PETFlux: Taming Natural Foundation Models for Accurate MRI-to-PET Synthesis

Yuan Yin[1,2†], Honglin Xiong[1†], Haolin Huang[1], Yonghao Li[1], Yan Kong[4], Yitao Zhu[5], Zhuoxin Jiang[1], Kaicong Sun[1,3], Dinggang Shen[1,3], and Qian Wang[1,3*]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
qianwang@shanghaitech.edu.cn
[2] College of Biomedical Engineering, Fudan University, Shanghai, China
[3] Shanghai Clinical Research and Trial Center, Shanghai, China
[4] School of Intelligence Science and Technology, Nanjing University, Nanjing, China
[5] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong

Abstract. Synthesizing PET from MRI offers a radiation-free route to functional biomarkers but remains fundamentally ill-posed due to the weak coupling between anatomy and metabolism. Existing medical generative models trained on limited molecular imaging data often overfit or generate artifacts, while natural-image foundation models cannot be directly transferred because their priors and semantics are misaligned with radiological domains. We introduce PETFlux, a task-aligned adaptation framework that repurposes a pretrained natural-image backbone for the specific demands of MRI-to-PET synthesis. The framework addresses the key barriers of domain mismatch, structural-functional ambiguity, and spatial over-smoothing through three synergistic strategies: (i) Low-Rank Adaptation (LoRA) via rectified flow matching to realign natural priors with PET distributions; (ii) LLM-augmented prefix tuning grounded in an evidence-based clinical knowledge base to inject patient-specific metabolic semantics; and (iii) a log-scale focal frequency loss to preserve clinically critical functional detail. Experiments on the ADNI dataset demonstrate that PETFlux achieves state-of-the-art synthesis fidelity and yields substantial gains in downstream diagnostic classification, highlighting a principled pathway for repurposing natural foundation models in molecular imaging by effectively integrating anatomical constraints and patient metadata.

Keywords: Generative Foundation Model · MRI-to-PET Synthesis · Rectified Flow

1 Introduction

Positron Emission Tomography (PET) is pivotal for diagnosing and staging of neurodegenerative disorders such as Alzheimer's disease (AD) by quantifying glu-

[†]These authors contributed equally to this work. [*]Corresponding author.

cose metabolism or amyloid deposition before structural atrophy [12]. However, its high cost, radioactive exposure, and logistical constraints limit its accessibility in routine practice. Conversely, structural MRI is widely available, non-invasive, and routinely collected. Thus, MRI-to-PET synthesis holds substantial clinical value, offering a scalable, radiation-free pathway to functional biomarkers [25].

Despite its promise, MRI-to-PET synthesis is an exceptionally challenging problem. Beyond the severe scarcity of paired medical data, the mapping from structural anatomy to functional metabolism is inherently ill-posed: identical anatomical structures can exhibit markedly different metabolic patterns depending on disease stage, cognitive status, or tracer type. Traditional generative models (e.g., GANs [8], standard diffusion models [16]) trained on small datasets lack the inductive biases required to resolve this structural-functional ambiguity. Consequently, they often produce over-smoothed averages or artificial textures that are visually plausible but clinically misleading [29].

The rapid progress of large-scale foundation models offers a promising opportunity to overcome these limitations. Trained on billions of natural images, modern generation architectures such as Stable Diffusion [26] and Flux [5] encode rich generative priors over textures, spatial relationships, and global structures. Concurrently, Large Language Models (LLMs) have increasingly integrated domain-specific medical knowledge, offering sophisticated semantic guidance [28]. These priors exhibit strong generalization and capture expressive patterns that are difficult to learn from limited medical datasets, making such models a promising starting point for MRI-to-PET translation [19].

However, adapting natural foundation models to MRI-to-PET synthesis faces several fundamental obstacles. The first is the substantial domain gap: priors learned from natural photographs must be realigned to match the distribution of PET while preserving the expressive capacity. A second obstacle arises from the intrinsic ambiguity of the task: since MRI does not uniquely determine metabolic activity, the model requires patient-specific clinical metadata to guide the synthesis of plausible functional patterns. A third challenge concerns the emergence of artificial textures, as foundation models without appropriate constraints are prone to generating invalid details that compromise diagnostic reliability.

To systematically address these challenges, we introduce PETFlux, a task-aligned adaptation framework. To bridge the domain gap, we adapt the pre-trained Flux.1 Kontext [1] backbone using Low-Rank Adaptation (LoRA) within a rectified flow matching paradigm. To resolve structural-functional ambiguity, we incorporate patient-specific metadata and LLM-derived metabolic descriptions through a hybrid semantic conditioning mechanism. Crucially, to ensure these descriptions are physiologically rigorous rather than speculative, we guide the LLM using a curated medical knowledge base derived from established clinical guidelines (e.g., mapping cognitive scores to expected regional metabolism). To penalize these artificial textures, we introduce a log-scale focal frequency loss that explicitly emphasizes clinically relevant high-frequency components in the spectral domain. Together, these components form a coherent paradigm for taming natural foundation models for reliable medical image synthesis. Exten-

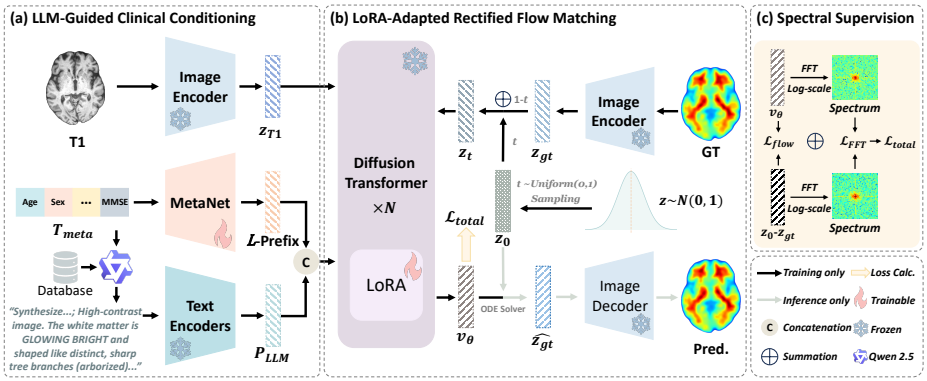

Fig. 1. Overview of the PETFlux framework for MRI-to-PET synthesis. (a) Clinical metadata is projected via MetaNet into soft tokens and concatenated with LLM embeddings (P_{LLM}) for hybrid semantic conditioning. (b) The pretrained foundation model is adapted using LoRA via rectified flow matching. (c) The training process is further guided by a Log-Scale Focal Frequency Loss to restore fine metabolic details.

sive experiments validate our framework — adapting the pretrained Flux alone outperforms most baselines, while integrating our LLM-guided conditioning and spectral supervision further maximizes synthesis fidelity and diagnostic reliability.

In summary, our main contributions are: (1) PETFlux, a novel paradigm that adapts billion-parameter priors for ill-posed MRI-to-PET synthesis via Rectified Flow matching; (2) a hybrid conditioning mechanism resolving structural-functional ambiguity by injecting patient-specific metabolic priors from clinical guidelines and LLMs; and (3) a log-scale focal frequency loss that explicitly preserves clinically valid metabolic signatures while penalizing artificial textures.

2 Method

2.1 Problem Formulation

Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^{H \times W}$ denote the domains of structural MRI slices and their corresponding PET images, respectively. For each subject, we denote the input MRI slice as $x \in \mathcal{X}$ and the target PET slice as $y \in \mathcal{Y}$. To capture subject-specific metabolic and clinical context, we associate each sample with a clinical metadata vector T_{meta} (e.g., age, sex, CDR-SB, MMSE). Our objective is to learn a conditional generative model G_θ that synthesizes a PET estimate $\hat{y}$ consistent with both the anatomical structure in x and the hybrid semantic patient context derived from T_{meta} and its corresponding clinical narrative: $\hat{y} = G_\theta(x, T_{meta})$.

The overall pipeline of the proposed framework is illustrated in Fig. 1. We detail the specific implementations of each component in the subsequent sections.

2.2 LLM-Guided Clinical Conditioning

MRI-to-PET synthesis is fundamentally ill-posed because anatomical structure alone does not uniquely determine metabolic function. To constrain the solution space toward physiologically plausible metabolic patterns without causing data leakage via diagnostic labels (e.g., NC, MCI, AD), we introduce a hybrid semantic conditioning mechanism. First, we construct a curated medical knowledge base $\mathcal{D}$ that maps clinical indicators (e.g., MMSE, CDR-SB) to expected regional metabolic alterations. Using $\mathcal{D}$, the tracer type t, and non-categorical metadata T_{meta} (e.g., age, sex), we query an LLM (Qwen 2.5) with a decoding temperature of zero to generate a deterministic clinical narrative. Crucially, the LLM acts as a semantic bridge, translating abstract clinical metadata into visual-descriptive prompts (e.g., specifying contrast, intensity, and spatial patterns) that the natural foundation model can inherently interpret. This narrative is then encoded by frozen text encoders into a fixed semantic embedding P_{LLM}.

While P_{LLM} provides robust qualitative priors, natural language lacks the continuous, quantitative precision of raw numerical metadata. To preserve this fine-grained precision, we complement the textual embeddings with learnable soft tokens, L-Prefix. A lightweight MetaNet projector $\Psi(\cdot)$, implemented as a simple MLP, maps the normalized T_{meta} into $N = 8$ continuous tokens:

$$L\text{-Prefix} = \Psi(T_{meta}) \in \mathbb{R}^{N \times d}.$$

We construct the final semantic condition by concatenating these representations: $[\, L\text{-Prefix} \; ; \; P_{LLM} \,]$. The MetaNet's final projection layer is near-zero initialized, allowing the tokens to initially act as neutral padding. As the MetaNet is jointly optimized with the LoRA modules during training, these prefix tokens progressively learn to encode subtle, quantitative clinical variations. This hybrid conditioning explicitly prepares the tailored semantic inputs (P_{LLM} and L-Prefix) required to guide the subsequent rectified flow matching process.

2.3 LoRA-Adapted Rectified Flow Matching

To bridge the natural-to-medical domain gap while preserving Flux's structural priors, we formulate MRI-to-PET translation via Rectified Flow (RF) matching. Unlike standard diffusion, RF learns a deterministic, straight-line trajectory from noise to the target data distribution via ordinary differential equations.

A frozen image encoder first projects the input T1-weighted MRI and ground-truth PET image into latents z_{T1} and z_{gt}. Semantic conditioning is provided by the LLM text embeddings P_{LLM} and MetaNet-generated soft tokens (L-Prefix). During training, we sample noise $z_0 \sim \mathcal{N}(0, I)$ and compute the intermediate latent z_t at a uniformly sampled time step $t \sim \text{Uniform}(0, 1)$:

$$z_t = (1 - t)z_{gt} + tz_0$$

The target velocity driving this straight-line flow is naturally defined as $v = z_0 - z_{gt}$.

To prevent catastrophic forgetting of pre-trained priors, we freeze the Flux Diffusion Transformer (DiT) backbone and achieve structural-functional alignment by injecting trainable Low-Rank Adaptation (LoRA) matrices into the Transformer blocks. The adapted network predicts the velocity field v_θ conditioned on the noisy latent, time step, and the hybrid anatomical-semantic inputs:

$$v_\theta = DiT_{ada}(z_t, z_{T1}, P_{LLM}, L\text{-Prefix}, t)$$

The flow matching objective minimizes the mean squared error between the predicted and target velocities:

$$\mathcal{L}_{flow} = \mathbb{E}\left[\|v_\theta - (z_0 - z_{gt})\|_2^2\right]$$

This parameter-efficient formulation smoothly warps the pre-trained natural generative manifold to the radiological PET distribution with minimal computational overhead. In practice, we apply LoRA with a rank of $r = 128$ within the Flux Transformer blocks to efficiently bridge the domain shift.

2.4 Spectral Supervision

While the spatial-domain flow matching objective $\mathcal{L}_{flow}$ aligns global structure, MSE-based losses often lead to over-smoothed results lacking fine metabolic details. To mitigate this, we introduce a latent-space spectral regularization that ensures both memory efficiency and stable optimization.

Let $A(v)$ and $A(v_\theta)$ denote the amplitude spectra of the target velocity $v = z_0 - z_{gt}$ and the predicted velocity v_θ computed via 2D Fast Fourier Transform (FFT). To prevent low-frequency dominance, we apply a logarithmic transform $L(\cdot) = \log(A(\cdot) + 1)$ and define the spectral discrepancy as $\Delta = |L(v) - L(v_\theta)|$. Inspired by the focal frequency loss [15], we employ a focal weighting mechanism to explicitly penalize hard-to-learn high-frequency bands:

$$\mathcal{L}_{FFT} = \frac{1}{HW}\sum_{u,v}(\Delta_{u,v})^\gamma \cdot \Delta_{u,v},$$

where γ controls the focusing degree. The final objective is $\mathcal{L}_{total} = \mathcal{L}_{flow} + \alpha(s)\mathcal{L}_{FFT}$, where the balancing weight $\alpha(s) = \lambda_{\max}(1 - e^{-\beta s})$ follows an exponential warm-up schedule based on the training step s. This curriculum allows the model to establish global structure early on, gradually increasing the spectral constraint to refine intricate metabolic details.

3 Experiments

3.1 Dataset and Implementation Details

Dataset. We evaluate paired MRI/PET data from ADNI [14,24] for two tracers: ^{18}F-FDG (1,549 subjects) and ^{18}F-AV45 (1,124 subjects). We use a strict

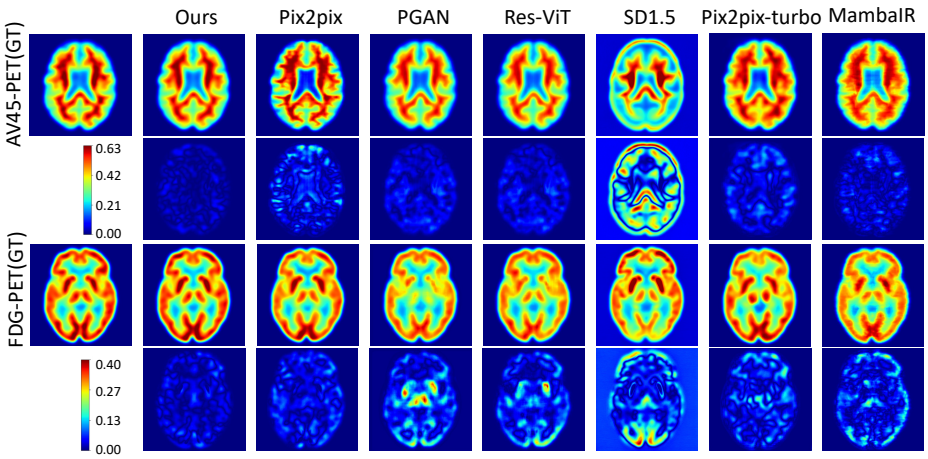

Fig. 2. Qualitative comparison of PET image synthesis. The top two rows display the synthesized $A\beta$-PET images and their corresponding error maps from a representative subject from the test set. The bottom two rows present the FDG-PET synthesis results and error maps under the same configuration.

8:1:1 subject-independent train/validation/test split without leakage. All volumes are skull-stripped, registered to MNI152, normalized to $[0, 1]$, and extracted as 512×512 2D slices to match Flux's native resolution.

LLM Prompting. The medical knowledge base $\mathcal{D}$ maps cognitive assessments [13,24,11,2,22] (e.g., MMSE, CDR-SB) to expected regional metabolic alterations based on established clinical guidelines [3,6,20,21]. Querying the Qwen with $\mathcal{D}$ and patient metadata synthesizes a specific narrative p_{desc} (e.g., severe cognitive decline with MMSE < 15 yields "widespread bilateral temporoparietal hypometabolism" for FDG, or "elevated cortical amyloid deposition" for AV45). This mitigates the inherent structural-functional ambiguity by providing a robust, clinically grounded semantic baseline for our dynamic prefix tuning.

Training Details. Built upon the Flux backbone, we train the network with AdamW on 4 NVIDIA A100 GPUs for 30,000 iterations using a batch size of 4 and learning rate 1×10^{-4}. The 12.2B-parameter backbone is frozen, and LoRA updates only 298.8M parameters (2.45%).

3.2 Comparisons with State-of-the-Arts

To validate our approach, we compare PETFlux with strong baselines across GANs (pix2pix, PGAN), Transformers (ResViT), state-space models (MambaIR), and diffusion-based models (SD v1.5, pix2pix-turbo).

Table 1 and Fig. 2 present the comparisons. While traditional GANs yield reasonable structural scores, they suffer from inferior perceptual quality. ResViT achieves strong metrics but produces over-smoothed textures that obscure fine pathological details. Although advanced models (MambaIR, pix2pix-turbo) offer

Table 1. Quantitative comparison of MRI-to-PET synthesis and downstream classification performance. MAE_S denotes the Mean Absolute Error of SUVR. Best results are **bold**, second-best results are underlined.

Method	Mod.	Synthesis Metrics				Classification	
		PSNR↑	SSIM↑	LPIPS↓	MAE_S ↓	ACC↑	AUC↑
pix2pix [10]	AV45	21.13	0.8831	0.1526	0.1399	0.7432	0.5824
	FDG	26.31	0.9615	0.0410	0.0558	0.8251	0.6693
PGAN [17]	AV45	23.89	0.9571	0.0613	0.1526	0.8074	0.5412
	FDG	24.87	0.9152	0.0654	0.2188	0.7935	0.5068
ResViT [4]	AV45	24.66	0.9654	0.0521	0.2125	0.8314	0.6537
	FDG	25.04	0.9638	0.0683	0.0631	0.7672	0.5461
SD v1.5 [26,31]	AV45	15.46	0.3025	0.2958	0.2382	0.6548	0.6023
	FDG	18.08	0.4251	0.1832	0.1115	0.7226	0.5765
pix2pix-turbo [23]	AV45	23.10	0.9217	0.1205	0.1506	0.8236	0.7194
	FDG	24.85	0.9453	0.1137	0.0545	0.8003	0.4731
MambaIR [7]	AV45	23.89	0.9309	0.1983	0.1506	0.8231	0.7215
	FDG	25.93	0.9452	0.1866	0.0600	0.8129	0.6991
PETFlux **(Ours)**	AV45	**26.43**	**0.9736**	**0.0412**	**0.0889**	**0.8415**	**0.7946**
	FDG	**27.37**	**0.9772**	**0.0335**	**0.0500**	**0.8572**	**0.7684**
Real PET	AV45	-	-	-	-	0.8412	0.8583
	FDG	-	-	-	-	0.8266	0.8427

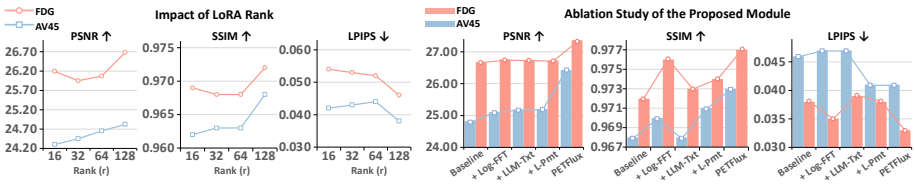

Fig. 3. Ablation studies of PETFlux. Left: Impact of LoRA rank. Right: Component-wise analysis of Log-FFT, LLM-Txt, and L-Pmt; PETFlux denotes the full model.

competitive fidelity, they struggle with complex structural-functional mapping. Furthermore, standard diffusion (SD v1.5) suffers severe performance degradation due to the profound domain gap.

In contrast, PETFlux achieves state-of-the-art overall synthesis performance. By leveraging foundation priors alongside strict spectral supervision, it significantly outperforms all baselines in both perceptual quality and structural fidelity. To rigorously evaluate clinical utility, we compute the Mean Absolute Error of the Standardized Uptake Value Ratio (MAE_S) across an established cortical composite ROI comprising five AD-core regions (frontal, cingulate, parietal, temporal, and precuneus) [18]. Crucially, PETFlux achieves the lowest MAE_S, demonstrating its exceptional ability to synthesize clinically accurate metabolic distributions rather than merely visually plausible textures.

Downstream Diagnostic Utility. We evaluate practical utility by feeding the synthesized images into a diagnostic classifier pre-trained on real PET data to distinguish AD from NC/MCI [9]. As shown in Table 1, existing baselines perform inconsistently across tracers and yield suboptimal AUC scores. Conversely, PETFlux achieves the highest ACC and AUC among all synthesis methods, confirming that our LLM-augmented conditioning effectively imbues the generated images with accurate, disease-specific metabolic signatures.

3.3 Ablation Studies

Impact of LoRA Rank. We evaluate the adaptation rank r in Fig. 3 (left). Performance improves steadily up to $r = 128$, indicating that the MRI-to-PET domain shift requires sufficient adaptation capacity rather than very low-rank updates. Although larger ranks may bring marginal gains, we cap r at 128 to balance synthesis quality, VRAM consumption, and overfitting risk on limited medical data.

Component Analysis. We validate the individual contributions of our key modules: the Log-Scale Focal Frequency Loss (Log-FFT), LLM-derived textual conditioning (LLM-Txt), and learnable soft prompts (L-Pmt). As shown in Fig. 3, the baseline Flux backbone with standard spatial loss yields suboptimal structural fidelity. Integrating the Log-FFT Loss significantly boosts PSNR and SSIM across both tracers by adaptively penalizing high-frequency synthesis errors and suppressing artifacts.

Furthermore, the explicit clinical narrative (LLM-Txt) effectively translates medical indices into visual semantics (as illustrated in Fig. 1), aligning clinical expectations with the backbone's natural generative priors to provide a stable semantic anchor. The most notable enhancement in perceptual quality occurs with the addition of the dynamically learnable prefix tokens (L-Pmt). This ablation supports our hypothesis that LLM text sets the general metabolic expectation, while learnable soft prompts provide the precise mapping needed to resolve structural-functional ambiguity and synthesize tracer-specific details.

4 Conclusion

We presented PETFlux, a framework adapting natural foundation models for inherently ill-posed MRI-to-PET synthesis. By combining LLM-guided clinical metadata conditioning with an empirical log-scale focal frequency loss, PETFlux mitigates structural-functional ambiguity and spatial over-smoothing. Evaluations on ADNI demonstrate state-of-the-art synthesis fidelity and superior downstream diagnostic utility across multiple tracers, while external validation and blinded expert assessment remain necessary before clinical translation. Inspired by recent LLM-assisted diagnostic prompting and cross-modality fusion studies [27,30], future work could explore multi-step semantic prompts, region-aware metabolic guidance, and MoE-style conditioning for heterogeneous tracers, cohorts, and acquisition protocols.

Acknowledgements. This work was partially supported by National Natural Science Foundation of China (82394432), the AI4S Initiative and HPC Platform of ShanghaiTech University.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Black Forest Labs, Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., et al.: FLUX.1 Kontext: Flow matching for in-context image generation and editing in latent space. arXiv preprint arXiv:2506.15742 (2025)
2. Bohnen, N.I., Djang, D.S., Herholz, K., et al.: Effectiveness and safety of 18F-FDG PET in the evaluation of dementia: a review of the recent literature. Journal of Nuclear Medicine **53**(1), 59–71 (2012)
3. Clark, C.M., Schneider, J.A., Bedell, B.J., Beach, T.G., et al.: Use of florbetapir-PET for imaging β-amyloid pathology. JAMA **305**(3), 275–283 (2011)
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. Esser, P., Kulal, S., Blattmann, A., Entezari, R., et al.: Scaling rectified flow transformers for high-resolution image synthesis (2024)
6. Foster, N.L., Heidebrink, J.L., Clark, C.M., et al.: FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. Brain **130**(10), 2616–2635 (2007)
7. Guo, H., Li, J., Dai, T., et al.: MambaIR: A simple baseline for image restoration with state-space model. In: European conference on computer vision. pp. 222–241 (2024)
8. Hu, S., Lei, B., Wang, S., et al.: Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. IEEE Transactions on Medical Imaging **41**(1), 145–157 (2021)
9. Huang, H., Shen, Z., Wang, J., et al.: MetaAD: Metabolism-aware anomaly detection for Parkinson's disease in 3d 18F-FDG PET. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 291–301 (2024)
10. 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)
11. Jack, C.R., Knopman, D.S., Jagust, W.J., et al.: Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology **12**(2), 207–216 (2013)
12. Jack Jr, C.R., Andrews, J.S., Beach, T.G., et al.: Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's association workgroup. Alzheimer's & Dementia **20**(8), 5143–5169 (2024)
13. Jack Jr, C.R., Bennett, D.A., Blennow, K., et al.: NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia **14**(4), 535–562 (2018)
14. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging **27**(4), 685–691 (2008)

15. Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 13919–13929 (2021)
16. Jiang, Y., Jin, Y., Chen, Q., et al.: A text-information-controlled diffusion model for multimodal whole-body PET synthesis from MRI. In: Fifth International Conference on Biomedicine and Bioinformatics Engineering (ICBBE 2025). vol. 13971, pp. 102–107 (2025)
17. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)
18. Landau, S.M., Breault, C., Joshi, A.D., et al.: Amyloid-β imaging with pittsburgh compound b and florbetapir in Alzheimer's disease. Annals of Neurology **72**(4), 578–586 (2012)
19. Mao, Y., Li, H., Pang, W., Papanastasiou, G., Yang, G., Wang, C.: Selora: Self-expanding low-rank adaptation of latent diffusion model for medical image synthesis. arXiv preprint arXiv:2408.07196 (2024)
20. McKhann, G.M., Knopman, D.S., Chertkow, H., et al.: The diagnosis of dementia due to Alzheimer's disease: recommendations from the NIA-AA workgroups. Alzheimer's & dementia **7**(3), 263–269 (2011)
21. Minoshima, S., Giordani, B., Berent, S., Frey, K.A., Foster, N.L., Kuhl, D.E.: Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. Annals of Neurology **42**(1), 85–94 (1997)
22. Morris, J.C.: The clinical dementia rating (CDR) current version and scoring rules. Neurology **43**(11), 2412–2412 (1993)
23. Parmar, G., Park, T., Narasimhan, S., Zhu, J.Y.: One-step image translation with text-to-image models. arXiv preprint arXiv:2403.12036 (2024)
24. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., et al.: Alzheimer's disease neuroimaging initiative (ADNI) clinical characterization. Neurology **74**(3), 201–209 (2010)
25. Rao, V.M., Hla, M., Moor, M., et al.: Multimodal generative AI for medical image interpretation. Nature **639**(8056), 888–896 (2025)
26. 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)
27. Shi, J., Li, C., Gong, T., Wang, C., Fu, H.: CoD-MIL: chain-of-diagnosis prompting multiple instance learning for whole slide image classification. IEEE Transactions on Medical Imaging **44**(3), 1218–1229 (2025)
28. Shool, S., Adimi, S., Saboori Amleshi, R., et al.: A systematic review of large language model (LLM) evaluations in clinical medicine. BMC Medical Informatics and Decision Making **25**(1), 117 (2025)
29. Vaccaro, M., Rosa, E., Placidi, E., et al.: Deep learning for synthetic PET imaging: a systematic mapping review of techniques, metrics, and clinical relevance. European Radiology Experimental **10**(1), 12 (2026)
30. Yang, S., Zhou, H., Yang, Y., Wang, W., Chen, S., Yang, G., Fu, H., Zhu, L.: LCM-Net: LLM-driven cross-modality MoE feature fusion network for cancer survival analysis. IEEE Transactions on Medical Imaging (2026)
31. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3836–3847 (2023)