This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Physics-driven Signal Regularization in Diffusion Models for Multi-contrast MR Image Synthesis

Yejee Shin[1 [0009−0008−0678−6354]], Yunsu Byeon[1 [0009−0009−4263−6414]], Geonhui Son[1 [0009−0009−5946−6706]], Hanbyol Jang[1 [0000−0001−9573−2586]], Dosik Hwang[1,2,3,4,* [0000−0002−2217−2837]], and Sewon Kim[5,6,* [0000−0002−3893−252X]]

[1] School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea
{yejeeshin, dosik.hwang}@yonsei.ac.kr
[2] Artificial Intelligence and Robotics Institute, Korea Institute of Science and Technology, 5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul, Republic of Korea
[3] Department of Oral and Maxillofacial Radiology, Yonsei University College of Dentistry, Seoul, Republic of Korea.
[4] Department of Radiology and Center for Clinical Imaging Data Science (CCIDS), Yonsei University College of Medicine, Seoul, Republic of Korea
[5] NAVER Digital Healthcare Lab, NAVER Cloud, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea
[6] NAVER AI Lab, NAVER Cloud, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea
se1.kim@navercorp.com

Abstract. To achieve accurate diagnostic outcomes, it is often necessary to acquire multiple series of magnetic resonance imaging (MRI) with varying contrasts. However, this process is time-consuming and imposes a significant burden on patients and healthcare providers. While diffusion models have emerged as a highly effective tool for image synthesis, they face challenges in handling the complexities of real-world clinical data and may distort vital information during medical image synthesis. To address these issues, we propose MRDiff, a novel diffusion model for multi-contrast MR image synthesis. MRDiff leverages the intrinsic relationship between different contrast images to derive shared anatomical information based on MR physics equations. Our approach integrates MR physics-based signal regularization for proper content feature generation and employs self-content consistency training to capture accurate anatomical structures. Experimental results demonstrate that MRDiff outperforms existing methods by generating diagnostically valuable images, highlighting its potential for clinical applications in MR image synthesis.

Keywords: Multi-contrast imaging · Image synthesis · Diffusion models.

* Corresponding authors

1 Introduction

Magnetic resonance imaging (MRI) is a preferred non-invasive diagnostic tool essential for patient diagnosis and treatment. Providing a range of MRI modalities offers additional information that characterizes patients and improves patient management. However, acquiring multi-contrast MR images is more time-consuming and expensive compared to other imaging modalities such as Computed Tomography (CT) and X-ray imaging. This poses a significant burden on both patients, particularly those who are uncomfortable lying still for long periods, and healthcare facilities [11, 7].

The advent of MR image synthesis has opened the possibility of significant reductions in scan time. As deep learning-based methods have evolved, numerous studies have been conducted to improve MR image synthesis. Recent studies leveraging generative adversarial networks (GANs) have demonstrated successful performance based on multi-constrast MR images [14, 23, 24, 26, 9, 3]. However, GAN-based models are challenged by the instability of training in achieving convergence [21]. Although these models demonstrate remarkable efficiency in processing clean images, their performance also significantly declines when faced with noise or other disturbances, making them challenging to use in practical applications.

Diffusion models (DMs) [6, 18, 4] have achieved state-of-the-art performance in synthesizing images. DMs are proficient in training and generating complex and diverse images, which reduces the risk of modality collapse and enables more stable training processes [10]. However, most studies have been conducted in noise-free environments [2] and therefore do not fully cover real-world scenarios where noise is often present in MR images. This noise affects various textures depending on the imaging equipment and scanning parameters used. DMs are introduced in the field of medical image synthesis to generate high-quality images [8, 16, 15, 13, 17]. Despite these advancements, the preservation of anatomical integrity during the synthesis of high-quality images presents an ongoing challenge.

In this study, we propose a novel diffusion model for MR image synthesis from multi-contrast images with MR physics-based content regularization (MRDiff). MRDiff focuses on generating MR images with high-frequency details effectively while considering anatomical structure. First, we leverage the principles of MR physics to extract salient features essential for the synthesis of target contrast images. These fundamental tenets of MR physics are operationalized in the latent space. The hypothetical quantitative image features derived through MR physics are designed to be combined with the properties of the diffusion model to regularize the content features. Second, we introduce a self-consistency training mechanism to further refine the content features. This mechanism ensures that the anatomical information remains stable across various perturbed generations, thereby enhancing the model's capacity to capture the underlying physical and anatomical properties.

Our model is evaluated on two datasets: an in-house spine dataset, which includes T1-weighted (T1-w), T2-weighted (T2-w), and fat saturation T2-weighted

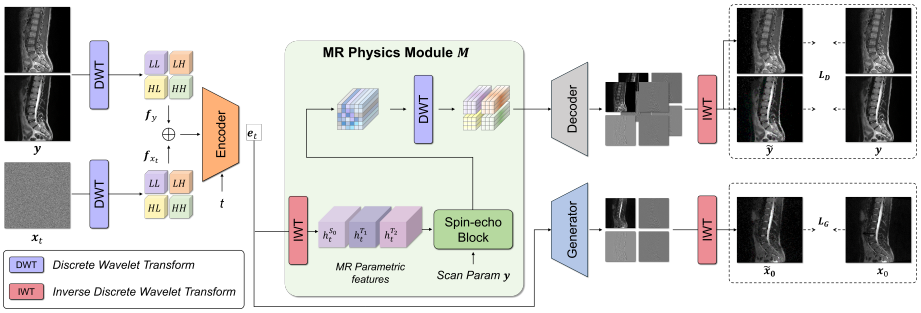

Fig. 1. MRP-Diff for MR image synthesis. $\boldsymbol{x}_t$ is a noisy MR contrast image at randomly sampled noise levels. $\boldsymbol{y}$ is clean multi-contrast images.

(T2 FS) images, and the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [1], which consists of Proton Density (PD), T1-w, and T2-w MR images. Experimental results demonstrate that MRDiff outperforms existing methods in multi-contrast MR image synthesis, effectively bridging the gap between theoretical MR physics and practical image generation.

2 Methods

This section presents the proposed architectural framework, namely MRDiff, which is integrated into the diffusion model to learn the relationship between multi-contrast images and to generate a target MR image from given multi-contrast inputs. Our approach leverages the decomposition module for its superior ability to capture high-frequency textures, and the MR physics to recognize the intrinsic relationship between multi-contrast images, improving the quality of the reconstructed and synthesized images. The overall framework of our method is illustrated in Fig. 1.

2.1 Conditional Diffusion Model

For a training sample $\boldsymbol{x}_0 \in \mathbb{R}^{h\times w\times 1}$, where h and w denote the height and width of the image, the subscript 0 indicates that the sample represents the original data. We set the multi-contrast images $\boldsymbol{y} \in \mathbb{R}^{h\times w\times 2}$ as the condition. We aim to model the distribution of $\boldsymbol{x}_0$, which is $p(\boldsymbol{x}_0 \mid \boldsymbol{y})$. In the forward diffusion process, $\boldsymbol{x}_0$ is gradually perturbed by adding Gaussian noise, resulting in $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T$, which approximates the normal distribution $\boldsymbol{x}_T \sim \mathcal{N}(0, \mathbf{I})$. This process is described by a Markov chain, where $q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ is defined according to a predetermined variance schedule $\beta_1, \beta_2, ..., \beta_T$ as:

$$q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t \mathbf{I}\right). \tag{1}$$

Due to the properties of Gaussian distribution, $\boldsymbol{x}_t$ is directly sampled by:

$$q(\boldsymbol{x}_t \mid \boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right), \tag{2}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{T} \alpha_i$.

We desire to synthesize image through sampling from $p(\boldsymbol{x}_0 \mid \boldsymbol{y})$, achieved by iteratively sampling from $p(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y})$. During the training, encoder E_ϕ, decoder D_ξ, and generator G_θ are trained to predict $\boldsymbol{x}_0$ from $\boldsymbol{x}_t$ and restore $\boldsymbol{y}$ from $\boldsymbol{y}$ as an autoencoder by optimizing the objective with l_2 loss, leading to :

$$L_G = \mathbb{E}_{\boldsymbol{x}_0, \boldsymbol{y}, \epsilon, t} \left[\| \boldsymbol{x}_0 - G_\theta(E_\phi(\boldsymbol{x}_t, \boldsymbol{y}, t)) \|^2 \right], \tag{3}$$

and

$$L_D = \mathbb{E}_{\boldsymbol{x}_0, \boldsymbol{y}, \epsilon, t} \left[\| \boldsymbol{y} - D_\xi(E_\phi(\boldsymbol{x}_t, \boldsymbol{y}, t)) \|^2 \right]. \tag{4}$$

Encoder E_ϕ projects both $\boldsymbol{x}_t$ and $\boldsymbol{y}$, allowing for understanding anatomical information. Decoder D_ξ focuses on reconstructing given multi-contrast data $\boldsymbol{y}$. Generator G_θ receive features from the encoder and creates the target contrast image. Our model focuses on predicting a denoised image denoted by $\boldsymbol{x}_0$.

2.2 Frequency-aware Decomposition

To effectively capture high-frequency details, our model employs two types of operations: the discrete wavelet transform (DWT) and the inverse discrete wavelet transform (IWT) [5]. Given an input image $\boldsymbol{x}_t$ and $\boldsymbol{y}$ are decomposed into low and high subbands through Haar wavelet decomposition. Each matrix is decomposed into four distinct subbands: $\boldsymbol{f}_{x_t} = \{\boldsymbol{f}_{x_t}^{LL}, \boldsymbol{f}_{x_t}^{LH}, \boldsymbol{f}_{x_t}^{HL}, \boldsymbol{f}_{x_t}^{HH}\} \in \mathbb{R}^{\frac{h}{2} \times \frac{w}{2} \times c}$ for $\boldsymbol{x}_t$, and $\boldsymbol{f}_y = \{\boldsymbol{f}_y^{LL}, \boldsymbol{f}_y^{LH}, \boldsymbol{f}_y^{HL}, \boldsymbol{f}_y^{HH}\} \in \mathbb{R}^{\frac{h}{2} \times \frac{w}{2} \times c}$ for $\boldsymbol{y}$. The final output of the decoder represents a decomposed feature for low- and high-frequency components. This particular structure derived from frequency components aids in the reconstruction of $\hat{\boldsymbol{y}}$, where $\hat{\boldsymbol{y}}$ is the prediction through IWT. Similarly, the generator synthesizes the target contrast image leveraging a hypothetical decomposed feature with IWT. As depicted in Eq. (3) and Eq. (4), the objective for optimizing focuses on $\boldsymbol{x}_0$ and $\boldsymbol{y}$, guiding each decomposed feature to effectively learn the information corresponding to both low and high-frequency components.

2.3 MR Physics-based Architecture Design

By utilizing the decomposition and feeding into the encoder block, latent feature $\boldsymbol{e}_t$ is yielded to subband-specific features. Therefore, the IWT operation followed by convolution block is used to generate the quantitative map features as follows:

$$\boldsymbol{h}_t = I_\sigma(\boldsymbol{e}_t) = \mathrm{Conv}(\mathrm{IWT}(\boldsymbol{e}_t)), \tag{5}$$

where $\boldsymbol{h}_t = \{\boldsymbol{h}_t^{T_1}, \boldsymbol{h}_t^{T_2}, \boldsymbol{h}_t^{S_0}\}$. To guide $\boldsymbol{h}_t$ to the quantitative map as content features, we leverage the MR physics, which can facilitate the accurate modeling of the physical properties underlying MR imaging, leading to:

$$S_0(1 - \exp^{-\frac{TR}{T_1}}) \exp^{-\frac{TE}{T_2}}, \tag{6}$$

where TR and TE denote specific repetition time (TR) and echo time (TE) parameters used within MRI scanning protocols. The signal follows the MR physics of the spin-echo pulse sequence. S_0 is the initial signal intensity, indicating the signal strength when fully aligned with the external magnetic field. $\exp^{-\frac{TR}{T_1}}$ accounts for the attenuation due to the relaxation process T_1, where T_1 is the longitudinal relaxation time. Similarly, $\exp^{-\frac{TE}{T_2}}$ captures the attenuation due to the T_2 relaxation process, where T_2 is the transverse relaxation time. This equation allows for analyzing tissue characteristics.

The MR physics is exploited in latent space to capture essential structural and physical information relevant to multi-contrast MR images. The MR physics aims to reflect style features (i.e., contrast) with content features (i.e., anatomical structure), providing a basis for the reconstruction of multi-contrast images. To further enhance the feature extraction process for each frequency band, DWT is employed for separating the features obtained from MR physics into four discrete subband features. This process results in the generation of features corresponding to each subband. The MR physics modulation process is represented as

$$M_\sigma(\boldsymbol{e}_t; \boldsymbol{TR}, \boldsymbol{TE}) = \mathrm{DWT}(\mathrm{Conv}(\boldsymbol{h}_t^{S_0}(1 - \exp^{-\frac{\boldsymbol{TR}}{\boldsymbol{h}_t^{T_1}}}) \exp^{-\frac{\boldsymbol{TE}}{\boldsymbol{h}_t^{T_2}}})), \quad (7)$$

where $\boldsymbol{TR} = \{TR_1, TR_2\}$ and $\boldsymbol{TE} = \{TE_1, TE_2\}$ from MR pulse sequence parameters $\boldsymbol{y}$. The subband features are concatenated and subsequently fed into the decoder.

The generator G_θ receives only the content features that contain information on the anatomical structure, which are guided by the characteristics of the quantitative map. The decoder D_ξ receives the modified features injected in style, which contain anatomical and contrast information. These features are guided by the MR physics equations. The objective function for reconstruction only is rewritten as follows:

$$L_D = \mathbb{E}_{\boldsymbol{x}_0, \boldsymbol{y}, \epsilon, t}\left[\|\boldsymbol{y} - D_\xi(M_\sigma(E_\phi(\boldsymbol{x}_t, \boldsymbol{y}, t)))\|^2\right]. \quad (8)$$

2.4 Self-content Consistency Training

By the random sampling inherent to the diffusion model, the input data is represented as $\boldsymbol{x}_t = \{\boldsymbol{x}_{t_1}, \boldsymbol{x}_{t_2}, ..., \boldsymbol{x}_{t_N}\}$ with a consistent level of noise. Simultaneously, the corresponding encoder features are denoted as $\boldsymbol{e}_t = \{\boldsymbol{e}_{t_1}, \boldsymbol{e}_{t_2}, ..., \boldsymbol{e}_{t_N}\}$. It is indicated that the set of $\boldsymbol{e}_t$ exhibits variability due to the variations in $\boldsymbol{x}_t$. However, since the given multi-contrast images $\boldsymbol{y}$ have the same anatomical structure information, they inherently converge to a single point that represents the shared anatomical structure. We define a global content feature $\bar{\boldsymbol{e}}_t$ to represent the anatomical information shared across multiple variants, leading to:

$$\bar{\boldsymbol{e}}_t = \frac{1}{N}\sum_{i=1}^{N} E_\phi\left(\boldsymbol{x}_{ti}, \boldsymbol{y}, t\right). \quad (9)$$

Each local content feature $\boldsymbol{e}_t$ and global content feature $\bar{\boldsymbol{e}}_t$ is trained with l_2 loss to ensure that the generated features converge toward a consistent representation of the anatomy. Our content consistency loss is defined as follows:

$$L_{CCT} = \mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{y},\epsilon,t,N}\left[\|E_\phi\left(\boldsymbol{x}_t,\boldsymbol{y},t\right) - \bar{\boldsymbol{e}}_t\|_2^2\right]. \tag{10}$$

As a result, the total loss function of the encoder, MR physics module, decoder, and generator, L_{train} is defined as the accumulation of the respective loss functions as follows:

$$L_{train} = \lambda_1 L_G + \lambda_2 L_D + \lambda_3 L_{CCT}, \tag{11}$$

where λ_1, λ_2, and λ_3 are the parameters of each loss function, respectively.

3 Experiments and Results

3.1 Setup

Dataset. The ADNI dataset [1] consists of a total of 737 MRI scans from non-cognitively impaired (NC) individuals and Alzheimer's Disease (AD) patients. There are three different contrasts: T1-weighted Magnetization Prepared Rapid Gradient Echo (MP-RAGE), T2-weighted Turbo Spin Echo (TSE), and Proton Density (PD) TSE. For T1-weighted MP-RAGE, the TR ranges from 6.608 ms to 10.4 ms, TE varies between 2.84 ms and 4.436 ms, and the inversion time (TI) is set at 1000 ms. The T2-weighted TSE shows TR ranging from 2700 ms to 5650 ms, with TE values spanning 95.22 ms to 101.84 ms. For PD TSE, the TR values range from 2700 ms to 5650 ms, and the TE values range from 9.78 ms to 11.016 ms. For training, 70% of the dataset was randomly selected, while for testing, 100 subjects with 1000 slices were randomly chosen.

The spine dataset consists of two subsets collected from Gangnam Severance Hospital in Seoul, Korea. The first spine subset (Subset 1) includes MRI scans from 240 subjects. These scans encompass subjects without specific disease considerations. The second spine subset (Subset 2) comprises scans from 70 subjects who were examined for bone metastasis or red marrow hyperplasia. Each subject provided seventeen slices for three different contrasts: T1-weighted TSE, T2-weighted TSE, and T2 FS TSE. The pulse sequence parameters were set for each contrast: T1-weighted TSE had a TR of 450 ms and TE of 9.8 ms. T2-weighted TSE parameters were TR of 3760 ms and TE of 100 ms. Lastly, for T2 FS TSE, the TR was 4580 ms, and TE was 113 ms. Subset 1, containing 2040 slices, was used for training, while Subset 2, with 1003 slices, was used for testing.

All datasets were approved in advance by the institutional review board. Datasets were only available to authorized researchers for research purposes.

Metrics. In evaluating the performance of MRDiff, we employ widely recognized metrics: Peak Signal-to-Noise Ratio (PSNR) (dB) [20] and Structural Similarity Index (SSIM). Feature Similarity (FSIM) [25] is also used to evaluate structure

Table 1. Quantitative results of the comparison study on two datasets (in-house and ADNI dataset). The best PSNR, SSIM, FSIM, and GMSD values are in bold. Paired t-tests were utilized to evaluate the statistical difference between the proposed method and other comparative methods (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Dataset	Methods	PSNR(dB) ↑	SSIM ↑	FSIM ↑	GMSD ↓
In-house	HiNET[26]	27.33±2.34***	0.7938±0.0606***	0.8765±0.0316***	0.1035±0.0250***
	ResViT[3]	28.12±2.72**	0.8056±0.0707***	0.8826±0.0331***	0.0964±0.0266***
	BlochGAN[9]	28.24±2.70*	0.8136±0.0695***	0.8868±0.0332	0.0940±0.0263*
	CM[19]	25.98±2.28***	0.7945±0.0564***	0.8526±0.0287***	0.1173±0.0240***
	cDPM[16]	28.36±3.05	0.8051±0.0802***	**0.8909±0.0372**	0.0925±0.0288*
	M2DN[13]	25.00±1.94***	0.7835±0.0643***	0.8468±0.0280***	0.1358±0.0237***
	MRDiff	**28.54±3.41**	**0.8228±0.0624**	0.8893±0.0357	**0.0913±0.0292**
ADNI	HiNET[26]	19.47±1.73***	0.7350±0.0552***	0.8657±0.0328***	0.1842±0.0245***
	ResViT[3]	19.62±1.90***	0.7335±0.0618***	0.8658±0.0372***	0.1844±0.0289***
	BlochGAN[9]	19.67±2.17***	0.7351±0.0694***	0.8680±0.0429***	0.1841±0.0315***
	CM[19]	18.35±2.08***	0.7510±0.0690***	0.8600±0.0341***	0.1857±0.0451***
	cDPM[16]	20.07±2.30***	0.7568±0.0839**	0.8739±0.0434*	0.1797±0.0331***
	M2DN[13]	18.77±1.25***	0.7576±0.0451***	0.8620±0.0233***	0.1875±0.0165***
	MRDiff	**20.78±2.33**	**0.7715±0.0757**	**0.8804±0.0416**	**0.1727±0.0313**

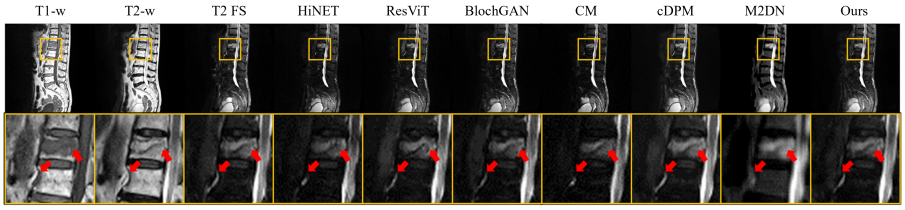

Fig. 2. Qualitative results of the comparison study on the in-house spine dataset.

similarity. Gradient Magnitude Similarity Deviation (GMSD) [22] is utilized to evaluate the texture similarity of the generated images. To assess statistical differences in the assessment results, paired t-tests [12] are used.

Implementation Details. We utilize a UNet-based ADM [4] as our MR image synthesis model. The model is trained on the in-house and ADNI datasets for 300k and 350k iterations, respectively. Training is conducted using an Adam optimizer with a learning rate of 2×10^{-4} and a batch size of 8. The forward process is set to 1000 timesteps during training. During inference, MRDiff implements the reverse diffusion process to synthesize MR images in 50 sampling steps. We set N, λ_1, λ_2, and λ_3 as 3, 1, 1, and 0.01.

3.2 Results

We selected six recent state-of-the-art (SOTA) image synthesis methods, including HiNet [26], ResViT [3], BlochGAN [9], CM [19], cDPM [16], and M2DN [13]. The quantitative comparisons are presented in table 1, where our method outperforms existing approaches across most evaluation metrics. These improve-

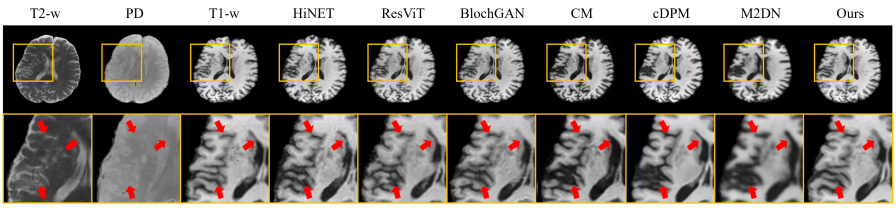

Fig. 3. Qualitative results of the comparison study on the ADNI dataset.

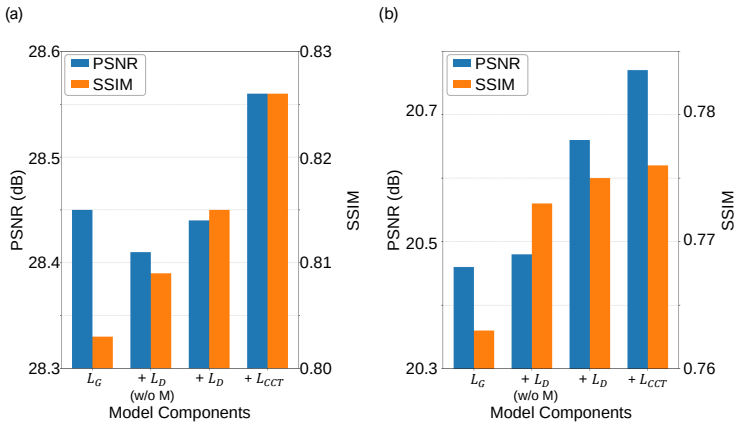

Fig. 4. Effect of model components. (a) represents the results for the in-house spine dataset, and (b) represents the results for the ADNI dataset.

ments highlight our method's capability to effectively capture complex details and structural information in MR images. The qualitative results of the comparison study are presented in Fig. 2 and Fig. 3. The images in the second row, which are enlarged versions of the areas marked with yellow boxes in the first row, respectively, highlight specific regions for detailed evaluation. As depicted in Fig. 2, each comparison method correctly reproduces contrast and texture and closely matches the reference images. However, regions of edema within the vertebral bodies, as indicated by the red arrows in the enlarged images, are not accurately captured by most of the methods evaluated. MRDiff correctly depicts these areas and the region is bright enough. Fig. 3 illustrates that comparative methods are successful in achieving adequate levels of contrast in relation to reference images. However, within the axial views, the comparison methods inaccurately represent specific regions related to the ventricles and cortex, highlighted by red arrows in the enlarged views. In contrast, MRDiff accurately describes these regions, ensuring fidelity without significant distortion.

Ablation studies were conducted to determine the significance of each component within our method. Fig. 4 shows the results of the effect of the model components. It shows that utilizing L_G outperforms other existing methods. The

addition of other proposed components, such as L_D without modulation of MR physics M, L_D, and L_{CCT}, leads to improved performance. Although there is a slight reduction in PSNR values, an analysis of both PSNR and SSIM metrics across diverse datasets reveals an overall upward trend in performance. When all components are integrated, our model demonstrates superior performance in the synthesis of MR images.

4 Conclusion

In this study, we propose a multi-contrast MR image synthesis method by integrating an encoder-decoder-generator architecture with MR physics-based and CCT loss. The experimental results indicate that the MRDiff outperforms existing networks in the multi-contrast datasets. These results demonstrate the potential of our approach to revolutionize medical imaging by providing more reliable and detailed visualizations to improve patient diagnosis and treatment.
Acknowledgements. This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (RS-2025-02215070, RS-2025-02217919); in part by Artificial Intelligence Graduate School Program at Yonsei University (RS-2020-II201361); in part by the KIST Institutional Program (2E33801, 2E33800); in part by the Yonsei Signature Research Cluster Program of 2024 (2024-22-0161).
Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alzheimer's disease neuroimaging initiative (adni). https://adni.loni.usc.edu/ (2024), accessed: 2024-03-07
2. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. arXiv preprint arXiv:2209.14687 (2022)
3. 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)
4. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
5. Graps, A.: An introduction to wavelets. IEEE computational science and engineering **2**(2), 50–61 (1995)
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
7. Hollingsworth, K.G.: Reducing acquisition time in clinical mri by data undersampling and compressed sensing reconstruction. Physics in Medicine & Biology **60**(21), R297 (2015)
8. Jiang, L., Mao, Y., Wang, X., Chen, X., Li, C.: Cola-diff: Conditional latent diffusion model for multi-modal mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 398–408. Springer (2023)

9. Kim, S., Jang, H., Hong, S., Hong, Y.S., Bae, W.C., Kim, S., Hwang, D.: Fat-saturated image generation from multi-contrast mris using generative adversarial networks with bloch equation-based autoencoder regularization. Medical Image Analysis **73**, 102198 (2021)
10. Lee, J.R., Shin, Y., Son, G., Hwang, D.: Diffusion bridge: Leveraging diffusion model to reduce the modality gap between text and vision for zero-shot image captioning. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 4050–4059 (2025)
11. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **58**(6), 1182–1195 (2007)
12. McDonald, J.H.: Handbook of Biolological Statistics. New York• (2014)
13. Meng, X., Sun, K., Xu, J., He, X., Shen, D.: Multi-modal modality-masked diffusion network for brain mri synthesis with random modality missing. IEEE Transactions on Medical Imaging (2024)
14. Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering **65**(12), 2720–2730 (2018)
15. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging (2023)
16. Peng, W., Adeli, E., Bosschieter, T., Park, S.H., Zhao, Q., Pohl, K.M.: Generating realistic brain mris via a conditional diffusion probabilistic model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 14–24. Springer (2023)
17. Shin, Y., Lee, Y., Jang, H., Son, G., Kim, H., Hwang, D.: Anatomical consistency and adaptive prior-informed transformation for multi-contrast mr image synthesis via diffusion model. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30918–30927 (2025)
18. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
19. Song, Y., Dhariwal, P., Chen, M., Sutskever, I.: Consistency models. In: International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 202, pp. 32211–32252. PMLR (2023)
20. 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)
21. Xiao, Z., Kreis, K., Vahdat, A.: Tackling the generative learning trilemma with denoising diffusion gans. arXiv preprint arXiv:2112.07804 (2021)
22. Xue, W., Zhang, L., Mou, X., Bovik, A.C.: Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE transactions on image processing **23**(2), 684–695 (2013)
23. Yurt, M., Dar, S.U., Erdem, A., Erdem, E., Oguz, K.K., Çukur, T.: mustgan: multi-stream generative adversarial networks for mr image synthesis. Medical image analysis **70**, 101944 (2021)
24. Yurt, M., Özbey, M., Dar, S.U., Tinaz, B., Oguz, K.K., Çukur, T.: Progressively volumetrized deep generative models for data-efficient contextual learning of mr image recovery. Medical Image Analysis **78**, 102429 (2022)

25. Zhang, L., Zhang, L., Mou, X., Zhang, D.: Fsim: A feature similarity index for image quality assessment. IEEE transactions on Image Processing **20**(8), 2378–2386 (2011)
26. Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L.: Hi-net: hybrid-fusion network for multi-modal mr image synthesis. IEEE transactions on medical imaging **39**(9), 2772–2781 (2020)