T1 and T2 Mapping Reconstruction Based on Conditional DDPM

Yansong Li, Lulu Zhao, Yun Tian(✉), and Shifeng Zhao

School of Artificial Intelligence, Beijing Normal University, Beijing, China
tianyun@bnu.edu.cn

Abstract. Cardiac magnetic resonance imaging (CMR) has emerged as a crucial imaging modality for the diagnosis of cardiac diseases. T1 and T2 mapping are essential techniques for detecting cardiomyopathies. However, the imaging speed is noticeably slow and conventional mapping models often struggle to produce accurate results when the imaging process is compromised. To overcome this limitation, accelerated mapping techniques have been developed to reduce motion artifacts and enhance image quality. In this study, we propose a novel reconstruction method based on a conditional denoising diffusion probabilistic model (CDDPM). By utilizing accelerated mapping as a conditioning factor and iteratively applying a denoising process, we generate refined T1 and T2 maps from initially corrupted data consisting of pure Gaussian noise. The experimental results of the CMR Reconstruction Challenge demonstrate the effectiveness of our proposed method. Objective indicators show significant improvements, indicating enhanced image quality. Furthermore, our method successfully improves the texture quality of the images, providing more detailed and accurate information for cardiomyopathy diagnosis.

Keywords: CMR reconstruction · T1/T2 mapping · Diffusion model

1 Introduction

Cardiac magnetic resonance imaging (CMR) is a valuable non-invasive imaging technique with excellent soft tissue contrast for diagnosing heart diseases [1]. However, its inherent slow imaging speed and susceptibility to motion artifacts have been limitations. In recent years, there has been significant interest in accelerating CMR image acquisition and reconstruction.

One promising approach is the reconstruction of CMR images from highly undersampled k-space data, obtained during the scan. Advanced reconstruction algorithms that leverage sparsity in the Fourier domain can generate high-quality CMR images from fewer k-space samples. This reduces scan time, improves patient comfort, and mitigates motion artifacts, leading to more accurate diagnosis.

Accelerated CMR image reconstruction techniques, such as compressed sensing, parallel imaging, and deep learning-based methods, exploit image structure

O. Camara et al. (Eds.): STACOM 2023, LNCS 14507, pp. 303–313, 2024.
https://doi.org/10.1007/978-3-031-52448-6_29

and temporal redundancies to achieve high-fidelity reconstructions with minimal data acquisition. These advancements enable faster and more robust CMR examinations, facilitating early detection, accurate diagnosis, and treatment planning for cardiovascular conditions.

Parallel imaging (PI) is a CMR reconstruction technique that utilizes multiple coils to simultaneously obtain data and generate high-quality images in a shorter scanning time. PI can perform image reconstruction with fewer samples by reconstructing k-space undersampling using coil sensitivity. Traditional PI techniques, including SENSE [2], GRAPPA [3], SPIRiT [4], and ESPIRiT [5], have already been used clinically. However, these methods may cause image artifacts when k-space data is highly undersampled.

In the field of accelerating CMR image reconstruction, deep learning plays a crucial role and has achieved good results [6–8]. Among them, the generative model is more commonly applied [9,10]. Generative models are statistical models that learn the distribution of a dataset and generate new data that resembles the original dataset. By utilizing generative models, we can reconstruct high-quality CMR images from highly undersampled k-space data, thereby improving diagnostic accuracy and patient experience.

The choice of a specific generative model depends on the characteristics of the dataset and the requirements of the task at hand. In recent years, popular generative models include Variational Autoencoders (VAEs) [11], Generative Adversarial Networks (GANs) [12], and the highly successful Diffusion Models [13,14] in the field of Mediacal image generation.

Ho et al. [13] proposed the denoising diffusion probabilistic model (DDPM), which uses a variational inference approach to estimate the parameters of a diffusion process. Song et al. [14] proposed Noise-conditioned Score Networks (NCSNs), which rely on a maximum likelihood-based estimation approach. They utilize the score function of the loglikelihood of the data to estimate the parameters of the diffusion process. Jalal et al. [15] proposed Compressed Sensing with Generative Models (CSGM). CSGM trains the score-based generative models on CMR to utilize them as prior information for the inversion pathway in reconstructing realistic CMR data from undersampled CMR in a posterior sampling scheme. Chung et al. [16] proposed a score-based diffusion framework that solves the inverse problem for image reconstruction from accelerated CMR scans. They train a continuous time-dependent score function with denoising score matching. At the inference stage, they iterate between the numerical SDE solver and data consistency step to achieve reconstruction.

In this work, we propose a novel reconstruction method based on Conditional Denoising Diffusion Probability Model (CDDPM) to improve the accuracy and quality of T1 and T2 mapping in CMR. Our approach utilizes accelerated mapping as a conditional factor and incorporates an iterative denoising process to refine T1 and T2 mappings from the initially damaged data. By employing this method, we are able to enhance image quality, reduce motion artifacts, and provide more detailed and accurate information for the diagnosis of cardiomyopathy. We have demonstrated the effectiveness of the proposed method in the

dataset provided by the CMR Reconstruction Challenge [17], with significant improvements in both objective indicators and texture quality of the reconstructed images.

2 Method

The proposed framework is illustrated in Fig. 1. By adding noise to the original data x_0 (fully sampled CMR) and predicting noise through a conditional network with conditional y (accelerated CMR), the reconstructed image is restored from the noise x_T.

2.1 Conditional Denoising Diffusion Probability Model

DDPM uses Variational inference [18] to define a forward Markovian diffusion process q that gradually adds Gaussian noise to a high-resolution image x_0 over T iterations:

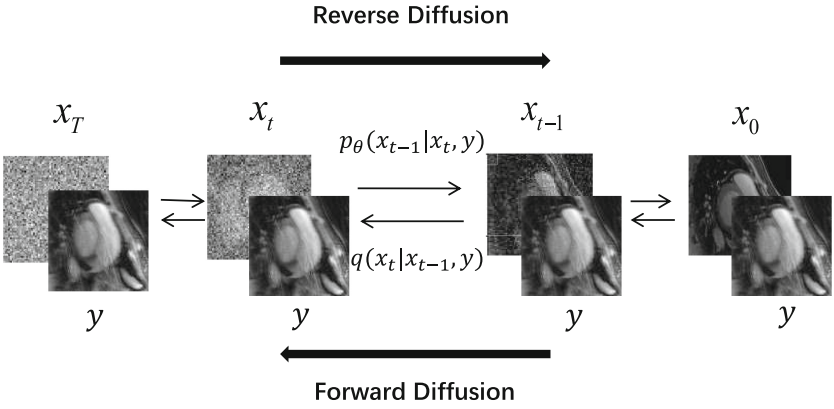

Fig. 1. Framework for CMR reconstruction based on CDDPM.

$$q\left(x_t \mid x_{t-1}\right) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t} \cdot x_{t-1}, \beta_t \cdot \mathbf{I}\right), \forall t \in \{1, 2, \ldots, T\} \tag{1}$$

where the parameters T, $\beta_1, \beta_2, \ldots, \beta_T \in [0, 1)$ represent the number of diffusion steps and the variance schedule across diffusion steps, respectively. $\mathbf{I}$ is the identity matrix and $\mathcal{N} \sim (x; \mu, \sigma)$ represents the normal distribution of mean μ and covariance σ. Considering $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=0}^{T} \alpha_s$ one can directly sample an arbitrary step of the noised latent conditioned on the input x_0 as follows:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \tag{2}$$

Leveraging the above definitions, approximating a reverse process to get a sample from $q(x_0)$. To this end, we can parameterize this reverse process by starting at $p_\theta(x_T) = \mathcal{N} \sim (x_T; \mu, \sigma)$ as follows:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta \left(x_{t-1} \mid x_t\right) \tag{3}$$

$$p_\theta \left(x_{t-1} \mid x_t\right) = \mathcal{N} \left(x_{t-1}; \mu_\theta \left(x_t, t\right), \mathbf{\Sigma}_\theta \left(x_t, t\right) \mathbf{I}\right) \tag{4}$$

To train this model in a way that allows $p(x_0)$ to learn the true data distribution $q(x_0)$, we can optimize the following variational bound on negative log-likelihood:

$$\mathbb{E}\left[-\log p_\theta \left(x_0\right)\right] \leq \mathbb{E}_q \left[-\log p\left(x_T\right) - \sum_{t \geq 1} \log \frac{p_\theta \left(x_{t-1} \mid x_t\right)}{q\left(x_t \mid x_{t-1}\right)}\right] = -L_{\mathrm{VL \cdot B}} \tag{5}$$

Ho et al. [34] found it better not to directly parameterize $\mu_\theta(x_t, t)$ as a neuralnetwork, but rather to train a model $\epsilon_\theta \left(x_t, t\right)$ to predict ϵ. Hence, by reparameterizing Eq. 6, they proposed a simplified objective as follows:

$$\mathrm{L}(\theta) := \mathrm{E}_{\mathrm{t}, x_0, \epsilon} \left[\left\|\epsilon - \epsilon_\theta \left(x_t, t\right)\right\|^2\right] \tag{6}$$

The CDDPM model generates a target image x_0 in T refinement steps. Starting with a pure noise image $x_T \sim \mathcal{N}\left(0; \mathbf{I}\right)$, the model iteratively refines the image through successive iterations$(x_{T-1}, x_{T-2}, \ldots, x_0)$ according to learned conditional transition distributions $p_\theta(x_{t-1} \mid x_t, y)$ such that $x_0 \sim p(x \mid y)$. Therefore, conditional information y (accelerated CMR) needs to be added in Eq. 6:

$$\mathrm{L}(\theta) := \mathrm{E}_{\mathrm{t}, x_0, \epsilon} \left[\left\|\epsilon - \epsilon_\theta \left(x_t, y, t\right)\right\|^2\right] \tag{7}$$

Once the neural network training is finished, start with random noise x_T to $\mathcal{N} \sim (0; \mathbf{I})$, and utilize the neural network to follow the current step T. Predict the noise at step $T-1$ using x_T and compute x_{T-1} based on the predicted mean and variance. Continue this iterative process until $t = 0$ to generate the data x_0 (reconstructed CMR).

2.2 Network Structure

As shown in Fig. 2, our method takes U-Net as backbone. U-Net is particularly well-suited for noise prediction in the diffusion process due to its compatibility with the image's dimensionality. The architecture of U-Net, with its encoder-decoder structure, allows it to capture and model the complex relationships between input images and their corresponding noise patterns. This capability is crucial in scenarios where noise characteristics depend on the specific features and structures within the image. By leveraging U-Net's hierarchical representation learning capabilities, it becomes proficient at learning and predicting noise

patterns that align with the image content, ultimately delivering accurate and meaningful noise predictions.

U-Net is composed of three main blocks: the residual blocks, the attention blocks, and the time embedding as illustrated in Fig. 2.

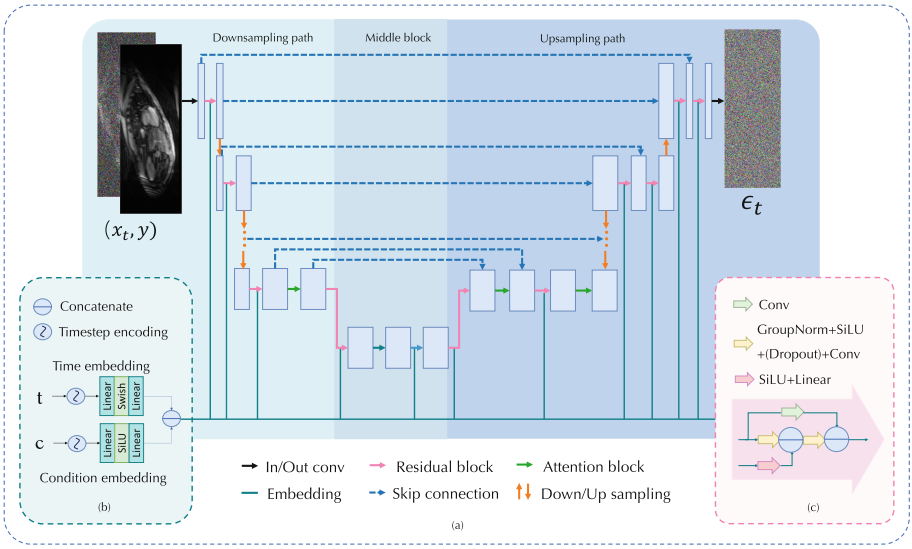

Fig. 2. Architecture of the U-Net used in CDDPM. (a) Overall U-Net structure. (b) Time and condition embedding module. (c) Detailed illustration of the Residual block.

Residual blocks: Inspired by residual learning, the residual blocks serve as a fundamental block of U-Net that help in capturing and preserving essential image features during the encoding and decoding phases. Comprising multiple convolutional layers and skip connections, these blocks facilitate the reuse of low-level features and enable the network to effectively learn residual representations.

Attention blocks: The attention blocks are designed to enhance the model's ability to focus on relevant image regions while suppressing irrelevant or noisy regions. They employ mechanisms like spatial or channel-wise attention to selectively emphasize or suppress certain parts of the feature maps. By incorporating attention mechanisms, U-Net can allocate more resources to crucial image features, thereby enhancing its ability to handle complex patterns and variations.

Time embedding: The Time embedding is specifically employed in scenarios involving input data with temporal dimensions, such as video or sequence data. It encodes the temporal information using specialized techniques. This module allows U-Net to capture temporal dependencies and exploit temporal context during the diffusion process, resulting in more accurate noise prediction over time.

Both the downsampling and upsampling path consist of 30 layers, with 5 of them containing the downsampling or upsampling modules. The downsampling path incorporates 4 attention blocks, the middle block has 1 attention block, and the upsampling path includes 4 attention blocks. The U-Net takes as input a noisy sample denoted as x_t., a conditional state represented as y, and a time step t. It produces an output in the shape of a sample. For the given time step t and condition y, they undergo processing in the embedding module, where they are adjusted and combined to match the size of the channel dimension of the input image features, as depicted in Fig. 2(b). Ultimately, the temporal feature vectors are broadcasted and added to the image features using residual blocks at each level.

3 Experiments

3.1 Datasets

A total of 300 healthy volunteers from a single center are included in the CMRxRecon challenge datasets, which includes T1 mapping and T2 mapping. Training data includes fully sampled k-space data, auto-calibration lines (ACS, 24 lines) and reconstructed images in the *.mat format will be provided. Validation data includes undersampled k-space data with acceleration factors of 4, 8, and 10, sampling mask, and auto-calibration lines (ACS, 24 lines) will be provided. The test dataset is similarly structured, with undersampled k-space data featuring acceleration factors of 4, 8, and 10, the sampling mask, auto-calibration lines (ACS, 24 lines), and the associated reconstructed images. The raw k-space data obtained from the scanner is pre-processed and is subsequently converted to the *.mat format using MATLAB. Furthermore, the multi-coil data is compressed to 10 virtual coils for standardization and storage efficiency.

3.2 Data Preprocessing

During the training phase, the raw k-space data is initially transformed into the image domain using the inverse fast Fourier transform (IFFT), resulting in data with dimensions represented as (kx, ky, sz, t). To create the training set for the neural network, both single-coil and multi-coil data are used, and acceleration factors of 4, 8, and 10 are combined into a single dataset.

We utilize function h to combine the multi-coil k-space data into a unified real-valued image. The merging procedure involves applying an inverse 2D Fourier transform to each coil and subsequently performing a sum-of-squares (SOS) operation. Through SOS calculation, all the coils are effectively combined into a singular real-valued image.

$$SOS\left(x_1, ..., x_N\right) = \sum_{n=1}^{N} |x_n|^2 \tag{8}$$

where x_1,..., x_N are the images from the N coils.

After combining the data, each image is split into individual images of size (kx, ky). To enhance the diversity of the training data, we apply data augmentation techniques to these two-dimensional images. One common augmentation technique is random cropping, where the images are randomly cropped into patches of size 256 × 256. By cropping to a smaller size, the network training process can be faster without affecting image quality.

The same process is also applied to the test set. The under-sampled k-space data in the test set, along with the corresponding sampling mask and auto-calibration lines, are transformed into the image domain using the IFFT. This results in smaller-sized images, which are then processed by the network to predict the reconstructed images.

After obtaining the predicted patch-level images, they are reassembled to their original size. This reassembly process involves stitching the overlapping patches together and handling any potential artifacts or inconsistencies at the patch boundaries. By reassembling the patches, we obtain the final reconstructed images for the test set. These reconstructed images can be evaluated and compared with the ground truth or reference images to assess the performance of the network in terms of CMR reconstruction. This patch-based approach allows for efficient processing and prediction on large images, while also enabling the evaluation of the network's performance on the entire image domain.

3.3 Training Details

Before employing random cropping augmentations, the input data undergoes a normalization process. The model is implemented using PyTorch and trained on a workstation equipped with an NVIDIA Tesla V100 GPU with 32 GB of memory. The model comprises a total of 52,090,177 parameters, and it was trained from scratch, without utilizing any pre-trained model. Training of the model is carried out for 4 * 24 h using the Adam optimizer, with diffusion step set to 1000.

3.4 Evaluation Metrics

The evaluation metrics primarily used in this competition are Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Root Mean Square Error (RMSE). These metrics provide quantitative measures to assess the performance of different methods in terms of image quality and similarity to ground truth.

During the competition phase, since the ground truth or reference images are not available for the test set, the evaluation of the methods can only be conducted using the validation set. The validation set serves as a proxy for assessing the generalization and effectiveness of the proposed approaches.

Participants are expected to submit their predictions for the test set, and these predictions will be evaluated based on the metrics mentioned above using the corresponding ground truth images from the validation set. This evaluation process allows for comparing and ranking the performance of different methods

in terms of noise reduction or image restoration without relying on the ground truth data.

The PSNR metric measures the difference between the predicted images and the ground truth images in terms of their signal power, thereby providing an indication of the reconstruction accuracy. Meanwhile, the SSIM metric evaluates the structural similarity between the predicted and ground truth images, considering factors including luminance, contrast, and structural content. Lastly, the RMSE metric quantifies the average difference between the predicted and ground truth images, providing an overall assessment of the reconstruction error.

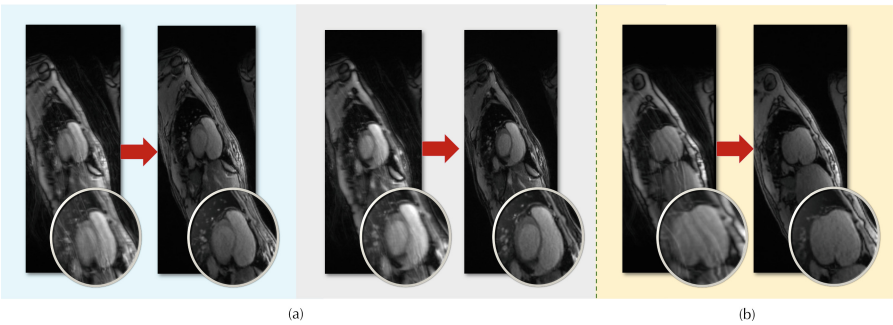

Fig. 3. 4× accelerated reconstruction for Validation Set P001. (a) T1 mapping. (b) T2 mapping.

4 Result

Table 1 presents the reconstruction outcomes from both the multi-coil and single-coil datasets T1/T2, considering undersampling factors of 4, 8, and 10. It's important to note that we utilize a U-Net directly on both the single-coil and combined multi-coil, rather than incorporating an additional step to estimate coil sensitivities. By omitting the use of coil sensitivities, our approach centers on the inherent spatial and temporal correlations present in the data to reconstruct high-quality images. The results demonstrate that the reconstruction of multi-coil data outperforms that of single-coil data. Remarkably, the neural network is trained to learn how to map the aliased coil images to the combined unaliased images [19]. This learning process empowers the exploit the sensitivity information embedded within the data for de-aliasing, without the need for explicit coil sensitivity maps. As a result, our strategy facilitates a more generalized and robust reconstruction process that is not reliant on specific coil sensitivity profiles.

The reconstruction results in Table 1 demonstrate the effectiveness of our method across different undersampling factors and datasets. The achieved image quality is evaluated based on metrics such as PSNR, SSIM, and RMSE, providing

Table 1. The quantitative metric results of the values for T1/T2 mapping of multi-coil and single-coil with different acceleration factors.

Modality	Coil	Acc	PSNR	SSIM	RMSE
T1 mapping	single	4×	30.1389	0.8476	0.1393
		8×	28.8884	0.8361	0.1335
		10×	27.1265	0.7911	1.3116
	multi	4×	32.4824	0.8609	0.1307
		8×	28.1413	0.8453	0.1427
		10×	27.1353	0.8223	0.1528
T2 mapping	single	4×	30.3452	0.8325	0.1331
		8×	29.3465	0.8102	0.1448
		10×	27.7631	0.7891	0.1591
	multi	4×	32.7234	0.8631	0.1343
		8×	29.4661	0.8497	0.1282
		10×	28.8552	0.8314	0.1522

quantitative measures of the reconstruction accuracy, structural similarity, and overall error.

The absence of coil sensitivity information in our approach simplifies the reconstruction pipeline and reduces computational complexity. This makes our method more practical and applicable to various imaging scenarios, especially those where access to coil sensitivity maps may be limited or unavailable.

Overall, the results presented in Table 1 highlight the potential of our coil-insensitive approach in achieving accurate and high-fidelity reconstructions, even under challenging undersampling conditions.

Figure 3 show cases the reconstruction details of different regions in T1 mapping using 4× accelerated sampling. Despite the accelerated sampling, the reconstructed T1 mapping exhibit remarkable details of various anatomical regions.

5 Conclusion

To enhance the accuracy and image quality of T1 and T2 mapping in CMR for the detection of cardiomyopathies, we propose a novel reconstruction method based on CDDPM. By leveraging accelerated mapping as a conditioning factor and employing an iterative denoising process, the proposed method effectively generates refined T1 and T2 maps from initially corrupted data containing pure Gaussian noise. The results from our experiments, conducted as part of the CMR Reconstruction Challenge, demonstrate the effectiveness of the proposed method. Objective indicators reveal significant improvements, indicating a notable enhancement in image quality. Importantly, the proposed method not only improves the accuracy of T1 and T2 mapping but also enhances the texture quality of the images. This enhancement provides more detailed and accurate

information for the diagnosis of cardiomyopathies. A drawback associated with this method is its relatively slow sampling during the training phase. However, considering that our approach strategically decouples the computationally intensive training steps from the more efficient inference steps, the extended training process is typically acceptable. This strategy enables the rapid acquisition of high-quality images immediately after scanning, making it highly practical for real-world applications. In the future, we will focus on accelerating the training process and ensuring image quality.

References

1. Hundley, W.G., Bluemke, D.A., Finn, J.P., et al.: ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents. J. Am. Coll. Cardiol. **55**(23), 2614–2662 (2010)
2. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., et al.: SENSE: sensitivity encoding for fast MRI. Magn. Reson. Med. Official J. Int. Soc. Magn. Reson. Med. **42**(5), 952–962 (1999)
3. Griswold, M.A., Jakob, P.M., Heidemann, R.M., et al.: Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med. Official J. Int. Soc. Magn. Reson. Med. **47**(6), 1202–1210 (2002)
4. Lustig, M., Pauly, J.M.: SPIRiT: iterative self-consistent parallel imaging reconstruction from arbitrary k-space. Magn. Reson. Med. **64**(2), 457–471 (2010)
5. Uecker, M., Lai, P., Murphy, M.J., et al.: ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. Magn. Reson. Med. **71**(3), 990–1001 (2014)
6. Qin, C., Schlemper, J., Caballero, J., et al.: Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging **38**(1), 280–290 (2018). https://doi.org/10.1109/TMI.2018.2863670
7. Qin, C., Duan, J., Hammernik, K., et al.: Complementary time-frequency domain networks for dynamic parallel MR image reconstruction. Magn. Reson. Med. **86**(6), 3274–3291 (2021). https://doi.org/10.1002/mrm.28917
8. Lyu, J., Sui, B., Wang, C., et al.: DuDoCAF: dual-domain cross-attention fusion with recurrent transformer for fast multi-contrast MR imaging. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 474–484. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_45
9. Lyu, J., Li, G., Wang, C., et al.: Region-focused multi-view transformer-based generative adversarial network for cardiac cine MRI reconstruction. Med. Image Anal. **85**, 102760 (2023). https://doi.org/10.1016/j.media.2023.102760
10. Lyu, J., Tong, X., Wang, C.: Parallel imaging with a combination of SENSE and generative adversarial networks (GAN). Quant. Imaging Med. Surg. **10**(12), 2260–2273 (2020). https://doi.org/10.21037/qims-20-518
11. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: Proceedings of International Conference on Learning Representations, ICLR, pp. 1–14. Springer, Cham (2014)
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural. Inf. Process. Syst. **33**, 6840–6851 (2020)

14. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
15. Jalal, A., Arvinte, M., Daras, G., et al.: Robust compressed sensing MRI with deep generative priors. Adv. Neural. Inf. Process. Syst. **34**, 14938–14954 (2021)
16. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated MRI. Med. Image Anal. **80**, 102479 (2022)
17. Wang, C., Lyu, J., Wang, S., et al.: CMRxRecon: an open cardiac MRI dataset for the competition of accelerated image reconstruction. arXiv preprint arXiv:2309.10836 (2023)
18. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: a review for statisticians. J. Am. Stat. Assoc. **112**(518), 859–877 (2017)
19. Kwon, K., Kim, D., Park, H.W.: A parallel MR imaging method using multilayer perceptron. Med. Phys. **44**(12), 6209–6224 (2017)