Lesion-Aware Post-Training of Latent Diffusion Models for Synthesizing Diffusion MRI from CT Perfusion

Junhyeok Lee[1⋆], Hyunwoong Kim[2⋆], Hyungjin Chung[3], Heeseong Eom[1], Joon Jang[4], Chul-Ho Sohn[2], and Kyu Sung Choi[2†]

[1] College of Medicine, Seoul National University, Seoul, Republic of Korea
[2] Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea
[3] Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[4] Department of Biomedical Sciences, Seoul National University, Seoul, Republic of Korea
{ent1127@snu.ac.kr}

Abstract. Image-to-Image translation models can help mitigate various challenges inherent to medical image acquisition. Latent diffusion models (LDMs) leverage efficient learning in compressed latent space and constitute the core of state-of-the-art generative image models. However, this efficiency comes with a trade-off, potentially compromising crucial pixel-level detail essential for high-fidelity medical images. This limitation becomes particularly critical when generating clinically significant structures, such as lesions, which often occupy only a small portion of the image. Failure to accurately reconstruct these regions can severely impact diagnostic reliability and clinical decision-making. To overcome this limitation, we propose a novel post-training framework for LDMs in medical image-to-image translation by incorporating lesion-aware medical pixel space objectives. This approach is essential, as it not only enhances overall image quality but also improves the precision of lesion delineation. We evaluate our framework on brain CT-to-MRI translation in acute ischemic stroke patients, where early and accurate diagnosis is critical for optimal treatment selection and improved patient outcomes. While diffusion MRI is the gold standard for stroke diagnosis, its clinical utility is often constrained by high costs and low accessibility. Using a dataset of 817 patients, we demonstrate that our framework improves overall image quality and enhances lesion delineation when synthesizing DWI and ADC images from CT perfusion scans, outperforming existing image-to-image translation models. Furthermore, our post-training strategy is easily adaptable to pre-trained LDMs and exhibits substantial potential for broader applications across diverse medical image translation tasks.

Keywords: Latent Diffusion · CT-to-MRI Translation · Ischemic Stroke

[⋆] These authors contributed equally to this paper.

1 Introduction

Medical imaging plays a crucial role in modern medicine, providing spatially resolved information of organs and tissues. Various imaging modalities offer unique clinical insights with distinct advantages and limitations based on their underlying physical principles. In the context of acute stroke management where "time is brain", rapid imaging for diagnosis is crucial as timely intervention directly impacts patient outcomes [22,19]. Computed tomography (CT) is frequently utilized due to its widespread availability, short acquisition times, and low cost. Although CT can detect early signs of acute ischemic stroke, these indicators are often subtle or absent within the initial hours following stroke onset, leading to suboptimal sensitivity and inter-rater agreement [3]. In contrast, diffusion-weighted imaging (DWI) on magnetic resonance imaging (MRI) offers superior sensitivity for detecting acute ischemic stroke and distinguishing stroke mimics [8,4]. However, MRI has several limitations compared to CT, including higher costs, restricted accessibility, longer scan durations, and challenges related to patient intolerance or contraindications.

With the recent explosion of image generative models, there has been wide interest in medical image-to-image translation model development to help overcome challenges in medical image acquisition [1,12]. In recent years, diffusion models have surpassed generative adversarial networks (GANs) as state-of-the-art image generation models [6]. Latent diffusion models (LDMs) have emerged as a particularly efficient approach, operating within a compressed latent space to improve both computational efficiency and generative performance [20]. Moreover, LDMs demonstrated promising results in various medical imaging applications, including image synthesis [18], super-resolution [16], and image-to-image translation [12]. However, LDMs learn the diffusion process only in the latent space and often freeze the decoder, potentially overlooking high-frequency image details. Existing methods have limitations in addressing this challenge effectively. Only recently, few studies have explored post-training techniques for LDMs incorporating image space objectives for sharper and more realistic image generation [2,25]. In medical imaging, this challenge intensifies when generating clinically significant structures, such as lesions, which often show low spatial occupancy. Deficiencies in this reconstruction can substantially degrade diagnostic reliability and subsequent clinical decision-making.

In this study, we propose a novel post-training framework for LDMs in medical image-to-image translation with lesion-aware medical image space objectives. Our method incorporates medical image space loss to generate more realistic medical images. In addition to the pixel loss for overall image quality, we introduce a task-specific objective for ischemic lesion areas to enhance the accuracy of lesion delineation. Evaluation on a diffusion MRI-CT perfusion paired dataset from 817 acute ischemic stroke patients demonstrate that our LDM post-training framework outperforms state-of-the-art models in both qualitative and quantitative evaluations. Moreover, we apply our framework to other diffusion models that utilizes the latent space, demonstrating its adaptability and potential for broad utility in diverse medical image generative tasks.

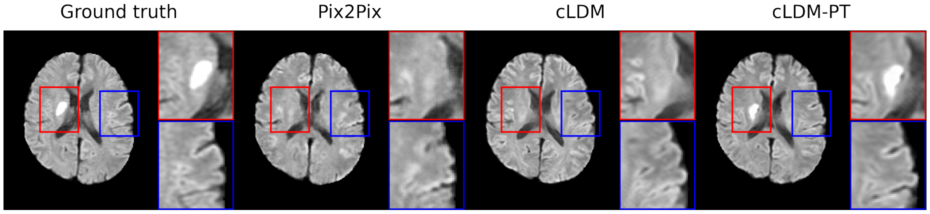

Fig. 1. Previous models fail to accurately depict ischemic stroke lesion in diffusion MRI synthesized from CT perfusion. Our model (right) shows higher lesion conspicuity (red) and enhanced image fidelity, highlighted with grey-white matter differentiation (blue).

2 Method

2.1 Base Conditional Latent Diffusion Model

Given an axial slice of target MRI images $\boldsymbol{x} \in \mathbb{R}^{H\times W\times m}$ with m modalities concatenated into channels, our base model is a conditional LDM with the VQGAN [7] framework where the encoder $\mathcal{E}$ encodes the images into latent representations $\boldsymbol{z} = \mathcal{E}(\boldsymbol{x})$. The decoder $\mathcal{D}$ learns to reconstruct latent representations back into MRI images as $\boldsymbol{x} = \mathcal{D}(\boldsymbol{z})$. The corresponding axial slice of the CTP image $\boldsymbol{c} \in \mathbb{R}^{H\times W\times n}$ with n time points is used as the input condition for the CTP-to-MRI translation diffusion process. To align the CTP image $\boldsymbol{c}$ with the latent representation $\boldsymbol{z}$ of MRI images, we follow the standard procedure for semantic synthesis with LDMs [20]. A bilinear interpolator combined with 1×1 convolutions is used as a spatial rescaler $\mathcal{F}$ to downsample $\boldsymbol{c}$ into $\tilde{\boldsymbol{c}} = \mathcal{F}(\boldsymbol{c})$, then $\tilde{\boldsymbol{c}}$ is fed into the diffusion process by channel-wise concatenation.

The conditional LDM learns the latent distribution of the MRI images $p_\theta(\boldsymbol{z})$ by learning denoising conditional autoencoders $\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}})$ from the sequence of noisy images $\{\boldsymbol{z}_0, ..., \boldsymbol{z}_T\}$. The forward process that diffuses the latent input $\boldsymbol{z}_0 = \boldsymbol{z}$ with pre-defined Gaussian noise schedules $\{\beta_1, ..., \beta_T\}$ is a Markov process formulated as:

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

which allows sampling during training via:

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

where $\alpha_t = 1-\beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t}\alpha_i$. The reverse process that denoises noisy images is formulated with a time-conditional UNet [21] with parameters θ as:

$$p(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t, \tilde{\boldsymbol{c}}) = \mathcal{N}(\boldsymbol{z}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}}), \boldsymbol{\Sigma}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}})). \tag{3}$$

After parameterizing $\boldsymbol{\mu}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}}) = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{z}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}}))$ and simplifying $\boldsymbol{\Sigma}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}}) = \sigma_t^2\mathbf{I}$, the latent objective for the conditional LDM trained to predict $\boldsymbol{\epsilon}$ is given as:

$$\mathcal{L}_{latent} = \mathbb{E}_{\mathcal{E}(\boldsymbol{x}),\boldsymbol{\epsilon},t}[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}})\|_2^2]. \tag{4}$$

Detailed illustration of the model is shown in Figure 1.

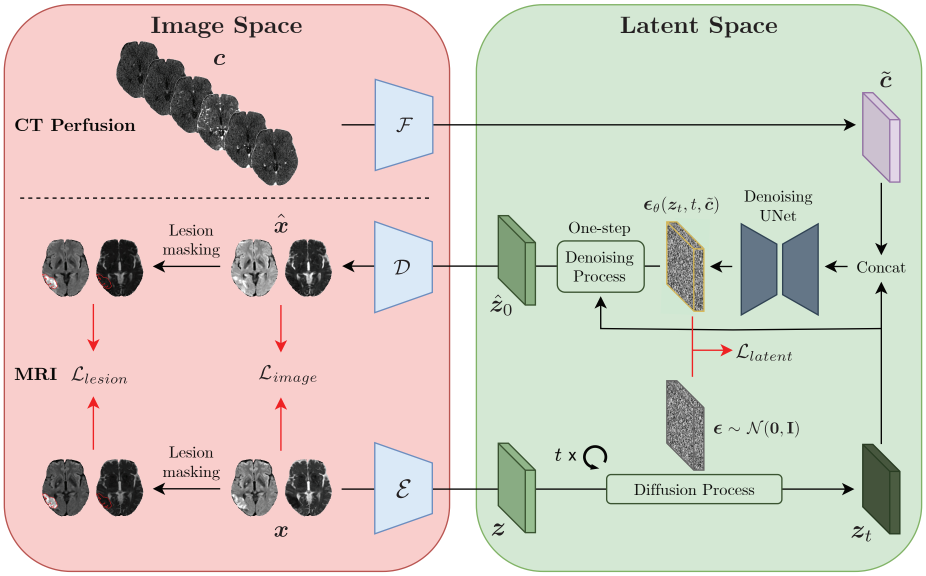

Fig. 2. Overview of our post-training framework. During post-training, in addition to the latent objective for the conditional LDM $\mathcal{L}_{latent}$, we introduce medical image space objectives $\mathcal{L}_{image}$, $\mathcal{L}_{lesion}$ to enhance overall image quality and lesion conspicuity.

2.2 Lesion-Aware Image Space Objectives for LDM Post-Training

We introduce medical image space objectives for LDM post-training to improve overall image quality and allow ischemic lesion-aware diffusion MRI generation. Recent studies show post-training LDMs with image space objectives can lead to sharper and more realistic natural images [2,25]. For image generation in domains where high image detail is essential such as medical imaging [1], remote sensing [23], or face generation [13], post-training LDMs with task-specific image space objectives can be particularly beneficial.

Given a MRI image $\boldsymbol{x}$ and its latent representation $\boldsymbol{z} = \mathcal{E}(\boldsymbol{x})$, the noisy version of $\boldsymbol{z}$ is sampled as $\boldsymbol{z}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{z} + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$ during training. We then project the estimated noise free latent $\hat{\boldsymbol{z}}_0$ given as:

$$\hat{\boldsymbol{z}}_0 = \frac{\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, \tilde{\boldsymbol{c}})}{\sqrt{\bar{\alpha}_t}}, \tag{5}$$

back to the image space using the decoder $\mathcal{D}$ to get $\hat{\boldsymbol{x}} = \mathcal{D}(\hat{\boldsymbol{z}}_0)$. This formulation allows fast and efficient one-step inference of $\hat{\boldsymbol{z}}_0$ and thus $\hat{\boldsymbol{x}}$ during training instead of the iterative denoising steps of the standard inference process [2,10]. The image space objective is then defined as:

$$\mathcal{L}_{image} = \mathbb{E}_{\mathcal{E}(\boldsymbol{x}),\boldsymbol{\epsilon},t}[\|\boldsymbol{x} - \mathcal{D}(\hat{\boldsymbol{z}}_0)\|_2^2]. \tag{6}$$

One of the main challenges of CTP-to-MRI translation is training the model to accurately generate ischemic lesions that constitute only a small portion of the voxels in the dataset. We believe standard objectives covering the entire image or latent representations are insufficient to train the model to precisely generate lesions because the gradients to guide the model to generate lesions is diluted with signals from other brain parenchyma or even less important background voxels. To boost accurate ischemic lesion generation, we designed a new image space objective focusing only on ischemic lesion regions. The lesion-aware objective is formulated as:

$$\mathcal{L}_{lesion} = \mathbb{E}_{\mathcal{E}(\boldsymbol{x}),\boldsymbol{\epsilon},t}[\|M_{lesion}(\boldsymbol{x}) \odot (\boldsymbol{x} - \mathcal{D}(\hat{\boldsymbol{z}}_0))\|_2^2], \tag{7}$$

where $M_{lesion}(\boldsymbol{x}) \in \mathbb{R}^{H \times W}$ is a binary mask for the ischemic lesion. The final objective function combining the latent and image space losses with hyperparameters $\lambda_{image}, \lambda_{lesion}$ is given as:

$$\mathcal{L}_{total} = \mathcal{L}_{latent} + \lambda_{image} \cdot \mathcal{L}_{image} + \lambda_{lesion} \cdot \mathcal{L}_{lesion}. \tag{8}$$

3 Experiments

3.1 Experimental Setup

Dataset All images used in this study were collected from the Seoul National University Hospital (SNUH) with approval from its Institutional Review Board. MRI scans including diffusion-weighted imaging (DWI) were acquired using 3.0T scanners with voxel size of 0.9375×0.9375 mm to 1×1 mm (in-plane) and 4–5 mm (axial slice thickness). The images were skull-stripped then resampled to a uniform voxel size of 1×1×5 mm^3.

Apparent diffusion coefficient (ADC) maps were derived from the DWI image (b=1000), and was used for ischemic lesion segmentation by medical experts to create ground truth lesion masks. CTP scans were acquired using Aquilion 64 CT scanner (TOSHIBA) with voxel size of 0.47×0.47×1 mm^3 and spanning 15 time points. To align CTP images with MRI images, CTP images were skull-stripped then registered into the DWI space using ANTs [24].

The final dataset comprised of paired DWI, ADC, and CTP images with ischemic lesion masks from 817 patients. The dataset were randomly divided into training (571 patients; 14083 axial slices), validation (81 patients; 1948 axial slices), and test (165 patients; 4015 axial slices) sets. Across the training, validation, and test sets, slices containing lesions constituted 10.9% (1542/14083), 11.3% (220/1948), and 11.0% (441/4015), respectively, with mean (±SD) lesion volumes of 15.17±41.17 ml, 12.46±32.94 ml, and 14.66±36.24 ml.

Implementation Details Implementation of our model was based on the LDM [20] framework with VQGAN [7] that compresses the image space into the latent space by a factor of 4 in both coronal and sagittal directions. The model was trained with the AdamW [17] optimizer with base learning rate of 2×10^{-6}.

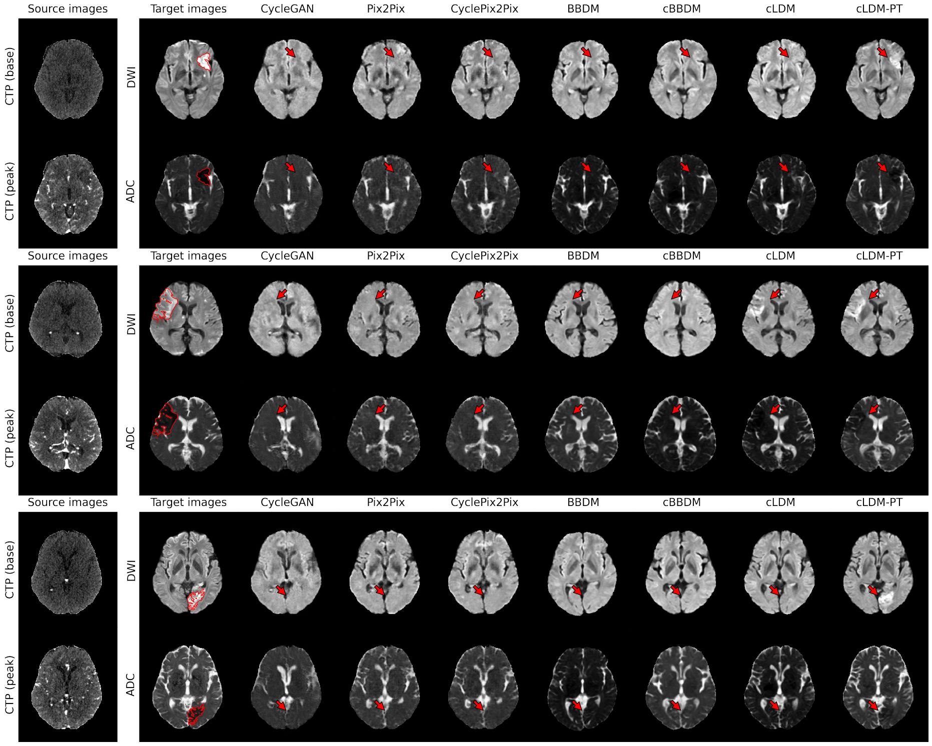

Fig. 3. Visualization of synthesized diffusion MRI images from CTP images in acute ischemic stroke patients. Our model with post-training (cLDM-PT) excels in lesion delineation (red arrows), accurately depicting ischemic stroke lesions with restricted diffusion (red contour) based on hypo-perfused regions in source CTP images. (Top) A case with infarct core in the left inferior frontal area. (Middle) A case of acute ischemic stroke by large vessel occlusion in the right middle cerebral artery. (Bottom) A case of acute infarction in the left occipital lobe.

The model was trained with $T = 1000$ diffusion steps, with 200 DDIM sampling steps used during inference. For LDM post-training, $\lambda_{image} = 0.01, 0.05, 0.1$ were tested and several λ_{lesion} values were selected accordingly. We selected $\lambda_{image} = 0.01, \lambda_{lesion} = 0.02$ as the hyper-parameters of our best model. All experiments were conducted using the NVIDIA A6000 GPU with a batch size of 48. Our code is publicly available at https://github.com/snuh-rad-aicon/Diffusion-LAPT.

Evaluation Metrics To evaluate diffusion MRI synthesis, we used both distortion and perception measures to compare ground truth $\boldsymbol{x}$ with synthesized image $\hat{\boldsymbol{x}}$. The mean absolute error (MAE) measures accuracy of image reconstruction, and to evaluate lesion delineation we additionally defined lesion MAE as: Lesion MAE $= (\sum M_{lesion}(\boldsymbol{x}) \odot |\boldsymbol{x} - \hat{\boldsymbol{x}}|)/\sum M_{lesion}(\boldsymbol{x})$. Peak signal-to-noise ratio (PSNR) assesses reconstruction quality. The multi-scale structural sim-

Table 1. Quantitative results of DWI and ADC synthesis from CTP.

	DWI					**ADC**				
Model	MAE↓	Lesion MAE↓	PSNR↑	MS-SSIM↑	FID↓	MAE↓	Lesion MAE↓	PSNR↑	MS-SSIM↑	FID↓
CycleGAN	0.143	0.235	20.04	0.803	40.62	0.321	0.151	12.20	0.602	65.95
Pix2Pix	0.083	0.215	26.25	0.859	46.90	0.073	0.127	26.52	0.872	70.64
PairedCycleGAN	0.123	0.223	21.86	0.865	47.96	0.081	0.124	25.55	**0.883**	70.34
BBDM	0.104	0.225	24.19	0.787	31.84	0.103	0.136	23.10	0.764	66.81
cBBDM	0.093	0.226	25.21	0.800	32.32	0.098	0.139	23.63	0.777	66.98
cLDM	0.073	0.222	**27.94**	0.855	30.82	0.072	0.125	27.08	0.866	66.71
cLDM-PT(ours)	**0.072**	**0.199**	27.78	**0.867**	**29.95**	**0.052**	**0.105**	**31.49**	0.876	**61.91**

ilarity index measure (MS-SSIM) measures the similarity between images at multiple resolutions. The perceptual quality of the synthesized MRI images was measured using the Fréchet Inception Distance (FID) [9].

Baselines We compared the performance of our model with state-of-the-art models, including models based on generative adversarial networks (Pix2Pix [11], CycleGAN [26], PairedCycleGAN [5]) and latent diffusion (conditional LDM [20], BBDM [15], conditional BBDM [14]).

3.2 Experimental Results

Quantitative Results We observe a clear improvement from conventional methods in our refined latent diffusion model, cLDM-PT, which was optimized via our proposed post-training framework (Table 1). For both DWI and ADC generation, our cLDM-PT model achieved the lowest MAE, highest MS-SSIM, and lowest FID, indicating marked improvements in accuracy, clarity, and structural fidelity. Furthermore, our model achieved the lowest lesion MAE of 0.199 for DWI and 0.105 for ADC images, which underlines its enhanced capability for precise lesion delineation. Overall, our model showed 14.5% deduction in image MAE and 12.4% deduction in lesion MAE after post-training.

Qualitative Evaluation Figure 3 visualizes synthesized DWI and ADC from CTP of acute ischemic stroke patients with lesions in various brain regions. Due to low signal-to-noise ratio of CTP, it is difficult to accurately estimate ischemic core volumes. Small infarcts such as lacunar infarcts are also poorly visualized in CTP. These factors make it challenging for generative models to accurately reconstruct ischemic lesions in synthesized MRI. While the diffusion model series generates more realistic images compared to GAN-based models, they encounter difficulty in lesion delineation. Our model, cLDM-PT, excels in lesion delineation and demonstrates exceptional ability to generate accurate and detailed images.

Table 2. Effect of image space loss weights on DWI and ADC synthesis from CTP.

Loss Weights		**DWI**					**ADC**				
λ_{image}	λ_{lesion}	MAE↓	Lesion MAE↓	PSNR↑	MS-SSIM↑	FID↓	MAE↓	Lesion MAE↓	PSNR↑	MS-SSIM↑	FID↓
0	0	0.073	0.222	**27.94**	0.855	30.82	0.072	0.125	27.08	0.866	66.71
0.01	0	0.074	0.202	27.48	**0.867**	31.08	0.054	0.109	30.89	0.876	61.84
0.01	0.01	0.078	0.198	26.70	**0.867**	30.50	0.054	0.106	30.76	0.877	64.00
0.01	**0.02**	**0.072**	0.199	27.78	**0.867**	**29.95**	**0.052**	0.105	**31.49**	0.876	61.91
0.01	0.05	0.077	0.196	27.15	0.865	30.74	0.058	0.093	29.33	**0.878**	**61.52**
0.05	0	0.077	0.212	27.05	0.863	30.66	0.058	0.116	29.73	0.874	64.62
0.05	0.1	0.077	**0.195**	27.00	0.853	30.06	0.060	0.093	29.70	0.865	62.90
0.05	0.2	**0.072**	0.197	27.87	0.859	32.42	0.086	**0.083**	24.37	0.876	63.28
0.1	0	0.077	0.210	27.15	0.861	31.27	0.061	0.111	29.27	0.873	64.71
0.1	0.1	0.073	0.197	27.33	0.862	31.09	0.071	0.091	26.88	0.876	63.44
0.1	0.2	0.073	0.197	27.42	0.855	33.70	0.083	0.090	24.75	0.868	62.42

Table 3. Evaluating the application of our post-training framework on cBBDM.

	DWI					**ADC**				
Model	MAE↓	Lesion MAE↓	PSNR↑	MS-SSIM↑	FID↓	MAE↓	Lesion MAE↓	PSNR↑	MS-SSIM↑	FID↓
cBBDM	0.093	0.226	25.21	0.800	32.32	0.098	0.139	23.63	0.777	66.98
cBBDM-PT	**0.078**	**0.221**	**27.40**	**0.857**	**31.55**	**0.074**	**0.120**	**26.62**	**0.866**	**66.67**

Impact of Loss Weights We test the effect of varying the weights of image space objectives during post-training, defined as hyper-parameters λ_{image} and λ_{lesion} (Table 2). Increasing λ_{image} improves overall image accuracy and structural consistency, but this effect diminishes as λ_{image} is further increased. As we increase λ_{lesion}, lesion MAE decreases sharply before reaching a plateau. Further increasing λ_{lesion} results in overall distortion of synthesized images. We observe that small weights for image space objectives is sufficient to achieve the best balance between global image quality and lesion conspicuity.

Generalization to Other Latent Diffusion Models To showcase the broad applicability of our framework, we apply it to cBBDM which learns stochastic Brownian bridge process in the latent space. Post-training cBBDM by our framework with $\lambda_{image} = 0.01$ and $\lambda_{lesion} = 0.02$ led to improvements across all metrics, indicating better image quality and lesion delineation (Table 3).

4 Conclusion

In this study, we present a novel post-training framework for LDMs in medical images to improve image quality and lesion delineation, generating more realistic and clinically accurate images. By integrating medical image space objectives, our method addresses the challenge of capturing high-frequency details in LDMs. Evaluation on a diffusion MRI-CTP paired dataset of acute ischemic stroke patients demonstrates that our framework surpasses state-of-the-art models in both overall image fidelity and ischemic lesion conspicuity. These results underscore the effectiveness of our framework in enhancing diagnostic reliability and its potential to support clinical decision-making. Moreover, its consistent performance across various LDMs suggests its broad applicability to diverse medical image translation tasks.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00251022) (K.S.C.); the Phase III (Postdoctoral fellowship) grant of the SPST (SNU-SNUH Physician Scientist Training) Program (K.S.C.); the SNUH Research Fund (No. 04-2024-0600) (K.S.C.); and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) grant funded by the Ministry of Health&Welfare (No. RS-2024-00439549) (K.S.C.).

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

References

1. Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B.: Medgan: Medical image translation using gans. Computerized Medical Imaging and Graphics **79**, 101684 (2020)
2. Berrada, T., Astolfi, P., Hall, M., Havasi, M., Benchetrit, Y., Romero-Soriano, A., Alahari, K., Drozdzal, M., Verbeek, J.: Boosting latent diffusion with perceptual objectives. In: The Thirteenth International Conference on Learning Representations (2025)
3. Bryan, R.N., Levy, L.M., Whitlow, W.D., Killian, J.M., Preziosi, T.J., Rosario, J.A.: Diagnosis of acute cerebral infarction: comparison of ct and mr imaging. American Journal of Neuroradiology **12**(4), 611–620 (1991)
4. Chalela, J.A., Kidwell, C.S., Nentwich, L.M., Luby, M., Butman, J.A., Demchuk, A.M., Hill, M.D., Patronas, N., Latour, L., Warach, S.: Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. The Lancet **369**(9558), 293–298 (2007)
5. Chang, H., Lu, J., Yu, F., Finkelstein, A.: Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 40–48 (2018)
6. Dhariwal, P., Nichol, A.Q.: Diffusion models beat GANs on image synthesis. In: Advances in Neural Information Processing Systems (2021)
7. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12873–12883 (2021)
8. Fiebach, J., Schellinger, P., Jansen, O., Meyer, M., Wilde, P., Bender, J., Schramm, P., Jüttler, E., Oehler, J., Hartmann, M., Hähnel, S., Knauth, M., Hacke, W., Sartor, K.: Ct and diffusion-weighted mr imaging in randomized order. Stroke **33**(9), 2206–2210 (2002)

9. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems. pp. 6626–6637 (2017)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020), https://arxiv.org/abs/2006.11239
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5967–5976 (2017)
12. Kim, J., Park, H.: Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. In: IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 7589–7598 (2024)
13. Kim, M., Liu, F., Jain, A., Liu, X.: Dcface: Synthetic face generation with dual condition diffusion model. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12715–12725 (2023)
14. Kim, S.H., Chung, D.W.: Conditional brownian bridge diffusion model for vhr sar to optical image translation. arXiv preprint arXiv:2408.07947 (2024)
15. Li, B., Xue, K., Liu, B., Lai, Y.K.: Bbdm: Image-to-image translation with brownian bridge diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1952–1961 (2023)
16. Li, G., Rao, C., Mo, J., Zhang, Z., Xing, W., Zhao, L.: Rethinking diffusion model for multi-contrast mri super-resolution. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 11365–11374 (2024)
17. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
18. Müller-Franzes, G., Niehues, J.M., Khader, F., Arasteh, S.T., Haarburger, C., Kuhl, C., Wang, T., Han, T., Nolte, T., Nebelung, S., Kather, J.N., Truhn, D.: A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. Scientific Reports **13**(1), 12098 (2023)
19. Puig, J., Shankar, J., Liebeskind, D., Terceño, M., Nael, K., Demchuk, A.M., Menon, B., Dowlatshahi, D., Leiva-Salinas, C., Wintermark, M., Thomalla, G., Silva, Y., Serena, J., Pedraza, S., Essig, M.: From "time is brain" to "imaging is brain": A paradigm shift in the management of acute ischemic stroke. Journal of Neuroimaging **30**(5), 562–571 (2020)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models . In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10674–10685 (2022)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing (2015)
22. Saver, J.L.: Time is brain—quantified. Stroke **37**(1), 263–266 (2006)
23. Sebaq, A., ElHelw, M.: Rsdiff: remote sensing image generation from text using diffusion model. Neural Computing and Applications **36**(36), 23103–23111 (2024)
24. Tustison, N.J., Cook, P.A., Holbrook, A.J., Johnson, H.J., Muschelli, J., Devenyi, G.A., Duda, J.T., Das, S.R., Cullen, N.C., Gillen, D.L., Yassa, M.A., Stone, J.R., Gee, J.C., Avants, B.B.: The ANTsX ecosystem for quantitative biological and medical imaging. Scientific Reports **11**(1) (2021)
25. Zhang, C., Motwani, S., Yu, M., Hou, J., Juefei-Xu, F., Tsai, S., Vajda, P., He, Z., Wang, J.: Pixel-space post-training of latent diffusion models. arXiv preprint arXiv:2409.17565 (2024)

26. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 IEEE International Conference on Computer Vision (ICCV) pp. 2242–2251 (2017)