Unsupervised Polyaffine Transformation Learning for Echocardiography Motion Estimation

Yingyu Yang, Maxime Sermesant

Unsupervised Polyaffine Transformation Learning for Echocardiography Motion Estimation

Yingyu Yang, Maxime Sermesant

Université Côte d'Azur, Inria Epione Team, Sophia Antipolis, France
Contact: {firstname}.{lastname}@inria.fr

Abstract. Echocardiography plays an important role in the diagnosis of cardiac dysfunction. In particular, motion estimation in echocardiography is a challenging task since ultrasound images suffer largely from low signal-to-noise ratio and out-of-view problems. Current deep learning-based models for cardiac motion estimation in the literature estimate the dense motion field with spatial regularization. However, the underlying spatial regularization can only cover a very small region in the neighborhood, which is not enough for a smooth and realistic motion field for the myocardium in echocardiography. In order to improve the performance and quality with deep learning networks, we propose applying polyaffine transformation for motion estimation, which intrinsically regularizes the myocardium motion to be polyaffine. Our thorough experiments demonstrate that the proposed method not only presents better evaluation metrics on the registration of cardiac structures but also shows great potential in abnormal wall motion detection.

Keywords: Motion Estimation · Echocardiography · Polyaffine Transformation.

1 Introduction

Echocardiography is one of the most widely used modalities for cardiac dysfunction diagnosis. It's radiation-free, non-invasive, and real-time, making echocardiography very suitable for portable analysis, such as myocardial motion evaluation. However, ultrasound images generally have poorer quality than other modalities, such as MRI and CT, due to their low signal-to-noise ratio. Additionally, limitations in acquisitions and patient variability may cause the myocardium to occasionally be outside the imaging field-of-view. These burdens make motion estimation of the myocardium in echocardiography a very challenging task.

Traditional methods for motion estimation in echocardiography, such as block matching [3,5] or optical flow [3,6,11], are typically time-consuming and not suitable for real-time or portable analysis. However, recent advances in deep learning (DL) have improved both the time efficiency and tracking performance of motion estimation algorithms. DL-based models can generally be classified into two types: optical flow-based models and registration-based models. Optical flow-based DL models estimate dense displacement fields between consecutive image

frames and typically require ground truth displacement for supervised training [10,17]. Since obtaining ground truth displacement from real-world echocardiography images can be difficult, synthetic echocardiography datasets are often used for supervised training of these models [2,10]. Registration-based DL models typically register all other frames to end-diastole either in an unsupervised manner or with weak supervision using segmentation masks, enabling them to work on real-world datasets [1,25]. U-net-like architectures are often used as the core design of such methods [1,25]. Many other studies on cardiac MRI motion estimation have also utilized the registration approach with different temporal smoothness strategies [13,21]. Recently, researchers have also incorporated biomechanical modeling knowledge into deep learning networks with the aim of improving the generalizability of motion estimation performance [22,26].

The motion of the myocardium is not arbitrary, and various spatial regularization techniques have been explored in the literature, such as the divergence penalty [15] for enforcing incompressibility, the rigidity penalty for smoothness [24], and the elastic strain energy for mechanical correctness [19], among others. Many deep learning models have incorporated these regularization techniques into their work [1, 8, 26]. However, these regularization techniques are usually based on first- or second-order derivatives of the displacement vector field, which are applied at the pixel level in discrete implementation and are insufficient for studying the myocardium in echocardiography. The high noise-to-signal ratio in echocardiography degrades the quality of the estimated motion field, despite the pixel-wise constraints in deep learning methods. Other methods tackle the smoothness issue at the transformation level, for instance using regionally affine deformations [16].

In this work, we introduce a deep learning-based motion model leveraging such approach and tailored for 2D echocardiography motion estimation, called the PolyAffine Motion model (PAM). It incorporates global-level smoothness regularization for the myocardium through polyaffine motion fusion. Our contributions include:

- Proposing a comprehensive pipeline for using PAM in a weakly-supervised manner on a large public echocardiography dataset, demonstrating the effectiveness of our approach for unsupervised learning of myocardial motion.
- Conducting extensive experiments on various real-world datasets to show that PAM outperforms other popular deep learning-based motion estimation methods, thereby highlighting its potential for clinical applications.

2 Methodology

2.1 Polyaffine motion model

Given source image I_S and target image I_T, we would like to estimate a dense motion field $\mathcal{T}_{S\leftarrow T} : \mathrm{R}^2 \to \mathrm{R}^2$ from target image I_T to source image I_S such that

$$I_S(\mathcal{T}_{S\leftarrow T}(\cdot)) = I_T(\cdot). \tag{1}$$

Inspired by the polyaffine motion fusion framework proposed by Arsigny et al. [4], we adapted the motion estimation module from [23] to develop our proposed method for cardiac motion estimation in echocardiography.

The Polyaffine motion model consists of two steps. Firstly, the motion of the left ventricle is approximated through the sparse motion of several key points. An encoder-decoder network is used to output the location of key points and their local affine mapping for both I_S and I_T separately. Secondly, from the sparse motion, we obtain the final dense motion field through polyaffine motion fusion. The proposed method is illustrated in Fig. 1.

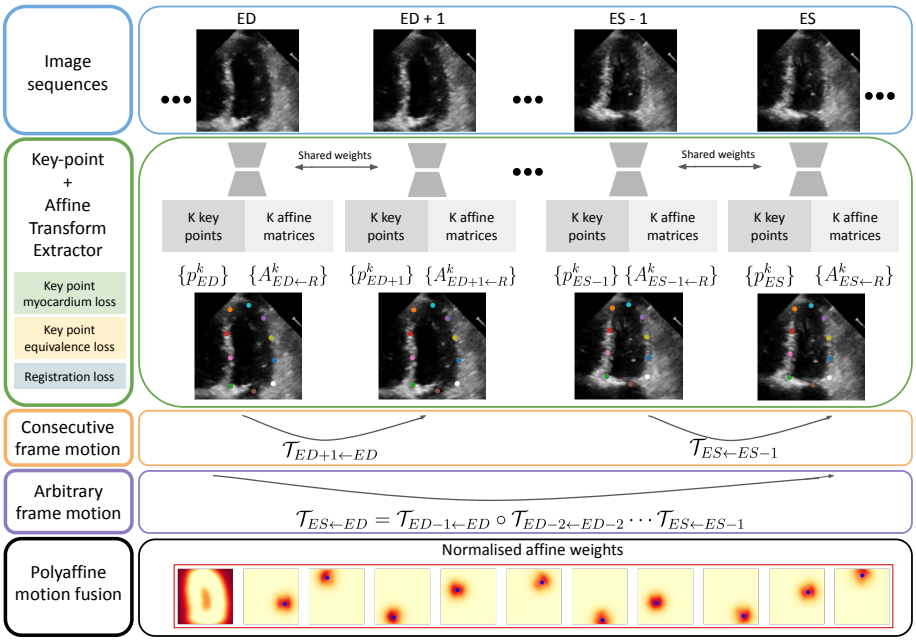

Fig. 1: Method overview.

Key point and affine transformation estimation We adopt the encoder-decoder architecture for key point extraction as presented in [23] and provide a brief review of the method from the point of view of affine transformation. In order to process each image separately, we assume there exists an abstract reference frame R. Given an image X, the encoder-decoder network outputs the estimated key points $p^k_X, k = 1, 2, ...K$ as well as the corresponding linear mappings $A^k_{X\leftarrow R} \in \mathcal{R}^{2\times 2}, k = 1, 2, ...K$. The local affine transformation from target image I_T to source image I_S is then computed using the following equation

$$\mathcal{T}^k_{S\leftarrow T}(z) = \underbrace{\bar{A}^k_{S\leftarrow T} \cdot z}_{\text{Linear mapping}} + \underbrace{(p^k_S - \bar{A}^k_{S\leftarrow T} \cdot p^k_T)}_{\text{Translation}}, \tag{2}$$

where $z \in \mathrm{R}^2$ represents the coordinate in the target image, and $\bar{A}^k_{S\leftarrow T} = A^k_{S\leftarrow R}(A^k_{T\leftarrow R})^{-1}$. In order to capture motion around the myocardium, we in-

troduce a myocardium-related key-point prior (Section 3.2) and associated loss functions (Section 2.2). This encourages the network to output key points close to the myocardium, in contrast to [23], which employed a self-supervised approach for learning key-point positions.

Polyaffine motion fusion Once we obtain the local affine transformation, the dense motion field is computed through direct polyaffine motion fusion. For each local affine motion, a spatial weight $W_k(p_T^k, \sigma^2)$ controls its influenced region. It is a 2D isotropic Gaussian distribution centered at key point p_T^k with variance σ^2. W_0 represents the weight for the background region and the left ventricle cavity area. It is computed as follows:

$$W_0 = \mathbf{RELU}(1 - \sum_{k=1}^{K} W_k(p_T^k, \sigma^2)). \tag{3}$$

We normalise all weights into $\bar{W}_k = \frac{W_k}{\sum_{k=0}^{K} W_k}$, where $k = 0, 1, 2...K$. An example is shown in Fig. 1. We apply them to compute the polyaffine dense motion

$$\mathcal{T}_{S \leftarrow T}(z) = \bar{W}_0 z + \sum_{k=1}^{K} \bar{W}_k(p_T^k) \mathcal{T}_{S \leftarrow T}^k(z). \tag{4}$$

The original first-order motion model (FOMM) [23] utilised a second encoder-decoder network to estimate the normalised weights for each local affine transformation, which can not guarantee the center of weights close to the corresponding key point, thereby unstable for myocardial motion estimation.

Sequence motion estimation Considering a sequence of image frames $I_1, I_2, ...I_N$, the assumption of abstract reference frame enables a fast way to obtain the dense motion field between any arbitrary pair of frames in the sequence. Without loss of generality, we assume I_{ED} is the frame at end-diastole and I_{ES} is the frame at end-systole and that I_{ED} is ahead of I_{ES} in time. The local affine transformation from I_{ED} to I_{ES} can be calculated by composition:

$$\begin{aligned} \mathcal{T}_{ES \leftarrow ED}^k(z) &= \mathcal{T}_{ED+1 \leftarrow ED}^k \circ \mathcal{T}_{ED+2 \leftarrow ED+1}^k \cdots \mathcal{T}_{ES \leftarrow ES-1}^k \\ &= \bar{A}_{ES \leftarrow ED}^k \cdot z + (p_{ES}^k - \bar{A}_{ES \leftarrow ED}^k \cdot p_{ED}^k), \end{aligned} \tag{5}$$

where $\bar{A}_{ES \leftarrow ED}^k = A_{ES \leftarrow R}^k (A_{ED \leftarrow R}^k)^{-1}$. The final dense motion is computed by combining the local motion, as described in Equation 4.

2.2 Loss functions

The model is trained end-to-end using a combination of loss functions, which can be grouped into the following subsets for sequence motion estimation.

Keypoint myocardium losses We propose two loss functions to enforce the position of key points near the myocardium region. The first loss, denoted by $\mathcal{L}_{\text{kp\_prior}}$, penalizes the L_2 norm between the estimated key-point position and a prior position. We obtain the prior key-point position from the available training masks, as described in Section 3.2. The second loss, denoted by $\mathcal{L}_{\text{kp\_myo\_ED}}$, constrains key points at end-diastole (ED) to reside within the myocardium region by penalizing the distance between each key point and the myocardium.

$$\mathcal{L}_{\text{kp\_myo\_ED}} = -\sum_{k=1}^{K} H(p^k, \sigma_H^2) * (\text{Mask}_{\text{myo\_ED}} - 0.5), \tag{6}$$

where $H(p^k, \sigma_H^2)$ is the isotropic Gaussian heatmap centered at p^k with variance σ_H^2 and $\text{Mask}_{\text{myo\_ED}}$ the binary mask of myocardium at ED.

Keypoint equivalence losses Another two losses $\mathcal{L}_{\text{equi\_kp}}$ and $\mathcal{L}_{\text{equi\_affine}}$ impose an equivalence constraint to the detected key points [23]. They force the model to predict consistent key points and local linear mappings when applying a known transformation to the input image.

Registration losses The last four losses regularize the final dense motion field through image similarity and key-point similarity. First, for each pair of consecutive frames in the sequence, we constrain

$$\mathcal{L}_{\text{seq\_im}} = \sum_{seq} NCC(I_T, I_S(\mathcal{T}_{I_S \leftarrow I_T})), \tag{7}$$

$$\mathcal{L}_{\text{seq\_kp}} = \sum_{seq} \frac{1}{K} \sum_{k=1}^{K} |H(p_T^k, \sigma_H^2) - H(p_S^k, \sigma_H^2)(\mathcal{T}_{I_S \leftarrow I_T})|, \tag{8}$$

where NCC represents the normalised cross correlation. Second, in order to force the model to learn temporal motion, we apply a second pair of similarity loss between the image at end-diastole and all frames after the end-diastole frame, denoted as $\mathcal{L}_{\text{ED2any\_im}}$ and $\mathcal{L}_{\text{ED2any\_kp}}$ for image similarity and key-point similarity respectively.

The total loss for the PAM model becomes

$$\begin{aligned} \mathcal{L}_{\text{total}} =& \lambda_1 \mathcal{L}_{\text{kp\_prior}} + \lambda_2 \mathcal{L}_{\text{kp\_myo\_ED}} + \lambda_3 \mathcal{L}_{\text{equi\_kp}} + \lambda_4 \mathcal{L}_{\text{equi\_affine}} \\ &+ \lambda_5 \mathcal{L}_{\text{seq\_im}} + \lambda_6 \mathcal{L}_{\text{seq\_kp}} + \lambda_7 \mathcal{L}_{\text{ED2any\_im}} + \lambda_8 \mathcal{L}_{\text{ED2any\_kp}}. \end{aligned} \tag{9}$$

3 Experiments

3.1 Datasets

Three public datasets of echocardiography are included in our study. EchoNet[1] [18] contains 10030 echocardiography videos of long-axis 4 chamber view. Left-ventricle

[1] `https://echonet.github.io/dynamic/`

tracings at end-diastole and end-systole are available (see Fig. 2(a)). CAMUS dataset[2] [14] consists of apical 4-chamber view videos from 450 patients whose left heart segmentation masks are publicly accessible (see Fig. 2(e)). HMC-QU dataset[3] [9] contains 109 4-chamber view echocardiography recordings with the segmentation of the left myocardium along one cardiac cycle (see Fig. 2(d)). All

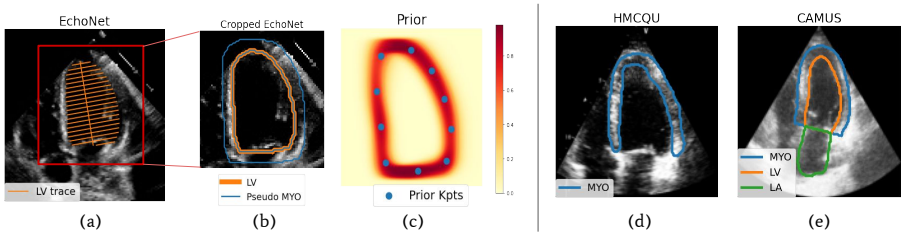

Fig. 2: Dataset overview. (a) EchoNet example and the given annotations of left ventricle tracing. (b) Left ventricle cropping with the generated pseudo myocardium contour (blue). (c) The mean mask from the EchoNet training set and the 10 prior key points. (d) HMC-QU example and the given annotation of the myocardium. (e) CAMUS example and the given annotation of different cardiac structures. (MYO: myocardium, LV: left ventricle, LA: left atrium)

image sequences are cropped around the left ventricle center according to the provided segmentation/tracings (see Fig. 2(b)). In this study, we follow the given data division of EchoNet dataset, with 7465 samples for training, 1288 samples for validation and 1277 samples for testing. CAMUS and HMC-QU datasets are used for evaluation during test phase.

3.2 Dataset preprocessing

Pseudo myocardium mask There is no ground truth of myocardium mask in EchoNet dataset. To provide guidance for key points, we generate a pseudo myocardium mask for ED and ES by applying a dilation operation using 13x13 and 17x17 structure element to the left ventricle mask at ED and ES frame respectively. The difference between the dilated mask and the original one is regarded as the pseudo myocardium mask (see Fig. 2(b)).

Key-point prior at end-diastole A mean myocardium mask is computed using all the pseudo myocardium masks at ED from the training set. Then, all the pixels of the mean mask are clustered into 10 groups using the KMeans function from the scikit-learn package [20], where the center of each group is considered as one key-point prior (see Fig. 2(c)).

3.3 Implementation

We compare our proposed model with the conditional variational autoencoder (CVAE) method proposed in [13], which has demonstrated effectiveness in sequential motion modelling of cardiac images. Its former registration version [12]

[2] `https://www.creatis.insa-lyon.fr/Challenge/camus/`
[3] `https://www.kaggle.com/datasets/aysendegerli/hmcqu-dataset`

has shown more regular motion field than other deep learning methods, including VoxelMorph [7]. We implemented the CVAE method following its description in [13] and used the same training hyper-parameters. However, we added a Dice loss between the transformed end-systolic (ES) mask and end-diastolic (ED) mask during CVAE training to keep it consistent with our PAM model, which used segmentation mask information during training (see Section 2.2).

We trained the PAM model using an Adam optimizer with a learning rate of $1e-4$. To determine the optimal hyperparameters, we conducted experiments on a randomly selected subset of 1000 training examples from the EchoNet dataset. The hyperparameters for the loss function were set to $\lambda_1 = 20, \lambda_2 = 0.1, \lambda_3 = 50, \lambda_4 = 50, \lambda_5 = \lambda_6 = \lambda_7 = \lambda_8 = 100$. The variance of the Gaussian heatmap was set to $\sigma^2 = 0.05$ for affine transformation weights (see Equation 3), and $\sigma_H^2 = 0.005$ for all other scenarios. We trained the model for a maximum of 100 epochs, and applied early stopping if the validation loss did not improve over 10 epochs.

4 Results

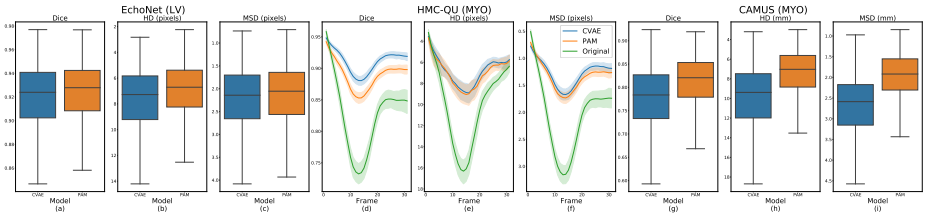

Fig. 3: Evaluation results on test datasets. (a-c) On EchoNet test split using available left ventricle masks of ED/ES. (d-f) On HMC-QU dataset using frame-wise myocardium masks. The original line represents the comparison between the ground truth ED frame and the ground truth mask of each frame. (g-i) On CAMUS 4-chamber view dataset using myocardium masks of ED/ES. (HD: Hausdorff distance, MSD: main surface distance)

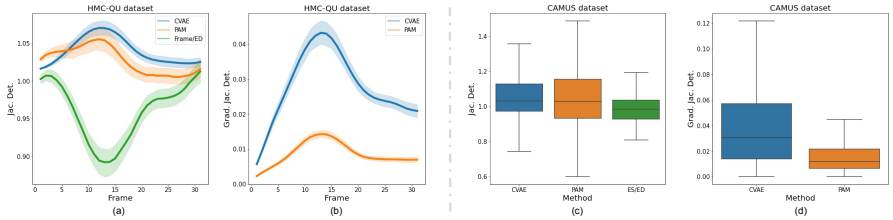

Fig. 4: Evaluation of Jacobian Determinant and its gradient in the myocardium region on the HMC-QU dataset (a-b) and on the CAMUS dataset (c-d). We also present the area change of the myocardium between the considered frame and end-diastole.

We first evaluated the motion estimation accuracy by assessing the registration performance using the available segmentation masks from all the test

datasets. For the EchoNet test split, we compared the ground truth left ventricle mask of end-diastole (ED) with that transformed from end-systole (ES). For the CAMUS dataset, the same evaluation was applied to the myocardium mask for ED/ES. For the HMC-QU dataset, the myocardium masks of one cardiac cycle were all transformed to ED using the estimated motion field. To enable group-level statistical analysis along the cardiac cycle, the temporal metric was interpolated to the same length. Our proposed PAM model demonstrated significant out-performance compared to the CVAE model on the test sets of EchoNet and CAMUS datasets. However, it showed slightly lower Dice scores on the HMC-QU dataset, which may be due to the fact that the pseudo-labelling process for generating the ground truth segmentation of the myocardium on the HMC-QU dataset [9] may be not very accurate. Furthermore, the PAM model exhibited more regular deformation fields compared to the baseline model (as shown in Fig. 4) when evaluating the gradient of Jacobian determinant. This smooth displacement field is advantageous for computing dense strain tensor, which is typically constructed using the second-order derivative of the displacement field. Additionally, as listed in Table 1, we demonstrated that our proposition of using myocardium prior and the polyaffine fusion mechanism greatly improved the registration performance, when compared with the original first-order motion model (FOMM) [23]. In particular, our explicit design of fusion weights helps the network to efficiently learn affine transformation locally.

Method	EchoNet-LV			CAMUS-MYO		
	Dice	HD *(pixels)*	MSD *(pixels)*	Dice	HD *(mm)*	MSD *(mm)*
All = PAM (ours)	**0.92**	**7.33**	**2.23**	**0.81**	**7.43**	**1.99**
FOMM+Prior+Polyaffine	0.91	7.53	2.36	0.80	7.95	2.08
FOMM+Prior	0.77	18.36	5.46	0.57	13.85	3.89
FOMM [23]	0.75	22.99	6.07	0.48	17.27	4.41
CVAE [13]	0.91	7.97	2.30	0.77	9.91	2.67

Table 1: Performance comparison of different methods on EchoNet and CAMUS datasets. FOMM: the original model in [23], without prior of any key points (Prior), without explicit design of polyaffine weights (Polyaffine), without registration penalty between ED and any other frames (Sequence). All = FOMM+Prior+Polyaffine+Sequence.

In addition to the good performance from various evaluation metrics, the proposed PAM model has the potential for abnormal wall motion detection. We provide examples of myocardial infarction (MI) from the HMC-QU dataset (see Supplementary video material) and show the abnormal strain values in different segments, highlighting the consistency of our findings with the ground truth diagnosis.

5 Discussion and conclusion

In this paper, we proposed a polyaffine motion model (PAM) for echocardiography motion estimation. The PAM model demonstrated excellent motion esti-

mation performance on real-world echocardiography datasets and showed good generalization to unseen datasets from other centers. Our explicit design of fusion weights enabled efficient learning of local affine transformation, and the intrinsic polyaffine structure improved the smoothness of the motion field, showing potential for abnormal wall motion detection. In the future, we will focus on integrating temporal regularization for the PAM model and conducting evaluations on synthetic datasets with known ground-truth displacement.

Acknowledgements 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) and by Inria PhD funding. The authors are grateful to the OPAL infrastructure from Université Côte d'Azur for providing resources and support.

References

1. Ahn, S.S., Ta, K., Lu, A., Stendahl, J.C., Sinusas, A.J., Duncan, J.S.: Unsupervised motion tracking of left ventricle in echocardiography. In: Medical imaging 2020: Ultrasonic imaging and tomography. vol. 11319, pp. 196–202. SPIE (2020)
2. Alessandrini, M., Chakraborty, B., et al.: 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**(3), 411–422 (2017)
3. Alessandrini, M., Heyde, B., et al.: Detailed evaluation of five 3d speckle tracking algorithms using synthetic echocardiographic recordings. IEEE transactions on medical imaging **35**(8), 1915–1926 (2016)
4. Arsigny, V., Commowick, O., Ayache, N., Pennec, X.: A fast and log-euclidean polyaffine framework for locally linear registration. Journal of Mathematical Imaging and Vision **33**(2), 222–238 (2009)
5. Azarmehr, N., Ye, X., et al.: An optimisation-based iterative approach for speckle tracking echocardiography. Medical and Biological Engineering and Computing **58**(6), 1309–1323 (2020)
6. Barbosa, D., Heyde, B., Dietenbeck, T., et al.: Fast Left Ventricle Tracking in 3D Echocardiographic Data Using Anatomical Affine Optical Flow. In: Ourselin, S., Rueckert, D., Smith, N. (eds.) Functional Imaging and Modeling of the Heart. pp. 191–199. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
7. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: MICCAI 2018: Granada, Spain, September 16-20, 2018, Proceedings, Part I. pp. 729–738. Springer (2018)
8. De Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis **52**, 128–143 (2019)
9. Degerli, A., Zabihi, M., Kiranyaz, S., Hamid, T., Mazhar, R., Hamila, R., Gabbouj, M.: Early detection of myocardial infarction in low-quality echocardiography. IEEE Access **9**, 34442–34453 (2021)
10. Evain, E., Sun, Y., Faraz, K., Garcia, D., et al.: Motion estimation by deep learning in 2d echocardiography: synthetic dataset and validation. IEEE transactions on medical imaging (2022)

11. Farneb, G.: Two-Frame Motion Estimation Based on Polynomial Expansion. Scandinavian conference on Image analysis **2749**(1), 363–370 (2003)
12. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE Transactions on Medical Imaging **38**(9), 2165–2176 (2019)
13. Krebs, J., Mansi, T., Ayache, N., Delingette, H.: Probabilistic motion modeling from medical image sequences: application to cardiac cine-mri. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 176–185. Springer (2020)
14. Leclerc, S., Smistad, E., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE transactions on medical imaging **38**(9), 2198–2210 (2019)
15. Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N.: ilogdemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. International journal of computer vision **92**, 92–111 (2011)
16. McLeod, K., Sermesant, M., Beerbaum, P., Pennec, X.: 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)
17. Østvik, A., Salte, I.M., Smistad, E., Nguyen, T.M., Melichova, D., et al.: Myocardial function imaging in echocardiography using deep learning. IEEE Transactions on Medical Imaging **40**(5), 1340–1351 (2021)
18. Ouyang, D., He, B., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020)
19. Papademetris, X., Sinusas, A.J., Dione, D.P., Duncan, J.S.: Estimation of 3d left ventricular deformation from echocardiography. Medical image analysis **5**(1), 17–28 (2001)
20. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
21. Qin, C., Bai, W., Schlemper, J., Petersen, S.E., Piechnik, S.K., Neubauer, S., Rueckert, D.: Joint learning of motion estimation and segmentation for cardiac mr image sequences. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 472–480. Springer (2018)
22. Qin, C., Wang, S., Chen, C., Qiu, H., Bai, W., Rueckert, D.: Biomechanics-informed neural networks for myocardial motion tracking in mri. In: International conference on medical image computing and computer-assisted intervention. pp. 296–306. Springer (2020)
23. Siarohin, A., Lathuilière, S., et al.: First order motion model for image animation. Advances in Neural Information Processing Systems **32** (2019)
24. Staring, M., Klein, S., Pluim, J.P.: A rigidity penalty term for nonrigid registration. Medical physics **34**(11), 4098–4108 (2007)
25. Ta, K., Ahn, S.S., Lu, A., Stendahl, J.C., Sinusas, A.J., Duncan, J.S.: 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). pp. 1734–1737. IEEE (2020)
26. Zhang, X., You, C., et al.: Learning correspondences of cardiac motion from images using biomechanics-informed modeling. In: Statistical Atlases and Computational Models of the Heart. pp. 13–25. Springer Nature Switzerland, Cham (2022)