavan Arafati, Daisuke Morisawa, Michael R. Avendi, M. Reza Amini, Ramin A. Assadi, Hamid Jafarkhani, and Arash Kheradvar. Generalizable fully automated multi-label segmentation of four-chamber view echocardiograms based on deep convolutional adversarial networks. *Journal of the Royal Society, Interface*, 17(169):20200267, 2020. ISSN 17425662. .

Vincent Arsigny, Olivier Commowick, Nicholas Ayache, and Xavier Pennec. A fast and log-euclidean polyaffine framework for locally linear registration. *Journal of Mathematical Imaging and Vision*, 33:222–238, 2009.

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

Djamal Boukerroui, J. Alison Noble, and Michael Brady. Velocity estimation in ultrasound images: A block matching approach. In Chris Taylor and J. Alison Noble, editors, *Information Processing in Medical Imaging*, pages 586–598, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. ISBN 978-3-540-45087-0.

Bidisha Chakraborty, Brecht Heyde, Martino Alessandrini, and Jan D'hooge. Fast myocardial strain estimation from 3d ultrasound through elastic image registration with analytic regularization. In *Medical Imaging 2016: Ultrasonic Imaging and Tomography*, volume 9790, pages 48–54. SPIE, 2016.

Maja Cikes and Scott D Solomon. Beyond ejection fraction: an integrative approach for assessment of cardiac structure and function in heart failure. *European heart journal*, 37 (21):1642–1650, 2016.

Bob D De Vos, Floris F Berendsen, Max A Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52:128–143, 2019.

Aysen Degerli, Morteza Zabihi, Serkan Kiranyaz, Tahir Hamid, Rashid Mazhar, Ridha Hamila, and Moncef Gabbouj. Early detection of myocardial infarction in low-quality echocardiography. *IEEE Access*, 9:34442–34453, 2021. .

Aysen Degerli, Serkan Kiranyaz, Tahir Hamid, Rashid Mazhar, and Moncef Gabbouj. Early myocardial infarction detection over multi-view echocardiography. *Biomedical Signal Processing and Control*, 87:105448, 2024.

Yinlong Deng, Peiwei Cai, Li Zhang, Xiongcheng Cao, Yequn Chen, Shiyan Jiang, Zhemin Zhuang, and Bin Wang. Myocardial strain analysis of echocardiography based on deep learning. *Frontiers in Cardiovascular Medicine*, 9:1067760, 2022.

Ewan Evain, Yunyun Sun, Khuram Faraz, Damien Garcia, Eric Saloux, Bernhard L Gerber, Mathieu De Craene, and Olivier Bernard. Motion estimation by deep learning in 2d echocardiography: synthetic dataset and validation. *IEEE transactions on medical imaging*, 41(8):1911–1924, 2022.

World Heart Federation. World heart report 2023: Confronting the world's number one killer. 2023. URL `https://heartreport23.world-heart-federation.org/`.

ED Folland, AF Parisi, PF Moynihan, D Ray Jones, Ch L Feldman, and DE Tow. Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. a comparison of cineangiographic and radionuclide techniques. *Circulation*, 60(4):760–766, 1979.

Daniel Grzech, Mohammad Farid Azampour, Ben Glocker, Julia Schnabel, Nassir Navab, Bernhard Kainz, and Loïc Le Folgoc. A variational bayesian method for similarity learning in non-rigid image registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 119–128, 2022.

Mu-Shiang Huang, Chi-Shiang Wang, Jung-Hsien Chiang, Ping-Yen Liu, and Wei-Chuan Tsai. Automated recognition of regional wall motion abnormalities through deep neural network interpretation of transthoracic echocardiography. *Circulation*, 142(16):1510–1520, 2020.

Mohammad H. Jafari, Hany Girgis, Amir H. Abdi, Zhibin Liao, Mehran Pesteie, Robert Rohling, Ken Gin, Terasa Tsang, and Purang Abolmaesumi. Semi-supervised learning for cardiac left ventricle segmentation using conditional deep generative models as prior. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 649–652, 2019. .

Shuman Jia, Antoine Despinasse, Zihao Wang, Hervé Delingette, Xavier Pennec, Pierre Jaïs, Hubert Cochet, and Maxime Sermesant. Automatically segmenting the left atrium from cardiac images using successive 3d u-nets and a contour loss. In *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*, pages 221–229. Springer, 2019.

Thierry Judge, Olivier Bernard, Woo-Jin Cho Kim, Alberto Gomez, Agisilaos Chartsias, and Pierre-Marc Jodoin. Asymmetric contour uncertainty estimation for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 210–220. Springer, 2023.

Mohammad Mahdi Kazemi Esfeh, Christina Luong, Delaram Behnami, Teresa Tsang, and Purang Abolmaesumi. A deep bayesian video analysis framework: towards a more robust estimation of ejection fraction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 582–590. Springer, 2020.

Kevin R King, Luanda P Grazette, Dina N Paltoo, John T McDevitt, Samuel K Sia, Paddy M Barrett, Fred S Apple, Paul A Gurbel, Ralph Weissleder, Hilary Leeds, et al. Point-of-care technologies for precision cardiovascular care and clinical research: National heart, lung, and blood institute working group. *JACC: Basic to Translational Science*, 1 (1):73–86, 2016.

Julian Krebs, Tommaso Mansi, Nicholas Ayache, and Hervé Delingette. Probabilistic motion modeling from medical image sequences: application to cardiac cine-mri. In *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers 10*, pages 176–185. Springer, 2020.

Kenya Kusunose, Takashi Abe, Akihiro Haga, Daiju Fukuda, Hirotsugu Yamada, Masafumi Harada, and Masataka Sata. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *Cardiovascular Imaging*, 13(2_Part_1): 374–381, 2020.

Sarah Leclerc, Erik Smistad, Thomas Grenier, Carole Lartizien, Andreas Ostvik, Frédéric Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, et al. Ru-net: A refining segmentation network for 2d echocardiography. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 1160–1163. IEEE, 2019a.

Sarah Leclerc, Erik Smistad, Joao Pedrosa, Andreas Østvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, et al. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE transactions on medical imaging*, 38(9):2198–2210, 2019b.

Hang Jung Ling, Damien Garcia, and Olivier Bernard. Reaching intra-observer variability in 2-d echocardiographic image segmentation with a simple u-net architecture. In *IEEE International Ultrasonics Symposium (IUS)*, 2022.

Bohan Liu, Hao Chang, Dong Yang, Feifei Yang, Qiushuang Wang, Yujiao Deng, Lijun Li, Wenqing Lv, Bo Zhang, Liheng Yu, et al. A deep learning framework assisted echocardiography with diagnosis, lesion localization, phenogrouping heterogeneous disease, and anomaly detection. *Scientific Reports*, 13(1):3, 2023.

Tommaso Mansi, Xavier Pennec, Maxime Sermesant, Hervé Delingette, and Nicholas Ayache. ilogdemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. *International Journal of Computer Vision*, 92:92–111, 2011.

Kristin McLeod, Maxime Sermesant, Philipp Beerbaum, and Xavier Pennec. Spatio-temporal tensor decomposition of a polyaffine motion model for a better analysis of pathological left ventricular dynamics. *IEEE transactions on medical imaging*, 34(7): 1562–1575, 2015.

Manuel A Morales, Maaike Van den Boomen, Christopher Nguyen, Jayashree Kalpathy-Cramer, Bruce R Rosen, Collin M Stultz, David Izquierdo-Garcia, and Ciprian Catana. Deepstrain: a deep learning workflow for the automated characterization of cardiac mechanics. *Frontiers in Cardiovascular Medicine*, 8:730316, 2021.

Akhil Narang, Shashank S Sinha, Bharath Rajagopalan, Nkechinyere N Ijioma, Natalie Jayaram, Aaron P Kithcart, Varsha K Tanguturi, and Michael W Cullen. The supply and demand of the cardiovascular workforce: striking the right balance. *Journal of the American College of Cardiology*, 68(15):1680–1689, 2016.

Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A Cook, Antonio De Marvao, Timothy Dawes, Declan P O'Regan, et al. Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. *IEEE transactions on medical imaging*, 37(2):384–395, 2017.

Hasmila A Omar, Arijit Patra, João S Domingos, Paul Leeson, and Alison J Noblel. Automated myocardial wall motion classification using handcrafted features vs a deep cnn-based mapping. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3140–3143. IEEE, 2018.

Andreas Østvik, Ivar Mjåland Salte, Erik Smistad, Thuy Mi Nguyen, Daniela Melichova, Harald Brunvand, Kristina Haugaa, Thor Edvardsen, Bjørnar Grenne, and Lasse Lovstakken. Myocardial function imaging in echocardiography using deep learning. *ieee transactions on medical imaging*, 40(5):1340–1351, 2021.

David Ouyang, Bryan He, Amirata Ghorbani, Matt P Lungren, Euan A Ashley, David H Liang, and James Y Zou. Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning. In *NeurIPS ML4H Workshop*, pages 1–11, 2019.

David Ouyang, Bryan He, Amirata Ghorbani, Neal Yuan, Joseph Ebinger, Curtis P Langlotz, Paul A Heidenreich, Robert A Harrington, David H Liang, Euan A Ashley, et al. Video-based ai for beat-to-beat assessment of cardiac function. *Nature*, 580(7802):252–256, 2020.

Nathan Painchaud, Nicolas Duchateau, Olivier Bernard, and Pierre-Marc Jodoin. Echocardiography segmentation with enforced temporal consistency. *IEEE Transactions on Medical Imaging*, 41(10):2867–2878, 2022.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

U Raghavendra, Hamido Fujita, Anjan Gudigar, Ranjan Shetty, Krishnananda Nayak, Umesh Pai, Jyothi Samanth, and U Rajendra Acharya. Automated technique for coronary artery disease characterization and classification using dd-dtdwt in ultrasound images. *Biomedical Signal Processing and Control*, 40:324–334, 2018.

Hadrien Reynaud, Athanasios Vlontzos, Benjamin Hou, Arian Beqiri, Paul Leeson, and Bernhard Kainz. Ultrasound video transformers for cardiac ejection fraction estimation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 495–505. Springer, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. First order motion model for image animation. *Advances in neural information processing systems*, 32, 2019.

Erik Smistad, Andreas Østvik, Ivar Mjåland Salte, Daniela Melichova, Thuy Mi Nguyen, Kristina Haugaa, Harald Brunvand, Thor Edvardsen, Sarah Leclerc, Olivier Bernard,

et al. Real-time automatic ejection fraction and foreshortening detection using deep learning. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 67(12): 2595–2604, 2020.

Marius Staring, Stefan Klein, and Josien PW Pluim. A rigidity penalty term for nonrigid registration. *Medical physics*, 34(11):4098–4108, 2007.

Asbjørn Støylen, Harald E. Mølmen, and Håvard Dalen. Relation between mitral annular plane systolic excursion and global longitudinal strain in normal subjects: The hunt study. *Echocardiography*, 35(5):603–610, 2018.

Kevinminh Ta, Shawn S. Ahn, Allen Lu, John C. Stendahl, Albert J. Sinusas, and James S. Duncan. A semi-supervised joint learning approach to left ventricular segmentation and motion tracking in echocardiography. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1734–1737, 2020.

Basma Touil, Adrian Basarab, Philippe Delachartre, Olivier Bernard, and Denis Friboulet. Analysis of motion tracking in echocardiographic image sequences: Influence of system geometry and point-spread function. *Ultrasonics*, 50(3):373–386, 2010.

Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. In Dimitris Metaxas, Leon Axel, Gabor Fichtinger, and Gábor Székely, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, pages 754–761, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. ISBN 978-3-540-85988-8.

Zihao Wang, Yingyu Yang, Maxime Sermesant, and Hervé Delingette. Unsupervised echocardiography registration through patch-based mlps and transformers. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 168–178. Springer, 2022.

Hongrong Wei, Heng Cao, Yiqin Cao, Yongjin Zhou, Wufeng Xue, Dong Ni, and Shuo Li. Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, pages 623–632. Springer, 2020.

Hongrong Wei, Junqiang Ma, Yongjin Zhou, Wufeng Xue, and Dong Ni. Co-learning of appearance and shape for precise ejection fraction estimation from echocardiographic sequences. *Medical Image Analysis*, 84:102686, 2023.

Yingyu Yang and Maxime Sermesant. Shape constraints in deep learning for robust 2d echocardiography analysis. In *FIMH*, pages 22–34. Springer, 2021.

Yingyu Yang and Maxime Sermesant. Unsupervised polyaffine transformation learning for echocardiography motion estimation. In *FIMH*, pages 384–393. Springer, 2023.

Anthony J Yezzi and Jerry L Prince. An eulerian pde approach for computing tissue thickness. *IEEE transactions on medical imaging*, 22(10):1332–1339, 2003.

Ling Zhang, Xiaosong Wang, Dong Yang, Thomas Sanford, Stephanie Harmon, Baris Turkbey, Bradford J. Wood, Holger Roth, Andriy Myronenko, Daguang Xu, and Ziyue Xu. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Transactions on Medical Imaging*, 39(7):2531–2540, 2020. .