Arrhythmia-Robust Cine-MRI via Latent Motion Artifact Characterization

Gaoning Ning[1*], Chengjin Yu[2*], Cailing Pu[3], Zhifan Gao[4], Chenchu Xu[2], and Huafeng Liu[1(✉)]

[1] Zhejiang University, Hangzhou, China
liuhf@zju.edu.cn
[2] Anhui University, Hefei, China
[3] Sichuan Provincial People's Hospital, Chengdu, China
[4] Sun Yat-sen University, Guangzhou, China

[*]These authors contributed equally to this work.

Abstract. Arrhythmias often violate the periodicity assumed in segmented cardiac cine-MRI, producing blurring and ghosting that degrade functional assessment. Due to ECG signal irregularities that cause temporal misalignment, arrhythmia-related artifacts impose more stringent requirements on denoising methods than conventional artifacts. We propose a Latent-Dictionary Encoding framework for Motion-Artifact Reduction (LDE-MAR). Using a latent image animator (LIA) formulation, LDE-MAR learns orthogonality-constrained sparse dictionaries to separate physiological motion into region and direction components, and adds a noise dictionary to capture artifact-related deviations in latent motion trajectories. Training is staged: motion dictionaries are pretrained on clean cine sequences, then only the noise dictionary is fine-tuned with paired clean-corrupted data to suppress irregular dynamics while preserving anatomy and plausible myocardial motion. This staged, multi-dictionary decomposition explicitly isolates arrhythmia-induced deviations from physiological dynamics, improving robustness under large R-R variability. We evaluate our method on a private dataset of 398 subjects and the public ACDC dataset. Our experimental results demonstrate that LDE-MAR outperforms state-of-the-art methods in artifact removal and in restoring motion and structural details. The code is available at `https://github.com/gaoningn/LDE-MAR`.

Keywords: Latent dictionary encoding · Cine-MRI artifact reduction

1 Introduction

Cardiovascular diseases (CVDs) are the leading cause of global mortality, making accurate diagnosis essential for risk stratification. Cardiac cine-MRI is the gold standard for quantifying ventricular function due to its high spatiotemporal resolution [8]. However, many patients who most need precise assessment-especially those with advanced heart failure-often have arrhythmias such as atrial fibrillation or frequent premature ventricular contractions [1]. These irregular rhythms disrupt standard acquisition and introduce severe artifacts in reconstructed images, obscuring anatomy and impairing measurements like ejection fraction and ventricular volumes. In severe cases, images become non-diagnostic, requiring repeat scans and increasing healthcare costs, which limits cine-MRI utility in critically ill patients.

Artifacts in conventional segmented cine-MRI stem from its acquisition principle [4]: k-space is collected over multiple ECG-gated cardiac cycles and concatenated under the assumption of periodic heartbeats [21]. Arrhythmias cause large R-R interval fluctuations, breaking temporal consistency across cycles and leading to incorrect mixing of k-space segments from different cardiac phases, i.e., mismatched trajectories. Unlike respiratory or bulk-motion artifacts that are mainly rigid/affine, arrhythmias often involve beat-to-beat, nonlinear nonrigid myocardial deformations [16]. As a result, artifacts include global blurring, and also complex, nonlocal signal voids and ghosting, obscuring anatomical boundaries and making true cardiac geometry recovery a severely ill-posed problem.

Considerable efforts have been devoted to mitigating arrhythmia-induced artifacts in cardiac cine-MRI. Clinical and engineering solutions, including compressed sensing [13, 6] and real-time cine imaging [5, 3], can alleviate artifacts but often at the cost of spatial resolution. More recently, deep learning approaches for denoising and accelerated reconstruction—such as convolutional networks [7, 10, 11], recurrent networks [14], and diffusion models [17, 12]—have been applied in the image domain; however, these pixel-wise mappings may under-utilize temporal motion coherence and over-smooth fine myocardial textures. Motion-aware iterative reconstruction methods [18, 22, 19] jointly estimate motion and images from under-sampled k-space, but are mainly tailored to under-sampling and mild perturbations under quasi-regular cardiac and respiratory patterns. With irregular rhythms and pronounced R-R variability, severe k-space trajectory inconsistencies violate these assumptions, limiting existing correction schemes. Consequently, robust cardiac motion modeling under complex rhythm conditions is essential for suppressing arrhythmia-related artifacts.

Building on the above observations, we propose LDE-MAR, a latent-space motion denoising framework that targets arrhythmia-induced artifacts through implicit motion representation learning rather than pixel-wise image mappings. Our framework learns orthogonality-constrained sparse dictionaries to disentangle latent cardiac motion into motion region and motion direction components, and further introduces a dedicated noise dictionary to capture artifact-related deviations in the latent motion trajectory. We adopt a two-stage training strategy: the motion dictionaries are first pretrained on clean cine sequences to model canonical physiological dynamics, and then the noise dictionary is fine-tuned using paired clean/corrupted data while keeping the motion priors fixed. This staged, multi-dictionary design suppresses irregular dynamics while preserving anatomical fidelity and kinematically plausible myocardial motion.

Our contributions are summarized as follows: 1) We propose a denoising method for arrhythmic cardiac cine-MRI to mitigate irregular rhythm-induced artifacts that hinder clinical assessment. 2) We introduce a latent-space mo-

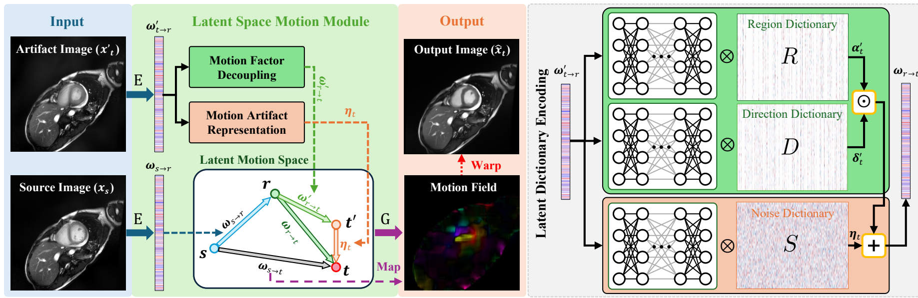

Fig. 1. Overview of the proposed LDE-MAR framework. The method extracts latent motion between a source and target cine frame, decomposes it with learned dictionaries (physiological motion vs. arrhythmia-related noise), and uses the refined motion to generate a motion field for warping-based reconstruction with reduced artifacts.

tion denoising framework with three dictionary modules: two sparse dictionaries encode the motion regions and motion directions, respectively, and the third dictionary models motion noise. Mapping the learned clean-motion dictionaries back to image space yields cine-MRI with preserved anatomical fidelity and motion authenticity. 3) We evaluate the method on a private dataset of 398 subjects and the public ACDC [2] dataset. Results show that LDE-MAR effectively reduces arrhythmia-related artifacts, outperforming comparable image-domain baselines in image quality (achieving a 1.2% lead in SSIM).

2 Methodology

2.1 Preliminary: Latent Image Animator

Latent Image Animator (LIA) [24, 25] is built upon an autoencoder architecture that represents the explicit motion deformation from a source image x_s to a target image x_t through a learnable linear trajectory $\boldsymbol{w}_{s\to t}$ in the latent space. LIA introduces a linear motion decomposition strategy, where the latent trajectory is formulated as a linear combination of learnable motion directions and their corresponding magnitudes. Specifically, an orthogonality-constrained learnable dictionary is employed to characterize the entire latent Euclidean space, with each orthogonal vector serving as a directional basis vector. The generator $\mathbf{G}$ decodes the latent motion trajectory $\boldsymbol{w}_{s\to t}$ to deform the source image x_s, thereby producing the target image x_t.

LIA shares one encoder $\mathbf{E}$ for latent motion and appearance, but its single-dictionary linear parameterization neither separates where motion occurs from how it moves nor models artifact deviations. We retain LIA's shared-reference coordinate while redesigning its trajectory model for cine-MRI: Sect. 2.2 factorizes region activation and signed motion with R and D, Sect. 2.3 introduces S for artifact deviations, and Sect. 2.4 learns these components sequentially.

2.2 Dual-Branch Dictionary-Driven Motion Factor Decoupling

Each frame is independently encoded as $\boldsymbol{w}_{in\to r} = \mathbf{E}(x_{in})$ relative to a shared, hypothesized reference x_r, whose latent code acts as an implicit normalization anchor rather than an explicitly generated image. Therefore, the latent motion path from the source image x_s to the target image x_t can be decomposed as:

$$\boldsymbol{w}_{s\to t} = \boldsymbol{w}_{s\to r} + \boldsymbol{w}_{r\to t} \tag{1}$$

The shared anchor avoids a pairwise encoder over (x_s, x_t) and provides a stable coordinate for separating motion from artifact deviations. Importantly, $\boldsymbol{w}_{r\to t}$ is not the inverse of $\boldsymbol{w}_{t\to r}$; instead, $\boldsymbol{w}_{t\to r} = \mathbf{E}(x_t)$ conditions a one-way dictionary recomposition. As illustrated in Fig. 1, we introduce a dual-branch dictionary to achieve fine-grained modeling of cardiac motion factors. Specifically, we construct a motion region dictionary $R = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_M\} \in \mathbb{R}^{L\times M}$ and a motion direction dictionary $D = \{\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N\} \in \mathbb{R}^{L\times N}$. The orthogonal basis vectors in R represent latent feature codes corresponding to different anatomical regions of the cardiac image, while D contains the motion direction bases in the latent space.

We employ parallel multilayer perceptrons (MLP) [23] to extract the magnitude coefficients of motion regions and motion directions from $\boldsymbol{w}_{t\to r} = \mathbf{E}(x_t)$, yielding:

$$U_t = f(x_t) = \texttt{softmax}(\texttt{MLP}_R(\boldsymbol{w}_{t\to r})),\ V_t = g(x_t) = \texttt{tanh}(\texttt{MLP}_D(\boldsymbol{w}_{t\to r})) \tag{2}$$

where $U_t = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_M\} \in \mathbb{R}^M$ and $V_t = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\} \in \mathbb{R}^N$. The `softmax` gives normalized, non-negative competitive weights over region bases, whereas the `tanh` gives bounded signed direction coefficients.

The coefficients are then multiplied with their corresponding dictionaries:

$$\boldsymbol{\alpha}_t = R \cdot U_t = \sum_{i=1}^{M} u_i \boldsymbol{r}_i,\ \boldsymbol{\delta}_t = D \cdot V_t = \sum_{i=1}^{N} v_i \boldsymbol{d}_i \tag{3}$$

Here, $\boldsymbol{\alpha}_t$ is a latent region-activation code that emphasizes motion-relevant feature components, whereas $\boldsymbol{\delta}_t$ specifies signed motion direction and magnitude. The two factors are complementary: $\boldsymbol{\alpha}_t$ determines where motion is expressed in the latent representation, while $\boldsymbol{\delta}_t$ determines how the activated components evolve. Their element-wise composition yields the latent motion trajectory $\boldsymbol{w}_{r\to t}$:

$$\boldsymbol{w}_{r\to t} = \boldsymbol{\alpha}_t \odot \boldsymbol{\delta}_t \tag{4}$$

where $\odot$ denotes element-wise multiplication. Thus, $\boldsymbol{\alpha}_t$ gates the latent components in which $\boldsymbol{\delta}_t$ is expressed; setting a component of $\boldsymbol{\alpha}_t$ to zero alone does not explicitly factor where motion can occur.

2.3 Noise Dictionary-Driven Motion Artifact Representation

The latent motion trajectory $\boldsymbol{w}'_{r\to t}$, obtained by processing the arrhythmic cine-MRI image x'_t using the module described in Sect. 2.2, exhibits a noise deviation relative to the latent code of the clean image x_t, due to motion artifacts. Their relationship can be linearly formulated in the latent Euclidean space as:

$$\boldsymbol{w}_{r\to t} = \boldsymbol{w}'_{r\to t} + \boldsymbol{\eta}_t = \boldsymbol{\alpha}'_t \odot \boldsymbol{\delta}'_t + \boldsymbol{\eta}_t \tag{5}$$

Similarly, we introduce a learnable dictionary $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\} \in \mathbb{R}^{L\times N}$ to model motion artifacts, where each basis vector represents a noise distribution direction in the latent space. The corresponding noise magnitude coefficients $P_t = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N\} \in \mathbb{R}^N$ are also extracted from $\boldsymbol{w}'_{t\to r}$, and the latent noise code $\boldsymbol{\eta}_t$ can then be expressed as:

$$\boldsymbol{\eta}_t = S \cdot P_t = \sum_{i=1}^{N} p_i \boldsymbol{s}_i,\ P_t = h(x'_t) = \texttt{tanh}(\texttt{MLP}_S(\boldsymbol{w}'_{t\to r})) \tag{6}$$

Because artifact corrections can have either sign, the `tanh` constrains P_t to bounded signed coefficients, limiting the magnitude of latent corrections.

Therefore, when an artifact-corrupted image x'_t is provided as input, we employ the noise-dictionary-based encoding scheme to obtain the clean latent motion trajectory $\boldsymbol{w}_{r\to t}$. The artifact-free image $\hat{\boldsymbol{x}}_t$ is subsequently generated through the generator $\mathbf{G}$.

2.4 Two-Stage Training Strategy

We first perform self-supervised pretraining using a normal cine-MRI sequence $X = \{x_1, x_2, \ldots, x_K\}$ to learn the representation pattern of clean cardiac motion. Specifically, the first frame x_1 and the k-th frame x_k in the sequence are treated as the source image and target image, respectively. The model outputs the predicted image $\hat{x}_k$:

$$\hat{x}_k = \mathbf{G}\{\mathbf{E}(x_1) + [R \cdot f(x_k)] \odot [D \cdot g(x_k)]\} \tag{7}$$

By computing the reconstruction loss between $\hat{x}_k$ and x_k, we conduct self-supervised training of the encoder $\mathbf{E}$, the generator $\mathbf{G}$, and the dual-branch dictionary module, thereby learning an accurate prior of cardiac motion.

After completing the pretraining stage, we perform supervised training using paired data from arrhythmic cine-MRI sequence $X' = \{x'_1, x'_2, \ldots, x'_K\}$ and normal sequences X to achieve motion artifact removal. Since the first frame x'_1 is minimally affected by arrhythmia, it is still used as the source image, while an artifact-corrupted frame x'_k is randomly selected as input. As shown in Fig. 1, the denoised output image $\hat{x}_k$ is formulated as:

$$\hat{x}_k = \mathbf{G}\{\mathbf{E}(x'_1) + [R \cdot f(x'_k)] \odot [D \cdot g(x'_k)] + S \cdot h(x'_k)\} \tag{8}$$

During this downstream stage, only the noise module is updated. This order makes S a residual relative to the fixed clean-motion coordinate learned by R and D; training S first or alternating unconstrained updates could let it absorb physiological variation and weaken motion/artifact separation.

Table 1. Quantitative comparison of LDE-MAR against other methods on the in-house and ACDC datasets.

Method	In-house (20f)			In-house (25f)			ACDC		
	MAE↓	SSIM↑	PSNR↑	MAE↓	SSIM↑	PSNR↑	MAE↓	SSIM↑	PSNR↑
ConvLSTM (TMI'21)	0.037	0.849	27.037	0.027	0.903	27.053	0.046	0.862	27.903
MoCo-Diff (MICCAI'24)	0.045	0.837	26.598	0.031	0.895	26.913	0.052	0.846	27.057
PFAD (AAAI'25)	0.035	0.852	27.305	0.026	0.906	27.112	0.043	0.879	28.274
Cinemo (CVPR'25)	0.027	0.873	28.402	0.023	0.914	29.971	0.034	0.892	29.011
DiTFlow (CVPR'25)	0.030	0.868	28.214	0.029	0.911	28.834	0.039	0.887	28.725
JDAC (MedIA'25)	0.024	0.881	29.037	0.021	0.916	30.229	**0.025**	0.901	29.318
LDE-MAR (Ours)	**0.019**	**0.887**	**29.435**	**0.016**	**0.938**	**31.700**	0.027	**0.906**	**29.847**

3 Experiments

3.1 Experimental Setup

Datasets. We use an in-house dataset of 398 subjects, comprising 175 subjects with 20-frame sequences and 223 subjects with 25-frame sequences, together with the public ACDC dataset. Clean-artifact cine pairs were generated by simulating ECG-gated segmented k-space acquisition (all frames share the same k-space line indices per heartbeat): a premature beat perturbs the corresponding k-space lines across frames. In bigeminy, we corrupt alternating lines with phase-dependent density (every 4 lines in early systole, every 2 lines in late phases) and sinusoidally modulated strength, applying phase errors focused on the heart mask and low-frequency-weighted amplitude attenuation.
Implementation Details. We optimize $\mathcal{L} = \mathcal{L}_1 + 0.5 * \mathcal{L}_{SSIM}$ with Adam (batch size 5) in PyTorch on one NVIDIA Quadro RTX 8000. In each training stage, learning rates 10^{-3}, 10^{-4} and 10^{-5} are used for 40 epochs each (120 epochs; approximately 60 h per stage). Outputs are single-channel 256×256 images, and the MLP hidden dimension is 512. Data are split into 70%/10%/20% for training/validation/testing; Mean Absolute Error (MAE), Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR) are averaged over all test subjects and frames.

3.2 Comparison with State-of-the-Art Methods

We conduct experiments on the in-house (20f/25f) and ACDC datasets to evaluate our method, comparing it with representative state-of-the-art approaches, including ConvLSTM [14], MoCo-Diff [9], PFAD [26], Cinemo [15], DiTFlow [20], and JDAC [27]. As shown in Table 1, our LDE-MAR achieves 0.887, 0.938, 0.906 on SSIM and 29.435, 31.700, 29.847 on PSNR, respectively. Compared with the strongest competing method (JDAC), LDE-MAR improves SSIM and PSNR by about 1.2% and 2.7% on average across the three benchmarks. Quantitative analysis demonstrates that LDE-MAR achieves superior image reconstruction quality in cine-MRI sequences affected by arrhythmia-induced artifacts.

We additionally perform a qualitative comparison across these methods. As shown in Fig. 2, we present the MR images and corresponding error maps

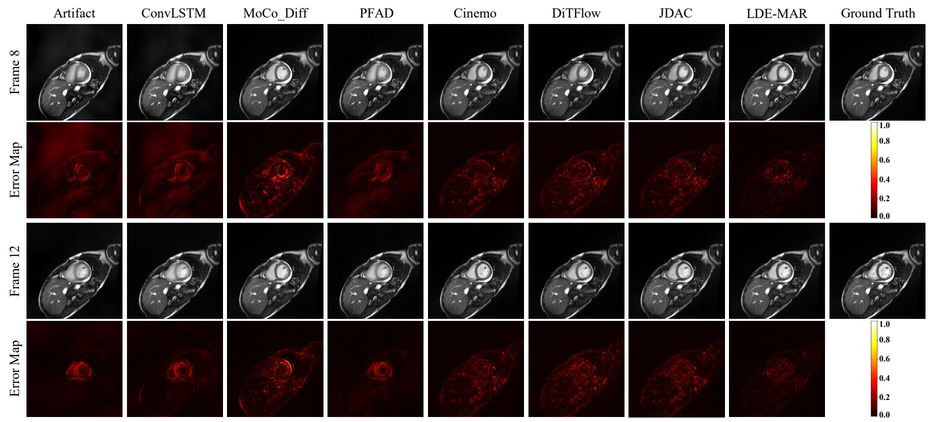

Fig. 2. Qualitative comparison of the LDE-MAR against other methods on one subject in the in-house (20f) dataset. The 8th and 12th frames in cine-MRI and their error maps are displayed.

of frames 8 and 12 from a 20-frame sequence in our private dataset. These frames represent challenging high-motion phases. Their error maps show improved extracardiac structure and smaller cardiac-region errors. These selected-frame results assess spatial reconstruction at challenging phases; full-cycle temporal preservation is not evaluated here.

3.3 Ablation Study

Effectiveness of Different Dictionaries. Table 2 uses D as the minimal motion branch because R produces region activations but cannot encode signed motion direction or magnitude; R-only and $R + S$ therefore cannot instantiate Eq. (4). We isolate the contribution of R through two matched comparisons. Adding R to D (Model 1 $\rightarrow$ Model 2) improves MAE/SSIM/PSNR from 0.034/0.875/28.419 to 0.029/0.889/29.528. Adding R to $D + S$ (Model 4 $\rightarrow$ the full model) further improves all three metrics: MAE decreases from 0.022 to 0.018, SSIM increases from 0.905 to 0.912, and PSNR increases from 30.073 to

Table 2. Hierarchical ablation of the dictionaries on the mixed in-house and ACDC dataset. Since R gates latent motion but cannot encode signed motion independently, D is used as the minimal motion branch.

Model	Dictionary Module			Index		
	Direction ($\mathcal{D}$)	Region ($\mathcal{R}$)	Noise ($\mathcal{S}$)	MAE↓	SSIM↑	PSNR↑
Baseline	—	—	—	0.037	0.864	27.515
Model 1	✓	—	—	0.034	0.875	28.419
Model 2	✓	✓	—	0.029	0.889	29.528
Model 3	—	—	✓	0.031	0.878	28.031
Model 4	✓	—	✓	0.022	0.905	30.073
LDE-MAR (Ours)	✓	✓	✓	**0.018**	**0.912**	**30.544**

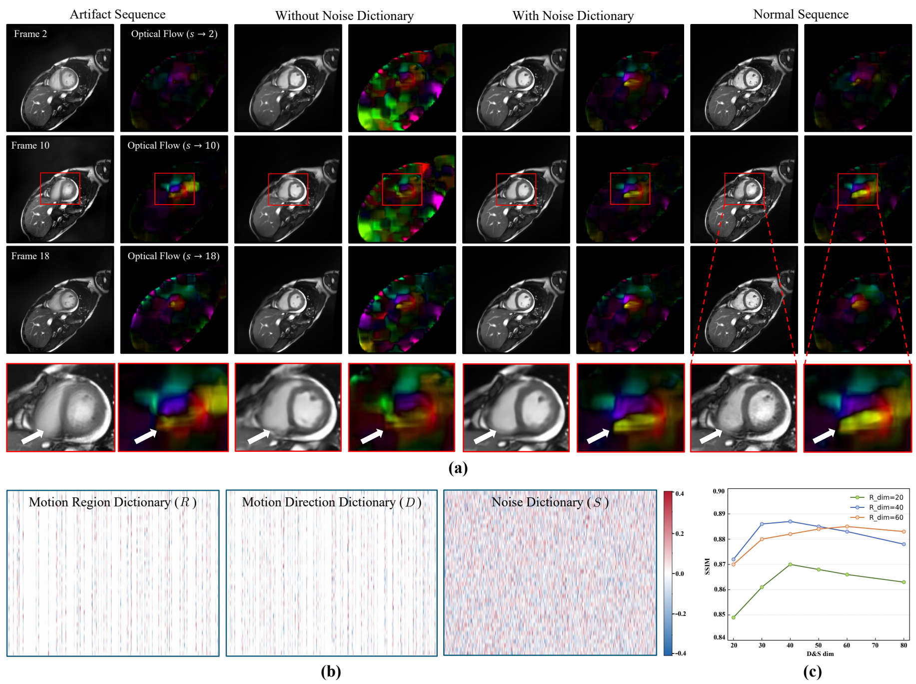

Fig. 3. Visualization of dictionary functions. The optical flow field in (a) demonstrates the suppression of motion artifacts by the noise dictionary. (b) shows the sparsity of the motion dictionaries and the randomness of the noise dictionary. (c) shows the influence of the dictionary dimension on the generated image quality.

30.544. Although the latter gain is modest, its consistency across all metrics indicates that R complements rather than duplicates D. Among single additions to D, S provides the largest improvement, whereas S alone (Model 3) remains inferior to $D + S$ (Model 4), showing that artifact correction benefits from a clean-motion prior. Figure 3(a) shows reduced flow corruption after adding S, while Fig. 3(b) contrasts the structured motion dictionaries with the less structured noise dictionary.

Influence of dictionary dimensionality. We further investigates the impact of dictionary dimensionality on model performance, as shown in Fig. 3(c). Owing to the consistency of the latent motion space, we assume that the motion-direction dictionary D and the noise dictionary S share the same dimensionality. By evaluating the SSIM of the reconstructed images under different dimensionality settings, we determine that the model achieves optimal performance when all three dictionaries are set to a dimensionality of 40.

4 Conclusion

LDE-MAR separates clean latent motion from arrhythmia-related artifact deviations using staged region, direction, and noise dictionaries, improving frame-wise reconstruction on the in-house and ACDC datasets. This separation is an inductive bias rather than a physiological guarantee: although the in-house cohort includes clinically acquired arrhythmic cases, paired evaluation relies on simulated acquisition artifacts, and performance under frequent or severe arrhythmias remains insufficiently established. The noise branch may attenuate pathological motion not represented by the pretrained clean-motion space. Evaluation is also limited to pixel metrics and selected frames, without full-cycle temporal, perceptual, or downstream EF/volume validation; comparisons omit dedicated dictionary-learning correction methods. These issues require prospective clinical testing and broader evaluation.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (62476001) 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. Axel, L., Bhatla, P., Halpern, D., Magnani, S., Stojanovska, J., Barbhaiya, C.: Correlation of mri premature ventricular contraction activation pattern in bigeminy with electrophysiology study-confirmed site of origin. The International Journal of Cardiovascular Imaging **39**(1), 145–152 (2023). https://doi.org/10.1007/s10554-022-02707-8
2. Bernard, O., Lalande, A., Zotti, C., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018). https://doi.org/10.1109/TMI.2018.2837502
3. Bishara, D., Hong, K., Davó-Jiménez, M., Collins, J.D., Topel, C., Fan, L., Sarnari, R., Knight, B.K., Lee, D.C., Kim, D.: Highly accelerated real-time cine mri pulse sequence for cardiac implantable electronic devices and arrhythmias. Radiology: Cardiothoracic Imaging **7**(6), e240554 (2025). https://doi.org/10.1148/ryct.240554
4. Contijoch, F., Han, Y., Kamesh Iyer, S., Kellman, P., Gualtieri, G., Elliott, M.A., Berisha, S., Gorman III, J.H., Gorman, R.C., Pilla, J.J., Witschey, W.R.T.: Closed-loop control of k-space sampling via physiologic feedback for cine mri. PLOS ONE **15**(12), e0244286 (2020). https://doi.org/10.1371/journal.pone.0244286
5. Feng, L., Srichai, M.B., Lim, R.P., Harrison, A., King, W., Adluru, G., DiBella, E.V.R., Sodickson, D.K., Otazo, R., Kim, D.: Highly accelerated real-time cardiac cine mri using k-t sparse-sense. Magnetic Resonance in Medicine **70**(1), 64–74 (2013). https://doi.org/10.1002/mrm.24440
6. Gao, Z., Guo, Y., Zhang, J., Zeng, T., Yang, G.: Hierarchical perception adversarial learning framework for compressed sensing mri. IEEE Transactions on Medical Imaging **42**(6), 1859–1874 (2023). https://doi.org/10.1109/TMI.2023.3240862

7. Kofler, A., Dewey, M., Schaeffter, T., Wald, C., Kolbitsch, C.: Spatio-temporal deep learning-based undersampling artefact reduction for 2d radial cine mri with limited training data. IEEE Transactions on Medical Imaging **39**(3), 703–717 (2020). https://doi.org/10.1109/TMI.2019.2930318
8. Leiner, T., Rueckert, D., Suinesiaputra, A., Baeßler, B., Nezafat, R., Išgum, I., Young, A.A.: Machine learning in cardiovascular magnetic resonance: Basic concepts and applications. Journal of Cardiovascular Magnetic Resonance **21**(1), 61 (2019). https://doi.org/10.1186/s12968-019-0575-y
9. Li, F., Zhou, Z., Fang, Y., Cai, J., Wang, Q.: Moco-diff: Adaptive conditional prior on diffusion network for mri motion correction. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Lecture Notes in Computer Science, vol. 15006, pp. 411–421. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72089-5_39
10. Liu, J., Kocak, M., Supanich, M., Deng, J.: Motion artifacts reduction in brain mri by means of a deep residual network with densely connected multi-resolution blocks (drn-dcmb). Magnetic Resonance Imaging **71**, 69–79 (2020). https://doi.org/10.1016/j.mri.2020.05.002
11. Liu, S., Thung, K.H., Qu, L., Lin, W., Shen, D., Yap, P.T.: Learning mri artefact removal with unpaired data. Nature Machine Intelligence **3**(1), 60–67 (2021). https://doi.org/10.1038/s42256-020-00270-2
12. Liu, Y., Diao, J., Zhou, Z., Qi, H., Hu, P.: Cardiac cine mri motion correction using diffusion models. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024). https://doi.org/10.1109/ISBI56570.2024.10635444
13. Longère, B., Allard, P.E., Gkizas, C.V., Coisne, A., Hennicaux, J., Simeone, A., Schmidt, M., Forman, C., Toupin, S., Montaigne, D., Pontana, F.: Compressed sensing real-time cine reduces cmr arrhythmia-related artifacts. Journal of Clinical Medicine **10**(15), 3274 (2021). https://doi.org/10.3390/jcm10153274
14. Lyu, Q., Shan, H., Xie, Y., Kwan, A.C., Otaki, Y., Kuronuma, K., Li, D., Wang, G.: Cine cardiac mri motion artifact reduction using a recurrent neural network. IEEE Transactions on Medical Imaging **40**(8), 2170–2181 (2021). https://doi.org/10.1109/TMI.2021.3073381
15. Ma, X., Wang, Y., Jia, G., Chen, X., Wong, T.T., Li, Y.F., Chen, C.: Consistent and controllable image animation with motion diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7288–7298 (2025)
16. Ming, Z., Pogosyan, A., Christodoulou, A.G., Finn, J.P., Ruan, D., Nguyen, K.L.: Dynamic regularized adaptive cluster optimization (draco) for quantitative cardiac cine mri in complex arrhythmias. Journal of Magnetic Resonance Imaging **61**(1), 248–262 (2025). https://doi.org/10.1002/jmri.29425
17. Oh, G., Jung, S., Lee, J.E., Ye, J.C.: Annealed score-based diffusion model for mr motion artifact reduction. IEEE Transactions on Computational Imaging **10**, 43–53 (2024). https://doi.org/10.1109/TCI.2023.3347917
18. Oh, G., Lee, J.E., Ye, J.C.: Unpaired mr motion artifact deep learning using outlier-rejecting bootstrap aggregation. IEEE Transactions on Medical Imaging **40**(11), 3125–3139 (2021). https://doi.org/10.1109/TMI.2021.3089708
19. Pan, J., Hamdi, M., Huang, W., Hammernik, K., Kuestner, T., Rueckert, D.: Unrolled and rapid motion-compensated reconstruction for cardiac cine mri. Medical Image Analysis **91**, 103017 (2024). https://doi.org/10.1016/j.media.2023.103017

20. Pondaven, A., Siarohin, A., Tulyakov, S., Torr, P., Pizzati, F.: Video motion transfer with diffusion transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22911–22921 (2025)
21. Sechtem, U., Pflugfelder, P.W., White, R.D., Gould, R.G., Holt, W., Lipton, M.J., Higgins, C.B.: Cine mr imaging: Potential for the evaluation of cardiovascular function. American Journal of Roentgenology **148**(2), 239–246 (1987). https://doi.org/10.2214/AJR.148.2.239
22. Sui, Z., Palaniappan, P., Paganelli, C., Kurz, C., Landry, G., Riboldi, M.: Imaging error reduction in radial cine-mri with deep learning-based intra-frame motion compensation. Physics in Medicine & Biology **69**(22), 225011 (2024). https://doi.org/10.1088/1361-6560/ad8dd0
23. Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., Dosovitskiy, A.: Mlp-mixer: An all-mlp architecture for vision. In: Advances in Neural Information Processing Systems. vol. 34, pp. 24261–24272 (2021)
24. Wang, Y., Yang, D., Bremond, F., Dantcheva, A.: Lia: Latent image animator. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 10829–10844 (2024). https://doi.org/10.1109/TPAMI.2024.3449075
25. Wang, Y., Yang, D., Chen, X., Bremond, F., Qiao, Y., Dantcheva, A.: Lia-x: Interpretable latent portrait animator. arXiv preprint arXiv:2508.09959 (2025)
26. Xu, J., Zhou, D., Hu, L., Guo, J., Yang, F., Liu, Z., Wang, N., Gao, X.: Motion artifact removal in pixel-frequency domain via alternate masks and diffusion model. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 8878–8886 (2025). https://doi.org/10.1609/aaai.v39i8.32960
27. Zhang, L., Wu, M., Wang, L., Steffens, D.C., Potter, G.G., Liu, M.: Iterative learning for joint image denoising and motion artifact correction of 3d brain mri. Medical Image Analysis **108**, 103872 (2026). https://doi.org/10.1016/j.media.2025.103872