Residual Diffusion Bridge in Wavelet Space for Medical Image Translation

Xiao Yang✉[1], Wanqing Zhao[1], Jérémie Nsengimana[2], and Jingjing Zhang[3]

[1] School of Computing, Newcastle University, Newcastle upon Tyne, UK
X.Yang64@newcastle.ac.uk
[2] Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
[3] School of Engineering, Newcastle University, Newcastle upon Tyne, UK

Abstract. Diffusion bridges provide a principled way to connect paired source and target images for medical image translation, but pixel-space bridge trajectories may still entangle global contrast transport with local texture synthesis. This makes frequency-specific modelling less explicit and can limit the interpretability of cross-contrast MRI translation. To address this issue, we propose Wavelet Residual Bridge (WRB), a frequency-separated paired-translation framework in complex-wavelet space. The proposed method first decomposes each image with a 3-level dual-tree complex wavelet transform (DTCWT). Low-pass coefficients are aligned by a deterministic monotone mapping to obtain a structure-aligned initialization, while modality-specific high-frequency residuals are sampled by Gaussian bridges at each DTCWT level in learned compact texture subspaces. A Dirichlet constraint anchors non-texture regions to the source image, and gradient-isolated multi-stage training reduces shortcut learning while keeping the low-frequency mapping, wavelet-bridge, and image-refinement outputs inspectable. On BraTS 2021, covering T1→T2, T1→FLAIR, and T2→FLAIR translation, WRB achieved the highest PSNR and SSIM among all baselines, including 27.74±2.56 dB PSNR and 92.69±2.14% SSIM for T1→T2. On Gold Atlas, WRB also achieved the highest PSNR on both MRI-to-CT synthesis tasks. Code is available at `https://github.com/xiao9672/RDBWS.git`

Keywords: Medical Image Synthesis · Diffusion Bridge · Wavelets · Cross-modality Image Translation

1 Introduction

Multi-contrast magnetic resonance imaging (MRI) provides complementary tissue contrasts for brain-tumor management. Missing contrasts affect radiotherapy planning and response assessment [18, 28], but acquiring all sequences increases scan time and the risk of motion artifacts. Scanner and protocol differences further lead to incomplete multi-contrast data [5, 25]. These factors motivate methods that synthesize missing contrasts (e.g. T2 from T1) for registration and downstream analysis, including patch-based and learning-based

pipelines [8, 9, 12]. For cross-modality image translation, another related application is MRI-to-CT synthesis for MRI-only radiotherapy planning [3, 15, 24].

Deep learning has seen progress in medical image synthesis. When aligned image pairs are available, supervised methods such as pix2pix learn a mapping with paired adversarial losses [11]. On the other hand, unpaired frameworks such as CycleGAN relax the registration requirement through cycle-consistency constraints [30]. However, adversarial training can suffer from mode collapse and training instability, and may introduce hallucinated structures or contrast drift, which is problematic when synthetic images are used for treatment planning.

Among these methods, diffusion models provide a non-adversarial alternative with stable training and high sample quality [4, 7, 23]. Conditional diffusion has been applied to both general image-to-image translation [20] and medical image synthesis [13, 17]. In cross-contrast synthesis, standard conditional diffusion uses the source image only as a condition, so the learned reverse trajectory is not explicitly tied to the paired target image. As a result, global contrast adjustment and local detail generation may be learned within the same image-space process, making frequency-specific modelling and interpretation less explicit.

Diffusion bridges couple source and target distributions by generating trajectories between paired endpoints [21]. Recent bridge-based work has improved medical image translation through self-consistent recursive diffusion bridges [1]. Nevertheless, pixel-space bridges still use a single trajectory schedule to model both global contrast transport and fine-grained detail synthesis. This can limit the learning of explicit controls for structure- and texture-related components. Moreover, when wavelet representations are used to separate frequency bands, inverse transforms can introduce artifacts near subband or anatomical boundaries. In this paper, we propose Wavelet Residual Bridge (WRB) for paired MRI translation, with the following contributions.

(1) Interpretable structure–texture separation. Cross-contrast translation is decomposed into deterministic low-frequency structure mapping and stochastic high-frequency texture synthesis in dual-tree complex wavelet transform (DTCWT) space. The low-frequency mapping, wavelet-bridge, and image-refinement stages produce inspectable intermediate outputs.

(2) Diffusion bridges in learned wavelet texture subspaces. Modality-specific texture is represented as masked DTCWT high-frequency residuals, projected onto compact texture subspaces at each DTCWT level, and sampled by Gaussian bridges.

(3) Gradient-isolated multi-stage training. The low-frequency mapper, DTCWT-level bridge denoisers, and image-domain refiner are trained with separate optimizers and stop-gradient boundaries. The model is trained stage-wise, learning iteratively from low-frequency structure alignment to high-frequency residual synthesis. This reduces shortcut learning and keeps the contribution of each component inspectable [6].

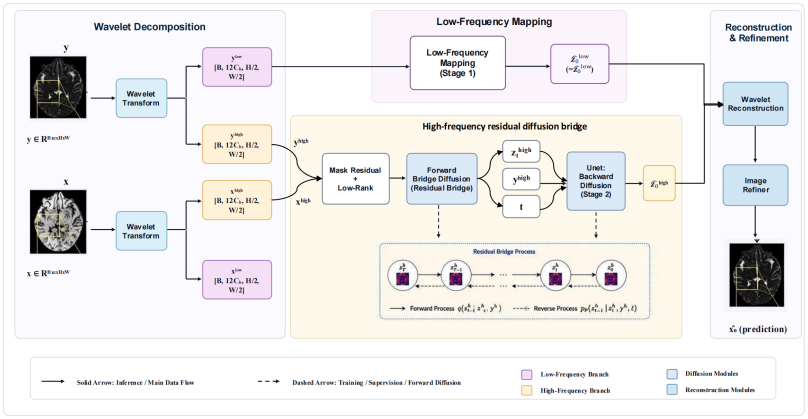

Fig. 1. Framework of WRB. DTCWT decomposes images into low-pass (LL) and high-pass subbands. Stage 1 learns a deterministic mapper to reconstruct x^{struct}, while Stage 2 learns a constrained Gaussian bridge to synthesize masked high-frequency residuals conditioned on x^{struct}, and reconstructs the final output $\hat{x}$ by inverse DTCWT.

2 Method

The proposed WRB addresses the frequency-coupling and boundary-artifact issues discussed above by separating cross-contrast translation into frequency-specific stages(Fig. 1). A DTCWT decomposes each image into low-pass (LL) and high-pass subbands; a deterministic mapper aligns the LL coefficients to produce x^{struct}, and per-level Gaussian bridges sample masked high-frequency residuals in learned subspaces, yielding $\hat{x}$ after inverse DTCWT.

2.1 Wavelet Residual Formulation

Let (y, x) denote a paired training sample, where $y \in \mathbb{R}^{C_s \times H \times W}$ is a source image and $x \in \mathbb{R}^{C_t \times H \times W}$ is the corresponding target image. The objective is to learn $p(x \mid y)$ and synthesize $\hat{x}$ given y. Stochastic modeling is confined to high-frequency residuals, while a deterministic low-frequency mapping is used to obtain a structure-aligned initialization:

$$\hat{x} \;=\; x^{\text{struct}} \;+\; \alpha\, r^{\text{hf}}, \qquad r^{\text{hf}} \;=\; (x^{\text{recon}} - x^{\text{struct}}) \odot M_{\text{fg}}, \tag{1}$$

where M_{fg} is a foreground mask obtained by thresholding a primary source channel, with the background optionally clamped to the source at inference. The blending coefficient α follows a schedule during training and may additionally be scaled by a residual weight.

DTCWT decomposition. Let $\mathcal{W}$ and $\mathcal{W}^{-1}$ denote the forward and inverse DTCWT. For $u \in \mathbb{R}^{C \times H \times W}$,

$$\mathcal{W}(u) = \big(c^{LL}(u), \{\, h^{\ell}(u) \,\}_{\ell=1}^{L}\big), \qquad h^{\ell}(u) \in \mathbb{R}^{(12C) \times H_{\ell} \times W_{\ell}}, \tag{2}$$

where each $h^\ell(u)$ contains the real and imaginary parts of six oriented complex high-pass coefficients, concatenated along the channel dimension. Let $D_s = 12C_s$ and $D_t = 12C_t$. The target-domain low-pass coefficients $\hat{c}^{LL}$ are predicted from the source low-pass coefficients by a deterministic mapper g_ϕ, realized as either a monotone spline OT map or a low-pass UNet:

$$\hat{c}^{LL} \;=\; g_\phi\big(c^{LL}(y),\, y,\, M_{\mathrm{fg}}\big), \qquad x^{\mathrm{struct}} \;=\; \mathcal{W}^{-1}\big(\hat{c}^{LL}, \{\, d^\ell_{\mathrm{struct}} \,\}_{\ell=1}^{L}\big)\,, \tag{3}$$

where $\{d^\ell_{\mathrm{struct}}\}$ are the detail coefficients used to reconstruct x^{struct}, either copied from the source primary channel or set to zero according to the chosen detail policy.

High-frequency residual target. A texture mask M^ℓ_{tex} is obtained by eroding M_{fg} at each wavelet scale, and $\overline{M}^{\ell,n}_{\mathrm{tex}}$ denotes its replication to n channels. The residual to be modeled at level ℓ is

$$r^\ell_\star \;=\; \Big(h^\ell(x) - h^\ell(x^{\mathrm{struct}})\Big) \odot \overline{M}^{\ell,D_t}_{\mathrm{tex}}\,. \tag{4}$$

2.2 Normalized Texture Subspace Representation

Because wavelet high-pass magnitudes vary across datasets and scales, each $r^\ell_\star$ is normalized by a channel-wise running scale $s^\ell \in \mathbb{R}^{D_t}$, maintained as an EMA of masked standard deviations, and projected into a low-rank texture subspace:

$$z^\ell_0 \;=\; (U^\ell)^\top (r^\ell_\star \oslash s^\ell), \qquad r^\ell_\star \oslash s^\ell \;\approx\; U^\ell z^\ell_0, \tag{5}$$

where $U^\ell \in \mathbb{R}^{D_t \times K_\ell}$ is a learnable basis of rank $K_\ell \ll D_t$, regularized toward orthonormality via $(U^\ell)^\top U^\ell \approx I$. The projection is applied pointwise over spatial locations.

The bridge is anchored at a source-derived endpoint in the same subspace. Let y_p denote the primary source channel used for structural anchoring. A source residual anchor in the normalized wavelet space is defined as:

$$\tilde{r}^\ell_T \;=\; \big(h^\ell(y_p) - h^\ell(x^{\mathrm{struct}})\big) \oslash s^\ell, \qquad z^\ell_T \;=\; (U^\ell)^\top \tilde{r}^\ell_T. \tag{6}$$

The target residual z^ℓ_0 and the source residual anchor z^ℓ_T are thus connected by the bridge. The condition c^ℓ is formed by concatenating the downsampled images $\mathcal{P}^\ell(x^{\mathrm{struct}})$ and $\mathcal{P}^\ell(y)$, the texture mask M^ℓ_{tex}, and the high-pass coefficients $h^\ell(x^{\mathrm{struct}})$ and $h^\ell(y)$ along the channel dimension.

2.3 Constrained Gaussian Bridge Diffusion and Reconstruction

For each level ℓ, a Gaussian bridge marginal is defined at discrete step $t \in \{0, 1, \ldots, T\}$ with $\tau = t/T$:

$$q\big(z^\ell_t \mid z^\ell_0, z^\ell_T\big) \;=\; \mathcal{N}\big(z^\ell_t;\; \mu_{0,t}\, z^\ell_0 + \mu_{T,t}\, z^\ell_T,\; \bar{\beta}_t I\big)\,, \tag{7}$$

where $\bar{\beta}_t$ is the marginal variance. A parabolic schedule is adopted: $\mu_{0,t} = 1{-}\tau$, $\mu_{T,t} = \tau$, $\bar{\beta}_t = \sigma_s^2\, \tau(1{-}\tau)$, so that both endpoints are deterministic, i.e. $\bar{\beta}_0{=}\bar{\beta}_T{=}0$, recovering the Brownian-bridge marginal.

To enforce spatially localized synthesis, a hard Dirichlet constraint is imposed at every step, anchoring the complement of the texture mask to the source endpoint:

$$z_t^\ell \leftarrow \overline{M}_{\text{tex}}^{\ell,K_\ell} \odot z_t^\ell + \left(1 - \overline{M}_{\text{tex}}^{\ell,K_\ell}\right) \odot z_T^\ell. \tag{8}$$

Coefficients outside the texture mask are thereby anchored to the source endpoint, ensuring consistent boundary conditions during both training and sampling. The clean coordinate z_0^ℓ is predicted from $(z_t^\ell, c^\ell, t/T)$ by a level-specific denoiser f_θ^ℓ: $\hat{z}_0^\ell = f_\theta^\ell([z_t^\ell, c^\ell], t/T)$. The timestep t is drawn from $\mathcal{U}\{1, \ldots, T\}$, and a masked L_1 regression objective is minimized:

$$\mathcal{L}_{\text{bridge}}^\ell = \mathbb{E}_{t,\epsilon}\left[\frac{\left\|(\hat{z}_0^\ell - z_0^\ell) \odot \overline{M}_{\text{tex}}^{\ell,K_\ell}\right\|}{\left\|\overline{M}_{\text{tex}}^{\ell,K_\ell}\right\| + \varepsilon}\right], \tag{9}$$

where $\epsilon \sim \mathcal{N}(0, I)$ is used to sample z_t^ℓ via Eq. (7), followed by the projection in Eq. (8). At inference, sampling proceeds from z_T^ℓ to z_0^ℓ over T DDIM-style steps with η=0 by default, though $\eta \in (0, 1]$ may be used to introduce stochasticity [22]. The Dirichlet projection (Eq. (8)) is reapplied after each step.

Wavelet reconstruction. After sampling, residual coefficients are reconstructed via synthesis in the learned subspace followed by de-normalization:

$$\hat{r}^\ell = s^\ell \odot \left(U^\ell \hat{z}_0^\ell\right), \qquad x^{\text{recon}} = \mathcal{W}^{-1}\left(\hat{c}^{LL}, \{h^\ell(x^{\text{struct}}) + \hat{r}^\ell\}_{\ell=1}^L\right). \tag{10}$$

The high-frequency residual $r^{\text{hf}} = (x^{\text{recon}} - x^{\text{struct}}) \odot M_{\text{fg}}$ is then computed, and $\hat{x}$ is obtained via Eq. (1).

Image refiner. We use a lightweight image refiner to absorb small pixel-domain errors caused by the low-rank residual representation and inverse DTCWT reconstruction. The refiner is a residual U-Net that takes $[\hat{x}, y]$ as input and predicts only a masked foreground correction. The sampled wavelet residuals are injected through zero-initialized FiLM adapters, and the output head is also zero-initialized so that the refiner starts as an identity mapping. To prevent the refiner from becoming a second generator, $\hat{x}$ and $\{\hat{r}^\ell\}_{\ell=1}^L$ are detached before entering the refiner. The refiner is trained with a masked L_1 and SSIM loss on the foreground region.

3 Experiments

3.1 Datasets and Implementation Details

We evaluated WRB on BraTS 2021 [2] and the Gold Atlas MRI–CT dataset [19]. For BraTS, 40 subjects were split at the case level into 25 training, 5 validation, and 10 test subjects. We extracted 100 fixed-index 2D axial slices from each subject and padded them to 256×256, yielding 2,500 training, 500 validation, and 1,000 test 2D images for each modality. These image data were used for three paired translation tasks: T1→FLAIR, T1→T2, and T2→FLAIR (Table 1).

Table 1. Quantitative comparison on BraTS 2021. Best results are marked with * and shown in bold. WRB uses T=10 steps per level (L=3). PSNR in dB, SSIM in %, MAE in 10^{-2} units. Paired Wilcoxon signed-rank tests confirm that all *-marked improvements are significant ($p < 0.05$).

	Method	PSNR (dB)↑	SSIM (%)↑	MAE (10^{-2})↓
T1→FLAIR	pix2pix [11]	23.15±2.07	82.73±3.76	6.23±2.35
	CycleGAN [30]	22.76±1.18	83.81±2.45	5.31±1.73
	BBDM [16]	23.22±1.84	83.21±4.47	6.91±1.20
	SelfRDB [1]	22.79±3.53	86.40±4.90	**3.82±3.38***
	DDPM [7]	22.88±2.24	83.87±1.87	4.28±2.74
	WRB (ours)	**24.41±3.16***	**90.36±2.09***	4.65±2.19
T1→T2	pix2pix	24.59±1.74	90.14±2.64	4.83±1.07
	CycleGAN	25.03±2.94	90.66±3.77	4.47±1.29
	BBDM	26.74±1.90	90.03±2.89	3.83±0.56
	SelfRDB	26.47±3.03	90.94±3.41	3.71±1.04
	DDPM	25.97±2.09	90.24±3.41	4.53±1.95
	WRB (ours)	**27.74±2.56***	**92.69±2.14***	**3.38±1.43***
T2→FLAIR	pix2pix	24.85±1.57	83.36±3.11	4.20±1.07
	CycleGAN	27.08±3.42	87.59±2.93	3.86±1.06
	BBDM	25.92±2.00	83.51±4.05	4.61±0.98
	SelfRDB	26.84±1.88	90.59±2.79	4.05±0.59
	DDPM	26.90±1.84	87.86±2.85	3.93±0.71
	WRB (ours)	**27.51±1.00***	**91.33±2.07***	**2.51±1.59***

For Gold Atlas, 19 patients were split at the patient level into 11 training, 4 validation, and 4 test cases. After CT-to-MR registration and preprocessing, the dataset contained 1,003 paired MRI–CT cross-sections, including 610 training, 195 validation, and 198 test paired images.

Implementation details. The framework was implemented in PyTorch 2.1.5 with CUDA 12.4 and trained on a single NVIDIA A800 GPU. Each BraTS slice was normalized to $[0, 1]$ and rescaled to $[-1, 1]$ during training. All metrics were computed within a foreground mask using a threshold of 0.5. We reported PSNR [26] in dB, SSIM [27] in %, and MAE [29] in 10^{-2} units as mean±standard deviation across test slices. MAE was computed within the foreground mask.

WRB used a 3-level DTCWT [14], a monotone spline map with K_s=8 knots, and wavelet-domain Gaussian bridges with subspace dimensions $(K_1, K_2, K_3) = (10, 8, 6)$. All components were optimized with Adam and cosine annealing, using a batch size of 8 and early stopping based on validation PSNR within 100 epochs. The image-domain refiner was a U-Net with 64 base channels and depth 4, trained with EMA at a decay rate of 0.999 using ℓ_1, SSIM, gradient, and FFT losses weighted by $(\lambda_1, \lambda_{\text{ssim}}, \lambda_{\text{grad}}, \lambda_{\text{fft}}) = (1.0, 0.5, 0.5, 0.1)$.

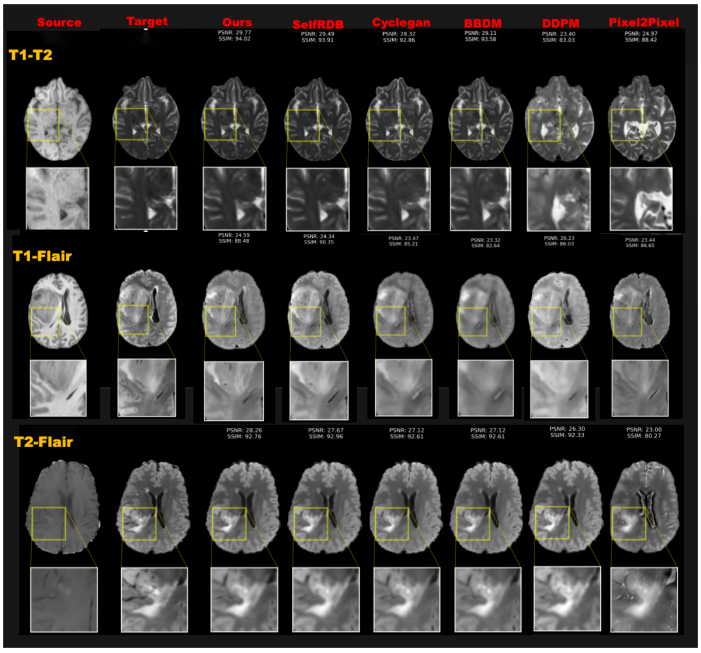

Fig. 2. Qualitative comparison on BraTS 2021 for T1→T2, T1→FLAIR, and T2→FLAIR. Columns show source, ground truth, WRB, SelfRDB, CycleGAN, BBDM, DDPM, and pix2pix. Yellow boxes highlight regions of interest.

3.2 Main Results

As shown in Table 1, WRB achieved the highest PSNR and SSIM on all three BraTS tasks. We compared WRB with pix2pix [11], CycleGAN [30], DDPM [7], BBDM [16], and SelfRDB [1] using the same datasets and evaluation protocol. The largest gains were observed on T1→FLAIR, where WRB exceeded BBDM by 1.19,dB in PSNR and SelfRDB by 3.96 percentage points in SSIM. WRB also achieved the lowest MAE on T1→T2 and T2→FLAIR, while SelfRDB performed best in MAE on T1→FLAIR.

Qualitative comparisons. Figure 2 presents visual comparisons for all three BraTS tasks. On T1→T2, DDPM and pix2pix lost fine structure, whereas WRB better preserved the lesion boundary. On T1→FLAIR, DDPM showed weaker texture recovery, and BBDM over-smoothed the lesion interior, whereas WRB retained the hyperintense rim more clearly. Table 2 reports MRI→CT results on Gold Atlas. On this dataset, WRB achieved the highest PSNR and the lowest MAE on both T1→CT and T2→CT, although it did not obtain the best SSIM in every translation direction.

We further conducted a preliminary BraTS segmentation experiment using a pre-trained nnU-Net [10]. The standard BraTS segmentation inputs were T1, T1ce, FLAIR, and either WRB-generated T2 or ground-truth T2. T1ce was

Table 2. Quantitative comparison for MRI→CT synthesis on Gold Atlas. Best results are marked with * and shown in bold. WRB uses T=10 steps per level. MAE is reported in 10^{-2} units.

Method	T1→CT			T2→CT		
	PSNR↑	SSIM↑	MAE↓	PSNR↑	SSIM↑	MAE↓
pix2pix	25.65±3.22	87.26±5.11	3.14±0.68	25.16±3.08	87.33±2.09	3.56±1.88
CycleGAN	26.08±2.87	91.00±2.35	3.20±1.06	26.88±1.96	90.18±2.03	2.97±0.64
BBDM	25.21±2.59	84.82±7.40	3.38±1.70	24.31±1.10	82.80±4.61	4.69±2.63
SelfRDB	26.55±2.13	**92.29±6.32***	2.71±0.56	27.12±2.18	90.46±2.15	3.13±2.51
DDPM	25.59±2.32	85.32±4.93	3.30±1.06	25.11±3.15	83.29±5.93	3.01±1.16
WRB(Ours)	**26.99±1.64***	90.05±2.65	**2.29±0.72***	**27.32±1.62***	**90.82±2.14***	**2.53±0.64***

Table 3. Ablation study on BraTS 2021 (T1→T2). Each row removes or replaces one component from the full model.

Variant	PSNR (dB)↑	SSIM (%)↑
(a) Pixel-space bridge (no wavelet factorization)	25.86	91.51
(b) Without subspace projection	24.99	88.83
(c) Without intensity map	24.78	87.14
(d) Without image refiner (bridge only)	27.42	92.29
(e) Without Dirichlet constraint	15.10	78.24
Full WRB	**27.74***	**92.69***

used only for the downstream segmentation experiment and was not used for translation training. The WT/TC/ET Dice scores were 0.886/0.857/0.847 using WRB-generated T2 and 0.916/0.854/0.845 using ground-truth T2. These results indicate that WRB-generated T2 images can provide useful downstream information when ground-truth T2 is unavailable in clinical practice.

3.3 Ablation Study

We evaluated the contribution of WRB on BraTS T1→T2 (Table 3). Removing the Dirichlet constraint (e) caused the largest performance degradation, reducing PSNR from 27.74,dB to 15.10,dB and SSIM from 92.69% to 78.24%. This result indicates that, without boundary anchoring, the bridge can assign large residuals to zero-valued background regions, leading to unstable synthesis. Removing the monotone intensity map (c) caused the largest stable drop, with PSNR and SSIM decreasing to 24.78,dB and 87.14%, respectively. Removing the subspace projection (b) also degraded performance, yielding 24.99,dB PSNR and 88.83% SSIM. Replacing the wavelet formulation with a pixel-space bridge (a) resulted in 25.86,dB PSNR and 91.51% SSIM. Removing the image-domain refiner (d) had the smallest effect, reducing PSNR by 0.32,dB, which suggests that the main performance gain comes from the wavelet residual bridge rather than the final refinement module.

4 Conclusion

In this paper, we proposed WRB (Wavelet Residual Bridge), which decomposes cross-contrast MRI synthesis into deterministic low-frequency mapping, stochastic wavelet-subspace bridges, and image-domain refinement. On BraTS 2021 and Gold Atlas, WRB achieves the highest PSNR on all five translation tasks and the highest SSIM on four of five. A limitation is that the current study focuses on 2D slice-wise image translation using a central-slice protocol, with fixed axial slices [27,127) selected to cover tumor voxels in the BraTS subset. Although the low-pass transport, high-frequency residual bridges, and DTCWT subspace projection are not inherently limited to 2D, extending the method to volumetric synthesis will require 3D DTCWT representations and 3D U-Net denoisers. Future work will also evaluate cross-site generalization on multi-center datasets.

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

References

1. Arslan, F., Kabas, B., Dalmaz, O., Ozbey, M., Çukur, T.: Self-consistent recursive diffusion bridge for medical image translation. Medical Image Analysis **106**, 103747 (2025). https://doi.org/10.1016/j.media.2025.103747
2. Baid, U., Ghodasara, S., Mohan, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. CoRR **abs/2107.02314** (2021), https://arxiv.org/abs/2107.02314
3. Boulanger, M., Nunes, J.C., Chourak, H., et al.: Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. Physica Medica **89**, 265–281 (2021). https://doi.org/10.1016/j.ejmp.2021.07.027
4. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. In: Advances in Neural Information Processing Systems (NeurIPS) (2021)
5. Ellingson, B.M., Bendszus, M., Boxerman, J., Barboriak, D., Erickson, B.J., Smits, M., et al.: Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. Neuro-Oncology **17**(9), 1188–1198 (2015). https://doi.org/10.1093/neuonc/nov095
6. Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R.S., Brendel, W., Bethge, M., Wichmann, F.A.: Shortcut learning in deep neural networks. Nature Machine Intelligence **2**(11), 665–673 (2020). https://doi.org/10.1038/s42256-020-00257-z
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 33, pp. 6840–6851 (2020)
8. Huynh, T., Gao, Y., Kang, J., Wang, L., Zhang, P., Lian, J., Shen, D.: Estimating CT image from MRI data using structured random forest and auto-context model. IEEE Transactions on Medical Imaging **35**(1), 174–183 (2016). https://doi.org/10.1109/TMI.2015.2461533
9. Iglesias, J.E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B.: Is synthesizing MRI contrast useful for inter-modality analysis? In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. LNCS, vol. 8149, pp. 631–638. Springer (2013). https://doi.org/10.1007/978-3-642-40811-3_79

10. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5967–5976 (2017). https://doi.org/10.1109/CVPR.2017.632
12. Jog, A., Carass, A., Roy, S., Pham, D.L., Prince, J.L.: MR image synthesis by contrast learning on neighborhood ensembles. Medical Image Analysis **24**(1), 63–76 (2015). https://doi.org/10.1016/j.media.2015.05.002
13. Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., et al.: Denoising diffusion probabilistic models for 3D medical image generation. Scientific Reports **13**, 7303 (2023). https://doi.org/10.1038/s41598-023-34341-2
14. Kingsbury, N.: Complex wavelets for shift invariant analysis and filtering of signals. Applied and Computational Harmonic Analysis **10**(3), 234–253 (2001). https://doi.org/10.1006/acha.2000.0343
15. Lei, Y., Harms, J., Wang, T., et al.: MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Medical Physics **46**(8), 3565–3581 (2019). https://doi.org/10.1002/mp.13617
16. 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 (CVPR). pp. 1952–1961 (2023). https://doi.org/10.1109/CVPR52729.2023.00194
17. Müller-Franzes, G., Niehues, J.M., Khader, F., et al.: A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. Scientific Reports **13**, 12098 (2023). https://doi.org/10.1038/s41598-023-39278-0
18. Niyazi, M., Brada, M., Chalmers, A.J., Combs, S.E., Erridge, S.C., Fiorentino, A., Grosu, A.L., Lagerwaard, F.J., Minniti, G., Mirimanoff, R.O., Ricardi, U., Short, S.C., Weber, D.C., Belka, C.: ESTRO-ACROP guideline "target delineation of glioblastomas". Radiotherapy and Oncology **118**(1), 35–42 (2016). https://doi.org/10.1016/j.radonc.2015.12.003
19. Nyholm, T., Svensson, S., Andersson, S., Jonsson, J., Sohlin, M., Gustavsson, C., Kjellén, E., Albertsson, P., Blomqvist, L., Zackrisson, B., Olsson, L.E., Gunnlaugsson, A.: Gold atlas – male pelvis – gentle radiotherapy (May 2017). https://doi.org/10.5281/zenodo.583096, dataset, version v1
20. Saharia, C., Chan, W., Chang, H., Lee, C.A., Ho, J., Salimans, T., Fleet, D.J., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings (2022). https://doi.org/10.1145/3528233.3530757
21. Shi, Y., De Bortoli, V., Campbell, A., Doucet, A.: Diffusion schrödinger bridge matching. In: Advances in Neural Information Processing Systems (NeurIPS) (2023)
22. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (ICLR) (2021)
23. 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 (ICLR) (2021)
24. Spadea, M.F., Maspero, M., Zaffino, P., Seco, J.: Deep learning based synthetic-CT generation in radiotherapy and PET: A review. Medical Physics **48**(11), 6537–6566 (2021). https://doi.org/10.1002/mp.15150

25. Thust, S.C., Heiland, S., Falini, A., Jäger, H.R., Waldman, A.D., Sundgren, P.C., et al.: Glioma imaging in europe: A survey of 220 centres and recommendations for best clinical practice. European Radiology **28**(8), 3306–3317 (2018). https://doi.org/10.1007/s00330-018-5314-5
26. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE Signal Processing Magazine **26**(1), 98–117 (Jan 2009). https://doi.org/10.1109/MSP.2008.930649
27. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (Apr 2004). https://doi.org/10.1109/TIP.2003.819861
28. Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., et al.: Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group. Journal of Clinical Oncology **28**(11), 1963–1972 (2010). https://doi.org/10.1200/JCO.2009.26.3541
29. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research **30**, 79–82 (2005). https://doi.org/10.3354/cr030079
30. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 2223–2232 (2017). https://doi.org/10.1109/ICCV.2017.244