TGH-DB: Template-Guided Heteroscedastic Diffusion Bridge for Brain MRI-to-PET Synthesis

Haowen Pang[1,2], Yitao Zhu[2], Yingji Fu[2], Chuyang Ye[1(✉)], and Anqi Qiu[2,3,4(✉)]

[1] School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China
chuyang.ye@bit.edu.cn
[2] Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong, China
[3] Mental Health Research Centre, Hong Kong Polytechnic University, Hong Kong, China
[4] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, USA
an-qi.qiu@polyu.edu.hk

Abstract. Amyloid *Positron Emission Tomography* (PET) plays a critical role in the early diagnosis and staging of neurodegenerative disorders; however, its widespread clinical use is constrained by high costs, radiation exposure, and limited accessibility. Synthesizing amyloid PET from routine structural *Magnetic Resonance Imaging* (MRI) represents a promising alternative, yet existing methods often yield unreliable, physiologically inconsistent results due to the lack of prior knowledge about amyloid plaque distribution. To address the problem, this work explores population-level amyloid pathology priors with explicit modeling of spatially varying voxel-wise variability and proposes a fully 3D *Template-Guided Heteroscedastic Diffusion Bridge* (TGH-DB) for brain MRI-to-PET synthesis. The proposed framework introduces a population-level amyloid pathology template, defined by voxel-wise mean and standard deviation estimated from the training cohort, to steer the diffusion process toward an empirically plausible target distribution. Unlike conventional diffusion models that drift toward an isotropic Gaussian prior, our forward process progressively guides PET signals toward the population-level amyloid pathology template with injection of spatially adaptive, heteroscedastic noise, thereby explicitly modeling regional variability in amyloid uptake. To effectively leverage structural information from various MRI modalities, we further develop the *Time-Aware Evolutionary Denoiser* (TAED) that integrates multimodal MRI information. TAED comprises parallel encoders, *Time-Aware Fusion* (TAF) modules, and a time-modulated decoding pathway so that different types of MRI information are adaptively exploited during the diffusion process. Extensive experiments on two public datasets demonstrate that the proposed method consistently outperforms representative existing approaches in both quantitative metrics and qualitative visual evaluations.

Keywords: Amyloid PET · MRI · Image synthesis · Diffusion model.

1 Introduction

Amyloid *Positron Emission Tomography* (PET) plays a critical role in the early diagnosis and staging of *Alzheimer's Disease* (AD) and related neurodegenerative disorders [5]. However, widespread clinical adoption of PET is limited by high acquisition costs, exposure to ionizing radiation, and restricted accessibility, particularly in large-scale screening and longitudinal follow-up settings [16]. In contrast, *Magnetic Resonance Imaging* (MRI) is routinely acquired in clinical practice and provides rich structural information without radiation risk [28]. These factors have motivated increasing interest in synthesizing PET images from MRI as a safe and cost-effective alternative for estimating amyloid burden [16].

Various MRI-to-PET synthesis methods have been proposed. Early MRI-to-PET synthesis approaches predominantly rely on encoder-decoder networks [22]. *Generative Adversarial Networks* (GANs) have also been widely adopted for MRI-to-PET synthesis [8,10,13,18,26,30]. More recently, diffusion models have shown better stability and synthesis fidelity in medical image synthesis than conventional generative models by modeling image generation as a continuous stochastic or deterministic dynamic process [1,14,27,29] .

Despite these advancements, MRI-to-PET synthesis remains inherently ill-posed due to the large cross-modal information gap [14]. While MRI structurally delineates anatomical morphology, amyloid PET reflects protein accumulation, which often exhibits complex nonlocal correlations with anatomy [3]. In particular, current diffusion models rely on a forward process that converges to non-informative isotropic Gaussian noise [7], which is sub-optimal for PET synthesis. This is because brain images across subjects share a high degree of topological consensus, which implicitly constrains the spatial organization of molecular pathology patterns. Forcing the model to diffuse toward pure noise discards this anatomy-informed amyloid pathology prior, increasing the risk of physiologically implausible hallucinations.

In this work, we explore the use of amyloid pathology prior knowledge by constructing a population-level amyloid pathology template and propose the *Template-Guided Heteroscedastic Diffusion Bridge* (TGH-DB) for MRI-to-PET synthesis. TGH-DB reformulates diffusion as a transition toward the population-level template. As shown in Fig. 1, unlike classical diffusion models, the forward process in TGH-DB steers latent states toward a mean template, while voxel-wise population variability is modeled through a heteroscedastic noise mechanism. The template-guided diffusion process is governed by a mean drift term, and this design preserves pathologically stable regions with high precision while allowing stochastic flexibility in regions exhibiting greater variability. To further ensure effective use of the structural information from various MRI modalities, we introduce the *Time-Aware Evolutionary Denoiser* (TAED) that adaptively integrates multimodal structural MRI information. TAED comprises parallel encoders for different MRI modalities and fuses their information with varying focus according to the diffusion time steps. TGH-DB was comprehensively validated on two public datasets, where it outperforms representative existing approaches to MRI-

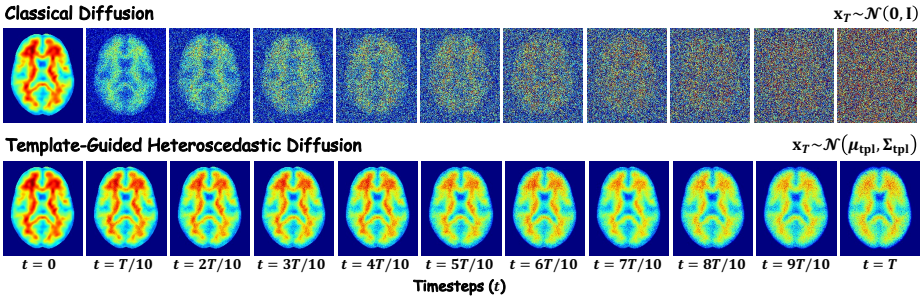

Fig. 1: Overview of TGH-DB diffusion and its comparison with classical diffusion. t denotes the diffusion time step, and T is the total number of diffusion steps. $\mathbf{x}_T \sim \mathcal{N}(\cdot, \cdot)$ represents Gaussian noise. $\boldsymbol{\mu}_{\text{tpl}}$ and $\boldsymbol{\sigma}_{\text{tpl}}$ are the population-level PET mean and standard deviation maps, respectively, and $\boldsymbol{\Sigma}_{\text{tpl}} = \text{diag}(\boldsymbol{\sigma}^2_{\text{tpl}})$.

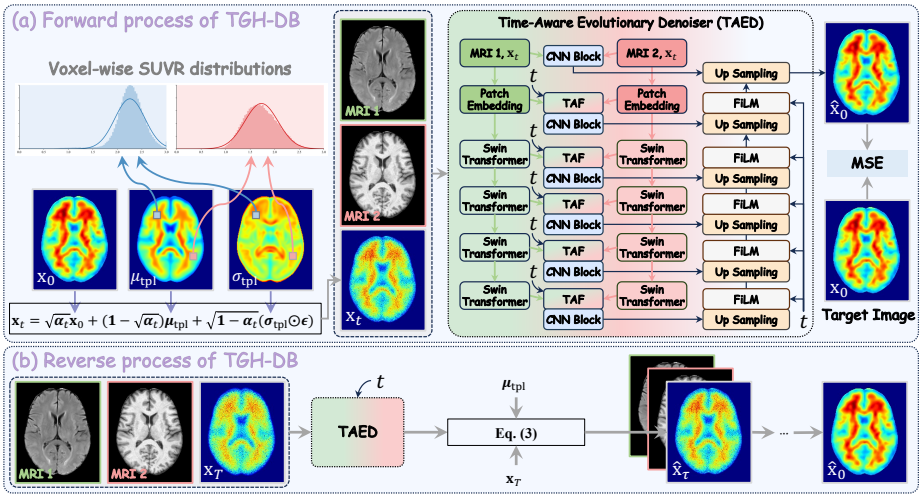

Fig. 2: Overview of TGH-DB: (a) The forward process for training. (b) The reverse process for inference. TAF denotes the Time-Aware Fusion module, and FiLM denotes Feature-wise Linear Modulation.

to-PET synthesis both qualitatively and quantitatively. The code of our method is available at https://github.com/PangHaowen-hub/TGH-DB.

2 Methods

Suppose we are given 3D multimodal MRI scans of a subject. Our goal is to synthesize the corresponding target 3D PET image from MRI scans, with the synthesis mapping learned with a 3D diffusion model from paired training data

of multimodal structural MRI scans and PET images. An overview of the proposed TGH-DB framework is shown in Fig. 2. In TGH-DB, a population-level amyloid pathology template is first constructed as the prior to guide the synthesis, where the template comprises the voxel-wise mean and standard deviation maps computed from the training data. The template then constrains the diffusion trajectory to a biologically plausible manifold, substantially reducing the solution space of the reverse generative process. It centers the latent variable around the template mean and modulates the diffusion noise in a voxel-wise heteroscedastic manner according to the template standard deviation, disentangling shared organizations from subtle pathological variations. Additionally, to enable effective integration of multimodal MRI information, we introduce TAED that performs adaptive information fusion as the denoising network.

2.1 Population-level Amyloid Pathology Template

To construct the population-level template, following [3,13], we spatially normalize training PET volumes to the MNI152 space [17] and aggregate PET intensities (*Standardized Uptake Value Ratio* (SUVR)) across subjects to characterize the population-level distribution. Since we have observed that the voxel-wise SUVR distributions follow Gaussian patterns, the population-level pathology characteristics are modeled by first- and second-order statistics. Specifically, the mean map $\boldsymbol{\mu}_{\text{tpl}} \in \mathbb{R}^{H\times W\times D}$ and standard deviation map $\boldsymbol{\sigma}_{\text{tpl}} \in \mathbb{R}^{H\times W\times D}$ (H, W, and D denote the image height, width, and depth, respectively) are estimated from the training cohort with isotropic Gaussian smoothing of a 1.0 mm kernel to suppress image noise, and $\boldsymbol{\mu}_{\text{tpl}}$ and $\boldsymbol{\sigma}_{\text{tpl}}$ form the final template. An example of the template and SUVR distributions is shown in Fig. 2a.

2.2 TGH-DB

We denote the multimodal MRI scans of a subject by $\mathbf{c} \in \mathbb{R}^{M\times H\times W\times D}$ and the corresponding real PET image by $\mathbf{x}_0 \in \mathbb{R}^{H\times W\times D}$, where M is the number of MRI modalities. Like classical diffusion models, TGH-DB comprises a forward process for model training and a reverse process for inference. In the forward process, the PET image $\mathbf{x}_0$ is progressively transformed into a noise-corrupted representation, and the model learns how to recover $\mathbf{x}_0$ from the noisy representation. The reverse process synthesizes the PET image through multiple steps by iteratively refining an initial noise sample based on the MRI scans $\mathbf{c}$. But unlike classical diffusion models that use standard Gaussian noise representation, TGH-DB incorporates prior knowledge about amyloid pathological organization into the noise distribution, which guides the diffusion process and simplifies the reverse inference process. The detailed design of TGH-DB is given below.

Forward Process. As shown in Fig. 2a, the forward process in TGH-DB is formulated as a template-guided diffusion process that gradually transforms PET data toward the population-level distribution by drifting toward the mean map $\boldsymbol{\mu}_{\text{tpl}}$ with spatially adaptive noise scaled by $\boldsymbol{\sigma}_{\text{tpl}}$, unlike conventional diffusion

models that drift toward a zero mean with homoscedastic variance. This ensures plausible diffusion and suppresses hallucinations.

At each time step $t \in \{0, \ldots, T\}$, the noisy PET image $\mathbf{x}_t$ is directly computed from $\mathbf{x}_0$ with a mean drift term and a heteroscedastic noise term as

$$\mathbf{x}_t = \underbrace{\sqrt{\alpha_t}\mathbf{x}_0 + (1 - \sqrt{\alpha_t})\boldsymbol{\mu}_{\text{tpl}}}_{\text{Mean Drift}} + \underbrace{\sqrt{1 - \alpha_t}(\boldsymbol{\sigma}_{\text{tpl}} \odot \boldsymbol{\epsilon})}_{\text{Heteroscedastic Noise}}, \tag{1}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\epsilon} \in \mathbb{R}^{H \times W \times D}$. All operations are performed voxel-wise in the 3D spatial domain. Inspired by the noise scheduling insights in [2], we define α_t as a scheduler that follows a linearly decreasing sequence from 1 to 0 over T time steps, controlling the interpolation between the original PET image and the mean template, as well as the progressive injection of heteroscedastic noise. The mean drift term deterministically steers the diffusion trajectory toward the population-level amyloid pathology center, while the heteroscedastic noise term introduces spatially adaptive stochasticity modulated by voxel-wise pathological variability. As a result, physiologically consistent regions are perturbed minimally to preserve pathological stability, whereas regions with higher inter-subject variability are granted increased generative flexibility.

Following [12], we use a denoiser $f_\theta(\mathbf{x}_t, t, \mathbf{c})$ parametrized by θ to produce the estimate $\hat{\mathbf{x}}_0$ of the PET image from $\mathbf{x}_t$. The denoiser model is learned by minimizing the *Mean Squared Error* (MSE) between the estimated and real PET images. The detailed denoiser architecture will be described later.

Reverse Process. The reverse process for inference is graphically summarized in Fig. 2b. According to Eq. (1), the reverse process begins from $\mathbf{x}_T \sim \mathcal{N}(\boldsymbol{\mu}_{\text{tpl}}, \boldsymbol{\Sigma}_{\text{tpl}})$, where $\boldsymbol{\Sigma}_{\text{tpl}} = \text{diag}(\boldsymbol{\sigma}^2_{\text{tpl}})$ represents the voxel-wise population variance. At the t-th reverse step (t starts from T and decreases to 0), the denoiser produces an estimate $\hat{\mathbf{x}}_0 = f_\theta(\mathbf{x}_t, t, \mathbf{c})$ of the clean PET image. Note that although $\hat{\mathbf{x}}_0$ is predicted at each step, like in classical diffusion models, it serves only as an intermediate variable to parametrize the reverse distribution. The actual generation requires iterative sampling of noisy PET images at different time steps, as the reverse process targets a complex real-data distribution that cannot be captured in a single step. This stepwise denoising enables progressive refinement of coarse structures and recovery of fine details.

To accelerate sampling, we derive a deterministic trajectory inspired by DDIM [23]. First, the noise term is analytically estimated by rearranging Eq. (1):

$$\hat{\boldsymbol{\epsilon}} = \frac{\mathbf{x}_t - \sqrt{\alpha_t}\hat{\mathbf{x}}_0 - (1 - \sqrt{\alpha_t})\boldsymbol{\mu}_{\text{tpl}}}{\sqrt{1 - \alpha_t}\,\boldsymbol{\sigma}_{\text{tpl}}}, \tag{2}$$

where $\hat{\boldsymbol{\epsilon}}$ is the noise estimate. We assume a non-Markovian forward process and construct a deterministic jump-step to any earlier time $\tau < t$. Specifically, by plugging the estimated noise $\hat{\boldsymbol{\epsilon}}$ into Eq. (1) at time step τ, we obtain the following

deterministic jump-step reverse process:

$$\begin{aligned}\mathbf{x}_\tau &= \sqrt{\alpha_\tau}\hat{\mathbf{x}}_0 + (1-\sqrt{\alpha_\tau})\boldsymbol{\mu}_{\text{tpl}} + \sqrt{\frac{1-\alpha_\tau}{1-\alpha_t}}\left(\mathbf{x}_t - \sqrt{\alpha_t}\hat{\mathbf{x}}_0 - (1-\sqrt{\alpha_t})\boldsymbol{\mu}_{\text{tpl}}\right) \\ &= \sqrt{\frac{1-\alpha_\tau}{1-\alpha_t}}\mathbf{x}_t + \left(\sqrt{\alpha_\tau} - \sqrt{\alpha_t}\sqrt{\frac{1-\alpha_\tau}{1-\alpha_t}}\right)\hat{\mathbf{x}}_0 + \left(1 - \sqrt{\frac{1-\alpha_\tau}{1-\alpha_t}} - \left(\sqrt{\alpha_\tau} - \sqrt{\alpha_t}\sqrt{\frac{1-\alpha_\tau}{1-\alpha_t}}\right)\right)\boldsymbol{\mu}_{\text{tpl}}.\end{aligned} \tag{3}$$

This formulation reveals that the deterministic reverse transition is essentially a guided interpolation among the current state $\mathbf{x}_t$, the estimated clean image $\hat{\mathbf{x}}_0$, and the population-level template $\boldsymbol{\mu}_{\text{tpl}}$. By explicitly anchoring the update rule with $\boldsymbol{\mu}_{\text{tpl}}$, the model encourages synthesized PET images to maintain pathological consistency. Notably, $\boldsymbol{\sigma}_{\text{tpl}}$ is not explicitly required during inference, as the denoiser implicitly captures spatial variability during training by handling amplified noise in high-variance regions. The iteration continues with a fixed time interval until $\tau = 0$, which produces the final synthesis result.

TAED. To estimate $\hat{\mathbf{x}}_0$ from $\mathbf{x}_t$, t, and $\mathbf{c}$, we develop TAED as a fully 3D denoiser, which is based on the SwinUNETR architecture [25] that captures long-range contextual dependencies and multi-scale structural representations. TAED also introduces time-aware flexibility that allows evolvable fusion of multimodal MRI information as the diffusion process progresses. The architecture of TAED is illustrated in Fig. 2a. For the t-th diffusion step, parallel Swin Transformer encoders are used, each independently extracting multi-scale features from an MRI modality concatenated with $\mathbf{x}_t$. The multimodal features are then fused using *Time-Aware Fusion* (TAF) modules. Specifically, a weight for each modality is computed by applying a softmax-activated linear projection to the time embedding, which is obtained by first encoding the diffusion time step with a sinusoidal positional embedding and then transforming it into a 96-dimensional representation with a multilayer perceptron [7]. The MRI features of different modalities are then integrated via weighted summation followed by $1 \times 1 \times 1$ convolution. This design enables adaptive emphasis on different anatomical cues at different time steps. Finally, TAED uses a U-shaped decoder [25] with skip connections to reconstruct the target signal from the fused representations. Time information is also injected into the decoder via *Feature-wise Linear Modulation* (FiLM) [19].

2.3 Implementation Details

MRI and PET volumes are spatially normalized to the MNI152 space. MRI intensities are rescaled to the range $[0, 1]$ based on the 0.5th and 99.5th percentiles. PET volumes are first converted to SUVR maps following [3] and subsequently linearly rescaled to $[0, 1]$. The diffusion steps in TGH-DB were set to $T = 1000$. To accelerate generation, a DDIM-like stride [23] of ten steps was applied. The model was trained using the MSE loss with AdamW [15] for 300 epochs.

3 Results

PET and MRI data were obtained from two public datasets: Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4) [24] and Health and Aging

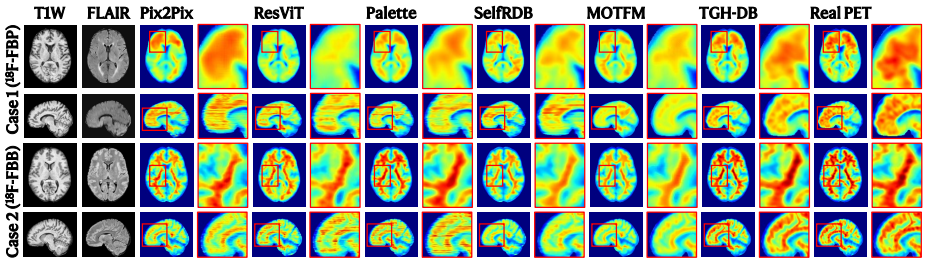

Fig. 3: Examples of synthesis results, shown together with the input multimodal MRI scans and real PET images for reference. Note the red boxes for comparison.

Brain Study: Health Disparities (HABS-HD) [20]. We focused on two clinically established amyloid PET tracers, fluorine-18 florbetapir (^{18}F-FBP) and fluorine-18 florbetaben (^{18}F-FBB). The final curated cohort comprised 1,785 ^{18}F-FBP scans from A4 and 3,798 ^{18}F-FBB scans from HABS-HD. In this work, we chose T1-weighted (T1W) and FLAIR scans as the MRI input for PET synthesis due to their routine availability in clinical practice and complementary characterization of brain tissue. All selected PET scans had corresponding MRI. Subjects were split into training/validation/testing sets, resulting in 1,250/178/357 scans for ^{18}F-FBP and 2,660/379/759 scans for ^{18}F-FBB, respectively.

3.1 Comparison with Existing Image Synthesis Methods

We compared our method with several representative image synthesis approaches, including Pix2Pix [9], ResViT [4], Palette [21], SelfRDB [1], and MOTFM [27]. All competing methods were applied with the same data splits as TGH-DB and their default hyperparameters to ensure fair comparison. Examples of synthesis results are shown in Fig. 3, where TGH-DB produces images that are more consistent with real images than those generated by competing methods. Next, we quantitatively compared TGH-DB with the competing methods by calculating the *Peak Signal-to-Noise Ratio* (PSNR) and *Structural Similarity Index Measure* (SSIM) between the synthesized and real images. The corresponding results are reported in Table 1, where statistical significance was also assessed with the Wilcoxon signed-rank test.

In addition, for the A4 dataset, we evaluated the clinical utility of our method using a downstream classification task, where binary classification was performed on the synthesized PET images to distinguish amyloid-positive from amyloid-negative subjects. The classification models were obtained from [5] and [11]. The classification result is also reported in Table 1, where the classification performance based on input MRI or real PET data is shown for reference. Compared with other methods, TGH-DB achieves the best *area under the receiver operating characteristic curve* (AUC) and balanced *accuracy* (ACC), and its performance approaches that obtained with real PET images. These results further demonstrate the superior synthesis quality and diagnostic relevance of TGH-DB.

Table 1: Comparison of synthesis quality between TGH-DB and competing methods. Asterisks indicate statistically significant differences between TGH-DB and each competing method based on the Wilcoxon signed-rank test: * $p < 0.05$.

Model	^{18}F-FBP				^{18}F-FBB	
	PSNR (dB)	SSIM (%)	AUC (%)	ACC (%)	PSNR (dB)	SSIM (%)
Input MRI	–	–	73.03	66.79	–	–
Pix2Pix [9]	$28.19 \pm 1.94^*$	$85.89 \pm 8.94^*$	71.12	64.03	$29.80 \pm 3.75^*$	$91.63 \pm 2.82^*$
ResViT [4]	$28.81 \pm 2.09^*$	$90.14 \pm 6.11^*$	73.29	65.60	$29.81 \pm 2.26^*$	$91.95 \pm 1.97^*$
Palette [21]	$29.79 \pm 2.32^*$	$91.10 \pm 2.43^*$	78.16	70.16	$27.65 \pm 1.67^*$	$92.24 \pm 1.34^*$
SelfRDB [1]	$29.98 \pm 2.49^*$	$91.92 \pm 1.96^*$	78.84	71.63	$30.14 \pm 3.26^*$	$92.08 \pm 2.04^*$
MOTFM [27]	$29.24 \pm 2.44^*$	$90.72 \pm 1.98^*$	82.28	76.54	$30.34 \pm 2.40^*$	$90.58 \pm 2.28^*$
TGH-DB	$\mathbf{31.55 \pm 2.24}$	$\mathbf{92.98 \pm 1.63}$	**88.79**	**81.28**	$\mathbf{31.93 \pm 2.45}$	$\mathbf{93.98 \pm 2.33}$
Real PET	–	–	96.60	90.45	–	–

Table 2: Ablation study. The means and standard deviations of PSNR and SSIM are presented and the best results are in bold.

Model	^{18}F-FBP		^{18}F-FBB	
	PSNR (dB)	SSIM (%)	PSNR (dB)	SSIM (%)
TGH-DB	$\mathbf{31.55 \pm 2.24}$	$\mathbf{92.98 \pm 1.63}$	$\mathbf{31.93 \pm 2.45}$	$\mathbf{93.98 \pm 2.33}$
w/o Mean Map	30.08 ± 2.38	91.03 ± 1.50	30.54 ± 2.42	92.42 ± 2.36
w/o Std Map	30.72 ± 2.59	91.78 ± 1.91	30.75 ± 2.64	92.66 ± 2.57
w/o Mean & Std Map	29.62 ± 2.55	90.86 ± 1.85	30.02 ± 2.59	91.37 ± 2.17
w/o TAF Modules	30.87 ± 2.69	91.93 ± 1.56	30.94 ± 2.48	92.97 ± 2.28
w/o TAED	30.45 ± 2.84	91.52 ± 1.93	30.66 ± 2.72	92.43 ± 2.64

3.2 Ablation Study

To assess the contribution of individual components in TGH-DB, we conducted ablation experiments, and the results are shown in Table 2. We first examined the role of the pathology template, including $\boldsymbol{\mu}_{\text{tpl}}$ and $\boldsymbol{\sigma}_{\text{tpl}}$. We removed $\boldsymbol{\mu}_{\text{tpl}}$ (referred to as "w/o Mean Map"), where the forward process was guided toward a zero-mean distribution while retaining the variance $\boldsymbol{\sigma}^2_{\text{tpl}}$. We also removed $\boldsymbol{\sigma}_{\text{tpl}}$ (referred to as "w/o Std Map"), where the noise variance was fixed to one while the mean drift toward $\boldsymbol{\mu}_{\text{tpl}}$ was preserved. We further reduced the diffusion process to a standard diffusion model formulation (referred to as "w/o Mean & Std Map"), where the terminal distribution was the standard Gaussian noise $\mathcal{N}(\mathbf{0}, \mathbf{I})$. Incorporating either component outperformed the vanilla diffusion model, while using both achieved the best synthesis quality, highlighting the benefit of leveraging prior knowledge of amyloid pathology distributions. To evaluate the effectiveness of the proposed denoiser TAED, we replaced the TAF modules with simple concatenation (referred to as "w/o TAF Modules") and further substituted TAED with a 3D U-Net from MAISI [6] (referred to as "w/o TAED"). Both alterations led to worse results, which indicates the important role of our TAED denoiser.

4 Conclusion

We have proposed TGH-DB that improves the quality of PET synthesis from multimodal MRI by leveraging population-level molecular pathology priors. TGH-DB models the PET synthesis as a directed transition from a population-level amyloid pathology template to the subject-specific distribution, thus enforcing physiological constraints within a diffusion model framework. Experimental results on public datasets show that our approach outperforms existing methods in synthesizing high-fidelity PET images and also benefits downstream tasks with the synthesized PET data.

Acknowledgments. This research is supported by Brain Science and Brain-like Intelligence Technology - National Science and Technology Major Project (2022ZD0209000). Additional support is provided by the University Grants Committee (GRF: 15201124), the Hong Kong Global STEM Scholar scheme, the National Natural Science Foundation of China (62576034), Beijing Municipal Natural Science Foundation (7242273), and Fundamental Research Funds for the Central Universities (2025CX01009).

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)
2. Chen, T.: On the importance of noise scheduling for diffusion models. arXiv preprint arXiv:2301.10972 (2023)
3. Chen, Z., Bi, S., Shan, Y., Wang, F., Wang, Y., Qi, Z., Wang, T., Li, X., Li, S., Xiao, H., et al.: MRI-to-PET synthesis via deep learning for amyloid-β quantification in Alzheimer's disease. European Radiology pp. 1–13 (2026)
4. Dalmaz, O., Yurt, M., Çukur, T.: ResViT: Residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)
5. Fan, S., Ponisio, M.R., Xiao, P., Ha, S.M., Chakrabarty, S., Lee, J.J., Flores, S., LaMontagne, P., Gordon, B., Raji, C.A., et al.: AmyloidPETNet: Classification of amyloid positivity in brain PET imaging using end-to-end deep learning. Radiology **311**(3), e231442 (2024)
6. Guo, P., Zhao, C., Yang, D., Xu, Z., Nath, V., Tang, Y., Simon, B., Belue, M., Harmon, S., Turkbey, B., et al.: MAISI: Medical AI for synthetic imaging. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4430–4441. IEEE (2025)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
8. Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., Shen, Y.: Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. IEEE Transactions on Medical Imaging **41**(1), 145–157 (2021)

9. 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)
10. Jin, Y., DuBois, J., Zhao, C., Zhan, L., Gabelle, A., Jahanshad, N., Thompson, P.M., Gafson, A., Belachew, S.: Brain MRI to PET synthesis and amyloid estimation in Alzheimer's disease via 3D multimodal contrastive GAN. In: International Workshop on Machine Learning in Medical Imaging. pp. 94–103. Springer (2023)
11. Kim, D., Ottesen, J.A., Kumar, A., Ho, B.C., Bismuth, E., Young, C.B., Mormino, E., Zaharchuk, G.: Deep learning-based prediction of PET amyloid status using MRI. American Journal of Neuroradiology **46**(12), 2590–2598 (2025)
12. Li, T., He, K.: Back to basics: Let denoising generative models denoise. arXiv preprint arXiv:2511.13720 (2025)
13. Li, Y., Li, H., Zhou, S.K.: Causal PETS: Causality-informed PET synthesis from multi-modal data. In: Medical Imaging with Deep Learning (2025)
14. Liu, T., Wang, H., Chang, H., Wang, F., Lian, C., Ma, J.: DIReCT: Domain-Informed rectified flow for controllable brain MRI to PET translation. In: International Conference on Information Processing in Medical Imaging. pp. 218–231. Springer (2025)
15. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
16. Manjooran, G.P., Malakkaran, A.J., Joseph, A., Babu, H.M., MS, M.: A review on cross-modality synthesis from MRI to PET. In: 2021 2nd International Conference on Secure Cyber Computing and Communications. pp. 126–131. IEEE (2021)
17. Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., et al.: A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **356**(1412), 1293–1322 (2001)
18. Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D.: Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 455–463. Springer (2018)
19. 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 (2018)
20. Petersen, M.E., Zhou, Z., Hall, J.R., Phillips, N., Meeker, K.L., Borzage, M.T., Braskie, M.N., Clark, A.L., Shi, Y., Rissman, R.A., et al.: Health and Aging Brain Study-Health Disparities (HABS-HD) methods and partner characteristics. Alzheimer's & Dementia: Translational Research & Clinical Interventions **11**(3), e70140 (2025)
21. 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)
22. Sikka, A., Peri, S.V., Bathula, D.R.: MRI to FDG-PET: Cross-modal synthesis using 3D U-Net for multi-modal Alzheimer's classification. In: International workshop on simulation and synthesis in medical imaging. pp. 80–89. Springer (2018)
23. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021)
24. Sperling, R.A., Rentz, D.M., Johnson, K.A., Karlawish, J., Donohue, M., Salmon, D.P., Aisen, P.: The A4 study: Stopping AD before symptoms begin? Science Translational Medicine **6**(228), 228fs13–228fs13 (2014)

25. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of Swin transformers for 3D medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20730–20740 (2022)
26. Vega, F., Addeh, A., MacDonald, M.E.: Amyloid-beta PET synthesis from structural MRI: A potential alternative method for Alzheimer's disease screening. Alzheimer's & Dementia **19**, e079258 (2023)
27. Yazdani, M., Medghalchi, Y., Ashrafian, P., Hacihaliloglu, I., Shahriari, D.: Flow matching for medical image synthesis: Bridging the gap between speed and quality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 216–226. Springer (2025)
28. Yu, B., Shan, Y., Ding, J.: A literature review of MRI techniques used to detect amyloid-beta plaques in Alzheimer's disease patients. Annals of Palliative Medicine **10**(9), 100620074–100610074 (2021)
29. Yu, M., Lalush, D.S., Monroe, D.C., Giovanello, K.S., Lin, W., Yap, P.T., Mihalik, J.P., Liu, M.: Distribution-guided multi-tracer brain PET synthesis from structural MRI with class-conditioned weighted diffusion. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 152–162. Springer (2025)
30. Zhang, J., He, X., Qing, L., Gao, F., Wang, B.: BPGAN: Brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis. Computer Methods and Programs in Biomedicine **217**, 106676 (2022)