Cardiac Motion Transfer with Adaptive Physics Constraint

Chengjin Yu[1], Puneng Xu[2], Cailing Pu[3], Mengmei Xi[1], Ruonan Xie[1], Yuanting Yan[2⋆], Dongsheng Ruan[4⋆], and Huafeng Liu[5]

[1] School of Big Data and Statistics, Anhui University, Hefei, China
[2] School of Computer Science and Technology, Anhui University, Hefei, China
[3] Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China
[4] School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China
[5] College of Optical Science and Engineering, Zhejiang University, Hangzhou, China

Abstract. Dynamic late gadolinium enhancement MRI (dLGE-MRI) synthesized from cine-MRI is clinically valuable for dynamic myocardial lesion assessment. Integrating physics information into LGE synthesis has become a research focus; however, existing methods overlook the potential physics information gap induced by inaccurate estimation of distorted myocardial deformation, leading the model to deviate toward erroneous physics losses. To address this, we propose a method that transfers motion fields extracted from cine-MRI to static LGE image for dynamic sequence generation with noise-adaptive weighted physics constraints. Unlike existing approaches that treat physics loss as a hard constraint the model must fully comply with, we first convert the physics loss into a probabilistic constraint loss based on virtual likelihood. When motion errors are substantial, we appropriately reduce the weight of the physics loss to prevent model non-convergence caused by excessive physics noise; meanwhile, when deformation estimation stabilizes in local regions, the physics loss further guides the deformation field (driven by cine-MRI pixel data) to align with the physics priors characterized in the target static LGE image. Experiments on myocardial infarction datasets demonstrate that our method enhances lesion fidelity while maintaining motion tracking accuracy. Reverse validation confirms high consistency between static LGE-MRI generated by a generator trained on the synthesized dLGE-MRI and real static LGE-MRI, verifying the effectiveness of the synthesized images.

Keywords: dynamic LGE-MRI · motion transfer · physics loss

1 Introduction

Cine-MRI captures cardiac motion over the full cardiac cycle with high temporal resolution and is widely used for quantifying myocardial motion and deforma-

⋆ Corresponding authors: Yuanting Yan (ytyan@ahu.edu.cn) and Dongsheng Ruan (ruandongsheng@zstu.edu.cn).

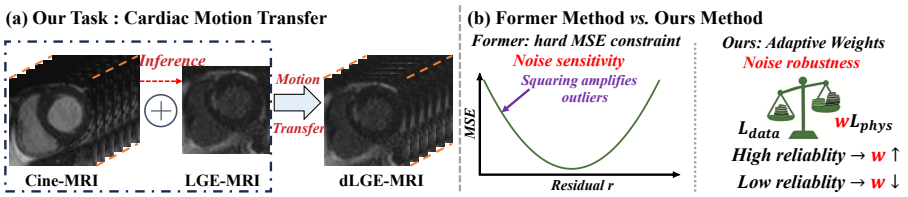

Fig. 1. (a) Our task: motion transfer from a driving cine-MRI sequence to a static LGE-MRI image to synthesize dLGE-MRI. (b) Comparison between former method (hard MSE constraint) and our method (adaptive weights).

tion [18], and functional assessment [14]. However, cine-MRI provides limited lesion-specific contrast [15]. LGE-MRI delineates scar with high conspicuity, but it is typically acquired as a static snapshot [10]. This creates a key gap: cine-MRI provides motion without lesions, whereas LGE-MRI provides lesions without dynamics. To bridge this gap, inspired by [7], we propose to transfer full-cycle motion from a driving cine sequence to a static LGE image, synthesizing dynamic LGE-MRI (dLGE-MRI) for lesion-aware motion analysis (Fig. 1a). However, in the field of motion transfer-based video synthesis, existing studies generally adopt a purely data-driven training paradigm [16, 21, 24]. Although this paradigm can statistically match pixel-level motion signal features, it lacks explicit modeling of physiological plausibility [19]. Recent studies have attempted to introduce physics-informed loss to enhance the interpretability [13, 29], but these methods can not be directly applicable to cardiac motion transfer task. The physics-informed loss to cardiac motion transfer [30] is hindered by noise in cine-MRI motion fields, compounded by physics information gaps from inaccurate myocardial deformation estimates [6]. Although some strategies have tried to dynamically adjust the weight of physics-informed constraints based on noise levels to alleviate such conflicts, they typically require prior estimation of noise intensity [2].

To address the physics loss gap caused by inaccurate myocardial motion estimation, we propose an adaptive weighted physics constrained method for dLGE-MRI synthesis. Unlike existing approaches that rigidly enforce physics loss, we introduce a probabilistic physics loss based on virtual residual likelihood. Specifically, as shown in Fig. 2, we first extract motion fields from cine-MRI using a data-driven model. Since the motion fields contain noise that could lead to erroneous physics guidance, we adaptively adjust the weight of physics constraints: when motion estimation errors are substantial, we weaken the physics constraints to prevent the model from being misled by inaccurate deformation priors; when deformation estimation becomes reliable, we strengthen the constraints to align the cine-driven motion fields with the physics prior embedded in static LGE images. This adaptive weighting strategy ensures that physics information is integrated as a flexible guide rather than an unconditional requirement. Finally, we warp the static LGE image using the refined motion field to synthesize dLGE-MRI. Our main contributions are summarized as follows:

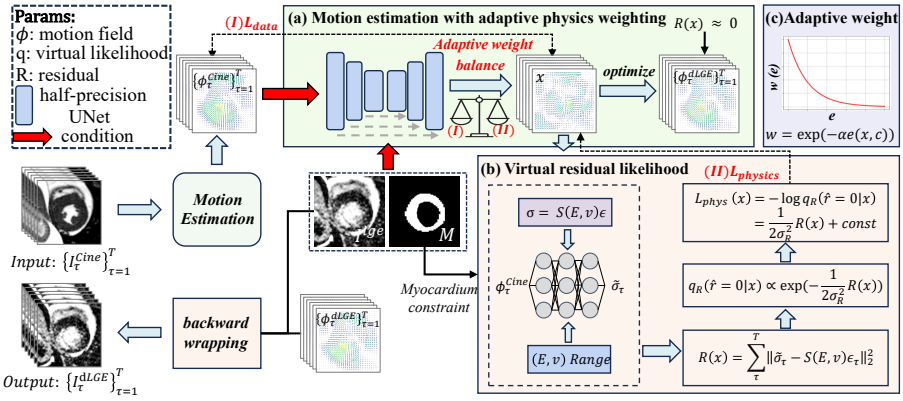

Fig. 2. Overview of our proposed method.

- We bridge the clinical gap between cine-MRI (motion without lesions) and LGE-MRI (lesions without dynamics) by first synthesizing dLGE-MRI that integrates both lesion visibility and motion dynamics.
- We propose a physics adaptive weighting mechanism that probabilistically adjusts physics constraints to alleviate optimization conflicts caused by uncertain gaps in physics information, thereby generating physiologically plausible dLGE-MRI consistent with LGE physics priors.
- We validate superior motion transfer accuracy on myocardial infarction dataset. we also propose a reverse validation protocol via verifying the effectiveness of models trained on synthetic dLGE-MRI.

2 Method

2.1 Problem Setup and Overview

Given a cine sequence $\{I_\tau^{\text{cine}}\}_{\tau=1}^T$, a static LGE image I^{lge}, and its binary myocardium mask M. Here, T denotes the number of frames (in this experiment, $T = 20$). We have $I_\tau^{\text{cine}}, I^{\text{lge}} \in \mathbb{R}^{160\times160}$ and $M \in \{0,1\}^{160\times160}$. We denote by $I_{\text{ref}}^{\text{cine}}$ the cine frame temporally matched to I^{lge}. Our goal is to synthesize a dynamic LGE sequence $\{I_\tau^{\text{dLGE}}\}_{\tau=1}^T$ by transferring the motion observed in the cine sequence to the LGE image, thereby preserving lesion appearance and maintaining physiological plausibility.

2.2 Physics Residual Constraint

Data-driven motion estimation. We estimate the motion fields $\{\phi_\tau^{\text{cine}}\}_{\tau=1}^T$ from cine-MRI using the registration framework SeqMorph [28]. Here, $\phi_\tau^{\text{cine}}(p)$ maps a pixel p in the reference frame $I_{\text{ref}}^{\text{cine}}$ to its corresponding location at phase

τ. For completeness, we include the reference phase in the sequence by defining an identity deformation $\phi^{\text{cine}}_{\text{ref}}(p) = p$ (equivalently, $\mathbf{u}^{\text{cine}}_{\text{ref}}(p) = \mathbf{0}$), so that the deformation sequence has length T. The displacement field is then defined as $\mathbf{u}^{\text{cine}}_{\tau}(p) = \phi^{\text{cine}}_{\tau}(p) - p$.

Construction of the physics residual constraint. Given a candidate motion field $\{\phi_\tau\}_{\tau=1}^{T}$, we compute displacement gradients per phase τ inside the myocardium mask M. For computational efficiency, we adopt a 2D isotropic linear-elastic model under plane strain. Under the infinitesimal-strain assumption, the strain tensor is

$$\boldsymbol{\epsilon}_\tau = \frac{1}{2}\left(\nabla \mathbf{u}_\tau + \nabla \mathbf{u}_\tau^\top\right), \qquad \mathbf{u}_\tau(p) = \phi_\tau(p) - p. \tag{1}$$

the strain is presented by three components $\boldsymbol{\epsilon}_\tau = [\epsilon_{xx}, \epsilon_{yy}, \epsilon_{xy}]^\top$, while ϵ_{xx} and ϵ_{yy} are the normal strain components along the x- and y-axes, and ϵ_{xy} is the in-plane shear strain. Thus, we can employ Hooke's law to calculate the stress [5]

$$\boldsymbol{\sigma}_\tau = \mathbf{S}(E, \nu)\boldsymbol{\epsilon}_\tau, \tag{2}$$

where the material matrix $\mathbf{S}$ is determined by Young's modulus E and Poisson's ratio ν [11] (fixed with $E \in [21.13, 99.8]$ and $\nu \in [0.26, 0.55]$ [12]).

Since boundary conditions and external forces are unavailable in clinical acquisitions, enforcing full equilibrium would require additional assumptions. Instead, we impose constitutive consistency by introducing an auxiliary stress field $\tilde{\boldsymbol{\sigma}}_\tau$ predicted by the network, where $\tilde{\boldsymbol{\sigma}}_\tau$ is produced by a lightweight stress head that shares features with the motion estimator. We define the physics residual energy as the discrepancy between the network-predicted stress $\tilde{\boldsymbol{\sigma}}_\tau$ and the stress $\boldsymbol{\sigma}_\tau$ computed from Eq. (2):

$$\mathcal{R}(\boldsymbol{\phi}) = \sum_{\tau=1}^{T} \left\| \tilde{\boldsymbol{\sigma}}_\tau - \boldsymbol{\sigma}_\tau \right\|_2^2, \tag{3}$$

where all terms are computed within the myocardium region specified by M. However, directly using $\mathcal{R}(\boldsymbol{\phi})$ as a hard constraint on the motion field is ineffective. The L2 hard penalty is sensitive to motion noise [3] and may misguide training, so we convert it into a probabilistic constraint (Sec. 2.3).

2.3 Physics-Guided Motion Estimation

Virtual residual likelihood. To effectively integrate $\mathcal{R}(\boldsymbol{\phi})$, we introduce a residual random variable $\hat{r}$ and treat the ideal physics-consistent state as the virtual observation $\hat{r} = \mathbf{0}$, which defines a residual likelihood $q_R(\hat{r} \mid x)$. Given a candidate motion x, we compute the physics residual energy $\mathcal{R}(x)$ following Eq. (3). Since cine-derived motion is noisy and the constitutive model is only approximately valid in vivo, we treat physics consistency as soft evidence rather

than a hard constraint and encode it probabilistically. We then define a soft residual likelihood as an energy-based model:

$$q_R(\hat{r} = \mathbf{0} \mid x) \;\propto\; \exp\!\left(-\frac{1}{2\sigma_R^2}\,\mathcal{R}(x)\right), \tag{4}$$

where σ_R^2 controls the tolerance of the residual constraint. Minimizing the negative log-likelihood yields the physics loss

$$\mathcal{L}_{\text{phys}}(x) = -\log q_R(\hat{r} = \mathbf{0} \mid x) = \frac{1}{2\sigma_R^2}\,\mathcal{R}(x) + \text{const.} \tag{5}$$

Motion estimation with adaptive physics weighting. We parameterize a stochastic conditional motion estimator with a half-precision UNet [20] backbone. Given $c = \big(\{\, I^{\text{lge}},\, M,\, x^{\text{obs}}\big)$, we draw $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and generate a motion sample via

$$x = f_\theta(c, z). \tag{6}$$

this implicitly defines a conditional distribution $p_\theta(x \mid c)$. Here $x^{\text{obs}} = \{\phi_\tau^{\text{cine}}\}_{\tau=1}^{T}$. Because $\mathcal{R}(\cdot)$ is nonlinear and contains squared residuals, evaluating physics consistency using a point estimate can mismatch the physics loss averaged over motion uncertainty (Jensen gap [6]), i.e. $\mathcal{L}_{\text{phys}}(\tilde{x}) \;\neq\; \mathbb{E}_{x\sim p_\theta(\cdot|c)}[\mathcal{L}_{\text{phys}}(x)]$, and the mismatch becomes more severe when the cine-derived proxy motion is noisy. In practice, outliers in cine motion fields can induce heavy-tailed physics residuals; due to the inherent L2 amplification, these outliers may dominate gradients, causing optimization to over-fit the physics term and drift away from the cine-anchor data term, leading to data collapse. We therefore optimize a cine-anchor data term together with a physics term, where an adaptive weight is applied to the physics term:

$$\mathcal{L}(\theta) = \mathbb{E}_{z\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\Big[\mathcal{L}_{\text{data}}(x, c) + w(x, c)\,\mathcal{L}_{\text{phys}}(x)\Big], \tag{7}$$

with the cine-anchor term defined by

$$\mathcal{L}_{\text{data}}(x, c) = \big\|x^{\text{obs}} - x\big\|_2^2. \tag{8}$$

Adaptive physics weight. We measure the reliability of x by the cine warping error it induces:

$$e(x, c) = \frac{1}{T}\sum_{\tau=1}^{T}\big\|W(I_{\text{ref}}^{\text{cine}}, \phi_\tau) - I_\tau^{\text{cine}}\big\|_2^2, \tag{9}$$

where $W(\cdot, \phi_\tau)$ denotes differentiable warping implemented via grid-based sampling. We define the adaptive physics weight as a monotone decreasing function of e:

$$w(x, c) = \exp(-\alpha\, e(x, c)), \qquad \alpha > 0. \tag{10}$$

This weighting can be interpreted as modulating the confidence of the virtual residual likelihood in Eq. (4). Intuitively, a large warping error indicates that the current motion estimate is unreliable; hence we down-weight physics when e is large and progressively tighten it as e decreases.

Table 1. Quantitative comparison of cardiac motion transfer on the clinical test set. The best performing results are highlighted in **bold**.

	Self-Reenactment			Segmentation Tracking		Cross-Reenactment	
Method	**SSIM↑**	**PSNR↑**	**LPIPS↓**	**Dice(Myo)↑**	**HD95(Myo)↓**	**IS(dLGE)↓**	**IQ(dLGE)↑**
OF [4]	0.772	26.35	0.131	0.705	5.279	0.283	40.36
VoxelMorph [1]	0.812	27.94	0.116	0.797	3.418	0.261	41.09
Deeptag [27]	0.819	28.03	0.114	0.796	3.504	0.258	41.13
SeqMorph [28]	0.822	28.20	0.110	0.805	3.143	0.241	41.95
FOMM [22]	0.664	22.40	0.270	0.720	5.201	0.262	41.85
DaGAN [8]	0.693	23.04	0.110	0.755	4.902	0.272	42.35
MC-Net [7]	0.722	23.73	0.176	0.775	4.449	0.252	42.90
LIA [25]	0.773	24.80	0.145	0.785	4.204	0.238	41.50
LIA-X [26]	0.810	25.66	0.128	0.800	3.802	0.206	44.07
w/o **Physics**	0.824	28.27	0.107	0.808	3.142	0.238	42.03
w/o **AdaW**	0.812	26.50	0.115	0.804	3.224	0.195	44.35
Ours (Full)	**0.830**	**29.03**	**0.103**	**0.815**	**3.124**	**0.175**	**45.06**

2.4 Motion Transfer for Dynamic LGE Synthesis

At inference, we draw $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and sample a full-cycle deformation sequence $x \sim p_\theta(\cdot \mid c)$ via $x = f_\theta(c, z)$. We then synthesize the dynamic LGE sequence by applying the sampled motion to the static LGE image.

For each phase τ, we use backward warping

$$I_\tau^{\mathrm{dLGE}}(p) = I^{\mathrm{lge}}\Big(\big(\phi_\tau\big)^{-1}(p)\Big), \qquad \tau = 1, \ldots, T, \tag{11}$$

so that each output pixel p samples its intensity from the corresponding pre-image location in I^{lge}.

3 Experiments and Results

Dataset. Our private dataset consists of a total of 850 MI subjects (13,304 cine sequences). The ethics committees of all participating hospitals approved the retrospective study and waived the requirement for informed consent. Due to differences across MRI scanners, the cine sequences contain 20–25 temporal frames; all sequences were uniformly resampled to 20 frames using linear interpolation to standardize the data while preserving the complete cardiac cycle. Myocardium and lesions were manually delineated on LGE images by an experienced radiologist and reviewed by a senior expert. Among them, the data are split into 70%/10%/20% for training/validation/testing.

Evaluation Metrics. We report Identity Similarity (IS), Image Quality (IQ), PSNR, SSIM, and LPIPS to quantify image quality. IS is defined as the average embedding distance between generated frames and the source image [26]. IQ is computed by a self-adaptive image quality assessment network [23]. For cardiac motion tracking, we use Dice and the 95th-percentile Hausdorff distance (HD95) (mm) [9].

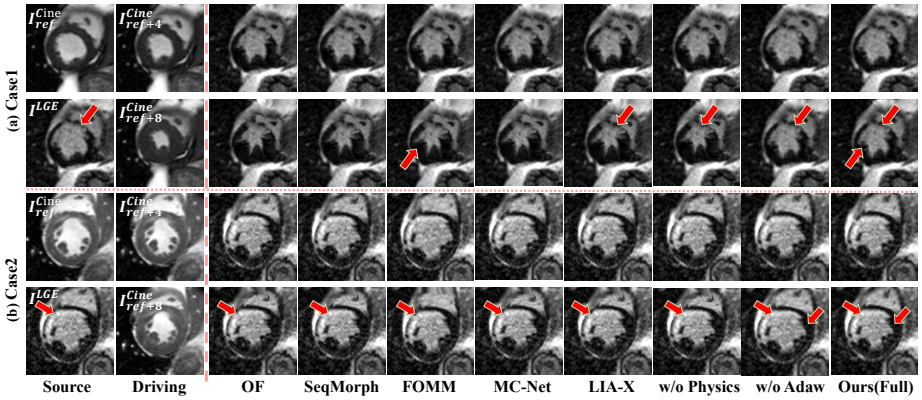

Fig. 3. Qualitative comparison of cross-reenactment on the clinical test set. **Source**: cine reference frame and static LGE. **Driving**: target cine frame providing motion. The remaining columns show the synthesized LGE warped to the driving frame by different methods.

Implementation Details. We implement our framework in PyTorch 2.2.1 (CUDA 12.0) and train on 2D short-axis slices. We adopt a half-precision UNet (5 scales, 2 ResBlocks/scale, channels $[128, 128, 256, 256, 512]$; self-attention with 4 heads at 32×32 and 16×16) to parameterize the stochastic motion estimator $x = f_\theta(c, z)$. Inputs are normalized to $[-1, 1]$, and the myocardium mask M is binary. During training, we sample $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and optimize Eq. (7) using AdamW (lr 10^{-4}, wd 10^{-2}, batch size 8) for 1500 epochs.

3.1 Comparison Studies

Outperforming mainstream methods. We compare our model with nine representative methods, including an optical flow method (OF) [4], image registration-based methods (VoxelMorph [1], Deeptag [27], SeqMorph [28]) and motion transfer methods (FOMM [22], DaGAN [8], MC-Net [7], LIA [25], and LIA-X [26]). As shown in Table 1, our full model achieves the best overall performance on the clinical test set, with SSIM/PSNR/LPIPS of 0.830/29.03/0.103 for self-reenactment, IS/IQ of 0.175/45.06 for cross-reenactment, and Dice/HD95 of 0.815/3.124 mm for myocardium segmentation tracking. Figure 3 qualitatively confirms these gains. Under the large-amplitude motion (case1), FOMM produces diastolic artifacts and incomplete diastole, while LIA-X appears less clear in fine anatomical details. In the constrained-motion (case2), although all methods capture global motion, our method better preserves local scar-involved deformation, indicating more faithful motion and structural transfer.

Higher consistency in reverse validation. We use the synthesized dLGE sequence as supervision to train a model [17] for generating static LGE, and evaluate each method using real static LGE frame as the ground truth. As shown in Fig. 4(b), we conduct reverse validation on four representative methods: OF,

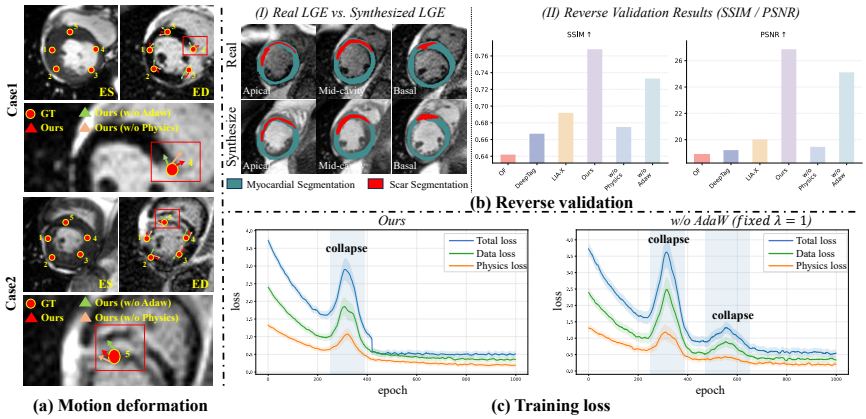

Fig. 4. Visualization of (a) motion deformation, (b) reverse validation: (I) comparison between the real LGE and the reverse-synthesized LGE and (II) validation results, and (c) Data loss, physics loss, and total loss during training.

Deeptag, LIA-X, and ours. Our full model achieves the best SSIM/PSNR/LPIPS scores, indicating a more accurate and robust Cine-to-LGE translation that better aligns with the real LGE distribution. Fig. 4(b) also presents visual comparisons between the synthesized and real registered LGE frame, showing that our method produces results with higher fidelity to real images. These results demonstrate the effectiveness and reliability of our method in generating clinically plausible LGE images.

3.2 Ablation studies

Physics-guided loss. As shown in Table 1, removing the physics constraint (w/o Physics) degrades nearly all quantitative metrics, including lower SSIM, PSNR, Dice, and IQ as well as higher LPIPS and IS, demonstrating clear performance drops in motion transfer, segmentation tracking, and dynamic-static consistency. Without the physics constraint, physically implausible noisy motions become more difficult to suppress. Qualitatively, the w/o Physics model produces less coherent local deformations and structural distortions in cross-reenactment, and introduces systematic bias in the motion field that deviates from physiological plausibility.

Adaptive weighting mechanism. Similarly, replacing the adaptive weight with a fixed coefficient (w/o AdaW) also leads to consistent performance degradation, revealing that a static trade-off between the cine-anchor term and physics term is suboptimal. The w/o AdaW model suffers from more severe structural distortion, larger loss fluctuations, and unstable convergence during training (Fig. 4(c)). For further validation, we perform reverse validation by comparing the synthesized LGE frame with the held-out real LGE image. As shown in Fig. 4(b), our full model achieves the best SSIM/PSNR/LPIPS among compet-

ing methods, generating more realistic LGE images with fewer artifacts. These results collectively validate the effectiveness of our physics guidance, adaptive weighting mechanism, and their contributions to accurate, stable, and clinically interpretable cardiac motion modeling.

4 Conclusion

We propose a physics-guided framework for synthesizing dLGE-MRI via cardiac motion transfer. The framework integrates a cine-derived motion observation to preserve subject-specific kinematics and a physics-guided strategy to enforce physics feasibility during training. By embedding a virtual residual likelihood based adaptive weighting mechanism, the model generates lesion-aware motion that is temporally coherent and mechanically consistent. Experimental results also demonstrate improvements in motion tracking, dLGE-MRI synthesis.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (62376002, 62101491, 62401016, and 62401508), the Natural Science Foundation of Zhejiang Province (LQN25F020003), the Foundation of Sichuan Provincial People's Hospital (2023QN25), and the Zhejiang Provincial "Jianbing Lingyan+X" Science and Technology Program (2025C01127).

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

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
2. Bastek, J.H., Sun, W., Kochmann, D.M.: Physics-informed diffusion models. arXiv preprint arXiv:2403.14404 (2024)
3. Black, M.J., Anandan, P.: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer vision and image understanding **63**(1), 75–104 (1996)
4. Carranza-Herrezuelo, N., Bajo, A., Sroubek, F., Santamarta, C., Cristóbal, G., Santos, A., Ledesma-Carbayo, M.J.: Motion estimation of tagged cardiac magnetic resonance images using variational techniques. Computerized Medical Imaging and Graphics **34**(6), 514–522 (2010)
5. Fung, Y.c.: Biomechanics: mechanical properties of living tissues. Springer Science & Business Media (2013)
6. Gao, X., Sitharam, M., Roitberg, A.E.: Bounds on the jensen gap, and implications for mean-concentrated distributions. arXiv preprint arXiv:1712.05267 (2017)
7. Hong, F.T., Xu, D.: Implicit identity representation conditioned memory compensation network for talking head video generation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 23062–23072 (2023)
8. Hong, F.T., Zhang, L., Shen, L., Xu, D.: Depth-aware generative adversarial network for talking head video generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3397–3406 (2022)

9. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. IEEE Transactions on pattern analysis and machine intelligence **15**(9), 850–863 (2002)
10. Jenista, E.R., Wendell, D.C., Azevedo, C.F., Klem, I., Judd, R.M., Kim, R.J., Kim, H.W.: Revisiting how we perform late gadolinium enhancement cmr: insights gleaned over 25 years of clinical practice. Journal of Cardiovascular Magnetic Resonance **25**(1), 18 (2023)
11. Liu, H., Shi, P.: State-space analysis of cardiac motion with biomechanical constraints. IEEE Transactions on Image Processing **16**(4), 901–917 (2007)
12. Liu, H., Shi, P., et al.: Maximum a posteriori strategy for the simultaneous motion and material property estimation of the heart. IEEE Transactions on Biomedical Engineering **56**(2), 378–389 (2008)
13. Mayo, P., Pirkl, C.M., Achim, A., Menze, B., Golbabaee, M.: Physics informed guided diffusion for accelerated multi-parametric mri reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 429–439. Springer (2025)
14. Menchón-Lara, R.M., Simmross-Wattenberg, F., Casaseca-de-la Higuera, P., Martín-Fernández, M., Alberola-López, C.: Reconstruction techniques for cardiac cine mri. Insights into imaging **10**(1), 100 (2019)
15. Ordovas, K.G., Higgins, C.B.: Delayed contrast enhancement on mr images of myocardium: past, present, future. Radiology **261**(2), 358–374 (2011)
16. Pondaven, A., Siarohin, A., Tulyakov, S., Torr, P., Pizzati, F.: Video motion transfer with diffusion transformers. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 22911–22921 (2025)
17. Qi, R., Tao, M., Xu, C., Li, X., Pan, S., Chen, J., Li, S.: Knowledge-driven interpretative conditional diffusion model for contrast-free myocardial infarction enhancement synthesis. Medical Image Analysis **105**, 103701 (2025)
18. Rajiah, P.S., Kalisz, K., Broncano, J., Goerne, H., Collins, J.D., François, C.J., Ibrahim, E.S., Agarwal, P.P.: Myocardial strain evaluation with cardiovascular mri: physics, principles, and clinical applications. Radiographics **42**(4), 968–990 (2022)
19. Rixner, M., Koutsourelakis, P.S.: A probabilistic generative model for semi-supervised training of coarse-grained surrogates and enforcing physical constraints through virtual observables. Journal of Computational Physics **434**, 110218 (2021)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E., Sebe, N.: Animating arbitrary objects via deep motion transfer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2377–2386 (2019)
22. Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E., Sebe, N.: First order motion model for image animation. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
23. Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., Zhang, Y.: Blindly assess image quality in the wild guided by a self-adaptive hyper network. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3667–3676 (2020)
24. Tu, S., Dai, Q., Cheng, Z.Q., Hu, H., Han, X., Wu, Z., Jiang, Y.G.: Motioneditor: Editing video motion via content-aware diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7882–7891 (2024)

25. Wang, Y., Yang, D., Bremond, F., Dantcheva, A.: Latent image animator: Learning to animate images via latent space navigation. arXiv preprint arXiv:2203.09043 (2022)
26. Wang, Y., Yang, D., Chen, X., Bremond, F., Qiao, Y., Dantcheva, A.: Lia-x: Interpretable latent portrait animator. arXiv preprint arXiv:2508.09959 (2025)
27. Ye, M., Kanski, M., Yang, D., Chang, Q., Yan, Z., Huang, Q., Axel, L., Metaxas, D.: Deeptag: An unsupervised deep learning method for motion tracking on cardiac tagging magnetic resonance images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7261–7271 (2021)
28. Ye, M., Yang, D., Huang, Q., Kanski, M., Axel, L., Metaxas, D.N.: Sequencemorph: A unified unsupervised learning framework for motion tracking on cardiac image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(8), 10409–10426 (2023)
29. Yuan, Y., Song, J., Iqbal, U., Vahdat, A., Kautz, J.: Physdiff: Physics-guided human motion diffusion model. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 16010–16021 (2023)
30. Zhang, J., Yan, R., Perelli, A., Chen, X., Li, C.: Phy-diff: Physics-guided hourglass diffusion model for diffusion mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 345–355. Springer (2024)