

Slice-Consistent 3D Volumetric Brain CT-to-MRI Translation with 2D Brownian Bridge Diffusion Model

Kyobin Choo¹, Youngjun Jun², Mijin Yun³, and Seong Jae Hwang²

¹ Department of Computer Science, Yonsei University, Seoul, Republic of Korea

² Department of Artificial Intelligence, Yonsei University, Seoul, Republic of Korea

³ Department of Nuclear Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

{chu, youngjun, seongjae}@yonsei.ac.kr; yunmijin@yuhs.ac

Abstract. In neuroimaging, generally, brain CT is more cost-effective and accessible imaging option compared to MRI. Nevertheless, CT exhibits inferior soft-tissue contrast and higher noise levels, yielding less precise structural clarity. In response, leveraging more readily available CT to construct its counterpart MRI, namely, medical image-to-image translation (I2I), serves as a promising solution. Particularly, while diffusion models (DMs) have recently risen as a powerhouse, they also come with a few practical caveats for medical I2I. First, DMs' inherent stochasticity from random noise sampling cannot guarantee consistent MRI generation that faithfully reflects its CT. Second, for 3D volumetric images which are prevalent in medical imaging, naïvely using 2D DMs leads to slice inconsistency, e.g., abnormal structural and brightness changes. While 3D DMs do exist, significant training costs and data dependency bring hesitation. As a solution, we propose novel style key conditioning (SKC) and inter-slice trajectory alignment (ISTA) sampling for the 2D Brownian bridge diffusion model. Specifically, SKC ensures a consistent imaging style (e.g., contrast) across slices, and ISTA interconnects the independent sampling of each slice, deterministically achieving style and shape consistent 3D CT-to-MRI translation. To the best of our knowledge, this study is the first to achieve high-quality 3D medical I2I based only on a 2D DM with no extra architectural models. Our experimental results show superior 3D medical I2I than existing 2D and 3D baselines, using in-house CT-MRI dataset and BraTS2023 FLAIR-T1 MRI dataset.

Keywords: Image-to-image translation · Volumetric image synthesis · Brownian bridge diffusion model · Brain CT-to-MRI

1 Introduction

Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans are crucial in diagnosing and treating various medical conditions [7,18], each with its own pros and cons. CT quickly delivers tissue conditions and attenuation

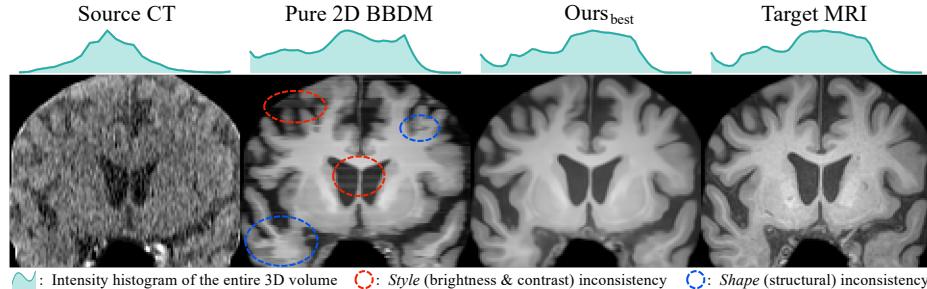


Fig. 1. Examples of slice inconsistency and resolved outcomes. This figure displays the coronal view of volumes synthesized by two 2D BBDMs trained on axial slices. The pure multi-slice 2D BBDM exhibits severe slice inconsistency, with noticeable discontinuities in both style and shape across slices. Our method produces slice-consistent volumes and can adjust the intensity histogram (i.e., style).

maps but struggles with detailed soft-tissue imaging [12]. Conversely, MRI offers outstanding soft-tissue clarity but comes with higher costs and longer acquisition times compared to CT [21]. This juxtaposition between CT and MRI sets the stage for exploring synthesized MRI from CT images in scenarios where only CT is available [17]. Obtaining MRI without additional scan enhances cost-efficiency and diagnostic accuracy, especially in practices like positron emission tomography/computed tomography (PET/CT) scan readings, where MRI’s clear anatomical details improves brain scan interpretations [10]. However, CT scans present some anatomical details subtly due to their lower soft-tissue contrast and higher noise levels [3], presenting a challenge in synthesizing faithful MRI from CT. This task can be defined as a image-to-image translation (I2I), which has key implications as it enables critical medical imaging tasks to be addressed, such as super-resolution [8], denoising [9], and multi-modal synthesis [19].

Recently, diffusion models (DMs), a type of deep generative model, have been increasingly favored in various medical imaging tasks [19,25,9], thanks to their ability to produce high-quality and diverse samples [11,32]. However, in medical I2I like CT-to-MRI translation, where it is essential for synthetic images to reliably reflect the structure of source images, DMs struggle to produce consistent outcomes due to their inherent stochasticity. Even using deterministic samplers like DDIM [28], their reliance on sampling from random noise introduces uncertainty, undermining reproducibility and medical reliability. Therefore, we performed medical I2I based on the Brownian Bridge Diffusion Model (BBDM [20]), which specifies the theoretical mapping between source and target data distributions through a diffusion process directly on the source image itself.

Meanwhile, medical imaging often involves 3D volumetric images. However, applying naïve 3D models to high-resolution volumes is impractical, due to the massive computational, memory, and data requirements [33]. As a solution, approaches based on latent diffusion models (LDMs), which compress the latent space using an autoencoder to train 3D LDMs, have been proposed [33,19].

However, the image quality of LDM relies on its 3D autoencoder, which faces challenges in achieving satisfactory performance on high-dimensional data and is still costly for ordinary GPUs in memory usage [23,13].

An alternative of using 2D models for slice-wise generation exists. However, this leads to *slice inconsistency* within a volume, as seen in Fig. 1, which can be defined in two aspects: *style* (e.g., brightness and contrast) and *shape* (e.g., structural discontinuities in 3D views). In this paper, we aim to address this slice inconsistency issue in 2D models through two novel methods. **(1) Style Key Conditioning (SKC)**: We condition the DM with the MRI’s histogram to enable it to learn the mapping between the histogram and the actual imaging style. By controlling the style of the generated slices through the histogram, we not only ensure style consistency but also enable the free manipulation of the target MRI’s style. **(2) Inter-Slice Trajectory Alignment (ISTA)**: We ensure consistency in both style and shape by enabling adjacent slices to sample along a jointly agreed-upon trajectory, while still allowing for parallel slice-wise inference within each batch. Specifically, ISTA aggregates the multiple outputs from the model to *co-predict* the direction of the next time step for each slice and deterministically *corrects* the transition error from the co-prediction.

Contributions. Our main contributions are as follows: **(i)** Through our proposed SKC and ISTA, we performed slice-consistent 3D volumetric brain CT-to-MRI translation using a 2D DM. **(ii)** By combining our approach with BBDM, we achieved fully deterministic and reliable medical I2I, with the target volume’s structure and style dictated only by the source volume and histogram-based condition, respectively. **(iii)** Using in-house CT-MRI dataset and public FLAIR-T1 MRI dataset from BraTS2023, we demonstrate that proposed 2D-based method performs superior cross-modality I2I compared to existing 2D and 3D baseline models. To the best of our knowledge, this is the first study to achieve high-quality 3D medical I2I based only on a 2D DM with no extra architectural models like autoencoders. The code for our project can be accessed at <https://github.com/MICV-yonsei/CT2MRI>.

2 Methods

In this section, we first briefly explore BBDM (Sec. 2.1). Then, we detail the multi-slice training strategy with SKC for *global* style consistency (Sec. 2.2) and ISTA sampling technique for *local* consistency both in style and shape (Sec. 2.3).

2.1 Preliminaries: 2D Brownian Bridge Diffusion Model

BBDM redefines the diffusion model within the framework of a Brownian bridge to generalize the source distribution from Gaussian to any arbitrary distribution, making it suitable for I2I. Let (\mathbf{x}, \mathbf{y}) be any source and target image pair, then the forward Brownian bridge diffusion process at time step t is defined based on $\mathbf{x}_0 = \mathbf{x}$ (e.g., the target MRI slice) and $\mathbf{x}_T = \mathbf{y}$ (e.g., the source CT slice) as

$$q_{BB}(\mathbf{x}_t | \mathbf{x}_0, \mathbf{y}) = \mathcal{N}(\mathbf{x}_t; (1 - m_t)\mathbf{x}_0 + m_t \mathbf{y}, \delta_t \mathbf{I}), \quad (1)$$

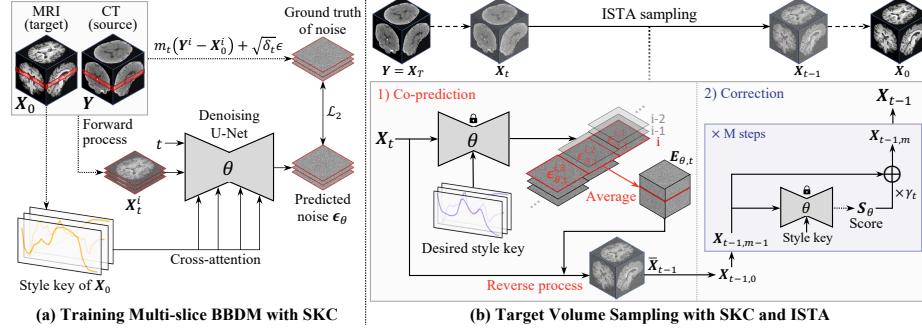


Fig. 2. Training and sampling scheme of the proposed methods. (a) During the multi-slice BBDM training, a target histogram-based style key is injected into the U-Net. (b) Target volume sampling proceeds in the manner of the Predictor-Corrector method. During the *co-prediction* phase, multiple $\epsilon_{\theta,t}^{i,k}$ are employed to establish connections among the predicted slices within $\tilde{\mathbf{X}}_{t-1}$. In the subsequent *correction* phase, the co-predicted volume is refined through a score-guided deterministic process.

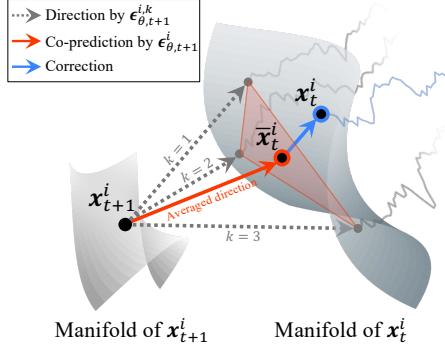
where $\delta_t = 2s(m_t - m_t^2)$ is the variance with scale factor s for the sampling diversity, and $m_t = t/T$. The Gaussian transition kernel of the forward Markov process $q_{BB}(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y})$, with variance $\delta_{t|t-1} = \delta_t - \delta_{t-1}(1 - m_t)^2 / (1 - m_{t-1})^2$, can be easily derived through Eq. (1). Similar to DDPM [14], the forward process posterior $q_{BB}(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0, \mathbf{y})$ can be derived in closed form of Gaussian with variance $\tilde{\delta}_t = \delta_{t|t-1} \cdot \delta_{t-1} / \delta_t$ and mean $\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0, \mathbf{y}) = c_{xt}\mathbf{x}_t + c_{yt}\mathbf{y} + c_{et}(m_t(\mathbf{y} - \mathbf{x}_0) + \sqrt{\delta_t}\epsilon)$ for some coefficient c_{xt} , c_{yt} , and c_{et} . The reverse process is defined as $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \tilde{\delta}_t \mathbf{I})$, where $\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$ is the predicted mean by the model, and the variance $\tilde{\delta}_t$ uses the same value as $q_{BB}(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0, \mathbf{y})$. The model parameters θ are trained to make the distribution of $q_{BB}(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0, \mathbf{y})$ and $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y})$ identical, using a simplified evidence lower bound (ELBO) as

$$\mathbb{E}_{\mathbf{x}_0, \mathbf{y}, \epsilon} [c_{et} \| m_t(\mathbf{y} - \mathbf{x}_0) + \sqrt{\delta_t}\epsilon - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \|_2^2], \quad (2)$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is sampled random noise and $\boldsymbol{\epsilon}_\theta$ is the noise estimator model. A notable aspect in Eq. (2) is that, the prediction $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ includes not just random noise ϵ . We have also observed that incorporating the guiding term $m_t(\mathbf{y} - \mathbf{x}_0) + \sqrt{\delta_t}\epsilon$, which is the direction pointing to \mathbf{x}_t from the predicted \mathbf{x}_0 , improves performance over solely training the model to predict pure ϵ practically.

2.2 Training Multi-slice 2D BBDM with Style Key Conditioning

When the 2D BBDM is naively trained slice-wise, the uneven brightness or contrast is observed globally in the synthetic MRI (Fig. 1), which we call *global* style inconsistency. This occurs due to the target MRI in the training data exhibiting various styles, even for similar sources. To address global style inconsistency, we conditioned the model with a *style key* that uniquely matches each target style, which also allows us to freely control the style of the generated slice. We

**Algorithm 1** ISTA Sampling

```

1: sample input volume  $\mathbf{X}_T = \mathbf{Y} \sim q(\mathbf{Y})$ 
2: select desired style key  $\mathbf{c}_{SKC}$ 
3: for  $t = T - 1, \dots, 0$  do Co-prediction
4:    $\mathbf{E}_{\theta,t+1} = CP_{\theta}(\mathbf{X}_{t+1}, \mathbf{c}_{SKC}, t + 1)$ 
5:    $\mathbf{X}_t = ReverseProc_{BB}(\mathbf{X}_{t+1}, \mathbf{E}_{\theta,t+1}, \mathbf{Y})$ 
6:    $\mathbf{X}_{t,0} = \bar{\mathbf{X}}_t$ 
7:   for  $m = 1, \dots, M$  do Correction
8:      $\mathbf{S}_{\theta} = \mathbf{S}_{\theta}(\mathbf{X}_{t,m-1}, \mathbf{c}_{SKC}, t)$ 
9:      $\gamma_t = \lambda \delta_{t|t-1} (\sqrt{d}/\|\mathbf{S}_{\theta}\|_2)^2$ 
10:     $\mathbf{X}_{t,m} = \mathbf{X}_{t,m-1} + \gamma_t \mathbf{S}_{\theta}$ 
11:    $\mathbf{X}_t = \mathbf{X}_{t,M}$ 
return  $\mathbf{X}_0$ 

```

Fig. 3. Visualization of the latent space and algorithm for ISTA sampling. The trained U-Net produces inconsistent outputs for multi-slice inputs that include the i^{th} slice. The *co-prediction* unifies the direction of these independent inferences, while the *correction* aligns the co-predicted $\bar{\mathbf{x}}_t^i$ onto the manifold of \mathbf{x}_t^i .

discovered that the intensity histogram of an MRI volume can serve as an ideal style key because it vividly reflects imaging style while being invariant to an individual's anatomy. Consequently, we trained a 2D BBDM, conditioning on MRI histogram-based style keys (Fig. 2a), employing a multi-slice input-output approach (1) to utilize the information from adjacent slices and (2) to compile multiple predictions for a single slice at each ISTA sampling step.

From now on, we define the notations as follows: $\mathbf{X}, \mathbf{Y}, \mathbf{X}_t \in \mathbb{R}^{Z \times H \times W}$ represent complete MRI, CT, and latent volumes, $\mathbf{x}^i, \mathbf{y}^i, \mathbf{x}_t^i \in \mathbb{R}^{H \times W}$ are the i^{th} slices of these volumes, and $\mathbf{X}^i, \mathbf{Y}^i, \mathbf{X}_t^i \in \mathbb{R}^{(2N+1) \times H \times W}$ denote sub-volumes of the i^{th} index and its adjacent $2N$ slices (e.g., $\mathbf{X}^i = [\mathbf{x}^{i-N}, \mathbf{x}^{i-N+1}, \dots, \mathbf{x}^{i+N}]$). Our objective for training multi-slice 2D BBDM modifies Eq. (2) as follows:

$$\mathbb{E}_{\mathbf{X}_0^i, \mathbf{Y}^i, \epsilon} [c_{et} \|m_t(\mathbf{Y}^i - \mathbf{X}_0^i) + \sqrt{\delta_t} \epsilon - \epsilon_{\theta}(\mathbf{X}_t^i, \mathbf{c}_{SKC}, t)\|_2^2], \quad (3)$$

where $\epsilon, \epsilon_{\theta}(\cdot)$ always share the same dimensions as the volume or slice input and $\mathbf{c}_{SKC} \in \mathbb{R}^{3 \times B}$ is a set of three 1D histograms of target MRI volume \mathbf{X} (i.e., histogram, cumulative histogram, and histogram differential), each with B bins. For testing, the histogram of any MRI with desired style can be used as \mathbf{c}_{SKC} . We set $N = 1$ and $B = 128$ which suffice our task with minimal extra cost.

2.3 Target Volume Sampling with Inter-slice Trajectory Alignment

While SKC addresses global style inconsistencies, it does not directly combat the *local* slice inconsistency, which refers to style or shape discontinuities between adjacent slices. We resolve the local slice inconsistency by chaining all slices inside a volume exclusively within the 2D sampling process, without relying on any additional training beyond minimal multi-channel (i.e., 3-slice) burden. ISTA aligns the trajectories of adjacent slices during sampling to integrate independent inference outputs, ensuring it remains fully deterministic and parallel like

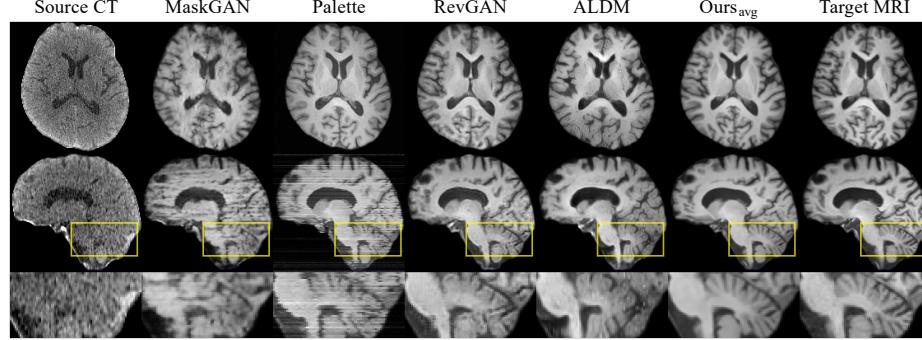


Fig. 4. Qualitative comparison with baselines (CT→MRI)

BBDM. Similar to the Predictor-Corrector method in [2], each ISTA sampling at time step t involves two steps (Fig. 2b and Fig. 3): *co-prediction* and *correction*.

Co-prediction. ISTA sampling utilizes all predicted $\epsilon_{\theta,t}^{i,k} := \epsilon_{\theta}(\mathbf{x}_t^{i,k}, \mathbf{c}_{SKC}, t) \in \mathbb{R}^{H \times W}$ for the i^{th} slice of a volume obtained by the overlapping inference at time step t ($k \in \{1, 2, \dots, 2N+1\}$). First, we aggregate the $2N+1$ predictions for the i^{th} slice into a single prediction $\epsilon_{\theta,t}^i := \mathbb{E}_k[\epsilon_{\theta,t}^{i,k}]$. Subsequently, we concatenate all $\epsilon_{\theta,t}^i$ to obtain the aggregated prediction volume $\mathbf{E}_{\theta,t}$. For an input volume \mathbf{X}_t , the process of co-predicting a single volume $\mathbf{E}_{\theta,t} \in \mathbb{R}^{Z \times H \times W}$ using multiple i^{th} slice's outputs from the trained multi-slice model will be defined as CP_{θ} :

$$\mathbf{E}_{\theta,t} = CP_{\theta}(\mathbf{X}_t, \mathbf{c}_{SKC}, t) := [\epsilon_{\theta,t}^1, \epsilon_{\theta,t}^2, \dots, \epsilon_{\theta,t}^Z]. \quad (4)$$

Finally, according to the reverse process of BBDM, we obtain the volume $\bar{\mathbf{X}}_{t-1} \in \mathbb{R}^{Z \times H \times W}$ using the averaged prediction $\mathbf{E}_{\theta,t}$, \mathbf{X}_t , and source volume \mathbf{Y} .

Correction. However, $\bar{\mathbf{x}}_{t-1}^i$, having moved a time step via the simple average of $\epsilon_{\theta,t}^{i,k}$ s (i.e., the averaged score at \mathbf{x}_t^i), does not guarantee its placement on the manifold of \mathbf{x}_{t-1}^i (Fig. 3 left). To mitigate this transition error, inspired by the Langevin MCMC-based corrector in [30], we propose a deterministic correction method with M steps:

$$\mathbf{X}_{t,m} = \mathbf{X}_{t,m-1} + \lambda \delta_{t|t-1} (\sqrt{d}/\|\mathbf{S}_{\theta}\|_2)^2 \mathbf{S}_{\theta}, \quad (5)$$

where $\mathbf{X}_{t,0} = \bar{\mathbf{X}}_t$, $\mathbf{X}_t = \mathbf{X}_{t,M}$, the step size λ is a hyperparameter, d is the dimensionality of \mathbf{x}_t^i , and $\mathbf{S}_{\theta} = \mathbf{S}_{\theta}(\mathbf{X}_{t,m-1}, \mathbf{c}_{SKC}, t)$ is the score function [30] at $\mathbf{X}_{t,m-1}$. We removed the random noise z of Langevin MCMC to ensure determinism, and applied an adaptive weighting for the magnitude of \mathbf{S}_{θ} and the variance. \mathbf{S}_{θ} can be derived similar to [14,31] on a volume basis as follows (Supp. C):

$$\mathbf{S}_{\theta}(\mathbf{X}_t, \mathbf{c}_{SKC}, t) = -\frac{1}{\delta_t} \{m_t(\mathbf{X}_t - \mathbf{Y}) + (1 - m_t)CP_{\theta}(\mathbf{X}_t, \mathbf{c}_{SKC}, t)\}. \quad (6)$$

Table 1. Quantitative comparison with baselines

Methods	CT→MRI (in-house)			FLAIR→T1 (BraTS)		
	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑
3D	RevGAN	0.0577	25.344	0.8925	0.0824	22.073
	ALDM (200 step)	0.0673	23.495	0.8474	0.0975	20.453
2D	MaskGAN	0.0910	21.328	0.7421	0.1112	19.311
	Palette (1000 step)	0.0811	21.877	0.4365	0.1495	16.974
	Ours (100 step)	0.0515	26.666	0.9199	0.0808	22.579
0.8837						

3 Experiments

Datasets. We evaluate our method and baselines on in-house CT-MRI (T1) dataset and public FLAIR-T1 MRI dataset from BraTS2023 [4,5,6,22]. CT and MRI images from the in-house dataset were coregistered using SPM12 [1] and skull-stripped with SynthStrip [16]. Both datasets were resampled to 1 mm isotropic voxels, cropped for background removal, and min-max normalized to a range of 0 to 1. Details of each dataset can be found in Supp. A.

Implementation Details. All 2D models used axial slices, and all DMs were trained with 1000 diffusion steps. Our method employed 100 DDIM sampling steps, and for ISTA sampling, we adjusted to 50 steps with $M = 1$ correction for a fair comparison. For testing, the default style key of SKC is the averaged histogram of the entire training dataset. More details can be found in Supp. A.

Baselines. For comparison, we reproduced 2D and 3D GANs and DMs baselines for I2I, using default settings unless specified otherwise: (1) MaskGAN [26], a cyclic GAN for unpaired 2D medical I2I, adapted for paired training; (2) Palette [27], a conditional DM for 2D I2I; (3) RevGAN [24], a GAN with a reversible 3D backbone, also adapted for paired training; (4) ALDM [19], a 3D LDM framework, with model channels reduced to fit our GPU memory capacity.

3.1 Evaluations

Quantitative Results. We confirmed the effectiveness of our method against baselines on the in-house CT-MRI and BraTS FLAIR-T1 datasets (Table 1). For quantitative evaluation, normalized root mean square error (NRMSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) were used. In examining baselines, we noted two points: DMs showed similar or lower metrics than GANs due to the diversity from their inherent stochasticity, and 3D models surpassed 2D models. Yet, our method, which is based on a 2D DM with small sampling step, outperformed all GANs and 3D models. This highlights the efficacy of (1) style uniformization through SKC, (2) shape and style consistency via ISTA, and (3) Brownian bridge-based stable mapping.

Table 2. Quantitative results of ablation studies.

Methods	CT→MRI (in-house)			FLAIR→T1 (BraTS)		
	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑
Pure BBDM	0.0575	25.359	0.8912	0.2232	16.639	0.7917
+SKC _{colin}	0.0619	25.285	0.9004	0.0855	22.223	0.8806
+SKC _{avg}	0.0517	26.443	0.9157	0.0820	22.409	0.8763
+ISTA	0.0525	26.551	0.9167	0.1787	17.652	0.8054
+SKC _{avg} , ISTA _{cp-only}	0.0515	26.664	0.9199	0.0809	22.578	0.8837
+SKC _{avg} , ISTA (Ours _{avg})	0.0515	26.666	0.9199	0.0808	22.579	0.8837
+SKC _{best} , ISTA (Ours _{best})	0.0454	27.589	0.9252	0.0625	24.644	0.8938

Qualitative Results. Our method generated high-quality MRI volumes that are very similar to the target MRI, outperforming baselines (Fig. 4 & Supp. B). Compared to 3D models, 2D baseline models showed severe slice inconsistency in sagittal views. However, our 2D-based method excelled in slice consistency and surpassed 3D models in anatomical clarity across all views. Furthermore, our method can produce volumes that reflect various desired styles (Supp. E).

Ablation Studies. We conducted experiments on various options for SKC and ISTA (Table 2 & Supp. D). Firstly, we set three options for SKC: the default averaged histogram (SKC_{avg}), the histogram of the public Colin 27 Average Brain [15] (SKC_{colin}), and the histogram similar to the target ground truth from the training data (SKC_{best}). The performance of SKC_{avg} over pure BBDM and the elimination of style inconsistency indicate that SKC enables slice generation with uniform style. The performance of SKC_{colin} may vary depending on the similarity between Colin’s style and the test data, but it outperformed the pure BBDM and produced volumes with the desired styles (Supp. E). The marked improvement with Ours_{best} highlights our method’s potential to accurately produce volumes close to the ground truth when provided with an exact style key for the target. Secondly, two options were set for ISTA: with correction (ISTA, 50 step, $M = 1$) and co-prediction only (ISTA_{cp-only}, 100 step, $M = 0$). ISTA, ISTA_{cp-only}, and without ISTA showed superior performance in that order, proving the effect of both local slice consistency via co-prediction and error reduction via correction. Ultimately, optimal performance was achieved with both SKC and ISTA.

4 Conclusion and Discussion

We have successfully achieved slice-consistent and high-quality 3D volumetric brain CT-to-MRI translation with 2D BBDM, enhancing medical reliability through fully deterministic generation. Our proposed methods, SKC and ISTA, tackle the slice inconsistency issue from different perspectives and hold broad applicability in medical volume synthesis tasks. These fundamental approaches have significant potential for future expansion. SKC, which allows for control

of the target imaging style based on the histogram, can be particularly utilized in addressing domain gap issues in medical imaging fields, such as multi-site MRI harmonization. Concurrently, ISTA can serve as a foundational method that enables iterative generative models, such as diffusion models, to establish connectivity in a higher-dimensional space than the input dimension.

Acknowledgments. This work was supported in part by the IITP 2020-0-01361 (AI Graduate School Program at Yonsei University), NRF RS-2024-00345806, NRF RS-2023-00262002 funded by Korean Government (MSIT), and a grant of the Korea Dementia Research Project through the Korea Dementia Research Center (KDRC), funded by the Ministry of Health & Welfare and Ministry of Science and ICT, Republic of Korea (RS-2022-KH127174).

Disclosure of Interests. The authors have no competing interests.

References

1. Statistical parametric mapping (spm). *Scholarpedia* **3**(4) (2008)
2. Allgower, E.L., Georg, K.: Numerical continuation methods: an introduction, vol. 13. Springer Science & Business Media (2012)
3. Babu Vimala, B., Srinivasan, S., Mathivanan, S.K., Mahalakshmi, Jayagopal, P., Dalu, G.T.: Detection and classification of brain tumor using hybrid deep learning models. *Scientific Reports* **13**(1), 23029 (2023)
4. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314* (2021)
5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The cancer imaging archive* **286** (2017)
6. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**(1), 1–13 (2017)
7. Chandarana, H., Wang, H., Tijssen, R., Das, I.J.: Emerging role of mri in radiation therapy. *Journal of Magnetic Resonance Imaging* **48**(6), 1468–1478 (2018)
8. Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D.: Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network. In: International conference on medical image computing and computer-assisted intervention. pp. 91–99. Springer (2018)
9. Chung, H., Lee, E.S., Ye, J.C.: Mr image denoising and super-resolution using regularized reverse diffusion. *IEEE Transactions on Medical Imaging* **42**(4), 922–934 (2022)
10. Decazes, P., Hinault, P., Veresezan, O., Thureau, S., Gouel, P., Vera, P.: Trimodality pet/ct/mri and radiotherapy: a mini-review. *Frontiers in Oncology* **10**, 614008 (2021)
11. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* **34**, 8780–8794 (2021)

12. Dowling, J., Lambert, J., Parker, J., Greer, P.B., Fripp, J., Denham, J., Ourselin, S., Salvado, O.: Automatic mri atlas-based external beam radiation therapy treatment planning for prostate cancer. In: International Workshop on Prostate Cancer Imaging. pp. 25–33. Springer (2010)
13. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: A papier-mâché approach to learning 3d surface generation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 216–224 (2018)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
15. Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C.: Enhancement of mr images using registration for signal averaging. Journal of computer assisted tomography **22**(2), 324–333 (1998)
16. Hoopes, A., Mora, J.S., Dalca, A.V., Fischl, B., Hoffmann, M.: Synthstrip: Skull-stripping for any brain image. NeuroImage **260**, 119474 (2022)
17. Jin, C.B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., Cui, X.: Deep ct to mr synthesis using paired and unpaired data. Sensors **19**(10), 2361 (2019)
18. Khoo, V., Padhani, A., Tanner, S., Finnigan, D., Leach, M., Dearnaley, D.: Comparison of mri with ct for the radiotherapy planning of prostate cancer: a feasibility study. The British journal of radiology **72**(858), 590–597 (1999)
19. Kim, J., Park, H.: Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7604–7613 (2024)
20. 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)
21. Mayerhoefer, M.E., Prosch, H., Beer, L., Tamandl, D., Beyer, T., Hoeller, C., Berzaczy, D., Raderer, M., Preusser, M., Hochmair, M., et al.: Pet/mri versus pet/ct in oncology: a prospective single-center study of 330 examinations focusing on implications for patient management and cost considerations. European journal of nuclear medicine and molecular imaging **47**, 51–60 (2020)
22. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
23. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4460–4470 (2019)
24. van der Ouderaa, T.F., Worrall, D.E.: Reversible gans for memory-efficient image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4720–4728 (2019)
25. Ozturkler, B., Liu, C., Eckart, B., Mardani, M., Song, J., Kautz, J.: Smrd: Sure-based robust mri reconstruction with diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 199–209. Springer (2023)
26. Phan, V.M.H., Liao, Z., Verjans, J.W., To, M.S.: Structure-preserving synthesis: Maskgan for unpaired mr-ct translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 56–65. Springer (2023)

27. Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10 (2022)
28. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2020)
29. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. Advances in neural information processing systems **32** (2019)
30. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2020)
31. Vahdat, A., Kreis, K., Kautz, J.: Score-based generative modeling in latent space. Advances in Neural Information Processing Systems **34**, 11287–11302 (2021)
32. Xiao, Z., Kreis, K., Vahdat, A.: Tackling the generative learning trilemma with denoising diffusion gans. In: International Conference on Learning Representations (2021)
33. Zhu, L., Xue, Z., Jin, Z., Liu, X., He, J., Liu, Z., Yu, L.: Make-a-volume: Leveraging latent diffusion models for cross-modality 3d brain mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 592–601. Springer (2023)

Supplementary Materials

A. Experimental Details

Table 3. Dataset details

	CT-MRI	FLAIR-T1
#total	219	1470
#train	197	1251
#test	22	219
shape	$160 \times 160 \times 160$	$176 \times 176 \times 160$
voxel size	$1 \times 1 \times 1 mm^3$	

Table 4. Implementation details

	CT→MRI	FLAIR→T1
batch size	16	8
#iterations	100,000	150,000
λ	0.5	0.05
time step		1000
sampling step		100
GPU	NVIDIA RTX A6000	

B. Additional Qualitative Comparison

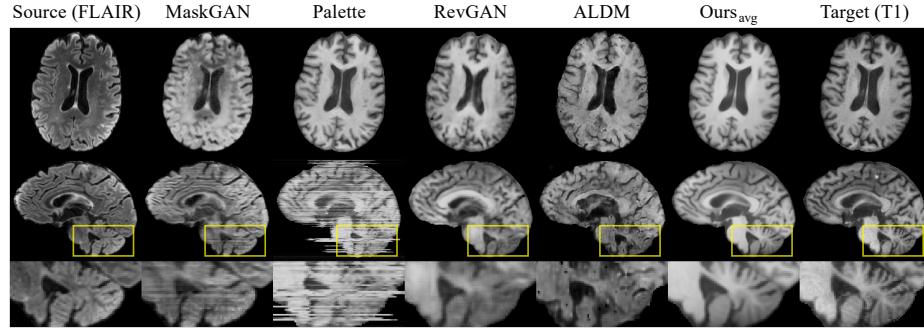


Fig. 5. Qualitative comparison with baselines (FLAIR→T1)

C. Derivation of Score Function $S_\theta(\mathbf{X}_t, \mathbf{c}_{SKC}, t)$ for θ Trained as $\theta = \theta^*$

Since $\mathbf{E}_{\theta^*, t} = CP_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t) = m_t(\mathbf{Y} - \mathbf{X}_0) + \sqrt{\delta_t} \boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t \in \mathbb{R}^{Z \times H \times W}$,

$CP_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t) = \mathbf{X}_t - \mathbf{X}_0$ ($\because \mathbf{X}_t = (1 - m_t)\mathbf{X}_0 + m_t\mathbf{Y} + \sqrt{\delta_t} \boldsymbol{\epsilon}_t$).

Then, $CP_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t) = m_t[\mathbf{Y} - \{\mathbf{X}_t - CP_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t)\}] + \sqrt{\delta_t} \boldsymbol{\epsilon}_t$

$$\iff \boldsymbol{\epsilon}_t = \frac{1}{\sqrt{\delta_t}} \{m_t(\mathbf{X}_t - \mathbf{Y}) + (1 - m_t)CP_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t)\}.$$

$$\begin{aligned} \text{Thus, } S_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t) &= \nabla_{\mathbf{X}_t} \log q_{BB}(\mathbf{X}_t \mid (1 - m_t)\mathbf{X}_0 + m_t\mathbf{Y}, \mathbf{c}_{SKC}) \\ &= -\frac{[\mathbf{X}_t - \{(1 - m_t)\mathbf{X}_0 + m_t\mathbf{Y}\}]}{\delta_t} = -\frac{\boldsymbol{\epsilon}_t}{\sqrt{\delta_t}} \quad (\text{cf. } \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}} \mid \mathbf{x}) = -\frac{\tilde{\mathbf{x}} - \mathbf{x}}{\sigma^2} \text{ in [29]}), \end{aligned}$$

since \mathbf{X}_t is the Gaussian perturbation of $(1 - m_t)\mathbf{X}_0 + m_t\mathbf{Y}$.

$$\therefore S_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t) = -\frac{1}{\delta_t} \{m_t(\mathbf{X}_t - \mathbf{Y}) + (1 - m_t)CP_{\theta^*}(\mathbf{X}_t, \mathbf{c}_{SKC}, t)\}.$$

D. Qualitative Results of Ablation Study

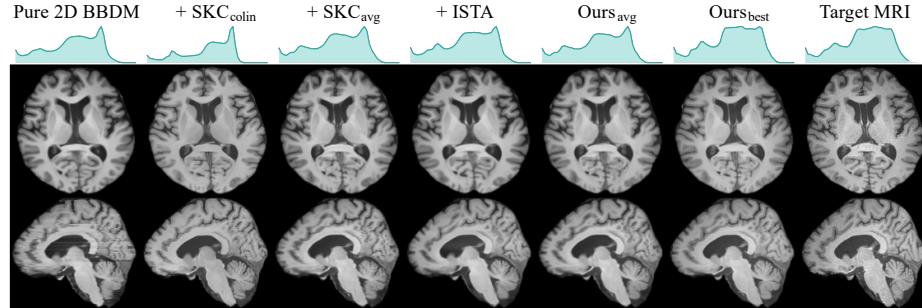


Fig. 6. Qualitative comparison with baselines (CT \rightarrow MRI)

E. Additional Study for Style Key Conditioning

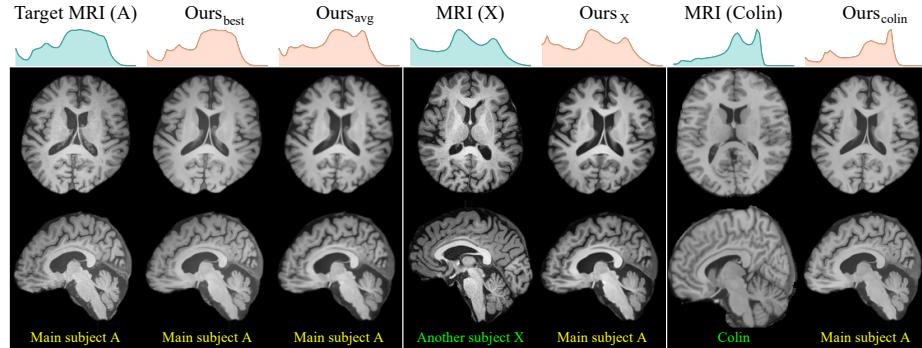


Fig. 7. Qualitative results for various style keys (CT \rightarrow MRI). Figure shows real MRIs of main subject A, another subject X, and Colin 27 Average Brain (public access), along with synthetic MRIs of subject A. Both $Ours_X$ and $Ours_{colin}$ present synthetic MRIs of subject A, each transformed into the respective styles of X and Colin. From the analysis of the intensity histogram presented above, we observe that $Ours_X$ closely matches the MRI histogram of subject X, while $Ours_{colin}$ aligns with that of Colin. This demonstrates that the style of MRIs can be effectively controlled using SKC.