Cardiac Physiology Knowledge-driven Diffusion Model for Contrast-free Synthesis Myocardial Infarction Enhancement

Ronghui Qi[1], Xiaohu Li[2], Lei Xu[3], Jie Zhang[2⋆], Yanping Zhang[1], and Chenchu Xu[1,4]

[1] School of Computer Science and Technology, Anhui University, Hefei, China
[2] First Affiliated Hospital of Anhui Medical University, Hefei, China
[3] Department of Radiology, Beijing AnZhen Hospital, Beijing, China
[4] Institute of Artifcial Intelligence, Hefei Comprehensive National Science Center, Hefei, China

Abstract. Contrast-free AI myocardial infarction enhancement (MIE) synthesis technology has a significant impact on clinics due to its ability to eliminate contrast agents (CAs) administration in the current MI diagnosis. In this paper, we propose a novel cardiac physiology knowledge-driven diffusion model (CPKDM) that, for the first time, integrates cardiac physiology knowledge into cardiac MR data to guide the synthesis of high-quality MIE, thereby enhancing the generalization performance of MIE synthesis. The combining helps the model understand the principles behind the data mapping between non-enhanced image inputs and enhanced image outputs, informing the model on how and why to synthesize MIE. CPKDM leverages cardiac mechanics knowledge and MR imaging atlas knowledge to respectively guide the learning of kinematic features in CINE sequences and morphological features in T1 sequences. Moreover, CPKDM proposes a kinematics-morphology diffusion integration model to progressively fuse kinematic and morphological features for precise MIE synthesis. Evaluation on 195 patients including chronic MI and normal controls, CPKDM significantly improves performance (SSIM by at least 4%) when comparing with the five most recent state-of-the-art methods. These results demonstrate that our CPKDM exhibits superiority and offers a promising alternative for clinical diagnostics.

Keywords: Contrast-free technology · Cardiac physiology · Diffusion model

1 Introduction

The contrast-free AI myocardial infarction enhancement (MIE) synthesis technology [1,2] has generated significant clinic impact for diagnosis. This technology

⋆ Corresponding Authors: Jie Zhang (jiezhang_ayfy@163.com) and Chenchu Xu (cxu332@gmail.com)

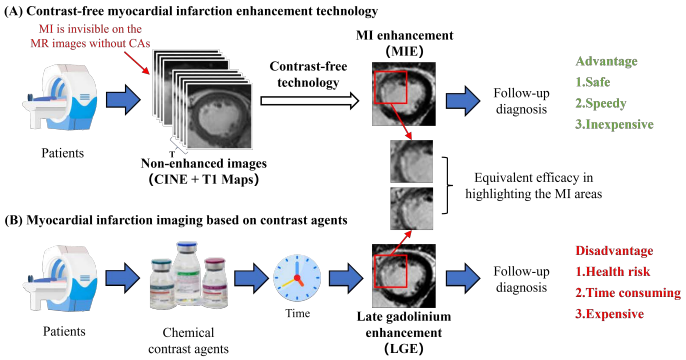

Fig. 1: The contrast-free AI MIE synthesis technology is a safe, fast, and cost-effective clinical alternative for the MI imaging based on CAs.

accurately synthesizes MIE images without contrast agents (CAs), offering effectiveness comparable to late gadolinium enhancement (LGE) images, such as in highlighting the MI areas. It eliminates the high risk [3–5] of using CAs, especially for patients with renal insufficiency. Moreover, it simplifies the clinical workflow by reducing the need for multiple scans [6, 7], saving costs, and preserving clinical resources. Therefore, this technology has great potential in significantly improving the safety and effectiveness of diagnosis [1, 2].

Currently, only two AI methods [1, 2] have attempted to synthesize MIE images without CAs. The principle behind these methods attempting to learn cardiac structures and identify MI areas is the integration of myocardial kinematics and morphology. Specifically, based on this principle, these methods utilize spatio-temporal networks to extract kinematic information from CINE sequences and morphological information from T1 sequences, and then employ generative adversarial networks [8] to generate an enhanced image. Finally, both methods have achieved clinically accepted results.

However, even these two state-of-the-art contrast-free MIE image synthesis technology [1,2] still struggles to demonstrate good performance when faced with highly variable individuals, especially in effectively synthesizing MI areas. The issue arises because these methods are purely data-driven, simply establishing a mapping between non-enhanced image inputs and enhanced image outputs. This simple mapping can mechanically instruct the model on what to do, but lacks the depth to comprehend the intricate correlations between inputs and outputs. When faced with highly variable individuals, this simple mapping relies solely on its learned patterns, unable to dynamically adapt or generalize and to represent the highly variable and complex features, resulting in synthesized images that significantly differ from the actual LGE images.

In this paper, we propose a novel cardiac physiology knowledge-driven diffusion model (CPKDM) to synthesize MIE images without CAs accurately. CPKDM innovatively combines cardiac physiology knowledge with the learning of

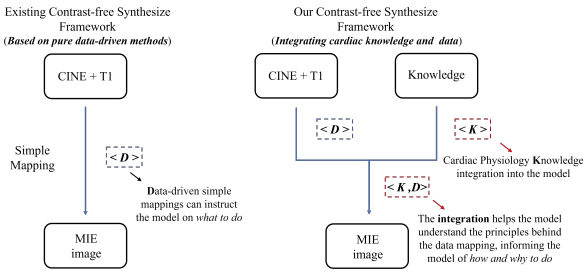

Fig. 2: A novel knowledge-and-data-driven method is proposed for contrast-free MIE image synthesis.

cardiac kinematics and morphological features from data. It helps the model understand the underlying principles behind the data mapping between non-enhanced image inputs and enhanced image outputs, informing the model on how and why to synthesize MIE images. This integration enhances the model's generalization capability when faced with highly variable individuals. Specifically, CPKDM integrates cardiac mechanics knowledge, interpreting the learning of kinematic features through stress calculations, thus better capturing abnormal areas from CINE sequences. Additionally, CPKDM integrates MR imaging atlas knowledge, guiding the learning of morphological features through myocardial imaging signal patterns, thus effectively capturing morphological details. Lastly, CPKDM proposes a kinematics-morphology diffusion integration (KMDI) model to efficiently fuse kinematic and morphological features through progressive diffusion process that incorporates a cross-attention mechanisms, thus precisely synthesizing MIE images.

The paper introduces the following contributions: **1)** For the first time, a knowledge-and-data-driven method is proposed for contrast-free MIE image synthesis, aimed at eliminating health risks associated with CAs and simplifying clinical workflows. **2)** A novel cardiac kinematics and morphology dual-stream framework that leverages myocardial strain and myocardial imaging signal patterns to respectively guide the learning of abnormal motion and structural details is proposed. **3)** A new state-of-the-art performance in the field of contrast-free MIE image synthesis is attained, and design choices and model variants are detailed analyzed.

2 Method

Our CPKDM is composed of a cardiac mechanics-guided kinematics interpretation (CM-KMI) module, a MR-enhanced spectral morphology perception (MR-ESMP) module and a kinematics-morphology diffusion integration (KMDI) model. The CM-KMI focuses on learning kinematic features derived from cardiac mechanics knowledge via CINE sequences. It separates myocardial contraction and rotation features using cardiac mechanics equations and incorporates these fea-

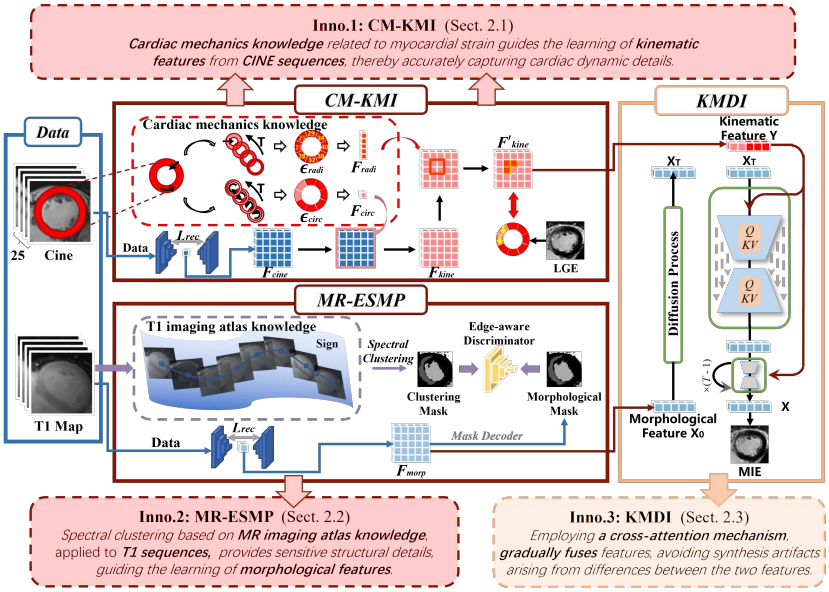

Fig. 3: Our CPKDM integrates a CM-KMI Module, a SA-CMMP Module and a KMDI Model. By incorporating cardiac physiology knowledge, CM-KMI learns kinematic features from CINE sequences, while MR-ESMP learns morphological features from T1 sequences. KMDI progressively fuses kinematic and morphological features, ultimately synthesizing a MIE image.

tures into kinematic features to explain abnormal myocardial motion. The MR-ESMP focuses on learning morphological features derived from MR imaging atlas knowledge via T1 sequences. It segments cardiac structures from multi-phase T1 sequences using clustering based on myocardial imaging signal patterns and integrates these clustering masks to identify sensitive structural features and subtle abnormalities. Subsequently, KMDI focuses on fusing kinematic and morphological features through a diffusion learning [9] process to generate MIE images. It employs a cross-attention mechanism [10] across multiple timesteps to iteratively refine the integration of features, thereby leveraging the fused features to accurately synthesize MIE images.

2.1 Cardiac mechanics-guided kinematics interpretation module

The CM-KMI utilizes myocardial strain during the cardiac cycle as cardiac mechanics knowledge to guide the learning of myocardial motion patterns from CINE sequences. It employs myocardial circumferential and radial strains to distinguish between myocardial contraction and rotation features in spatio-temporal dimensions, thereby accurately capturing cardiac dynamic details. It includes a brightness-based optical flow tracking, a myocardial strain calculation based

on cardiac mechanics, and a motion analysis network. The optical flow tracking enables to analyze changes in pixel brightness in CINE sequences. It utilizes brightness variations between adjacent frames to accurately estimate myocardial motion, thereby generating optical flow fields u. The strain calculation computes the displacement differences ϵ between each pixel and its adjacent pixels within the u.

$$\epsilon = \nabla u + (\nabla u)^T + \nabla u (\nabla u)^T \tag{1}$$

This process quantifies the relative deformation, which is then decomposed into radial components ϵ_{radi} and circumferential components ϵ_{circ} using the respective myocardial strain formulas to obtain circumferential and radial strains.

$$\epsilon_{radi}(x,y) = \boldsymbol{v}(x,y) \cdot \frac{\boldsymbol{p}-\boldsymbol{c}}{\|\boldsymbol{p}-\boldsymbol{c}\|}, \epsilon_{circ}(x,y) = \boldsymbol{v}(x,y) \cdot \begin{pmatrix} -e_{radi_y} \\ e_{radi_x} \end{pmatrix} \tag{2}$$

The calculation for $\epsilon_{\text{radi}}(x,y)$ involves the dot product of the displacement vector $\boldsymbol{v}(x,y)$ with the vector from the myocardial center $\boldsymbol{c}$ to a point $\boldsymbol{p}$, normalized by the magnitude of this vector. $\epsilon_{\text{circ}}(x,y)$ is determined by the dot product of $\boldsymbol{v}(x,y)$ with the unit vector perpendicular to the radial direction $\boldsymbol{e}_{radi}$. The ϵ_{circ} reflects myocardial wall shortening and thickening along its circumference, while the ϵ_{radi} represents the wall's radial thickening and transverse deformation, both crucial for evaluating myocardial mechanics and function. Both are mapped into feature vectors F_{circ} and F_{radi} through encoders. Finally, the motion analysis network sequentially introduces F_{circ} and F_{radi} into the feature flow F_{cine} from data, and then employs a recurrent neural network to process the fused features, obtaining F_{kine}.

2.2 MR-Enhanced spectral morphology perception module

The MR-ESMP utilizes myocardial imaging signal patterns in multiphase T1 sequences as MR imaging atlas knowledge to guide the learning of cardiac structural features from T1 sequences. It exploits signal variations across different time points in multiphase T1 sequences to identify spatio-temporal characteristics of cardiac tissues, thus enhancing the capture of cardiac morphological details. It includes a spectral clustering-based morphology segmentation and a segmentation-based morphological enhancement network. The morphology segmentation clusters eight native T1 inversion recovery–weighted images based on variations in T1 signals to achieve spatially compact clustering, thereby generating a cardiac clustering mask. For each pixel i in image I, it assigns its visual embedding i a three-dimensional sinusoidal positional encoding vector to capture signal variations at the same location across different time points t.

$$PE(i,t,2j) = \sin\left(\frac{t}{10000^{2j/d}}\right), PE(i,t,2j+1) = \cos\left(\frac{t}{10000^{2j/d}}\right) \tag{3}$$

In the morphological enhancement network, the autoencoder generates a morphological mask under the constraints of the clustering mask, directing the en-

coder to focus on morphological features F_{morp}. Meanwhile, the network introduces an edge-aware discriminator [11] that concentrates on the details of cardiac structure.

2.3 Kinematics-morphology diffusion integration model

The KMDI fuses cardiac morphological and kinematic features through a progressive diffusion process that incorporates a cross-attention mechanism, thereby synthesizing authentic image morphologies and precisely highlighting MI areas. KMDI generates a noisy version x_t of the input by adding T steps of noise after introducing cardiac morphological features as the input x_0. The synthesis process uses a conditional denoising autoencoder $\epsilon_\theta(x_t, t, y)$, where the KMDI incorporates cardiac kinematic features as the conditional input y. The corresponding objective can be simplified to

$$L_{DM} := \mathbb{E}_{\mathcal{E}(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta\left(x_t, t, y\right)\|_2^2\right], \tag{4}$$

with t uniformly sampled from $\{1, ..., T\}$. The neural backbone $\epsilon_\theta(\circ, t, \circ)$ is realized as a time-conditional UNet with the cross-attention mechanism. The final output, x, from the synthesis process can be decoded through a decoder in a single step into the image space, generating an MIE image. Meanwhile, KMDI introduces a discriminator, aiming to distinguish the final generated images from real images, thereby enhancing the overall quality of image generation.

3 Experiments

3.1 Experimental setup

Dataset CPKDM is trained and tested on a generalized dataset that consists of 195 MI cases including chronic MI (140), and normal controls (55). All patients completed CINE (14625 images), T1 maps(4680 images) and LGE (585 images) MR imaging scans. All images were obtained using 3-T MRI scanners using T1-weighted imaging. *For more imaging details, please see Supplementary Material.*
Implementation CPKDM used a 5-fold cross-validation (patient-wise) for training and independent testing. All codes are based on the PyTorch. ADAM optimizer [12] with a batch size of 1 and learning rate starts from 1e-4 (0.95 decays, tuned from [1e-3,1e-4,1e-5]). It required 38 hours for training, and 0.24 sec on average for a test image whose size is 128 × 128 on 4 × Nvidia 3090ti GPUs.
Comparative methods and evaluation metrics CPKDM is compared to five state-of-the-art methods, categorized into two groups: 1) Two recent contrast-free MIE synthesis technology, PSCGAN [2] and VNE [1, 13, 14]. 2) Three standard synthesis methods (Pix2Pix [15], BicycleGAN [16], and ResVit [17]) in computer vision. CPKDM and five comparative methods evaluate their performance in two aspects: imaging metrics and clinical metrics. For imaging metrics, our network employs two categories of indices: image-level metrics

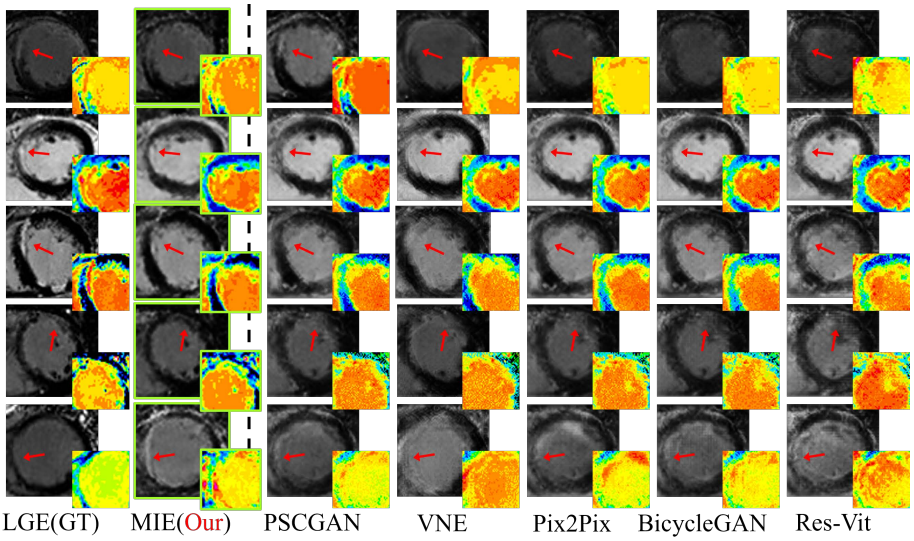

Fig. 4: Our CPKDM achieves the best visual results in experiments.

(SSIM [18], PSNR, LPIPS [19], and MSE) are used to assess the entire image quality, while region-level metrics (SSIM and PSNR) are utilized to evaluate the image quality of MI areas. For clinical metrics, our network assesses the visual spatial distribution and transmurality consistency of myocardial scarring [20].

3.2 Experimental results

Accurate contrast-free MIE synthesis Fig 4 presents that our CPKDM enables the accurate synthesis of MIE without CAs. Our CPKDM surpasses

Table 1: Our CPKDM achieves the best overall performance in comparative experiments and ablation experiments by four well-recognized metrics.

Experiments	Methods	SSIM↑	PSNR↑	LPIPS↓	MSE↓
Comparative methods	PSCGAN	0.74	21.45	0.096	0.0058
	VNE	0.66	19.78	0.145	0.0241
	Pix2Pix	0.63	18.34	0.162	0.0236
	BicycleGAN	0.56	19.52	0.168	0.0283
	Res-Vit	0.65	18.57	0.153	0.0163
Ablation Study	Only CINE	0.63	23.24	0.114	0.0042
	Only T1	0.67	20.34	0.117	0.0053
	Data-drive CINE	0.73	22.36	0.102	0.0063
	Data-drive T1	0.72	23.56	0.92	0.0039
	Pure data-drive	0.68	21.45	0.123	0.0069
	CPKDM	**0.78**	**26.97**	**0.083**	**0.0034**

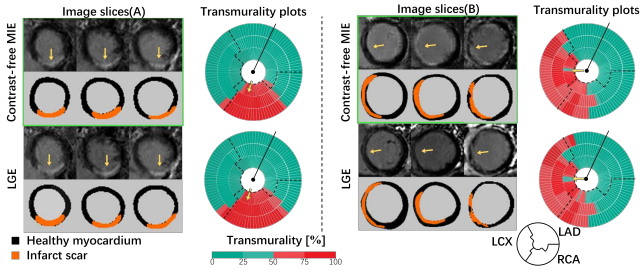

Fig. 5: MIE revealed a right coronary artery territory myocardial scar in patient A, and a left circumflex artery territory myocardial scar in patient B, all with MIE images closely matching the signals revealed by LGE.

existing methods in imageology metrics, achieving a SSIM of 0.78, a PSNR of 26.97, a LPIPS of 0.083, and a MSE of 0.0034, *at the image level*, as shown in Table 1; it attains an SSIM of 0.74 and a PSNR of 24.84, *at the region level*, as shown in Table 2. Moreover, the clinical metrics results demonstrate high concordance in visuospatial distribution and transmurality of myocardial scars between the MIE images and the LGE images, indicating the high performance of our CPKDM, as shown in Fig 5. Note that visualization results are enhanced from contrast-free cardiac MR images, where myocardial scars may be nearly invisible, while visible in LGE. These results proved our CPKDM has great potential to advance contrast-free MIE technology for clinical application and eliminate the high health risks associated with CA injection.

Outperformance of our CPKDM than all comparative methods Fig 4, Table 1 and Table 2 indicate that CPKDM outperforms all five comparative methods in visualization and across four evaluation metrics. In Table 1 and Table 2, comparing with the results of the five comparative methods, our CPKDM improved the SSIM by 0.04-0.22, the PSNR by 5.52-8.63, the LPIPS by 0.013-0.085 and the MSE by 0.0024-0.0249, *at the image level*; our CPKDM improved the SSIM by 0.03-0.14 and the PSNR by 4.52-7.11, *at the region level*. All the outperformance is because our CPKDM combine cardiac physiology knowledge with cardiac MR image data, enhancing the effectiveness of the synthesized MIE.

Table 2: Our CPKDM also achieves the best overall performance at the region level. *For more experiment details, please see Supplementary Material.*

Experiments	Comparative Methods			Ablation Study			
Methods	PSCGAN	VNE	Pix2Pix	Data-drive CINE	Data-drive T1	Prue data-drive	**CPKDM**
SSIM↑	0.71	0.68	0.60	0.70	0.71	0.66	**0.74**
PSNR↑	20.32	17.73	18.59	20.64	22.12	19.42	**24.84**

Superiority of our components in ablation experiments Table 1 and Table 2 show that the integration of cardiac physiology knowledge is the main reason for CPKDM to attain new state-of-the-art performance in this task. Comparing our CPKDM and five different versions (1. Input only CINE sequences; 2. Input only T1 sequences; 3. CINE sequences without knowledge; 4. T1 sequences without knowledge; 5. Both CINE and T1 sequences without knowledge), our CPKDM improved a SSIM by at least 0.05, a PSNR by at least 3.41, a LPIPS by at least 0.009 and a MSE by at least 0.0005, *at the image level*; it improved a SSIM by at least 0.03 and a PSNR by at least 2.72, *at the region level*.

4 Conclusion

In this paper, we propose a novel cardiac physiology knowledge-driven diffusion model for synthesizing MIE images without CAs. Our approach innovatively combines cardiac physiology knowledge with cardiac MR image data to enhance image synthesis accuracy. Our approach obtains the highest performance by evaluating our method on a generalization dataset (195 patients) and comparing with five state-of-the-art methods in four well-recognized metrics. Such results prove that our proposed approach provides contrast-free MIE comparable to CA-based imaging, presenting a safer, faster, and cost-effective alternative for clinical diagnostics.

Acknowledgments. This study was funded by The National Natural Science Foundation of China (62106001), The University Synergy Innovation Program of Anhui Province under Grant (GXXT-2021-007), The Anhui Provincial Natural Science Foundation (2208085Y19), and The Anhui Province Clinical Medical Research Transformation Special Project (202304295107020031).

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

References

1. Zhang, Q., Burrage, M.K., Shanmuganathan, M., Gonzales, R.A., Lukaschuk, E., Thomas, K.E., Mills, R., Leal Pelado, J., Nikolaidou, C., Popescu, I.A., et al.: Artificial intelligence for contrast-free MRI: Scar assessment in myocardial infarction using deep learning–based virtual native enhancement. Circulation **146**(20), 1492–1503 (2022)
2. Xu, C., Xu, L., Ohorodnyk, P., Roth, M., Chen, B., Li, S.: Contrast agent-free synthesis and segmentation of ischemic heart disease images using progressive sequential causal GANs. Medical Image Analysis **62**, 101668 (2020)
3. Xu, C., Xu, L., Brahm, G., Zhang, H., Li, S.: MuTGAN: simultaneous segmentation and quantification of myocardial infarction without contrast agents via joint adversarial learning. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, vol. 11, pp. 525-534. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_59

4. Schieda, N., Blaichman, J.I., Costa, A.F., Glikstein, R., Hurrell, C., James, M., Jabehdar Maralani, P., Shabana, W., Tang, A., Tsampalieros, A., et al.: Gadolinium-based contrast agents in kidney disease: a comprehensive review and clinical practice guideline issued by the Canadian Association of Radiologists. Canadian Journal of Kidney Health and Disease **5**, 2054358118778573 (2018)
5. Xu, C., Howey, J., Ohorodnyk, P., et al.: Segmentation and quantification of infarction without contrast agents via spatiotemporal generative adversarial learning. Medical Image Analysis **59**, 101568 (2020)
6. Xu, C., Xu, L., Gao, Z., Zhao, S., Zhang, H., Zhang, Y., Du, X., Zhao, S., Ghista, D., Li, S.: Direct detection of pixel-level myocardial infarction areas via a deep-learning algorithm. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III, vol. 20, pp. 240-249. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_28
7. Xu, C., Xu, L., Gao, Z., Zhao, S., Zhang, H., Zhang, Y., Du, X., Zhao, S., Ghista, D., Liu, H., et al.: Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. Medical Image Analysis **50**, 82–94 (2018)
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
9. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
11. Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P.: Ea-GANs: Edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Transactions on Medical Imaging **38**(7), 1750–1762 (2019)
12. Zhang, Z.: Improved adam optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), pp. 1–2. IEEE (2018). https://doi.org/10.1109/IWQoS.2018.8624183
13. Swoboda, P., Thompson, P., Zhang, Q., et al.: Gadolinium-Free Virtual Native Enhancement for Chronic Myocardial Infarction Assessment: Independent Blinded Validation and Reproducibility Between Two Centres. Journal of Cardiovascular Magnetic Resonance **26** (2024)
14. Zhang, Q., Burrage, M.K., Lukaschuk, E., et al.: Toward replacing late gadolinium enhancement with artificial intelligence virtual native enhancement for gadolinium-free cardiovascular magnetic resonance tissue characterization in hypertrophic cardiomyopathy. Circulation **144**(8), 589–599 (2021)
15. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
16. Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward multimodal image-to-image translation. Advances in Neural Information Processing Systems **30** (2017)
17. Dalmaz, O., Yurt, M., Çukur, T.: ResViT: Residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)

18. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
19. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018)
20. Kim, R.J., Wu, E., Rafael, A., Chen, E.-L., Parker, M.A., Simonetti, O., Klocke, F.J., Bonow, R.O., Judd, R.M.: The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. New England Journal of Medicine **343**(20), 1445–1453 (2000)