Learning from Good Neighbors: Slice-wise Quality-aware CT-to-Diffusion MRI Synthesis

You-Kyoung Na[1], Kang-Ho Choi[2], Hyung-Jeong Yang[1], Jahae Kim[1,2⋆], and Yeong-Jun Cho[1⋆]

[1] Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju, South Korea
{youkyoung, hjyang, jhbt0607, yj.cho}@jnu.ac.kr
[2] Chonnam National University Hospital, Gwangju, South Korea
{ckhchoikang, jhbt0607}@jnu.ac.kr

Abstract. Cross-modality synthesis from CT to MRI provides MRI-level soft-tissue detail without additional scans, yet existing approaches face three major limitations: the substantial modality gap between CT and MRI, slice-level quality variation without explicit estimation, and underutilization of 3D anatomical continuity. While prior work has primarily focused on MRI-to-CT synthesis, CT-to-MRI generation remains challenging. We propose LGN (Learning from Good Neighbors), a framework for synthesizing MRI from CT, consisting of three components: (1) a Latent Diffusion Model (LDM) for draft MRI generation, (2) a slice-level Quality Predictor trained via LPIPS-based teacher–student learning, and (3) a 3D context-aware Cross-Attention Refiner that selectively improves suboptimal slices using high-quality neighboring slices. Experiments on 398 patients demonstrate consistent improvements in LPIPS, PSNR, and MAE with reduced artifacts and enhanced anatomical consistency, highlighting the effectiveness of quality-guided cross-slice refinement for CT-to-MRI synthesis. The source code is available at https://github.com/YouKyoung-Na/LGN.

Keywords: Medical Imaging · CT-to-MRI Synthesis · Cross-Modality.

1 Introduction

Cross-modality synthesis in medical imaging generates images of one modality from another, reducing the need for additional scans [1]. This is particularly useful in clinical settings where acquiring multiple imaging modalities is challenging. For example, in emergency care, Computed Tomography (CT) is widely used as first-line imaging due to its speed and accessibility. However, CT provides limited soft-tissue contrast compared to Magnetic Resonance Imaging (MRI), which is more sensitive for neuroimaging, including stroke and hemorrhage detection. Despite these advantages, MRI is often impractical due to high cost, claustrophobia, long scan times, and contraindications in patients with certain implants

⋆ Corresponding authors

or devices (e.g., pacemakers, neurostimulators). Therefore, generating MRI from fast, widely available CT is clinically advantageous, as it can provide MRI-level soft-tissue detail without requiring MRI acquisition.

However, existing CT-to-MRI synthesis approaches face several limitations. First, the physical gap between CT and MRI is significant: CT measures X-ray attenuation while MRI captures magnetic resonance signals from hydrogen nuclei. Consequently, many prior works have focused on MRI-to-MRI translation (e.g., T1-to-T2) [2, 3, 5] or MRI-to-CT synthesis [6–9], leaving CT-to-MRI synthesis relatively less studied. Recent diffusion-based models [4] have demonstrated strong performance in cross-modal synthesis [5, 10], yet CT-to-MRI synthesis remains imperfect in practice. Second, generated MRI quality can vary substantially across slices in a 3D volume, but most methods do not quantify slice-level quality. Therefore, a mechanism to mitigate slice-wise quality inconsistency is needed. Medical images exhibit strong anatomical continuity between adjacent slices, yet many methods process slices independently and fail to leverage this 3D context.

To address these limitations, we propose LGN (Learning from Good Neighbors), a framework for generating diffusion-weighted imaging (DWI) from CT. DWI is a widely used MRI sequence for neuroimaging, particularly for assessing acute stroke and related hemorrhagic lesions. DWI includes multiple acquisitions at different b-values; we use b=0 as an anatomical reference and b=1000 to capture diffusion restriction. For brevity, we hereafter use "MRI" to denote DWI (b=0 and b=1000) throughout this paper.

Our framework consists of three components. First, we employ a Latent Diffusion Model (LDM) [16] for CT-to-DWI synthesis. By modeling the diffusion process in a compressed latent space, LDM enables computationally efficient learning of complex cross-modal mappings [5, 10]. Second, we introduce a Quality Predictor to estimate slice-level quality for the LDM-generated MRI. Trained via LPIPS-based teacher–student learning, it predicts slice-level quality scores to guide selective refinement. Third, we design a Cross-Attention Refiner to leverage 3D context. It aggregates information from high-quality neighboring slices to refine the low-quality target slice, explicitly exploiting anatomical continuity across the volume.

To validate the effectiveness of LGN, we evaluate it on 398 patients with paired CT–MRI data from *** Hospital. Results show consistent improvements over existing methods in SSIM, PSNR, and LPIPS. Our method produces higher-quality synthesis than an LDM-only baseline [16], with qualitative results showing reduced artifacts. The Cross-Attention Refiner improves anatomical continuity and slice-level consistency, suggesting that leveraging high-quality neighboring slices is beneficial for CT-to-MRI synthesis. Our main contributions are as follows: (1) We propose LGN, a slice-wise quality-aware CT-to-DWI synthesis framework that integrates latent diffusion with explicit slice-level quality estimation. (2) We introduce a quality-guided cross-slice refinement mechanism that exploits 3D anatomical continuity to mitigate slice-wise inconsistency. (3) We

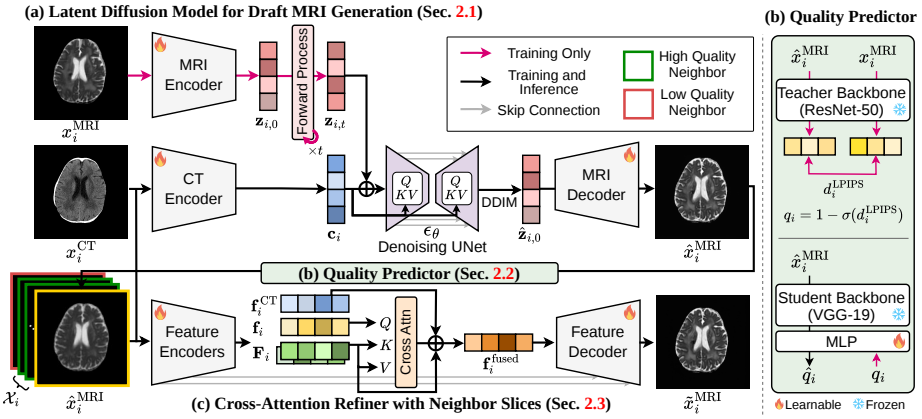

Fig. 1. Overview of the proposed LGN for CT-to-MRI synthesis. (a) CT-conditioned LDM for draft MRI generation. (b) A teacher–student Quality Predictor estimates quality. (c) A Cross-Attention Refiner leverages good neighbors to refine the draft.

demonstrate consistent improvements over strong baselines on a clinical dataset of 398 patients.

2 Proposed Methods

We propose a framework for generating Diffusion-Weighted Imaging (DWI) from Computed Tomography (CT). DWI is a Magnetic Resonance Imaging (MRI) sequence; for brevity, we refer to DWI as MRI throughout this paper as mentioned in Sec. 1. Figure 1 illustrates the overall architecture, which consists of three components: (a) Latent diffusion model (LDM), (b) Quality predictor, and (c) Cross-attention refiner. The LDM first generates draft MRI from CT image, then the quality predictor estimates the quality of each generated draft slice. Finally, the cross-attention refiner selectively refines low-quality slices by leveraging context from neighboring slices.

2.1 Latent Diffusion Model for Draft MRI Generation

Latent Diffusion Model (LDM) [16] performs diffusion in a low-dimensional latent space rather than the high-dimensional image space, achieving both computational efficiency and high generation quality. For cross-modality tasks like CT-to-MRI, LDM effectively bridges large domain gaps while preserving fine anatomical structures [5, 10]. Therefore, we adopt LDM to generate draft MRI from CT image. Given the i-th $H \times W$ MRI slice $x_i^{\mathrm{MRI}} \in \mathbb{R}^{H \times W}$ from a patient's 3D volume, we encode it into a d-dimensional latent representation $\mathbf{z}_{i,0} = E_{\mathrm{MRI}}(x_i^{\mathrm{MRI}}) \in \mathbb{R}^d$ using a pretrained MRI encoder E_{MRI}.

Diffusion models learn data distributions by modeling a noise-adding forward process and a noise-removing reverse process. The forward process gradually

adds Gaussian noise to the clean latent $\mathbf{z}_{i,0}$ over T steps, transforming it into pure noise $\mathbf{z}_{i,T}$ as follows:

$$q(\mathbf{z}_{i,t}|\mathbf{z}_{i,t-1}) = \mathcal{N}(\mathbf{z}_{i,t}; \sqrt{1-\beta_t}\mathbf{z}_{i,t-1}, \beta_t\mathbf{I}), \tag{1}$$

where $t \in \{1, 2, \cdots, T\}$ denotes the diffusion step, β_t controls the noise level at each step, and $\mathbf{I}$ is the identity matrix indicating independent Gaussian noise across dimensions.

To generate MRI corresponding to input CT image, we perform conditional generation using CT image as the condition. The CT image x_i^{CT} is encoded into a conditioning vector $\mathbf{c}_i = E_{\mathrm{CT}}(x_i^{\mathrm{CT}}) \in \mathbb{R}^d$ via a separate CT encoder E_{CT}. The reverse process learns to denoise the noisy latent $\mathbf{z}_{i,t}$ conditioned on $\mathbf{c}_i$, recovering the original latent $\mathbf{z}_{i,0}$. Then, a U-Net-based denoising network ϵ_θ is trained to predict the added noise at each step by $\hat{\epsilon}_{i,t} = \epsilon_\theta(\mathbf{z}_{i,t}, t, \mathbf{c}_i)$, where $\mathbf{z}_{i,t}$ is the noisy latent at step t, and $\mathbf{c}_i$ is the CT conditioning vector. The predicted noise $\hat{\epsilon}$ is used to iteratively recover the clean latent $\mathbf{z}_{i,0}$ from the noisy $\mathbf{z}_{i,T}$.

During inference, we use DDIM [19], which replaces stochastic diffusion sampling with a deterministic process, enabling reproducible results from the initial noise $\mathbf{z}_{i,T}$ and condition $\mathbf{c}_i$, while significantly accelerating inference with fewer sampling steps. At each timestep t, DDIM first predicts the clean latent $\hat{\mathbf{z}}_{i,0}$:

$$\hat{\mathbf{z}}_{i,0} = \big(\mathbf{z}_{i,t} - \hat{\epsilon}_{i,t}\sqrt{1-\bar{\alpha}_t}\big)/\sqrt{\bar{\alpha}_t}, \tag{2}$$

where $\bar{\alpha}_t = \prod_{n=1}^{t}(1-\beta_n)$ is the cumulative noise schedule. Then, the next latent $\mathbf{z}_{i,t-1}$ is computed as: $\mathbf{z}_{i,t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\mathbf{z}}_{i,0} + \hat{\epsilon}_{i,t}\sqrt{1-\bar{\alpha}_{t-1}}$. This process repeats from $t = T$ to $t = 1$, transforming pure noise $\mathbf{z}_{i,T} \sim \mathcal{N}(0, \mathbf{I})$ into the final clean latent $\hat{\mathbf{z}}_{i,0}$. Finally, a pretrained decoder D_{MRI} generates the draft MRI $\hat{x}_i^{\mathrm{MRI}} = D_{\mathrm{MRI}}(\hat{\mathbf{z}}_{i,0})$, which serves as input for subsequent quality evaluation and selective refinement.

2.2 Quality Predictor

Draft MRI $\hat{x}^{\mathrm{MRI}}$ generated by the LDM exhibits inconsistent quality, with some slices containing artifacts or distortions. To address this issue, we introduce a quality estimation module that predicts quality of the draft MRI. To this end, we employ a teacher–student framework, where the teacher provides pseudo-labels, and the student learns to predict quality. We employ ResNet-50 [22] as the backbone of the teacher model, whereas a more lightweight VGG-19 [23] is used for the student model, as illustrated in Fig. 1(b).

The teacher model ϕ_{teacher} is a feature extractor pretrained on a large-scale 2D image dataset (ImageNet [21]). Given an image pair $(\hat{x}_i^{\mathrm{MRI}}, x_i^{\mathrm{MRI}})$ as input, the LPIPS [20] distance between the extracted features by the teacher model is computed by $d_i^{\mathrm{LPIPS}} = \mathrm{LPIPS}(\phi_{\mathrm{teacher}}(\hat{x}_i^{\mathrm{MRI}}, x_i^{\mathrm{MRI}}))$. To obtain a pseudo-label q_i normalized to [0,1], the perceptual distance is transformed as $q_i = 1 - \sigma(d_i^{\mathrm{LPIPS}})$, where $\sigma(\cdot)$ denotes the sigmoid function. Then, the student model ϕ_{student} is trained on the pseudo-labels q_i and predicts quality scores $\hat{q}_i$ for draft MRI.

The teacher model is used solely for training the student model, and during inference with the proposed LGN, only the student model is employed as the quality predictor.

2.3 Cross-Attention Refiner with Neighbor Slices

Finally, we design a cross-attention-based refiner to refine low-quality draft MRI slices using high-quality neighbors. For each draft MRI $\hat{x}_i^{\text{MRI}}$, we select neighboring slices with higher quality based on q_i than the target slice to form a good neighbors set defined by $\mathcal{X}_i = \{\hat{x}_j^{\text{MRI}} \mid i-n \le j \le i+n,\ j \ne i,\ \hat{q}_j > \hat{q}_i\}$, where n controls the number of neighboring slices considered around the i-th slice. This excludes low-quality neighbors and retains only those that can provide useful information for refinement.

The refiner comprises feature encoders, a cross-attention fusion module, and a decoder. The target slice i, its corresponding CT image, and neighboring slices are encoded into feature representations using separate encoders E_{cen}, E_{CTfeat}, E_{neigh}, respectively, as follows:

$$\mathbf{f}_i = E_{\text{cen}}(\hat{x}_i^{\text{MRI}}),\ \ \mathbf{f}_i^{\text{CT}} = E_{\text{CTfeat}}(x_i^{\text{CT}}),\ \ \mathbf{F}_i = \{\mathbf{f}_j \mid \mathbf{f}_j = E_{\text{neigh}}(\hat{x}_j^{\text{MRI}}),\ j \in \mathcal{X}_i\}. \quad (3)$$

High-quality neighbors contain useful anatomical information for refining the target slice i. To this end, we apply cross-attention with the center as query and neighbor information as key and value as follows:

$$\mathbf{f}_i^{\text{cross}} = \text{CrossAttn}(Q = \mathbf{f}_i,\ K = V = \mathbf{F}_i). \quad (4)$$

The feature $\mathbf{f}_i^{\text{cross}}$ is then concatenated with the center feature $\mathbf{f}_i$ and CT condition $\mathbf{f}_i^{\text{CT}}$ as $\mathbf{f}_i^{\text{fused}} = \text{Concat}(\mathbf{f}_i, \mathbf{f}_i^{\text{cross}}, \mathbf{f}_i^{\text{CT}})$. Through this refinement, low-quality slices are improved by leveraging information from high-quality neighbors and CT image. The fused feature $\mathbf{f}_i^{\text{fused}}$ is then passed through a decoder to produce final MRI $\tilde{x}_i^{\text{MRI}}$ with reduced artifacts and enhanced structural consistency.

2.4 Loss Functions

The proposed LGN consists of three modules, (a), (b), and (c) (see Fig. 1), which are trained independently and then merged in a unified framework. The loss functions used to train each module are defined as follows.

LDM loss. The denoising network ϵ_θ is trained to predict the added noise ϵ at each timestep t. During training, Gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ is added to the latent $\mathbf{z}_{i,0}$ to produce $\mathbf{z}_{i,t}$, which is fed into the network along with the CT condition $\mathbf{c}_i$. The loss is the mean squared error between predicted $\hat{\epsilon}_{i,t}$ and ground-truth noise ϵ as follows:

$$\mathcal{L}_{\text{LDM}} = \mathbb{E}_{\mathbf{z}_{i,0},t,\mathbf{c}_i,\epsilon\sim\mathcal{N}(0,\mathbf{I})}\left[\|\epsilon - \hat{\epsilon}_{i,t}\|_2^2\right]. \quad (5)$$

Quality Predictor loss. The student model ϕ_{student} is trained to regress pseudo-labels q_i generated by the teacher model ϕ_{teacher}. Given a draft MRI $\hat{x}_i^{\text{MRI}}$, the

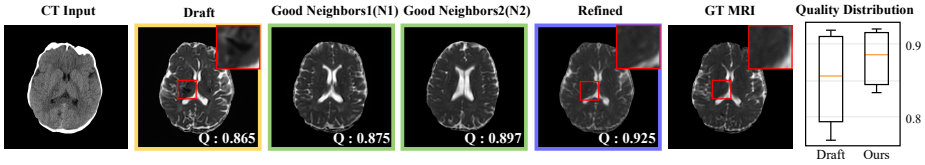

Fig. 2. Examples of CT and MRI images in the refinement process. Red boxes highlight artifact regions. Q denotes the quality score predicted by the Quality Predictor.

student model predicts a quality score and minimizes the mean squared error:

$$\mathcal{L}_{\text{quality}} = \mathbb{E}\left[\left(\phi_{\text{student}}(\hat{x}_i^{\text{MRI}}) - q_i\right)^2\right]. \tag{6}$$

Cross-Attention Refiner loss. We train the refiner using two losses. L_1 loss enforces pixel-wise accuracy between the refined and the ground-truth MRI, while LPIPS loss encourages perceptual similarity in feature space. The final loss is a weighted combination as follows:

$$\mathcal{L}_{\text{refiner}} = \frac{1}{2}\mathbb{E}\left[\|\tilde{x}_i^{\text{MRI}} - x_i^{\text{MRI}}\|_1\right] + \frac{1}{2}\mathbb{E}\left[\text{LPIPS}(\tilde{x}_i^{\text{MRI}}, x_i^{\text{MRI}})\right]. \tag{7}$$

3 Experiments

3.1 Datasets and Settings

Datasets. This study uses paired CT and DWI (b = 0 and 1000 s/mm^2) scans from 398 patients (male 56%, mean age 70, range 26–96) collected at Chonnam National University Hospital under IRB approval. CT was acquired on a GE Discovery CT750 HD and MRI on a Philips Ingenia CX 3T. Each CT–MRI pair was affine-registered and skull-stripped using HD-BET [11] to focus analysis on intracranial tissue. The dataset was split into 375 patients (5,878 slices) for training and 23 patients (513 slices) for evaluation.

Settings. All experiments were conducted on an NVIDIA L40S GPU with the AdamW optimizer. LDM: 300 epochs, batch size 16, learning rate 1×10^{-4}. Quality Predictor: 50 epochs, batch size 32, learning rate 5×10^{-4}. Cross-Attention Refiner: 500 epochs, batch size 8, learning rate 1×10^{-4}.

3.2 Experimental Results

Quality Predictor. To validate the Quality Predictor, we measured the Pearson correlation between the predicted scores $\hat{q}_i$ and the LPIPS-based pseudo-ground truth q_i, achieving 0.986. This high correlation indicates that the student model reliably approximates the teacher's quality estimates, enabling accurate quality prediction of the draft images $\hat{x}_i^{\text{MRI}}$.

Effects of Refiner with Neighboring Slices. As shown in Fig. 2, we first use the Quality Predictor to identify neighboring slices with higher predicted

Table 1. Quantitative ablation study of LGN variants for MRI synthesis (DWI b=0 and 1000). Best results are in **bold**, and second-best results are underlined.

Configuration	b=0				b=1000			
	LPIPS↓	SSIM↑	PSNR↑	MAE↓	LPIPS↓	SSIM↑	PSNR↑	MAE↓
LDM Only	0.087	0.863	24.82	0.029	0.081	0.879	32.12	0.010
Refiner Only	0.082	0.865	25.08	0.024	0.078	0.888	32.86	0.010
LDM + Refiner (w/o CT)	0.075	0.855	24.65	0.025	0.078	0.884	32.40	0.010
LDM + Refiner (w/o neigh-)	0.066	0.864	25.55	0.021	0.070	0.891	34.10	**0.009**
Full (Ours)	**0.061**	**0.871**	**25.74**	**0.020**	**0.068**	**0.892**	**34.55**	**0.009**

Table 2. Effect of the number of good neighbor slices on refinement performance.

# Neighbors	b=0				b=1000			
	LPIPS↓	SSIM↑	PSNR↑	MAE↓	LPIPS↓	SSIM↑	PSNR↑	MAE↓
0 (w/o neighbors)	0.066	0.864	25.55	0.021	0.070	0.891	34.10	**0.009**
1	0.063	0.869	25.68	**0.020**	0.069	**0.892**	34.45	**0.009**
3	**0.061**	**0.871**	**25.74**	**0.020**	**0.068**	**0.892**	**34.55**	**0.009**
5	0.064	0.867	25.62	0.021	0.071	0.891	34.48	**0.009**
All (w/o selection)	0.065	0.865	25.60	0.021	0.072	0.891	34.20	0.010

quality than the draft image and use them as references for cross-attention refinement. This refiner suppresses artifacts present in the draft MRI and better recovers fine anatomical details, producing an output that is visually closer to the ground-truth MRI. The predicted quality score also improves markedly, increasing from $Q = 0.865$ (draft) to $Q = 0.925$ (refined), outperforming even the selected high-quality neighbors N1 and N2 ($Q = 0.875$ and $Q = 0.897$). Furthermore, the quality distribution also shows reduced variance compared to draft, indicating more consistent synthesis as shown in the last column of Fig. 2.

Table 1 shows that adding the Refiner improves all metrics over LDM only even without neighbors (using only the target slice and CT), and incorporating neighboring slices in Full (Ours) further reduces LPIPS from 0.066 to 0.061 for b=0 and from 0.070 to 0.068 for b=1000, demonstrating the benefit of anatomical context from high-quality neighbors. In contrast, LDM+Refiner (w/o CT) underperforms Full, indicating that CT conditioning and neighboring-slice context are complementary. Table 2 further shows that LGN performs best with three neighboring slices. Using too few neighbors (0 or 1) provides limited contextual benefit, whereas using too many neighbors (5 or all) can dilute useful context and degrade refinement performance.

Performance Comparison. Table 3 provides quantitative comparison against existing state-of-the-art methods. LGN achieves the best LPIPS and MAE (lower is better) and the highest PSNR for both b=0 and b=1000. In terms of LPIPS, LGN achieves 0.061 for b=0 and 0.068 for b=1000, corresponding to 30% and 16% improvements over the previous best method, Latent Diffusion Model (LDM) [16] (also used as our baseline), respectively. LGN also achieves the best PSNR and MAE among all compared methods. However, CycleGAN [13] and Pix2Pix [14] achieve higher SSIM than LGN. As noted in prior studies [17, 18], SSIM

Table 3. Comparison with state-of-the-art methods on DWI (b=0 and 1000) synthesis. Best results in **bold**, second best underlined.

Method	b=0				b=1000			
	LPIPS↓	SSIM↑	PSNR↑	MAE↓	LPIPS↓	SSIM↑	PSNR↑	MAE↓
GAN (2014) [12]	0.106	0.802	17.21	0.056	0.145	0.823	29.09	0.016
CycleGAN (2017) [13]	0.096	0.842	20.51	0.040	0.103	0.864	30.06	0.014
Pix2Pix (2017) [14]	0.088	**0.882**	24.11	0.027	0.095	0.889	32.05	0.011
CUT (2020) [15]	0.089	0.874	23.08	0.029	0.090	**0.898**	32.67	0.011
LDM (2022–2025) [5, 10, 16]	0.087	0.863	24.82	0.029	0.081	0.879	32.12	0.010
LGN(Ours)	**0.061**	0.871	**25.74**	**0.020**	**0.068**	0.892	**34.55**	**0.009**

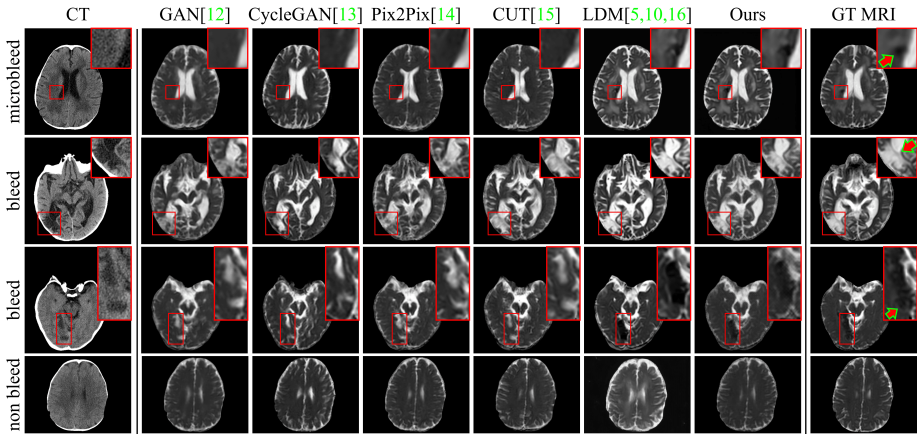

Fig. 3. Qualitative comparison with other methods for CT-to-MRI synthesis. Red boxes and arrows highlight hemorrhagic regions (e.g., microbleeds and bleeds).

has inherent limitations: it is based on local pixel statistics and overall intensity distributions, which can assign high scores to images that preserve coarse structures even when fine details are blurred.

Figure 3 shows qualitative results for each method. Note that our goal is anatomically faithful CT-to-MRI synthesis, not lesion detection, and we do not quantitatively evaluate lesion detectability. Nevertheless, preserving clinically meaningful fine details (e.g., hemorrhage) is crucial for practical use. Previous methods generally capture the overall structure but often miss subtle details. In the first row (microbleed), other methods fail to reproduce microhemorrhages, whereas LGN successfully synthesizes them (red arrows). In the second and third rows (bleed), LGN more accurately reconstructs both acute and chronic hemorrhages. Overall, these results suggest that LGN better preserves anatomically faithful MRI appearance, including clinically important lesion details, despite the substantial modality gap between CT and MRI.

4 Conclusions and Future Works

We proposed LGN, a framework for CT-to-MRI synthesis from non-contrast CT, consisting of three components: (1) a Latent Diffusion Model for draft MRI generation, (2) a slice-level Quality Predictor trained via LPIPS-based teacher–student learning, and (3) a 3D context-aware Cross-Attention Refiner that selectively refines suboptimal slices using high-quality neighboring slices. Experiments on 398 patients demonstrate consistent improvements in LPIPS, PSNR, and MAE over strong baselines. These findings highlight that quality-guided cross-slice refinement is an effective strategy for improving anatomical consistency in CT-to-MRI synthesis. Future work includes extending the framework to other MRI sequences and exploring the use of synthesized MRI for downstream clinical tasks such as stroke and hemorrhage analysis.

Acknowledgments. This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. IITP-2026-RS-2024-00437718); the Artificial Intelligence Convergence Innovation Human Resources Development grant funded by the Korea government (MSIT) (No. IITP-2023-RS-2023-00256629); and the Regional Innovation System & Education (RISE) Glocal University 30 Program through the Gwangju RISE Center, funded by the Ministry of Education (MOE) and the Gwangju Metropolitan City, Republic of Korea (No. 2026-RISE(glocal university 30)-05-011).

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

References

1. DAYARATHNA, Sanuwani, et al. Deep learning based synthesis of MRI, CT and PET: Review and analysis. Medical image analysis, 2024, 92: 103046.
2. CHARTSIAS, Agisilaos, et al. Multimodal MR synthesis via modality-invariant latent representation. IEEE transactions on medical imaging, 2017, 37.3: 803-814.
3. KAWAHARA, Daisuke; NAGATA, Yasushi. T1-weighted and T2-weighted MRI image synthesis with convolutional generative adversarial networks. reports of practical Oncology and radiotherapy, 2021, 26.1: 35-42.
4. HO, Jonathan; JAIN, Ajay; ABBEEL, Pieter. Denoising diffusion probabilistic models. Advances in neural information processing systems, 2020, 33: 6840-6851.
5. ÖZBEY, Muzaffer, et al. Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging, 2023, 42.12: 3524-3539.
6. FLORKOW, Mateusz C., et al. Deep learning–based MR-to-CT synthesis: the influence of varying gradient echo–based MR images as input channels. Magnetic resonance in medicine, 2020, 83.4: 1429-1441.
7. LEI, Yang, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Medical physics, 2019, 46.8: 3565-3581.

8. PENG, Yinglin, et al. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. Radiotherapy and Oncology, 2020, 150: 217-224.
9. DAR, Salman UH, et al. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE transactions on medical imaging, 2019, 38.10: 2375-2388.
10. LEE, Junhyeok, et al. Lesion-Aware Post-training of Latent Diffusion Models for Synthesizing Diffusion MRI from CT Perfusion. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2025. p. 282-291.
11. Isensee F. et al. Automated brain extraction of multi-sequence MRI using artificial neural networks. Hum Brain Mapp. 2019; 1–13. https://doi.org/10.1002/hbm.24750
12. GOODFELLOW, Ian J., et al. Generative adversarial nets. Advances in neural information processing systems, 2014, 27.
13. ZHU, Jun-Yan, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2223-2232.
14. ISOLA, Phillip, et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1125-1134.
15. PARK, Taesung, et al. Contrastive learning for unpaired image-to-image translation. In: European conference on computer vision. Cham: Springer International Publishing, 2020. p. 319-345.
16. ROMBACH, Robin, et al. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 10684-10695.
17. Renieblas, Gabriel Prieto, et al. "Structural similarity index family for image quality assessment in radiological images." Journal of medical imaging 4.3 (2017): 035501-035501.
18. Mudeng, Vicky, Minseok Kim, and Se-woon Choe. "Prospects of structural similarity index for medical image analysis." Applied Sciences 12.8 (2022): 3754.
19. Song, Jiaming, Chenlin Meng, and Stefano Ermon. "Denoising diffusion implicit models." arXiv preprint arXiv:2010.02502 (2020).
20. Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
21. Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.
22. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
23. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).