LaMoD: Latent Motion Diffusion Model For Myocardial Strain Generation

Jiarui Xing[1*], Nivetha Jayakumar[1], Nian Wu[1], Yu Wang[3], Frederick H. Epstein[3], and Miaomiao Zhang[1,2]

[1] Department of Electrical and Computer Engineering, University of Virginia, USA
[2] Department of Computer Science, University of Virginia, USA
[3] Department of Biomedical Engineering, University of Virginia Health System, USA
`jx8fh@virginia.edu`

Abstract. Motion and deformation analysis of cardiac magnetic resonance (CMR) imaging videos is crucial for assessing myocardial strain of patients with abnormal heart functions. Recent advances in deep learning-based image registration algorithms have shown promising results in predicting motion fields from routinely acquired CMR sequences. However, their accuracy often diminishes in regions with subtle appearance changes, with errors propagating over time. Advanced imaging techniques, such as displacement encoding with stimulated echoes (DENSE) CMR, offer highly accurate and reproducible motion data but require additional image acquisition, which poses challenges in busy clinical flows. In this paper, we introduce a novel Latent Motion Diffusion model (LaMoD) to predict highly accurate DENSE motions from standard CMR videos. More specifically, our method first employs an encoder from a pre-trained registration network that learns latent motion features (also considered as deformation-based shape features) from image sequences. Supervised by the ground-truth motion provided by DENSE, LaMoD then leverages a probabilistic latent diffusion model to reconstruct accurate motion from these extracted features. Experimental results demonstrate that our proposed method, LaMoD, significantly improves the accuracy of motion analysis in standard CMR images; hence improving myocardial strain analysis in clinical settings for cardiac patients. Our code is publicly available at `https://github.com/jr-xing/LaMoD`.

1 Introduction

Motion and deformation analysis of CMR videos provides valuable measurements to quantify myocardial strain in patients with heart disease, offering clinically significant data for disease assessment, diagnosis, and treatment planning [17,3,24,29,30]. CMR feature tracking (FT) is widely used in clinical practice to evaluate myocardial motion, deformation, and strain functions [25,15,18,19]. Such a technique employs optical flow-based algorithms [10] to track image features or patterns within the myocardium throughout the cardiac cycle. While FT is convenient and integrates well into routine clinical workflows, this technique is generally less accurate due to its limited motion tracking accuracy [1,32].

With recent advancements in deep learning, several research groups have utilized image registration-based networks to predict myocardial motion and strain from CMR images [15,18,31]. Existing methods incorporated new regularization techniques, such as enforcing temporal consistency [18], or introducing prior knowledge of cardiac biomechanics [21,35,20], to achieve improved motion prediction quality. However, these approaches still partially address the challenges of detecting motion in regions with subtle image appearance changes (i.e., the myocardium mid-wall motion) [1,6,16], leading to compromised accuracy of myocardial strain measurements. To alleviate this issue, another research line has utilized "ground-truth" data from the advanced imaging technique DENSE, which provides highly accurate and reproducible myocardial motion data to supervise the motion learning process [27,28]. In contrast to conventional CMR techniques, DENSE directly encodes tissue displacement within the imaging data, allowing for precise quantification of myocardial motion throughout the cardiac cycle. Despite promising progress in predicting myocardial motion for improved strain analysis under the guidance of DENSE [27,28], current approaches heavily rely on features extracted from segmented myocardial contours rather than the underlying cardiac motions, potentially compromising strain accuracy. Furthermore, the network training involved only DENSE contours that may not fully generalize to standard CMR images.

To address this problem, we propose to develop a novel Latent Motion Diffusion model (LaMoD) to further improve the current deep networks to predict highly accurate DENSE motions from standard CMR videos. In particular, our method, LaMoD, first employs an encoder from a pre-trained registration network that learns latent motion features of myocardial deformations derived from both cine and DENSE image sequences. Supervised by the ground-truth motion provided by DENSE, LaMoD then leverages a probabilistic latent diffusion model to reconstruct accurate motion from these extracted features. Once our model is trained, the DENSE data is no longer required in the testing phase. Our contributions are threefold:

(i) Develop a new framework, LaMoD, that generates time-sequential myocardial deformation fields from a learned latent space of motion features.
(ii) Our method is the first to leverage latent diffusion models in the motion space to produce highly accurate myocardial strain from standard CMR videos.
(iii) Opens promising research avenues for transferring knowledge from advanced strain imaging to routinely acquired CMR data; hence maximizing benefits for patients with cardiac diseases.

2   Background: Registration-based Motion Deformation Estimation

In this section, we briefly review the concept of image registration, which is a fundamental technique for estimating motion deformation between images. We will apply this concept to motion tracking in CMR video sequences.

Given a source image S and a target image I, the problem of diffeomorphic image registration is typically formulated as an energy minimization over a time-dependent deformation fields, $\{\phi_t : t \in [0,1]\}$, i.e.,

$$E(\phi_1) = \frac{1}{2\sigma^2} \operatorname{Dist}(S \circ \phi_1^{-1}, I) + \operatorname{Reg}(\phi_1). \tag{1}$$

Here, $\circ$ denotes an interpolation operator, which deforms a source image S to match the target I. The $\operatorname{Dist}(\cdot, \cdot)$ is a distance function that measures the dissimilarity between images weighted by a positive parameter σ, and $\operatorname{Reg}(\cdot)$ is a regularization term to enforce the smoothness of transformation fields. In this paper, we use a commonly used sum-of-squared intensity differences (L_2-norm) [4] as the distance function.

In our implementation, we adopt the large diffeomorphic deformation metric mapping (LDDMM) framework [4,34] to generate diffeomorphic deformations, parameterized by time-dependent velocity fields, $v_t : t \in [0,1]$, i.e.,

$$\frac{d\phi_t}{dt} = v_t \circ \phi_t. \tag{2}$$

A geodesic shooting algorithm [13,26] has shown that the geodesic path of deformation, ϕ_t, with a given initial condition, v_0, can be uniquely determined through integrating the Euler-Poincaré differential equation (EPDiff) [2,14] as

$$\frac{dv_t}{dt} = -K \left[(Dv_t)^{\mathrm{T}} m_t + Dm_t \, v_t + m_t \operatorname{div} v_t \right], \tag{3}$$

where D denotes a Jacobian matrix and div is the divergence operator. The K denotes an inverse operator of $L : V \to V^*$, which is a symmetric, positive-definite differential operator that maps a tangent vector $v \in V$ into the dual space $m \in V^*$. Numerous studies and applications have used the initial velocities, v_0, to represent diffeomorphic deformations in the context of motion/deformation analysis [33,13].

3 Our Model: LaMoD

Our model consists of two main components: (i) a pre-trained registration network based on the LDDMM framework [4] to extract latent motion features from CMR sequences, and (ii) a motion prediction model leveraging latent diffusion models to reconstruct realistic and highly accurate myocardial motion, supervised by DENSE data. An overview of our proposed framework, LaMoD, is illustrated in Fig. 1.

3.1 Latent Motion Feature Learning From Image Videos

Given a image sequence, $\{I^\tau\}_{\tau=0}^T$, that includes $T+1$ time frames covering an entire motion cycle of myocardium. We pre-train a LDDMM-based registration
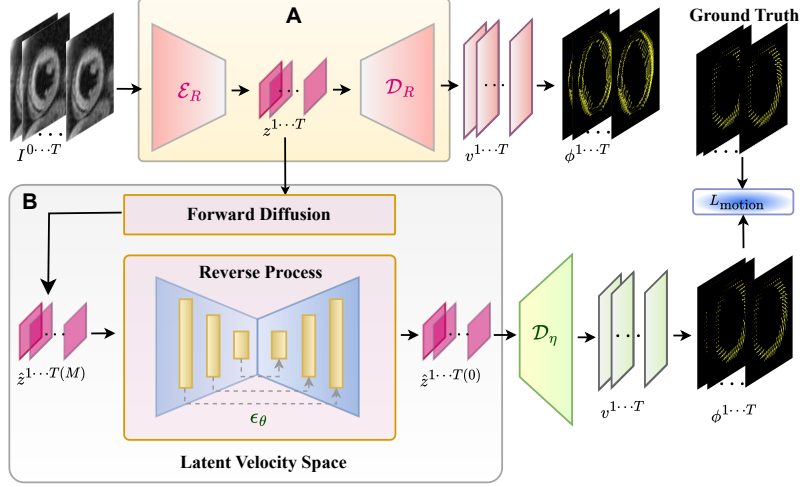
Fig. 1: An overview of our proposed network framework. (A) Registration-based network to learn latent motion features represented by initial velocity fields. (B) Diffusion model in latent motion spaces.

network [8] to learn the deformation fields between the initial frame I^0 and each subsequent frame $\{I^\tau\}$. This results in a number of T pairs of images, i.e., $\{(I^0, I^1), (I^0, I^2) \cdots, (I^0, I^T)\}$. To utilize the intrinsic spatial connections and motion consistency across the sequence of time frames, we stack the T pairwise images into a 3D volume and predict the deformation fields simultaneously in the implementation. We employ an UNet architecture [23] as our network backbone, featuring an encoder $\mathcal{E}_R$ that maps the input image pairs $\{(I^0, I^\tau)\}$ to the corresponding latent velocity features $\{z^\tau\}$, and a decoder $\mathcal{D}_R$ that project $\{z^\tau\}$ back to the input image space, i.e., initial velocity fields $\{v_0^\tau\}$. The corresponding final transformation fields, $\{\phi_1^\tau\}$, are computed through Eq. (2) and Eq. (3). For a simplified notation, we will drop the time index in following sections, i.e., $\phi_1^\tau \overset{\Delta}{=} \phi^\tau$ and $v_0^\tau \overset{\Delta}{=} v^\tau$.

3.2 Diffusion Model In Latent Motion Spaces

We introduce a new latent motion diffusion module that captures the complex distribution of latent motion features, resulting in higher quality motion reconstruction under the supervision of DENSE ground-truth. More specifically, the registration-based latent velocity features are first refined through a diffusion process and then fed into a motion reconstruction network to predict the final, highly accurate myocardial motion. This approach ensures a more precise and realistic depiction of myocardial dynamics, demonstrating significant improvements over existing methods.

Inspired by the Denoising Diffusion Probabilistic Models (DDPM)[9,22], we formulate the model as a latent Markov chain with M steps. The framework has a forward and reverse diffusion process. The forward process iteratively adds random Gaussian noise to the input latent features over $m \in [1, 2, ..., M]$ steps. Similar to [11], we employ smoothed Gaussian noise ϵ' in the diffusion process to facilitate faster optimization convergence compared to normal Gaussian noise ϵ, i.e. $\epsilon' = \mathcal{K}(\epsilon)$, where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ and $\mathcal{K}(\cdot)$ is a Gaussian smoothing kernel.

Defining $\boldsymbol{z}^{(m)} \triangleq \{z^{1(m)}, z^{2(m)}, \ldots, z^{T(m)}\}$ as the learned latent motion features (represented by initial velocities) for all frames at step m, our **forward diffusion process** is defined as:

$$q(\boldsymbol{z}^{(m)}|\boldsymbol{z}^{(m-1)}) = \mathcal{N}(\boldsymbol{z}^{(m)}; \sqrt{1 - \beta_m}\boldsymbol{z}^{(m-1)}, \beta_m \mathbf{I}) \tag{4}$$

where β_m is a time-dependent variance schedule used to parameterize the probabilistic transitions. The forward process starts from the registration-based latent feature, i.e. $\boldsymbol{z}^{(0)} = \{z^\tau\} = \{z^1, z^2, \ldots, z^T\}$. Alternatively, the forward diffusion process can be formulated as a single step process:

$$q(\boldsymbol{z}^{(m)}|\boldsymbol{z}^{(0)}) = \mathcal{N}(\boldsymbol{z}^{(m)}; \sqrt{\bar{\alpha}_m}\boldsymbol{z}^{(0)}, (1 - \bar{\alpha}_m)\mathbf{I}) \tag{5}$$

$$\boldsymbol{z}^{(m)} = \sqrt{\bar{\alpha}_m}\boldsymbol{z}^{(0)} + \sqrt{1 - \bar{\alpha}_m}\,\mathcal{K}(\epsilon), \tag{6}$$

where $\bar{\alpha}_m = \prod_{i=1}^m \alpha_i$ and $\alpha_m = 1 - \beta_m$.

The reverse diffusion process aims to restore the latent feature with a deep neural network. It takes place over m steps as:

$$p_\theta(\hat{\boldsymbol{z}}^{(m-1)}|\hat{\boldsymbol{z}}^{(m)}) = \mathcal{N}(\hat{\boldsymbol{z}}^{(m-1)}; \mu_\theta(\hat{\boldsymbol{z}}^{(m)}, m), \Sigma_\theta(\hat{\boldsymbol{z}}^{(m)}, m)), \tag{7}$$

where $\hat{\boldsymbol{z}}^{(M)}$ are the latent velocity features obtained from the reverse process, $\mu_\theta = \frac{1}{\sqrt{\alpha_m}}(\hat{\boldsymbol{z}}^{(m)} - \frac{\beta_m}{\sqrt{1-\bar{\alpha}_m}}\epsilon_\theta(\hat{\boldsymbol{z}}^{(m)}, m))$ and $\Sigma_\theta(\hat{\boldsymbol{z}}^{(m)}, m) = \beta_m \mathbf{I}$ are the mean and variance of the transitions of the reverse process respectively.

The **reverse process** is implemented by training a noise-prediction network ϵ_θ that takes the noisy motion feature $\hat{\boldsymbol{z}}^{(m)}$ and the step m as input. For each step m, the sampling in the reverse process is defined as:

$$\hat{\boldsymbol{z}}^{(m-1)} = \frac{1}{\sqrt{\alpha_m}}(\hat{\boldsymbol{z}}^{(m)} - \frac{1 - \alpha_m}{\sqrt{1 - \bar{\alpha}_m}}\epsilon_\theta(\hat{\boldsymbol{z}}^{(m)}, m)) + \sigma_m \mathcal{K}(\gamma), \tag{8}$$

where $\gamma \sim \mathcal{N}(0, \mathbf{I})$.

Noting the shared diffusion step as m, and latent features as well as the additive noise of the n-th input sequence as $\boldsymbol{z}_n^{(m)}$ and ϵ'_n, respectively, the corresponding loss function over the whole training dataset with N sequences is defined as

$$l_{\text{diffusion}} = \sum_{n=1}^N E_{m,z,\epsilon}\left[\|\epsilon'_n - \epsilon_\theta(\boldsymbol{z}_n^{(m)}, m))\|_2\right] + \lambda_\epsilon \text{reg}(\theta), \tag{9}$$

where the function $\text{reg}(\cdot)$ represents network L_2 weight decay regularity weighted by λ_ϵ.

Algorithm 1 LaMoD: Latent Motion Diffusion Model For Myocardial Strain Generation

---

Input: DENSE CMR videos, $\{I^\tau\}$, with directly encoded motion, $\{\Phi^\tau\}$;
Output: Reconstructed myocardial motion, $\{\hat{\Phi}^\tau\}$;

    Latent Motion Feature Learning
1: $\boldsymbol{z}^{(0)} = \{z^1, \ldots, z^T\} = \mathcal{E}_R(\{I^0, \ldots, I^T\})$

    Diffusion Model Inference
2: $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
3: $\boldsymbol{z}^{(M)} = \sqrt{\bar{\alpha_M}}\boldsymbol{z}^{(0)} + \sqrt{1 - \bar{\alpha_M}}\,\mathcal{K}(\epsilon)$
4: **for** $m = M, M-1, \ldots, 1$ **do**
5:    $\hat{\boldsymbol{z}}^{(m-1)} = \frac{1}{\sqrt{\alpha_m}}(\hat{\boldsymbol{z}}^{(m)} - \frac{1-\alpha_m}{\sqrt{1-\alpha_m}}\epsilon_\theta(\hat{\boldsymbol{z}}^{(m)}, m)) + \sigma_m \mathcal{K}(\gamma)$
6: **end for**

7: $\{\hat{\Phi}^\tau\} = \mathcal{D}_\eta(\hat{\boldsymbol{z}}^{(0)});$
8: **return** $\{\hat{\Phi}^\tau\}$

The refined latent features $\boldsymbol{z}^{(0)}$ are fed into to the motion decoder, $\mathcal{D}_\eta$, to reconstruct the accurate myocardial motion of original data resolution, i.e., $\hat{\boldsymbol{\Phi}}_n = \{\hat{\Phi}_n^1, \hat{\Phi}_n^2, \ldots, \hat{\Phi}_n^T\} = \mathcal{D}_\eta(\boldsymbol{z}_n^{(0)})$, where η is the motion decoder and $n \in \{1, \cdots, N\}$ is the data index. With the supervision of DENSE ground truth, $\boldsymbol{\Phi}_n$, our motion reconstruction loss is defined as

$$l_{\text{motion}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=1}^{T}\|\Phi_n^\tau - \hat{\Phi}_n^\tau\|_2^2 + \lambda_M \text{reg}(\eta),$$

$$= \frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=1}^{T}\|\Phi_n^\tau - \mathcal{D}_\eta(\boldsymbol{z}_n^{(0)})\|_2^2 + \lambda_M \text{reg}(\eta), \quad (10)$$

where λ_M is the weighting parameter of weight decay regularization term $\text{reg}(\cdot)$.

We jointly train the loss functions of diffusion model (Eq. (9)) and the motion reconstruction (Eq. (10)) till its convergence. We define the total loss as

$$l_{\text{total}} = l_{\text{diffusion}} + \alpha l_{\text{motion}},$$

where α is the loss weighting term. We summarize our model in Algorithm 1.

4   Experimental Evaluation

We first demonstrate the effectiveness of LaMoD on DENSE CMR videos and then test its performance on standard cine CMRs, highlighting its clinical potential. Both quantitative and visualization results are presented. Note that our network is trained solely on the DENSE dataset and then tested on both DENSE and CINE datasets.

The training of all our experiments is implemented on an server with AMD EPYC 7502 CPU of 126GB memory and Nvidia GTX 3090Ti GPUs. We train

our networks using Adam optimizer [12] with maximal 2000 epochs with the early stop strategy. The batch size is set to 32 and weight decay weights are set to $\lambda_\epsilon = \lambda_M = 1E - 4$. The hyper-parameters are optimized with grid search strategy. The optimal learning rate is $1E - 4$ and the optimal loss weight is $\alpha = 1E - 2$.

4.1   Dataset

DENSE CMR videos with directly encoded motion data. We utilize 741 DENSE CMR videos of the left ventricular (LV) myocardium collected from 284 subject, including 124 healthy volunteers and 160 patients with various types of heart disease. Data are collected from eight centers (University of Virginia, Charlottesville; University Hospital, Saint-Etienne, France; University of Kentucky, Lexington; University of Glasgow, Scotland; St Francis Hospital, New York; the Royal Brompton Hospital, London, England; Emory University, Atlanta, Georgia; and Stanford University, Palo Alto, California). Each DENSE scan was performed in 4 short-axis planes at the basal, two midventricular, and apical levels (with temporal resolution of 17 ms, pixel size of 2.65^2 mm^2, and slice thickness=8mm). Other parameters include displacement encoding frequency $k_e = 0.1$ cycles/mm, flip angle 15°, and echo time = 1.08 ms.

Standard cine CMR videos. We tested our proposed model on 105 short-axis cine CMRs slices of 40 subjects from the DENSE dataset mentioned above, including 14 patients and 26 volunteers. All scans were acquired during repeated breath hold covering the left ventricle (LV) (field of view, 320×320 to 380×380 mm2; temporal resolution, 30–55 msec, depending on heart rate). Each selected cine slice corresponds to a DENSE scan of the same patient at the same spatial location (within ±2mm), allowing the DENSE displacement field to serve as the ground-truth motion for the cine slices.

Data Pre-processing. All cine and DENSE CMR sequences were temporally and spatially aligned for efficient network training. In particular, the standard cine sequences were temporally resampled to 40 frames to match the DENSE temporal resolution. All images were resampled to a 1 mm^2 resolution and cropped to 128×128. We ran all experiments on binary LV myocardium segmentation from both cine and DENSE sequences to avoid appearance gaps between the magnitude images, using masks manually labeled by clinical experts.

4.2   Experimental Design

We used all standard CMR and DENSE videos to train our registration network that can effectively learn latent motion features. However, we only use DENSE motion to train the diffusion model to reconstruct myocardial deformations. In particular, the DENSE data set was divided into 538 samples for training, 101 for validation and 102 for testing. We employed a site-balanced splitting strategy, maintaining consistent proportions from each site across all sets to ensure

representative sampling and mitigate site-specific biases. Paired DENSE data with cine CMR videos were used for regional segmental strain comparison in our experimental evaluation.

Evaluate motion and strain error on DENSE data. We first evaluate the quality of predicted motions for DENSE sequences, using pixel-wise motion field error (i.e., end-point error (EPE)) defined as the Euclidean distance between the predicted and ground truth motion vectors. We compare the performance of our proposed model with three state-of-the-art deep learning-based motion/strain prediction models — StrainNet [27], UNetR [7], 3D TransUNet [5]. All methods are trained on the same dataset, and their best performances are reported.

We then generate strain maps from the predicted motion fields for the DENSE input and compare it with all baseline algorithms. We performed a quantitative analysis of strain errors on regional segmental strains. Consistent with previous studies [27,28], we divided the myocardium in each DENSE slice into six segments, starting from the right ventricle insertion point and proceeding counterclockwise. We then calculated the average absolute error for strain evaluation in each segment.

Evaluate strain error on standard cine CMR videos. For cine sequences, our evaluation focuses exclusively on segmental strain error due to the significant differences between the paired DENSE contours and the collected cine CMR myocardium contours. This is mainly due to the spatial misalignment between the myocardium regions from input cine images and the DENSE-derived ground truth, which are collected from separate scans, making pixel-wise error computation impractical. We apply the same segmental strain approach as used for the DENSE data, dividing the myocardium into six segments and calculating the average absolute strain error for each. Furthermore, we compared strain values with those obtained using widely used commercial software (SuiteHeart version 5.0.4; NeoSoft) based on Feature Tracking (FT) in clinical settings.

4.3   Experimental Results

Fig. 2 (test on DENSE data) presents the visualizations of the predicted end-systolic displacement fields and the circumferential (Ecc) strain maps for all models compared to the ground truth from DENSE ground truth. To provide a comprehensive evaluation, we include examples from both healthy volunteers (top panel: ground A) and patients with heart failure (bottom panel: ground B), specifically those with left bundle branch block (LBBB). The predictions generated by our method consistently show a closer resemblance to the ground truth across all cases. These results collectively demonstrate that our method provides more accurate and robust performance compared to the baseline models.

Similarly, the Fig. 3 (test on standard cine CMRs) visualizes the predicted motion fields and the circumferential (Ecc) strain maps for all models compared to paired DENSE dataset. Note that the DENSE myocardium contours are slightly different from cine CMRs due to a different scanning time.
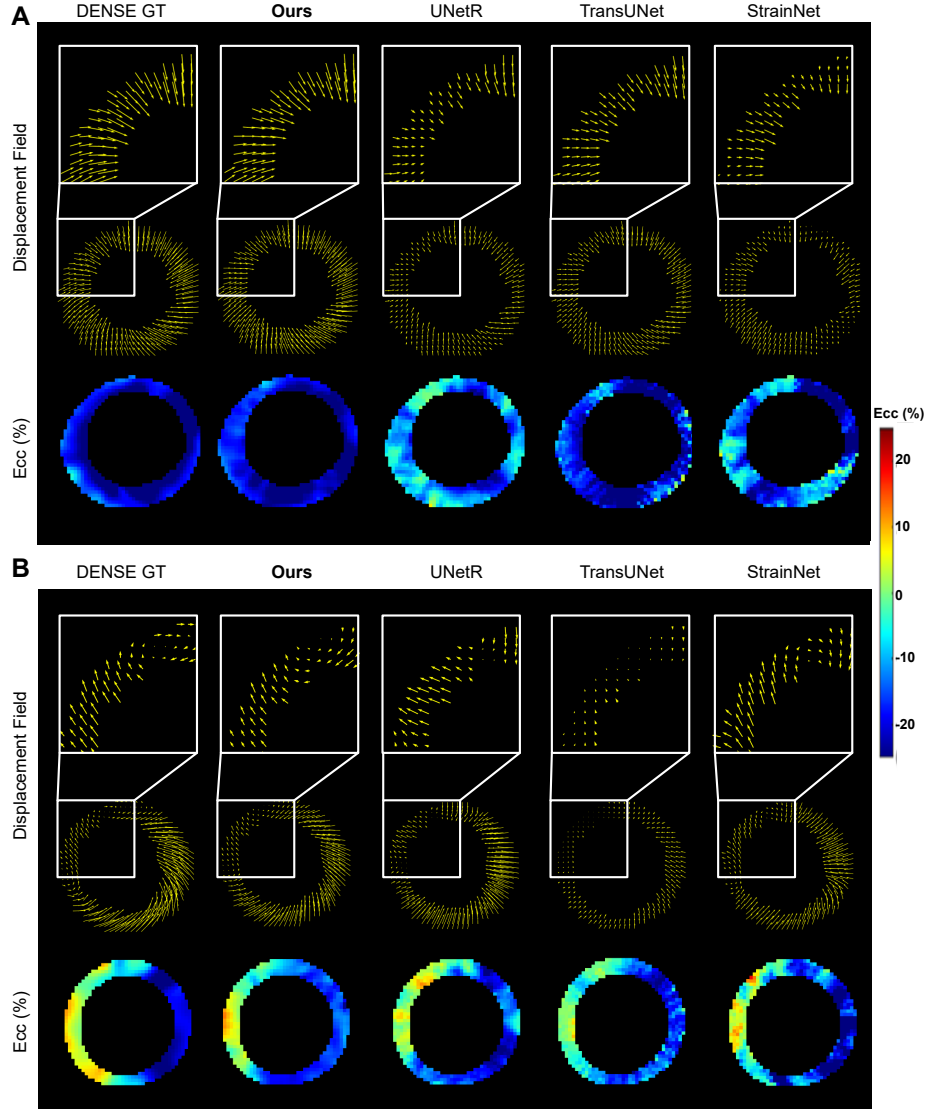
Fig. 2: Panel A (healthy volunteer) & B (heart failure patient with left bundle branch block): exemplary comparison of end-systolic displacement and circumferential strain maps (Ecc) derived from DENSE input across all methods. Left to right: DENSE ground truth and predictions from our model vs. baselines. Top to bottom: enlarged view of selected displacement region; full displacements; circumferential strain maps (contraction in blue vs. stretch in red).
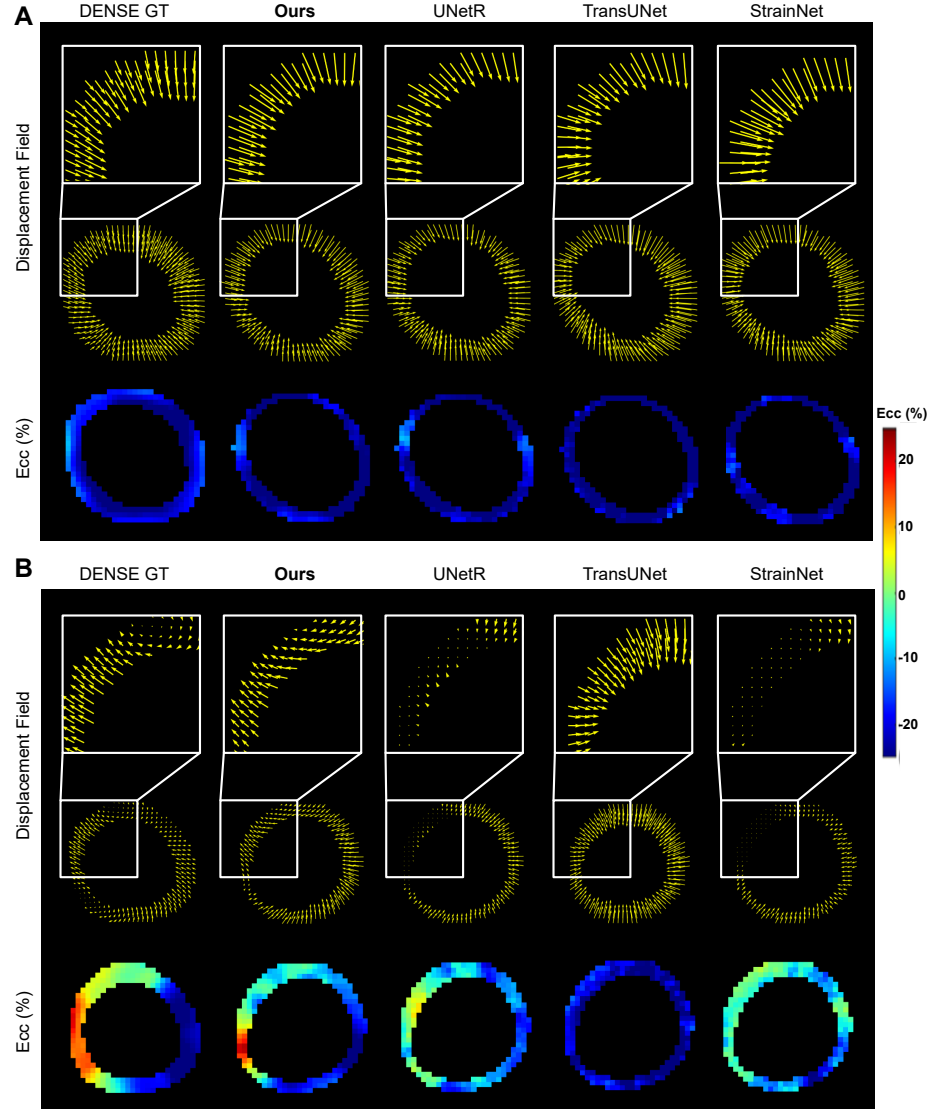
Fig. 3: Panel A (healthy volunteer) & B (heart failure patient with left bundle branch block): exemplary comparison of end-systolic displacement and circumferential strain maps (Ecc) derived from standard cine MRI videos.

Fig. 4 displays a quantitative comparison of displacement field EPE error and segmental circumferential strain error between our model and all baselines. The left two panels demonstrate the testing results on DENSE dataset, indicating that our method significantly outperforms the baselines in terms of both
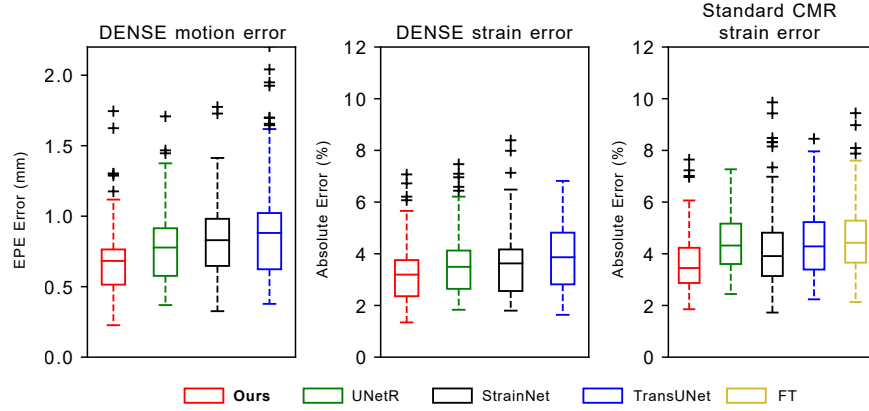
Fig. 4: Left to right: a comparison of displacement field EPE error on DENSE; predicted segmental circumferential strain error on DENSE; and predicted segmental circumferential strain error on standard cine CMRs from our model vs. all baselines.

displacement and circumferential strain error. The right panel shows the testing results on the cine CMRs dataset. It shows that our method consistently achieves superior performance of myocardial strain quality over all baselines.
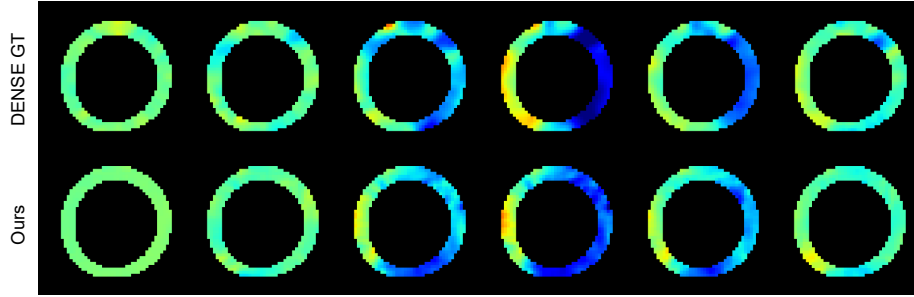

Fig. 5: A comparison of time-sequential myocardial strain generated by DENSE vs. our model throughout the cardiac cycle.

Fig. 5 illustrates a comparison between circumferential strain computed from DENSE and our predictions across evenly sampled time frames throughout the cardiac cycle. The visual similarity between the top and bottom rows demonstrates that our method effectively captures the strain patterns at various phases of cardiac motion.

5   Conclusion

In this paper, we introduced LaMoD, a novel Latent Motion Diffusion model that predicts highly accurate motions/strain from standard CMR videos. Our approach effectively addresses the challenges of detecting subtle myocardial movements, particularly in intramyocardial regions, by leveraging a pre-trained registration network and a probabilistic latent diffusion model in the latent motion space guided by DENSE CMRs. Experimental results demonstrate that LaMoD outperforms existing methods in motion prediction and strain generation accuracy. Our work has great potential to improve cardiac disease assessment based on strain, as well as treatment planning without requiring additional DENSE scans; hence ultimately improving patient care. Future work will focus on further validating the model's generalizability across diverse patient populations and clinical environments.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Amzulescu, M.S., De Craene, M., Langet, H., Pasquet, A., Vancraeynest, D., Pouleur, A.C., Vanoverschelde, J.L., Gerber, B.: Myocardial strain imaging: review of general principles, validation, and sources of discrepancies. European Heart Journal-Cardiovascular Imaging **20**(6), 605–619 (2019)
2. Arnold, V.: Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. In: Annales de l'institut Fourier. vol. 16, pp. 319–361 (1966)
3. Balter, J.M., Kessler, M.L.: Imaging and alignment for image-guided radiation therapy. Journal of clinical oncology **25**(8), 931–937 (2007)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision **61**(2), 139–157 (2005)
5. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: 3d transunet: Advancing medical image segmentation through vision transformers. arXiv preprint arXiv:2310.07781 (2023)
6. Claus, P., Omar, A.M.S., Pedrizzetti, G., Sengupta, P.P., Nagel, E.: Tissue tracking technology for assessing cardiac mechanics: principles, normal values, and clinical applications. JACC: Cardiovascular Imaging **8**(12), 1444–1460 (2015)
7. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
8. Hinkle, J.: jacobhinkle/lagomorph (5 2021), https://github.com/jacobhinkle/lagomorph

9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
10. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial intelligence **17**(1-3), 185–203 (1981)
11. Jayakumar, N., Hossain, T., Zhang, M.: Sadir: Shape-aware diffusion models for 3d image reconstruction. arXiv preprint arXiv:2309.03335 (2023)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Miller, M.I.: Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. NeuroImage **23**, S19–S33 (2004)
14. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision **24**(2), 209–228 (2006)
15. Morales, M.A., Van den Boomen, M., Nguyen, C., Kalpathy-Cramer, J., Rosen, B.R., Stultz, C.M., Izquierdo-Garcia, D., Catana, C.: Deepstrain: a deep learning workflow for the automated characterization of cardiac mechanics. Frontiers in Cardiovascular Medicine **8**, 730316 (2021)
16. Pedrizzetti, G., Claus, P., Kilner, P.J., Nagel, E.: Principles of cardiovascular magnetic resonance feature tracking and echocardiographic speckle tracking for informed clinical use. Journal of cardiovascular magnetic resonance **18**(1), 51 (2016)
17. Popović, Z.B., Kwon, D.H., Mishra, M., Buakhamsri, A., Greenberg, N.L., Thamilarasan, M., Flamm, S.D., Thomas, J.D., Lever, H.M., Desai, M.Y.: Association between regional ventricular function and myocardial fibrosis in hypertrophic cardiomyopathy assessed by speckle tracking echocardiography and delayed hyperenhancement magnetic resonance imaging. Journal of the American Society of Echocardiography **21**(12), 1299–1305 (2008)
18. Qiao, M., Wang, Y., Guo, Y., Huang, L., Xia, L., Tao, Q.: Temporally coherent cardiac motion tracking from cine mri: Traditional registration method and modern cnn method. Medical Physics **47**(9), 4189–4198 (2020)
19. Qin, C., Bai, W., Schlemper, J., Petersen, S.E., Piechnik, S.K., Neubauer, S., Rueckert, D.: Joint motion estimation and segmentation from undersampled cardiac mr image. In: Machine Learning for Medical Image Reconstruction: First International Workshop, MLMIR 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1. pp. 55–63. Springer (2018)
20. Qin, C., Wang, S., Chen, C., Bai, W., Rueckert, D.: Generative myocardial motion tracking via latent space exploration with biomechanics-informed prior. Medical Image Analysis **83**, 102682 (2023)
21. Qin, C., Wang, S., Chen, C., Qiu, H., Bai, W., Rueckert, D.: Biomechanics-informed neural networks for myocardial motion tracking in mri. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. pp. 296–306. Springer (2020)
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2021)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015)
24. Seo, D., Ho, J., Traverse, J.H., Forder, J., Vemuri, B.: Computing diffeomorphic paths with applications to cardiac motion analysis. In: 4th MICCAI Workshop on Mathematical Foundations of Computational Anatomy. pp. 83–94. Citeseer (2013)
25. Tee, M., Noble, J.A., Bluemke, D.A.: Imaging techniques for cardiac strain and deformation: comparison of echocardiography, cardiac magnetic resonance and car-

diac computed tomography. Expert review of cardiovascular therapy **11**(2), 221–231 (2013)

26. Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision **97**, 229–241 (2012)

27. Wang, Y., Sun, C., Ghadimi, S., Auger, D.C., Croisille, P., Viallon, M., Mangion, K., Berry, C., Haggerty, C.M., Jing, L., et al.: Strainnet: Improved myocardial strain analysis of cine mri by deep learning from dense. Radiology: Cardiothoracic Imaging **5**(3), e220196 (2023)

28. Wang, Y., Zhang, M., Bilchick, K., Epstein, F.: Transstrainnet: Improved strain analysis of cine mri with long-range spatiotemporal relationship learning. Journal of Cardiovascular Magnetic Resonance **26** (2024)

29. Xing, J., Ghadimi, S., Abdi, M., Bilchick, K.C., Epstein, F.H., Zhang, M.: Deep networks to automatically detect late-activating regions of the heart. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1902–1906. IEEE (2021)

30. Xing, J., Wang, S., Bilchick, K.C., Epstein, F.H., Patel, A.R., Zhang, M.: Multitask learning for improved late mechanical activation detection of heart from cine dense mri. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)

31. Xing, J., Wu, N., Bilchick, K., Epstein, F., Zhang, M.: Multimodal learning to improve cardiac late mechanical activation detection from cine mr images. arXiv preprint arXiv:2402.18507 (2024)

32. Young, A.A., Li, B., Kirton, R.S., Cowan, B.R.: Generalized spatiotemporal myocardial strain analysis for dense and spamm imaging. Magnetic resonance in medicine **67**(6), 1590–1599 (2012)

33. Zhang, M., Fletcher, P.T.: Bayesian principal geodesic analysis in diffeomorphic image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 121–128. Springer (2014)

34. Zhang, M., Fletcher, P.T.: Finite-dimensional lie algebras for fast diffeomorphic image registration. In: International Conference on Information Processing in Medical Imaging. pp. 249–260. Springer (2015)

35. Zhang, X., You, C., Ahn, S., Zhuang, J., Staib, L., Duncan, J.: Learning correspondences of cardiac motion from images using biomechanics-informed modeling. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 13–25. Springer (2022)