Uni-Brain: Anatomy-Informed Diffusion for Multi-Tracer Synthesis and Counterfactual Prognosis from Cross-Sectional Data

Hongjie Yu[1], Haoyue Yuan[1], Kaicong Sun[1], and Dinggang Shen[1,2,3]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200230, China
[3] Shanghai Clinical Research and Trial Center, Shanghai, 201210, China
Dinggang.Shen@gmail.com

Abstract. Alzheimer's disease (AD) is marked by the progressive evolution of amyloid, tau, and neurodegeneration biomarkers. While MRI provides accessible structural information, PET offers molecular specificity but remains invasive and costly, limiting frequent longitudinal monitoring. To address this, we propose Uni-Brain, a unified diffusion framework for multi-tracer PET synthesis and counterfactual prognosis. Here, counterfactual prognosis refers to individualized PET generation under fixed baseline anatomy and user-specified clinical covariates. Uni-Brain leverages a shared MRI-PET latent space and a conditional latent diffusion model to generate target-tracer PET volumes with anatomical fidelity and clinical controllability. The framework introduces (i) **anatomy-informed spatial conditioning**, which injects MRI latent codes as dense spatial guidance to mitigate structural drift, and (ii) **hybrid clinical conditioning**, which separately encodes continuous covariates (e.g., age, CDR-SB) and discrete attributes (e.g., APOEε4, sex) for smoother covariate control. Uni-Brain supports both longitudinal prognosis and counterfactual clinical intervention without requiring paired longitudinal training scans. Extensive experiments on the ADNI dataset demonstrate competitive cross-modal synthesis performance and indicate that the generated follow-up PET images faithfully reflect longitudinal disease progression, exhibiting strong agreement with real follow-up PET images and established clinical biomarkers. Code is available at https://github.com/hjYu-711/UniBrain.

Keywords: Alzheimer's Disease · Multi-tracer Synthesis · Counterfactual Prognosis · Latent Diffusion Models.

1 Introduction

Alzheimer's disease (AD) is described under the ATN framework, in which biomarkers of amyloid-β (Aβ) deposition (A), tau pathology (T), and neurodegeneration (N) define a continuous pathological spectrum rather than discrete stages [10]. Structural magnetic resonance imaging (MRI) provides accessible anatomical information and primarily reflects neurodegeneration, whereas

positron emission tomography (PET) offers molecular specificity for ATN-related pathology. However, PET is invasive and radioactive, limiting its use in large-scale screening and longitudinal monitoring. These limitations motivate a practical generative framework that infers PET-derived molecular states from baseline MRI, supports multi-tracer PET synthesis, and enables simulation of clinically meaningful disease trajectories without requiring repeated PET acquisitions.

Despite rapid progress in medical image translation, existing methods remain limited for unified and controllable PET simulation. Most cross-modal synthesis approaches, including adversarial-based methods with CNN [9,17,28] or transformer [4,5] architectures and recent diffusion-based [7,16] models, are typically formulated as static mappings at a single time point T_0. Consequently, they provide limited support for longitudinal prognosis and counterfactual simulation. In contrast, longitudinal prediction methods explicitly model temporal progression ($T_0 \rightarrow T_{\text{future}}$) but often require paired longitudinal scans of the same modality, which is difficult to obtain for PET and limits their applicability to MRI-to-PET prognosis [19,26]. Moreover, discretizing continuous clinical variables, such as age and clinical dementia rating sum of boxes (CDR-SB), into class embeddings may weaken smooth covariate control [10,22]. These limitations motivate a unified framework that connects multi-tracer synthesis with longitudinal prognosis and counterfactual PET simulation from cross-sectional training data.

To address this challenge, we formulate multi-tracer PET synthesis and clinically conditioned simulation from a shared latent-space perspective [27]. Since MRI and different PET tracers observe complementary aspects of the same subject, their latent representations are coupled through anatomical and biological factors [24]. We therefore consider a shared latent space in which different modalities and tracers occupy related manifolds, with anatomy and clinical conditions jointly determining the location of representations on these manifolds [6]. Under this formulation, baseline MRI serves as an anatomical anchor, whereas clinical covariates modulate tracer-specific pathological or metabolic changes. This perspective provides a unified interface for three related tasks: cross-modal synthesis by specifying the tracer indicator, longitudinal prognosis by adjusting the target age while holding anatomy fixed, and counterfactual simulation by intervening on clinically specified covariates.

Based on this formulation, we propose **Uni-Brain**, a unified diffusion framework for multi-tracer PET synthesis and counterfactual prognosis. Uni-Brain first establishes a shared MRI-PET latent space and then uses a conditional latent diffusion model (LDM) [20] to generate target-tracer PET images under anatomical and clinical conditions. Our contributions are summarized as follows: (1) an **anatomy-informed spatial conditioning** mechanism that injects MRI latent codes as dense spatial guidance to preserve subject-specific structure and reduce anatomical drift; (2) a **hybrid clinical conditioning** strategy that separately encodes continuous and discrete covariates, enabling smoother and more flexible covariate control; and (3) extensive validation on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[4], demonstrating competitive

[4] adni.loni.usc.edu

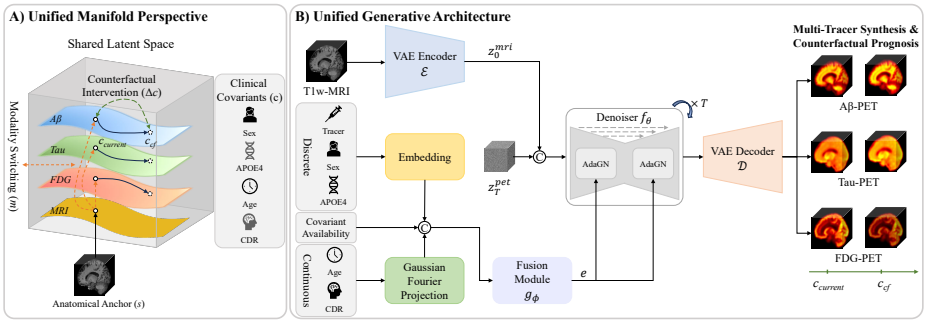

Fig. 1. Overview of Uni-Brain. (A) Manifold perspective: modalities are organized in a shared latent space $\mathcal{Z}$; anatomy s and clinical covariates c define subject-specific conditions, enabling cross-modal synthesis (switching m), longitudinal prognosis, and counterfactual simulation under clinically specified covariates c. (B) Unified generative architecture: a conditional LDM combines MRI conditioning (z_0^{mri}) and hybrid clinical conditioning (e) via AdaGN to synthesize multi-tracer PET.

multi-tracer synthesis performance and clinically consistent longitudinal prognosis against real follow-up PET scans.

2 Method

We propose **Uni-Brain**, a unified generative framework for synthesizing multi-tracer PET images from baseline structural MRI. Fig. 1(A) illustrates the manifold perspective, where different modalities are projected into corresponding manifolds within a shared latent space. In this space, anatomical information s anchors subject-specific identity, while clinical covariates c modulate PET patterns under specified disease-related conditions. Fig. 1(B) presents the overall architecture. A shared MRI-PET variational autoencoder (VAE) [13] first establishes the shared latent representation, upon which a conditional LDM generates target-tracer PET images conditioned on anatomy-preserving MRI latent codes and hybrid clinical embeddings.

2.1 Problem Formulation

Under the manifold hypothesis [6,24], we encode observations $x \in \mathcal{X}$ into a structured latent space $\mathcal{Z}$ where anatomy is subject-specific, and disease effects follow shared population regularities [15]. We represent each observation with an anatomical factor s capturing subject-specific anatomy, a clinical condition vector $c = (c_{\text{cont}}, c_{\text{disc}})$ collecting continuous covariates c_{cont} (e.g., age and CDR-SB) and discrete covariates c_{disc} (e.g., sex and APOEε4 status), and a tracer indicator m specifying the target PET tracer. In practice, we instantiate s with the MRI latent code z_0^{mri}. Given this factorization, we use a unified predictive function $\mathcal{F} : (s, c, m) \to x_{\text{pet}}$.

In this work, we employ counterfactual prognosis for two uses of $\mathcal{F}$: (1) **longitudinal prognosis**, where we fix anatomy and tracer, set age to the follow-up age, mask future CDR-SB, and evaluate against the real follow-up PET; and (2) **counterfactual simulation**, where we fix anatomy, tracer, and manually specified clinical covariates to study condition-dependent PET changes. This formulation supports individualized PET simulation from cross-sectional data.

2.2 Unified Multi-Tracer Generative Architecture

We operate in a VAE latent space shared across MRI and PET tracers. Given an MRI volume x_{mri} and a target-tracer PET volume x_{pet}, the encoder $\mathcal{E}$ maps them to latent codes $z_0^{\text{mri}} = \mathcal{E}(x_{\text{mri}})$ and $z_0^{\text{pet}} = \mathcal{E}(x_{\text{pet}})$, respectively. We pretrain the VAE and keep $(\mathcal{E}, \mathcal{D})$ frozen during diffusion training. We then train a conditional LDM to model the distribution of z_0^{pet} given z_0^{mri}, c, and m. During training, we sample $\epsilon \sim \mathcal{N}(0, I)$ and construct $z_t^{\text{pet}} = \sqrt{\bar{\alpha}_t}\, z_0^{\text{pet}} + \sqrt{1-\bar{\alpha}_t}\, \epsilon$.

Anatomy-Informed Spatial Conditioning To provide a subject-specific anatomical reference for multi-tracer synthesis, we incorporate the MRI latent code z_0^{mri} as a dense spatial condition. This acts as a structural anchor, guiding the diffusion model to distribute pathological signals along subject-specific anatomical structures. This design encourages the generated PET images to respect individual anatomical boundaries and heterogeneity, ensuring that the predicted pathological trajectories remain biologically consistent.

Hybrid Clinical Conditioning A key challenge is to encode heterogeneous clinical conditions $c = (c_{\text{cont}}, c_{\text{disc}})$ while preserving smooth covariate control in the conditional LDM. We adopt a hybrid conditioning scheme with separate pathways for continuous covariates c_{cont} and discrete covariates c_{disc}, while accounting for missing data and enabling classifier-free guidance (CFG) [8]. Continuous covariates c_{cont} are mapped via Gaussian Fourier projection [25] into a high-dimensional embedding that mitigates spectral bias, whereas discrete attributes c_{disc} and the tracer indicator m use learnable embeddings. To represent missingness and encourage information sharing across observed covariates, we introduce an availability indicator vector v_{av} that encodes covariate availability via a linear projection, enabling conditioning on partially observed profiles. During training, we set the conditioning signal to null with probability p to learn both conditional and unconditional score functions. The final condition vector e is produced by a fusion module g_ϕ that integrates the concatenated latents:

$$e = g_\phi([v_{\text{m}} \oplus v_{\text{sex}} \oplus v_{\text{apoe}} \oplus v_{\text{age}} \oplus v_{\text{cdr}} \oplus v_{\text{av}}]), \tag{1}$$

where $v_{\text{m}}, v_{\text{sex}}, v_{\text{apoe}}, v_{\text{age}}$, and v_{cdr} denote the embeddings of tracer type, sex, APOEε4, age, and CDR-SB, respectively, and $\oplus$ denotes the concatenation operator. We then inject the fused embedding e into the denoiser f_θ via adaptive group normalization (AdaGN) [18]. The diffusion timestep t is encoded with a

standard sinusoidal embedding and incorporated as in prior diffusion models [7]. For a feature map h, the modulation is applied as $h' = \text{GroupNorm}(h) \cdot (1 + \gamma(e)) + \beta(e)$, where $\gamma(\cdot)$ and $\beta(\cdot)$ are learned affine transforms of e.

Based on the conditioning mechanism specified above, we train the denoiser by minimizing a denoising objective in which f_θ predicts the clean latent z_0^{pet}:

$$\mathcal{L} = \mathbb{E}_{z_0^{\text{pet}}, z_0^{\text{mri}}, t, c, m} \left[\| z_0^{\text{pet}} - f_\theta(z_t^{\text{pet}}, z_0^{\text{mri}}, t, c, m) \|_2^2 \right]. \tag{2}$$

2.3 Inference

At inference time, we synthesize the target PET volume via an iterative reverse-diffusion process using a denoising diffusion implicit model (DDIM) sampler [23]. Starting from Gaussian noise $z_T^{\text{pet}} \sim \mathcal{N}(0, I)$, the model predicts the clean latent representation at each time step t. We apply CFG using a dual forward pass. The denoiser f_θ yields a guided estimate of the original latent $\hat{z}_0$ by combining conditional and unconditional predictions:

$$\hat{z}_0 = w \cdot f_\theta(z_t^{\text{pet}}, z_0^{\text{mri}}, t, c, m) + (1 - w) \cdot f_\theta(z_t^{\text{pet}}, z_0^{\text{mri}}, t, \emptyset, m), \tag{3}$$

where w denotes the guidance strength. Given $\hat{z}_0$, we compute z_{t-1} as follows:

$$z_{t-1}^{\text{pet}} = \sqrt{\bar{\alpha}_{t-1}} \hat{z}_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \cdot \frac{z_t^{\text{pet}} - \sqrt{\bar{\alpha}_t} \hat{z}_0}{\sqrt{1 - \bar{\alpha}_t}}. \tag{4}$$

This refinement proceeds until $t = 0$, after which, the final latent $\hat{z}_0^{\text{pet}}$ is mapped back to image space via the decoder, $\hat{x}_{\text{pet}} = \mathcal{D}(\hat{z}_0^{\text{pet}})$. This formulation provides a unified interface for different simulation tasks. Cross-modal synthesis is controlled by the tracer indicator m, longitudinal prognosis is controlled by the target age, and counterfactual simulation is realized by intervening on the clinical covariate vector c.

3 Experiments and Results

3.1 Experimental Setup

Dataset and Preprocessing. Data used in this study were obtained from ADNI. We curated a multimodal cohort of 2002 subjects with T1-weighted (T1w) MRI and three PET tracers: fluorodeoxyglucose (FDG), Aβ (AV45), and Tau (AV1451). In total, the dataset contains 5248 T1w-MRI scans, 3552 FDG-PET scans, 2824 Aβ-PET scans, and 1353 Tau-PET scans. To enable cross-modal synthesis and longitudinal evaluation, participants were required to have a T1w-MRI and at least one PET scan acquired at the same visit. The dataset was split at the subject level to prevent data leakage, with 1551 subjects for training, 50 for validation, and 401 for testing. For longitudinal prognosis, the test split contains 162, 322, and 359 follow-up AV1451, AV45, and FDG PET scans, respectively. Each synthesized follow-up PET scan is evaluated against the corresponding real follow-up PET scan from the same subject. Preprocessing used

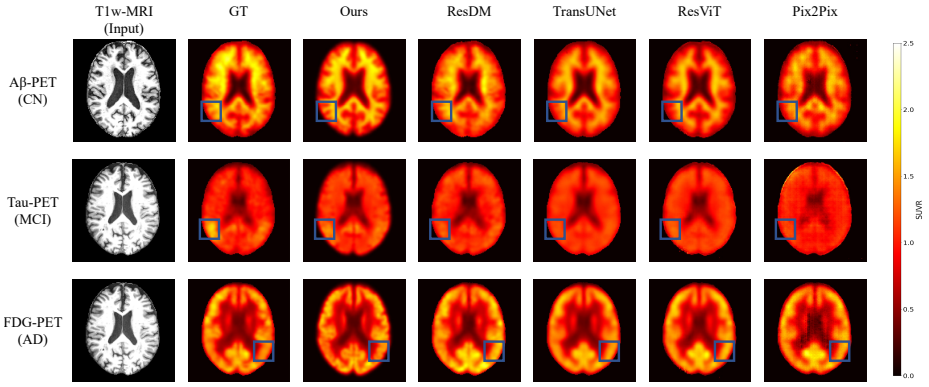

Fig. 2. Qualitative comparison of multi-tracer synthesis across disease stages.

Table 1. Quantitative evaluation of multi-tracer cross-modal synthesis and longitudinal prognosis. Results are reported as mean ± standard deviation (SD). *SCC* denotes Spearman rank correlation between synthesized and ground-truth SUVRs within tracer-specific MetaROIs. **Bold** marks the best performance. † indicates statistically significant improvement over all competing baselines under two-sided Wilcoxon signed-rank tests ($p < 0.05$) for PSNR and SSIM.

Method	**Aβ-PET** (Amyloid)			**Tau-PET** (Tau)			**FDG-PET** (Metabolism)		
	PSNR (dB) ↑	SSIM (%) ↑	SCC ↑	PSNR (dB) ↑	SSIM (%) ↑	SCC ↑	PSNR (dB) ↑	SSIM (%) ↑	SCC ↑
Task 1: Cross-Modal Synthesis ($T_0 \to T_0$)									
Pix2Pix [9]	$26.46_{\pm 1.92}$	$90.39_{\pm 1.62}$	0.248	$25.32_{\pm 1.51}$	$87.16_{\pm 1.66}$	0.173	$27.79_{\pm 1.79}$	$92.05_{\pm 0.99}$	0.453
ResViT [5]	$27.82_{\pm 2.65}$	$93.19_{\pm 1.59}$	0.394	$27.05_{\pm 2.40}$	$91.71_{\pm 2.26}$	0.309	$28.63_{\pm 2.01}$	$94.89_{\pm 1.19}$	0.493
TransUNet [4]	$28.78_{\pm 2.91}$	$95.15_{\pm 1.60}$	0.411	$27.84_{\pm 3.33}$	$93.09_{\pm 2.28}$	0.394	$29.89_{\pm 2.42}$	$96.34_{\pm 1.17}$	0.537
ResDM [16]	$29.14_{\pm 2.78}$	$95.37_{\pm 1.58}$	0.498	$28.01_{\pm 2.97}$	$93.54_{\pm 2.05}$	0.451	$30.07_{\pm 2.23}$	$96.65_{\pm 1.11}$	0.582
Ours	$\mathbf{29.69}^{\dagger}_{\pm 2.66}$	$\mathbf{95.94}^{\dagger}_{\pm 1.54}$	**0.587**	$\mathbf{28.53}^{\dagger}_{\pm 2.51}$	$\mathbf{94.13}^{\dagger}_{\pm 2.12}$	**0.558**	$\mathbf{30.45}^{\dagger}_{\pm 2.07}$	$\mathbf{97.11}^{\dagger}_{\pm 1.09}$	**0.665**
Task 2: Longitudinal Prognosis ($T_0 \to T_{\text{future}}$)									
ResDM [16]	$28.56_{\pm 2.89}$	$94.98_{\pm 1.61}$	0.391	$27.29_{\pm 3.41}$	$93.01_{\pm 2.12}$	0.343	$29.69_{\pm 2.30}$	$96.04_{\pm 1.12}$	0.471
Naive	$29.08_{\pm 2.79}$	$95.19_{\pm 1.55}$	0.461	$27.58_{\pm 2.82}$	$93.27_{\pm 2.24}$	0.387	$29.83_{\pm 2.11}$	$96.48_{\pm 1.12}$	0.525
Ours	$\mathbf{29.27}^{\dagger}_{\pm 2.51}$	$\mathbf{95.41}^{\dagger}_{\pm 1.54}$	**0.524**	$\mathbf{27.91}^{\dagger}_{\pm 2.62}$	$\mathbf{93.63}^{\dagger}_{\pm 2.20}$	**0.436**	$\mathbf{30.11}^{\dagger}_{\pm 2.04}$	$\mathbf{96.81}^{\dagger}_{\pm 1.13}$	**0.589**

`Clinica` [21] with the `t1-volume` and `pet-volume` pipelines. MRI volumes were processed with SPM for segmentation, DARTEL template construction, and MNI normalization [1,2]. Subsequent preprocessing applied skull stripping with FSL [12], followed by brain parcellation using SynthSeg [3]. PET volumes were rigidly aligned to each subject's MRI, normalized to MNI space via DARTEL [1], and intensity-normalized by reference uptake to compute standardized uptake value ratio (SUVR) maps (pons for FDG; cerebellum+pons for Aβ/Tau), after which regional uptake was computed. All images were resampled to isotropic 1.5 mm resolution and cropped $128 \times 128 \times 128$ to meet memory limits, with all intensities scaled to the range $[0, 3]$.

Implementation Details. Our framework was implemented using PyTorch and trained on a single NVIDIA A100 GPU (80GB). We optimized the model using

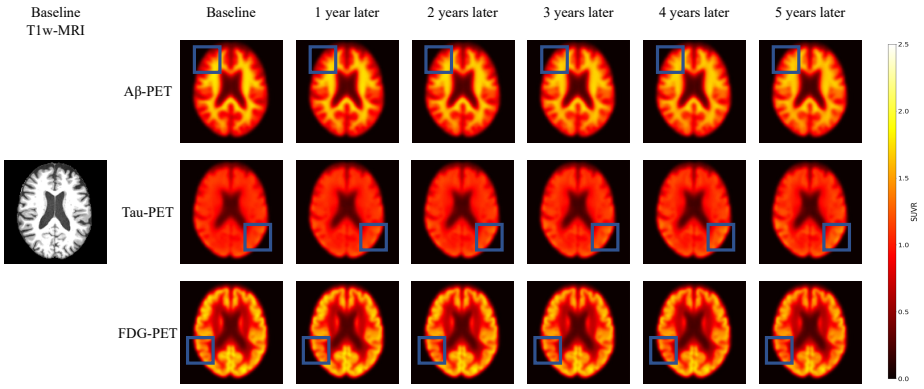

Fig. 3. Illustrative longitudinal prognosis over a 6-year horizon. Blue boxes highlight localized longitudinal PET changes captured by the model under age progression conditioned on the same baseline anatomy.

AdamW with a learning rate of 10^{-4} and a batch size of 2 for 300 epochs. The LDM employed 1,000 timesteps with a cosine noise schedule. We incorporated a conditioning dropout rate $p = 0.25$ to facilitate CFG. In inference, we utilized DDIM sampling with 50 steps and a guidance scale of $w = 7.5$.

Evaluation Metrics. Image fidelity was evaluated using peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM). Clinical validity was assessed by Spearman's correlation coefficient (SCC) between synthesized and ground-truth SUVRs within tracer-specific MetaROIs [11,14], which reflects the monotonic ranking of pathological burden. Statistical significance for subject-level PSNR and SSIM was tested using two-sided Wilcoxon signed-rank tests. Since SCC is a cohort-level metric, we estimated 95% confidence intervals (CIs) for SCC differences using subject-level paired bootstrap resampling between Uni-Brain and the best-performing competing method.

Competing Methods. For cross-modal synthesis, we compared Uni-Brain with four representative baselines, including Pix2Pix [9], ResViT [5], TransUNet [4], and ResDM [16]. Following their original implementations, Pix2Pix was trained separately for FDG, Aβ, and Tau synthesis, whereas the remaining methods were configured as unified multi-tracer models. For longitudinal prognosis, we benchmarked Uni-Brain against ResDM [16], which conditions on clinical text and thus provides partial controllability, and a Naive method that assumes no disease progression by setting $\widehat{PET}_{\text{future}} = \widehat{PET}_0$. Naive is non-trivial given the slow progression of AD and serves as a persistent reference to examine whether a prognosis model can capture meaningful longitudinal changes. BrLP [19] and SADM [26] were excluded from comparison because they are designed for same-modality longitudinal prediction, requiring paired longitudinal supervision, and are therefore not directly applicable to the MRI-to-PET synthesis setting.

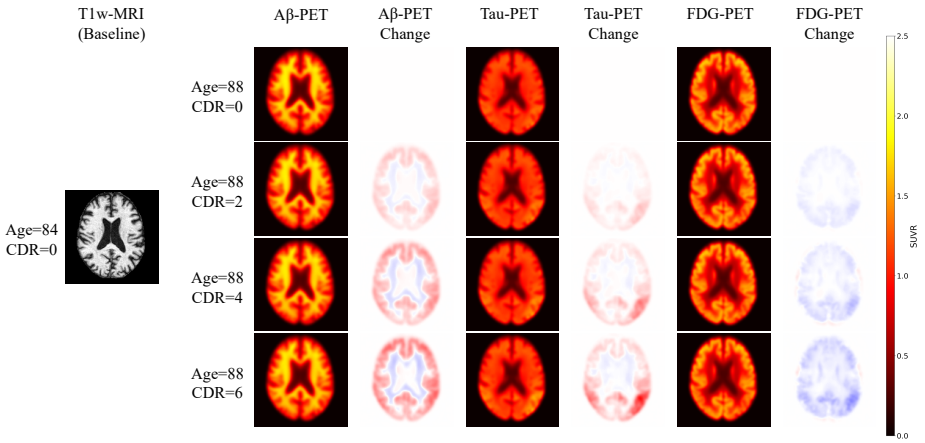

Fig. 4. Counterfactual simulation under clinically specified covariates. Systematic variation of CDR-SB ($0 \rightarrow 6$) reveals localized pathological accumulation and metabolic decline in AD-vulnerable regions, confirming a controllable PET response to changes in clinical condition. The change maps represent differences relative to CDR-SB$= 0$.

3.2 Results

Fidelity in Multi-Tracer Synthesis (Task 1). Table 1 shows Uni-Brain consistently outperforms all competing methods across FDG, Aβ, and Tau. Compared with the strongest diffusion-based method, ResDM, Uni-Brain yields higher SCC values across all tracers. Paired bootstrap analysis further shows positive SCC differences over the strongest baseline, with 95% CIs of [0.047, 0.144], [0.053, 0.172], and [0.041, 0.128] for Aβ-, Tau-, and FDG-PET, respectively. Figure 2 provides qualitative support: Pix2Pix and ResViT often over-smooth or mislocalize uptake, while Uni-Brain recovers tracer-specific signatures, including high-intensity Aβ burden and heterogeneous tau in mild cognitive impairment (MCI) subjects. For FDG-PET, Uni-Brain preserves sharp metabolic gradients such as posterior cingulate hypometabolism. Overall, the synthetic images reflect biologically meaningful distributions rather than superficial appearance translations.

Clinical Validity in Longitudinal Prognosis (Task 2). We assess Uni-Brain's ability to extrapolate future metabolic states from baseline MRIs. Table 1 shows the highest SCC among the evaluated methods (e.g., 0.524 for AV45 vs. 0.461 for Naive). Uni-Brain also achieves positive SCC differences over the strongest baseline, with paired bootstrap 95% CIs of [0.027, 0.112], [0.016, 0.092], and [0.018, 0.109] for AV45, AV1451, and FDG, respectively. Although Naive achieves relatively high PSNR/SSIM due to slow AD progression, it fails to reflect subject-specific longitudinal change. Figure 3 provides an illustrative example by simulating a 6-year PET evolution for a progressive MCI subject. By incrementally advancing the age variable while masking CDR-SB, Uni-Brain produces smooth PET changes with progressive Aβ accumulation in the frontal cortex, spreading

tau in temporal regions, and declining FDG metabolism in parietal lobes. These results support the utility of the proposed conditioning strategy for longitudinal PET simulation from cross-sectional training data.

Simulation of Disease Severity. To examine sensitivity to clinical biomarkers, we report counterfactual simulations in Fig. 4. We fix the baseline MRI (Age= 84, CDR-SB= 0), set a future age (Age= 88), and vary simulated CDR-SB from 0 to 6, yielding divergent PET patterns for the same individual. Difference maps show that higher CDR-SB selectively increases Aβ in the precuneus, produces tau hotspots in temporal cortex, and accentuates hypometabolism in parietal regions. This spatially specific response, rather than a global intensity shift, demonstrates variable-specific control enabled by hybrid clinical conditioning.

4 Conclusion

We propose **Uni-Brain**, a unified diffusion framework that preserves anatomy while controlling for tracer type and clinical covariates, enabling multi-tracer PET synthesis, follow-up PET generation at the target age, and counterfactual simulation under user-specified clinical conditions. By combining anatomy-aware spatial conditioning with hybrid clinical control, Uni-Brain achieves high-fidelity synthesis and improved consistency with follow-up PET and clinical biomarkers across Aβ, Tau, and FDG. These results suggest its potential as a practical foundation for cross-modal biomarker modeling and individualized PET simulation. Future work will incorporate temporal dynamics and generative diversity to enable more reliable multi-tracer PET synthesis.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers 62131015, U23A20295, 82441023, 82394432), the China Ministry of Science and Technology (STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), and HPC Platform of ShanghaiTech University.

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

References

1. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage **38**(1), 95–113 (2007)
2. Ashburner, J., Friston, K.J.: Unified segmentation. neuroimage **26**(3), 839–851 (2005)
3. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E., et al.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. Medical image analysis **86**, 102789 (2023)

4. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024)
5. 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)
6. Fefferman, C., Mitter, S., Narayanan, H.: Testing the manifold hypothesis. Journal of the American Mathematical Society **29**(4), 983–1049 (2016)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
8. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022)
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. Jack Jr, C.R., Andrews, J.S., Beach, T.G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., et al.: Revised criteria for diagnosis and staging of alzheimer's disease: Alzheimer's association workgroup. Alzheimer's & Dementia **20**(8), 5143–5169 (2024)
11. Jack Jr, C.R., Wiste, H.J., Weigand, S.D., Therneau, T.M., Lowe, V.J., Knopman, D.S., Gunter, J.L., Senjem, M.L., Jones, D.T., Kantarci, K., et al.: Defining imaging biomarker cut points for brain aging and alzheimer's disease. Alzheimer's & Dementia **13**(3), 205–216 (2017)
12. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**(2), 782–790 (2012)
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
14. Lee, J., Burkett, B.J., Min, H.K., Senjem, M.L., Dicks, E., Corriveau-Lecavalier, N., Mester, C.T., Wiste, H.J., Lundt, E.S., Murray, M.E., et al.: Synthesizing images of tau pathology from cross-modal neuroimaging using deep learning. Brain **147**(3), 980–995 (2024)
15. Mukherjee, S., Heath, L., Preuss, C., Jayadev, S., Garden, G.A., Greenwood, A.K., Sieberts, S.K., De Jager, P.L., Ertekin-Taner, N., Carter, G.W., et al.: Molecular estimation of neurodegeneration pseudotime in older brains. Nature communications **11**(1), 5781 (2020)
16. Ou, Z., Jiang, C., Pan, Y., Zhang, Y., Cui, Z., Shen, D.: A prior-information-guided residual diffusion model for multi-modal pet synthesis from mri. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. pp. 4769–4777 (2024)
17. Pan, Y., Liu, M., Xia, Y., Shen, D.: Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. IEEE transactions on pattern analysis and machine intelligence **44**(10), 6839–6853 (2021)
18. 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)
19. Puglisi, L., Alexander, D.C., Ravì, D., Initiative, A.D.N., et al.: Brain latent progression: Individual-based spatiotemporal disease progression on 3d brain mris via latent diffusion. Medical image analysis p. 103734 (2025)

20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
21. Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., et al.: Clinica: An open-source software platform for reproducible clinical neuroscience studies. Frontiers in neuroinformatics **15**, 689675 (2021)
22. Schmidt-Richberg, A., Ledig, C., Guerrero, R., Molina-Abril, H., Frangi, A., Rueckert, D., Initiative, A.D.N.: Learning biomarker models for progression estimation of alzheimer's disease. PloS one **11**(4), e0153040 (2016)
23. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=St1giarCHLP
24. Tafazoli, S., Bouchacourt, F.M., Ardalan, A., Markov, N.T., Uchimura, M., Mattar, M.G., Daw, N.D., Buschman, T.J.: Building compositional tasks with shared neural subspaces. Nature pp. 1–9 (2025)
25. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. Advances in neural information processing systems **33**, 7537–7547 (2020)
26. Yoon, J.S., Zhang, C., Suk, H.I., Guo, J., Li, X.: Sadm: Sequence-aware diffusion model for longitudinal medical image generation. In: International Conference on Information Processing in Medical Imaging. pp. 388–400. Springer (2023)
27. Zhou, T., Liu, M., Thung, K.H., Shen, D.: Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. IEEE transactions on medical imaging **38**(10), 2411–2422 (2019)
28. 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. pp. 2223–2232 (2017)