

Heterogeneity-Adaptive Diffusion Schrödinger Bridge for PET-Guided Whole-Body MRI Translation

Chengbo Wang^{1†}, Jiacheng Yu^{1†}, Linjie Bian^{2†}, Ming Qi², Xiaosheng Liu², Tongtong Che³, Jichang Zhang³, Shuyu Li³, and Shaoli Song^{2*} and Xiuying Wang^{1*}

¹ The University of Sydney, Sydney, Australia

² Department of Nuclear Medicine, Fudan University Shanghai Cancer Center, Shanghai, China

³ State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

Abstract. While whole-body multimodal medical imaging scanners have been increasingly recognized for more effective medical applications, the excessive long acquisition time in PET-MR scanning is a major obstacle in more efficient clinical practice. Deep learning-based MRI translation provides a potential solution to reduce scan duration. However, current models often focus on specific anatomical regions and face challenges for whole-body scans that consists of highly heterogeneous feature distributions mainly due to (1) different anatomical regions across whole-body, and (2) lesions or pathological tissues. This paper tackles the challenges through a novel Heterogeneity-Adaptive Diffusion Schrödinger Bridge (HA-DSB) framework. By explicitly modeling translation as stochastic transport between source and target distributions, HA-DSB incorporates region context embeddings derived from a vision-language model (VLM) to enable region-specific modeling. To enhance fidelity of the pathological tissue, lesion-aware metabolic prior from PET is integrated directly into the bridge dynamics through a dual-stage guidance mechanism. Specifically, a PET-guided noise modulation module adaptively scales spatial diffusion perturbations during the forward process, while PET features are leveraged during the reverse process to selectively amplify lesion-relevant structures via an attention mechanism. Experiments demonstrate the superiority of our method across different body regions in whole-body MRI translation and show improved translation quality in lesion areas under PET guidance. Our code is available at <https://github.com/xyw-medical-research/HADSB>.

Keywords: Medical Image Translation · PET/MR Imaging · Diffusion Schrödinger Bridge · Whole-body MR Translation

[†] contributed equally to this work.

^{*} corresponding authors. xiu.wang@sydney.edu.au, shaoli_song@fudan.edu.cn.

1 Introduction

Hybrid positron emission tomography combined with magnetic resonance imaging (PET/MR) is an emerging imaging modality that enables simultaneous acquisition of metabolic and high-resolution anatomical information, producing intrinsically co-registered PET and multi-sequence MR images while reducing radiation exposure compared with PET/CT [1]. Whole-body PET/MR supports comprehensive multi-bed imaging from the head and neck to the mid-thigh and has become increasingly important in clinical applications such as cancer metastasis evaluation [2]. However, whole-body PET/MR scan time typically ranges from 60–90 minutes, largely due to multiple MR sequences, which limits clinical efficiency and increases patient discomfort. Deep learning-based MR translation offers a practical solution by synthesizing time-consuming modalities from fast-acquired sequences, and recent diffusion-based approaches have demonstrated promising performance in MR image translation [3–6].

However, most MR translation focus on specific anatomical regions such as the brain in public datasets [7], and extending these models to whole-body MR translation in integrated PET/MR settings remains challenging. Whole-body scans exhibit strong cross-region heterogeneity, with substantial variations in tissue composition and signal characteristics across anatomical regions [8]. Such non-stationary and region-dependent intensity distributions significantly increase the difficulty of learning a unified modality mapping [9, 10]. For example, GANs are known to suffer from mode collapse in multi-modal distributions [11, 12], while diffusion models may be biased toward dominant feature modes [13]. Without explicit region conditioning, these models may struggle to capture region-specific structures in the whole body MR translation.

Another critical yet underexplored challenge in MR translation is the accurate reconstruction of pathological regions [14, 15], which are often of primary clinical importance. Compared to normal tissues, lesions exhibit highly heterogeneous appearances and complex cross-modality signal relationships. Most existing models do not explicitly differentiate pathological changes from normal tissues, causing lesion signals to be biased toward the dominant healthy distribution and leading to degraded pathological fidelity.

In this work, we propose a Heterogeneity-Adaptive Diffusion Schrödinger Bridge (HA-DSB) with PET guidance for whole-body MR translation (Fig. 1). Unlike conventional diffusion-based translation methods that rely on implicit noise prediction, HA-DSB explicitly optimizes a stochastic transport plan between source and target MR distributions [3, 16]. This explicit transport formulation is better suited for whole-body MR translation, where strong cross-region domain shifts make implicit alignment prone to over-smoothing and modality ambiguity. To address the pronounced region-dependent distribution shifts in whole-body MR translation, we further introduce region context embeddings derived from structured descriptions generated by the large vision-language model and fuse them with diffusion time embeddings, enabling fine-grained, region-specific bridge modeling and consistently improving translation performance across anatomical regions. Furthermore, to improve pathological fidelity, we ex-

exploit PET as a lesion-aware prior, which is naturally available in integrated PET/MR due to intrinsic co-registration. Since lesions typically manifest as focal high-uptake signals in PET, we introduce a PET-guided noise modulation mechanism that applies spatially adaptive corruption in the forward process. This spatially adaptive perturbation redistributes the stochastic transport difficulty, encouraging the model to learn pathology-sensitive mappings rather than being biased toward dominant healthy structures. A multi-scale PET-guided attention module is further incorporated into the reverse process to amplify lesion-relevant features, improving translation fidelity in pathological areas. Our main contributions include:

- We propose a Heterogeneity-Adaptive Diffusion Schrödinger Bridge (HADSDB) that models whole-body MR translation as an explicit distribution transport problem, achieving more accurate and stable cross-modality mapping across anatomically diverse regions compared to conventional noise-prediction-based diffusion models.
- We introduce region context embeddings generated by a large vision-language model and fuse them with diffusion time embeddings as conditioning signals, enabling the bridge to adapt its region-specific transport dynamics.
- We design a PET-guided lesion enhancement mechanism in which the forward corruption process is spatially modulated by PET uptake and region context, while multi-scale PET-aware attention improves lesion-relevant feature recovery during reverse denoising.

2 Methodology

2.1 Diffusion Schrödinger Bridge.

Diffusion Schrödinger bridge (DSB) [17, 18] constructs a stochastic process that transports samples between two endpoint distributions. In image-to-image translation, the bridge provides an endpoint-to-endpoint alternative to the usual Gaussian-prior diffusion models. The forward and backward process can be described as:

$$dX_t = [f + g^2 \nabla_x \log \Psi(t, X_t)] dt + g dW_t, \quad dX_t = [f - g^2 \nabla_x \log \hat{\Psi}(t, X_t)] dt + g dW_t. \quad (1)$$

where X_0 and X_1 denote the source and endpoint states, and Ψ and $\hat{\Psi}$ are time-dependent Schrödinger potentials [17]. We adopt the I²SB instantiation proposed in [18], assuming paired samples (X_0, X_1) drawn from the joint distribution $p(X_0, X_1) = p_A(X_0)p_B(X_1 | X_0)$ and setting the drift term $f = 0$. Under this formulation, the Schrödinger Bridge can be efficiently trained and inferred within the score-based generative modeling (SGM) framework. The analytic posterior of DSB given a boundary pair (X_0, X_1) can be expressed as $q(X_t | X_0, X_1) = \mathcal{N}(X_t; \mu_t(X_0, X_1), \Sigma_t)$.

$$\mu_t = \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} X_0 + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} X_1, \quad \Sigma_t = \frac{\sigma_t^2 \bar{\sigma}_t^2}{\sigma_t^2 + \bar{\sigma}_t^2} \mathbf{I} \quad (2)$$

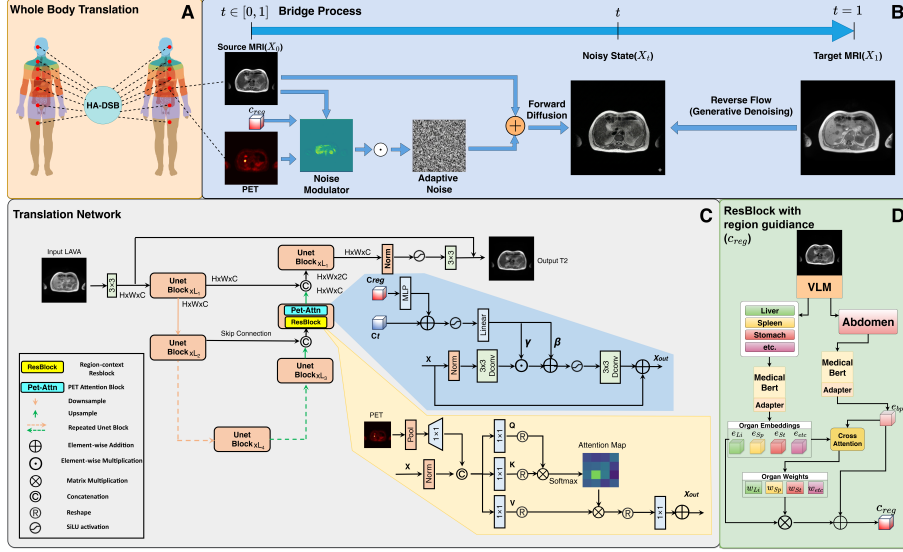


Fig. 1. Overview of HA-DSB. (a) Whole-body MR translation setting. (b) Diffusion Schrödinger bridge from LAVA to T2 with noise modulation jointly conditioned on PET and region context c_{reg} . (c) UNet translation backbone with region-conditioned residual blocks and multi-scale PET attention. (d) Construction of the region context embedding c_{reg} .

where $\sigma_t^2 := \int_0^t \beta_\tau d\tau$ and $\bar{\sigma}_t^2 := \int_t^1 \beta_\tau d\tau$ denote the variances accumulated on the two sides of time t . Unlike conventional image-to-image Schrödinger Bridge models, our method adopts a heterogeneity-adaptive conditional formulation that integrates region context embedding and PET-derived cues. The final training objective is defined as follows, where c represents the conditioning information.

$$\mathcal{L}(\theta) = \mathbb{E} \left[\left\| \varepsilon_\theta(X_t, c, t) - \frac{X_t - X_0}{\sigma_t} \right\|_2^2 \right]. \quad (3)$$

2.2 Region context embedding guidance.

To address anatomical heterogeneity in whole-body translation, we introduce a region context embedding $c_{reg} \in \mathbb{R}^d$ that adapts the bridge dynamics to region-dependent structural variability. For each slice, we assign a body-location label l_{bp} that partitions the body into $K = 11$ regions. A frozen PubMedBERT [19] encodes these labels into semantic features, which are linearly projected and combined with learnable embeddings to obtain the body-location representation $e_{bp} \in \mathbb{R}^d$.

To capture intra-region variability, each slice is additionally associated with a set of organ labels, which are encoded by the same frozen encoder to produce organ embeddings $\{e_{org}^i\}_{i=1}^M$, where $e_{org}^i \in \mathbb{R}^d$ and M is the number of organs

present in the slice. We fuse body-location and organ information via multi-head cross-attention, where $\mathbf{e}_{\text{bp}}$ serves as the query and organ embeddings act as the keys and values. The organ information is then added to the body-location representation as a residual correction, forming the final region context embedding $\mathbf{c}_{\text{reg}}$:

$$\alpha_i = \text{Softmax} \left(\frac{(\mathbf{W}_Q \mathbf{e}_{\text{bp}})^\top (\mathbf{W}_K \mathbf{e}_{\text{org}}^i)}{\sqrt{d_h}} \right), \quad \mathbf{e}_{\text{attn}} = \sum_{i=1}^M \alpha_i \mathbf{W}_V \mathbf{e}_{\text{org}}^i, \quad (4)$$

$$\mathbf{c}_{\text{reg}} = \mathbf{e}_{\text{bp}} + \text{MLP}(\text{LayerNorm}(\mathbf{e}_{\text{attn}})), \quad (5)$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{d_h \times d}$ are learnable projections, with $d_h = d/H$ and $H = 4$. The resulting $\mathbf{c}_{\text{reg}}$ encodes both coarse body-location and fine-grained organ information. It is added to the diffusion timestep embedding $\mathbf{e}_t \in \mathbb{R}^d$ and passed through an MLP to generate the scale and shift parameters in each residual block.

To avoid manual annotation, we use Google Gemini 3 Pro to generate body-location and organ labels. The reliability of body-location labels was empirically verified, achieving approximately 96% accuracy under professional validation. For organ labels, we constrain the VLM outputs to a predefined vocabulary of major MRI-visible organs to improve reliability and avoid unconstrained predictions.

2.3 PET-guided Noise Modulation

Standard diffusion bridges inject spatially uniform noise across all spatial locations [18], which is suboptimal for whole-body PET/MR translation because anatomical complexity and lesion-related variability are highly heterogeneous across regions. To address this, we introduce a PET-guided noise modulation mechanism that jointly conditions the forward corruption process on PET uptake and the region context embedding $\mathbf{c}_{\text{reg}}$. Specifically, we use a lightweight noise modulator $\mathcal{M}_\phi$ to predict a spatial scaling map $\mathbf{S} \in [s_{\min}, s_{\max}]^{H \times W}$. The modulator extracts spatial features from PET using a two-layer convolutional encoder, incorporates anatomical context via FiLM [20] conditioning with $\mathbf{c}_{\text{reg}}$, and outputs per-pixel scaling factors. We then replace the standard uniform-noise bridge forward process

$$\mathbf{x}_t = \mu_t(X_0, X_1) + \Sigma_t^{1/2} \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (6)$$

with the spatially modulated form, where $\odot$ denotes element-wise multiplication.

$$\mathbf{x}_t = \mu_t(X_0, X_1) + \Sigma_t^{1/2} (\mathbf{S} \odot \boldsymbol{\epsilon}), \quad (7)$$

This yields a heterogeneity-adaptive forward corruption process, allowing anatomically and pathologically complex regions to receive different perturbation strengths.

Table 1. Per-region quantitative results on the testset (mean $\pm$ std). Each cell shows SSIM (% , top) and PSNR (dB, bottom).

Method	Head/Neck <i>n</i> =275	Thorax <i>n</i> =498	Abdomen <i>n</i> =617	Pelvis/Hips <i>n</i> =472	Thighs <i>n</i> =225	Avg <i>n</i> =2,087
<i>General translation models</i>						
Pix2Pix[22]	86.5 $\pm$ 6.3 22.77 $\pm$ 1.64	79.9 $\pm$ 4.5 22.31 $\pm$ 1.89	80.9 $\pm$ 5.5 21.58 $\pm$ 1.75	83.5 $\pm$ 5.0 22.01 $\pm$ 2.12	85.7 $\pm$ 3.7 23.52 $\pm$ 1.93	82.5 $\pm$ 5.6 22.22 $\pm$ 1.97
Palette[23]	89.6 $\pm$ 6.9 25.83 $\pm$ 4.50	85.4 $\pm$ 4.0 23.83 $\pm$ 2.86	85.6 $\pm$ 5.3 23.61 $\pm$ 2.90	88.1 $\pm$ 5.1 24.59 $\pm$ 3.09	89.8 $\pm$ 4.2 25.44 $\pm$ 4.03	87.1 $\pm$ 5.4 24.37 $\pm$ 3.41
I^2SB [18]	88.0 $\pm$ 6.2 25.81 $\pm$ 1.91	82.2 $\pm$ 3.8 21.67 $\pm$ 1.74	84.0 $\pm$ 3.6 21.50 $\pm$ 1.31	85.5 $\pm$ 2.9 23.06 $\pm$ 1.27	85.7 $\pm$ 2.7 24.38 $\pm$ 2.13	84.6 $\pm$ 4.3 22.77 $\pm$ 2.20
<i>Medical translation models</i>						
PPT[24]	89.6 $\pm$ 4.4 23.91 $\pm$ 1.44	81.1 $\pm$ 3.5 21.42 $\pm$ 1.21	79.5 $\pm$ 4.1 20.43 $\pm$ 1.60	79.1 $\pm$ 3.2 20.93 $\pm$ 1.59	76.5 $\pm$ 2.8 22.37 $\pm$ 2.09	80.8 $\pm$ 5.2 21.45 $\pm$ 1.92
SelfRDB[6]	84.9 $\pm$ 7.2 22.12 $\pm$ 2.59	84.6 $\pm$ 7.1 22.02 $\pm$ 2.70	84.7 $\pm$ 6.6 21.95 $\pm$ 2.63	84.9 $\pm$ 6.5 22.10 $\pm$ 2.65	85.7 $\pm$ 6.4 22.27 $\pm$ 2.06	84.9 $\pm$ 6.8 22.06 $\pm$ 2.59
ResViT[5]	91.0 $\pm$ 5.5 26.35 $\pm$ 2.13	85.4 $\pm$ 3.8 23.76 $\pm$ 2.06	85.3 $\pm$ 4.1 22.89 $\pm$ 1.93	87.3 $\pm$ 4.5 23.74 $\pm$ 2.42	86.0 $\pm$ 3.3 24.83 $\pm$ 2.16	86.6 $\pm$ 4.7 23.95 $\pm$ 2.39
<i>Our method</i>						
HA-DSB(no PET)	94.0 [†] $\pm$ 3.1 28.59 [†] $\pm$ 2.36	89.0 [†] $\pm$ 4.3 24.87 [†] $\pm$ 2.44	88.1 [†] $\pm$ 4.8 24.06 [†] $\pm$ 2.81	93.1 [†] $\pm$ 3.9 27.42 [†] $\pm$ 2.95	93.3 [†] $\pm$ 3.0 28.51 [†] $\pm$ 2.44	90.8 [†] $\pm$ 4.8 26.09 [†] $\pm$ 3.24
HA-DSB(PET)	94.5 $\pm$ 3.5 28.88 $\pm$ 2.48	89.5 $\pm$ 4.2 24.92 $\pm$ 2.46	88.5 $\pm$ 4.8 24.15 $\pm$ 2.71	93.2 $\pm$ 4.0 27.59 $\pm$ 3.23	93.6 $\pm$ 3.2 28.89 $\pm$ 2.49	91.1 $\pm$ 4.8 26.25 $\pm$ 3.35

[†] Second-best result (SSIM and PSNR evaluated separately).

2.4 Multi-Scale PET Attention

Because the forward process applies spatially non-uniform corruption guided by PET and anatomical context, the reverse model must recover lesion-relevant details under the same setting. To this end, we introduce PET-aware self-attention at multiple UNet resolutions. For each feature map $\mathbf{x} \in \mathbb{R}^{B \times C \times H \times W}$, the PET image is aligned by adaptive average pooling and projected by a 1×1 convolution to obtain $\mathbf{p}'$. We concatenate $\mathbf{x}$ and $\mathbf{p}'$ to form $\mathbf{z} = [\mathbf{x}; \mathbf{p}']$, and apply PET-aware self-attention:

$$\mathbf{Q} = W_Q \tilde{\mathbf{z}}, \mathbf{K} = W_K \tilde{\mathbf{z}}, \mathbf{V} = W_V \tilde{\mathbf{z}}, \quad \mathbf{x}_{\text{out}} = \mathbf{x} + W_{\text{out}} \left[\text{Softmax} \left(\frac{\mathbf{QK}^\top}{\sqrt{d_h}} \right) \mathbf{V} \right], \quad (8)$$

where $\tilde{\mathbf{z}} \in \mathbb{R}^{B \times HW \times 2C}$ is the flattened form of $\mathbf{z}$. We use a zero-initialized W_{out} so that the block starts as an identity mapping and gradually incorporates PET-guided lesion cues during training.

3 Experiments

Dataset. All PET/MR examinations were performed on an integrated hybrid PET/MR system (SIGNATM PET/MR, GE HealthCare). This platform integrates a 3.0-T MRI scanner with time-of-flight PET capability. We prospectively collect whole-body imaging data from 246 subjects at an anonymous hospital.

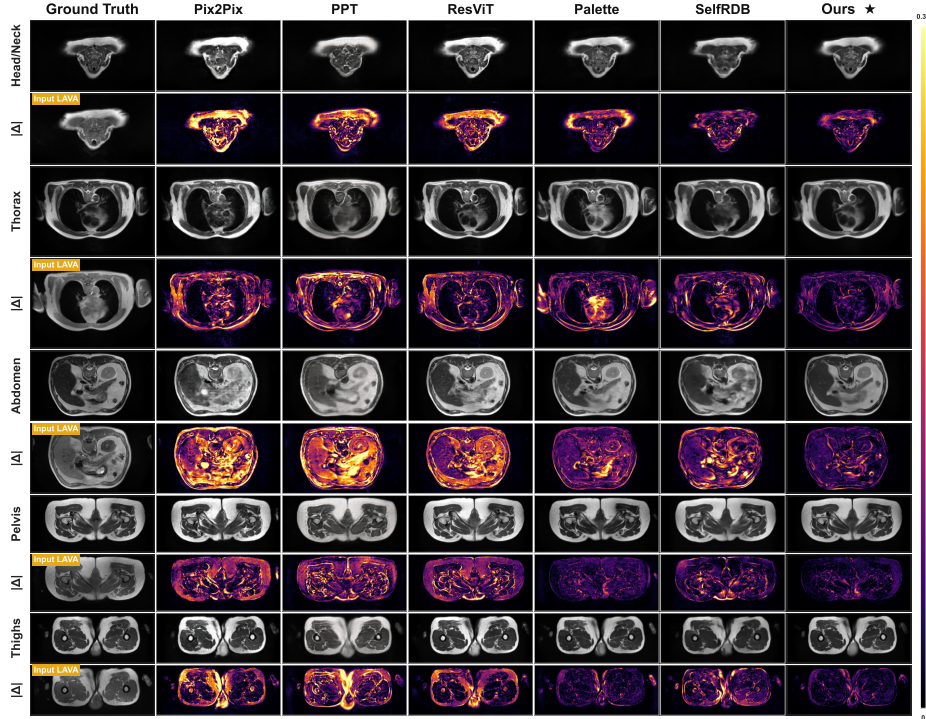


Fig. 2. Qualitative comparison across five anatomical regions. Top row: translated T2w results; bottom row: absolute error maps $|\hat{\mathbf{x}} - \mathbf{x}_{\text{GT}}|$.

Each subject contains three co-registered modalities (LAVA, T2, and PET), all resampled to 256×256 resolution. The dataset was split into a training set of 204 subjects (20,306 slices) and a held-out testset of 42 subjects (4,162 slices). The testset comprises 21 randomly sampled subjects and an additional 21 subjects with clinically confirmed lesions, allowing us to specifically assess synthesis fidelity in pathological regions.

Implementation Details. For the denoising network, we employ an ADM-style U-Net [21] featuring 128 base channels, channel multipliers of $[1, 1, 2, 2, 4, 4]$, and three residual blocks per stage. PET attention blocks are integrated at resolutions of 32×32 , 16×16 , and 8×8 within both the encoder and decoder. The diffusion bridge uses $T = 200$ training discretization steps with $\beta_{\text{max}} = 0.35$ and a symmetric variance schedule. Training was conducted on two NVIDIA RTX 4090 GPUs for 50k iterations. We utilized the AdamW optimizer with a learning rate of 10^{-4} , a total batch size of 16, and an EMA decay of 0.9998.

Comparison with SOTA methods. We compare HA-DSB against five representative baselines from GAN-based, diffusion-based, and bridge paradigms, all trained on whole-body data. As shown in Table 1, HA-DSB achieves the best average SSIM and PSNR, outperforming the strongest baseline, ResViT. Com-

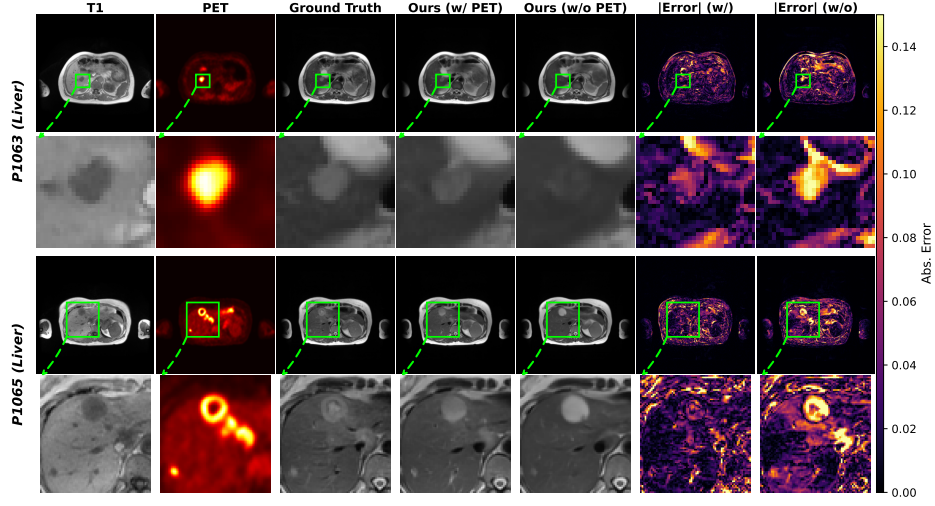


Fig. 3. Effect of PET guidance on pathological cases (liver lesions). From left to right: T1, PET, GT, ours (w/ PET), ours (w/o PET), and their absolute error maps; zoomed views correspond to the green boxes (shared color scale).

pared with the basic diffusion bridge baseline I²SB, HA-DSB achieves consistent improvements across all anatomical regions, clearly demonstrating the effectiveness of the proposed region-aware conditioning mechanism. Simpler regions, such as Head/Neck, Pelvis/Hips, and Thighs, achieve SSIM above 93%, while structurally complex regions such as the Abdomen also show clear gains. These results suggest that region guidance helps the bridge learn region-dependent transport paths. Qualitative comparisons in Fig. 2 further support this observation.

Effect of PET guidance. PET guidance (HA-DSB(pet) vs. HA-DSB(no pet)) brings only modest gains in overall SSIM and PSNR on the testset (21 patients, 2,087 slices). This is likely because lesion regions occupy only a small fraction of the total volume, and their improvements are diluted by the dominant healthy tissue where PET provides limited additional information. To better isolate the impact of PET guidance on pathological regions, we further evaluate performance on a separate lesion-confirmed cohort of 21 patients. A radiologist identified slices covering disease-bearing organs, resulting in 526 lesion-containing slices out of 2,075 (Table 2). On this pathological subset, PET guidance yields

Table 2. Effect of PET guidance on pathological cases. Δ denotes improvement from PET guidance. Best in **bold**.

Variant	PSNR (dB)	SSIM (%)
HA-DSB(no pet)	23.46 $\pm$ 2.25	86.4 $\pm$ 3.9
HA-DSB(pet)	24.26 $\pm$ 2.67	88.1 $\pm$ 3.9
Δ_{PET}	+0.80 (vs +0.16)	+1.7 (vs +0.3)

markedly larger improvements in both PSNR and SSIM than on the testset, indicating that the noise modulation and multi-scale attention primarily enhance lesion-related regions. Fig. 3 further illustrates two representative cases.

4 Conclusion

In this work, we proposed HA-DSB for PET-guided whole-body MR translation. By combining region context embeddings, PET-guided noise modulation, and multi-scale PET-aware attention, HA-DSB addresses anatomical and pathological heterogeneity in whole-body PET/MR. Experiments show improved translation quality across body regions, with additional gains in lesion-containing areas. These results indicate that heterogeneity-adaptive conditioning is effective for modeling region-dependent structural variability and recovering lesion-relevant details in challenging whole-body settings. More broadly, HA-DSB provides a practical bridge-based framework for robust whole-body medical image translation.

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

References

1. Masselli, G., Di Bella, C.: Will PET/MR Imaging Replace PET/CT for Pediatric Applications? *Diagnostics* 15(9), 1070 (2025)
2. Catalano, O.A.: *Clinical PET/MRI*. Academic Press, Cambridge (2022)
3. Özbey, M., et al.: Unsupervised Medical Image Translation with Adversarial Diffusion Models. *IEEE Transactions on Medical Imaging* 42(12), 3524–3539 (2023)
4. Xing, Z., et al.: Cross-Conditioned Diffusion Model for Medical Image to Image Translation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2024*. Springer, Cham (2024)
5. Dalmaz, O., Yurt, M., Cukur, T.: ResViT: Residual Vision Transformers for Multimodal Medical Image Synthesis. *IEEE Transactions on Medical Imaging* 41(10), 2598–2614 (2022). <https://doi.org/10.1109/TMI.2022.3167808>
6. Arslan, F., Kabas, B., Dalmaz, O., Ozbey, M., Cukur, 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>
7. Jiang, L., et al.: Cola-Diff: Conditional Latent Diffusion Model for Multi-Modal MRI Synthesis. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2023*. Springer, Cham (2023)
8. Dzyubachyk, O., et al.: Joint Intensity Inhomogeneity Correction for Whole-Body MR Data. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Springer, Berlin, Heidelberg (2013)
9. Zheng, J., et al.: Structure-Guided MR-to-CT Synthesis with Spatial and Semantic Alignments for Attenuation Correction of Whole-Body PET/MR Imaging. *Medical Image Analysis* 103, 103622 (2025)
10. Zheng, J., et al.: 3D Wavelet Latent Diffusion Model for Whole-Body MR-to-CT Modality Translation. *arXiv preprint arXiv:2507.11557* (2025)

11. Gong, Y., et al.: Testing Generated Distributions in GANs to Penalize Mode Collapse. In: Proceedings of the International Conference on Artificial Intelligence and Statistics. PMLR (2024)
12. Luo, Y., Yang, Z.: DynGAN: Solving Mode Collapse in GANs with Dynamic Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46(8), 5493–5503 (2024)
13. Um, S., Lee, S., Ye, J.C.: Don’t Play Favorites: Minority Guidance for Diffusion Models. In: International Conference on Learning Representations, ICLR (2023)
14. Zhang, X., et al.: Structure-Aware MRI Translation: Multi-Modal Latent Diffusion Model with Arbitrary Missing Modalities. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2025. Springer, Cham (2025)
15. Dayarathna, S., et al.: MU-Diff: A Mutual Learning Diffusion Model for Synthetic MRI with Application for Brain Lesions. *npj Artificial Intelligence* 1(1), 11 (2025)
16. Güngör, A., et al.: Adaptive Diffusion Priors for Accelerated MRI Reconstruction. *Medical Image Analysis* 88, 102872 (2023)
17. Chen, T., Liu, G.H., Theodorou, E.: Likelihood Training of Schrödinger Bridge Using Forward-Backward SDEs Theory. In: International Conference on Learning Representations, ICLR (2022)
18. Liu, G.H., et al.: I²SB: Image-to-Image Schrödinger Bridge. In: Proceedings of the 40th International Conference on Machine Learning. PMLR (2023)
19. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., Poon, H.: Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Transactions on Computing for Healthcare* 3(1), 1–23 (2022). <https://doi.org/10.1145/3458754>
20. Perez, E., Strub, F., de Vries, H., Dumoulin, V., Courville, A.: FiLM: Visual Reasoning with a General Conditioning Layer. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, pp. 3942–3951 (2018). <https://doi.org/10.1609/aaai.v32i1.11671>
21. Dhariwal, P., Nichol, A.: Diffusion Models Beat GANs on Image Synthesis. In: Advances in Neural Information Processing Systems, vol. 34, pp. 8780–8794 (2021)
22. 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, pp. 1125–1134 (2017). <https://doi.org/10.1109/CVPR.2017.632>
23. 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). <https://doi.org/10.1145/3528233.3530757>
24. Zhang, W., Hui, T.H., Tse, P.Y., Hill, F., Lau, C., Li, X.: High-Resolution Medical Image Translation via Patch Alignment-Based Bidirectional Contrastive Learning. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2024. LNCS, vol. 15004, pp. 178–188. Springer, Cham (2024)