# PASTA: Pathology-Aware MRI to PET CroSs-modal TrAnslation with Diffusion Models

Yitong Li[1,2][0009−0009−3874−6055], Igor Yakushev[3][0000−0003−4764−798X], Dennis M. Hedderich[4][0000−0001−8994−5593], and Christian Wachinger[0000−0002−3652−1874][1,2]

[1] Lab for Artificial Intelligence in Medical Imaging, Technical University of Munich (TUM), Germany
[2] Munich Center for Machine Learning (MCML), Germany
[3] Department of Nuclear Medicine, Klinikum rechts der Isar, TUM, Germany
[4] Department of Neuroradiology, Klinikum rechts der Isar, TUM, Germany
`yi_tong.li@tum.de`

**Abstract.** Positron emission tomography (PET) is a well-established functional imaging technique for diagnosing brain disorders. However, PET's high costs and radiation exposure limit its widespread use. In contrast, magnetic resonance imaging (MRI) does not have these limitations. Although it also captures neurodegenerative changes, MRI is a less sensitive diagnostic tool than PET. To close this gap, we aim to generate synthetic PET from MRI. Herewith, we introduce PASTA, a novel pathology-aware image translation framework based on conditional diffusion models. Compared to the state-of-the-art methods, PASTA excels in preserving both structural and pathological details in the target modality, which is achieved through its highly interactive dual-arm architecture and multi-modal condition integration. A cycle exchange consistency and volumetric generation strategy elevate PASTA's capability to produce high-quality 3D PET scans. Our qualitative and quantitative results confirm that the synthesized PET scans from PASTA not only reach the best quantitative scores but also preserve the pathology correctly. For Alzheimer's classification, the performance of synthesized scans improves over MRI by 4%, almost reaching the performance of actual PET. Code is available at https://github.com/ai-med/PASTA.

## 1 Introduction

To date, various tools support diagnosing brain disorders, including MRI, PET, and cognitive tests [1]. Structural MRI reveals brain regional atrophy, while fluorodeoxyglucose PET tracks glucose metabolism. In Alzheimer's disease (AD) and related brain disorders, glucose uptake drops severely in certain brain regions [15]. By sensitively reflecting functional disorders, PET has proven to have higher accuracy in early dementia detection and differential diagnosis [19, 3].

Despite its high diagnostic value, PET is commonly not offered in most medical centers worldwide due to its high cost and high-dose radiation exposure [13].
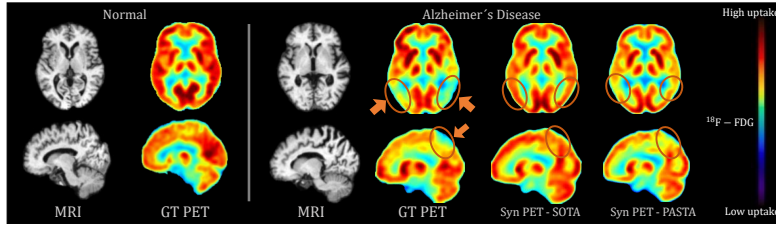
Fig. 1: For Alzheimer's disease, PET shows reduced glucose uptake in the temporoparietal lobe (bottom circles), mirroring atrophy on MRI with higher sensitivity. While state-of-the-art diffusion models fail to recover such pathology in the synthesized PET, PASTA achieves significant improvements.

While MRI is more accessible thanks to its non-invasive nature, it lacks the same functional insights as PET [7]. We aim to bridge this gap by converting MRI into synthetic PET, thereby extending the reach of functional brain imaging to facilitate AD diagnosis. Translating structural to functional imaging in medicine is challenging due to the possible inclusion of unrealistic pathology, leading to unreliable diagnoses. Yet, generating *pathology-preserved* images remains an underexplored area with significant potential for clinical impact.

Previous works on MRI to PET translation focus on adopting generative adversarial networks (GAN) [11, 34]. However, GANs are susceptible to mode collapse and unstable training [2], which can reduce the model performance and reliability. Recent advances in diffusion models (DM) [28, 8] introduced a new era in high-quality image generation with enhanced training stability, giving rise to advanced image translation frameworks [18, 25]. Despite these strides, existing DM-based translation methods primarily focus on preserving structural integrity, overlooking the critical aspect of pathology recovery, as shown in Fig. 1.

We address this gap by introducing PASTA, an efficient end-to-end DM-based framework for clinically valuable pathology-aware volumetric MRI to PET translation. It is based on a symmetric dual-arm architecture with adaptive conditional modules for multi-modal condition integration. PASTA also presents a memory-efficient approach for artifact-free volumetric generation and a cycle exchange consistency strategy, further lifting its generation quality.

In summary, we make the following contributions:

- A novel end-to-end cross-modal MRI to PET translation framework based on diffusion models with volumetric generation.
- Adaptive normalization layers for integrating multi-modal conditions to facilitate pathology awareness.
- Cycle exchange consistency for effective conditional DMs training.
- Quantitative and qualitative experiments show that PASTA achieves low reconstruction errors and preserves AD pathology to boost diagnosis accuracy.

**Related Work** Previous research on cross-modal MRI to PET translation mainly focused on GAN-based methods [32, 16, 33, 10, 26, 4, 27], with innovations
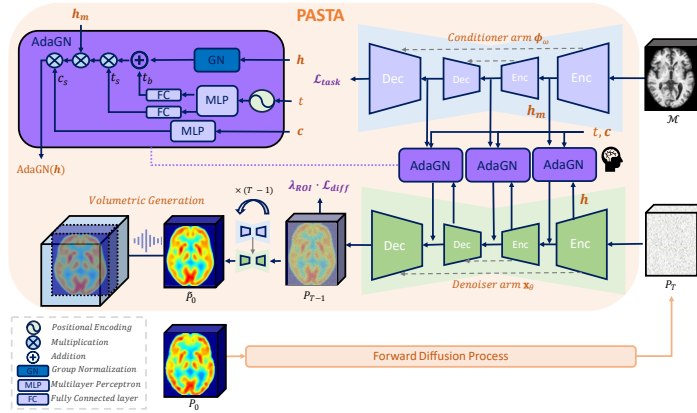
Fig. 2: PASTA conditions the feature maps $h$ from the denoiser arm on timestep $t$, clinical data $c$, and task representation $h_m$ from the conditioner arm.

like the sketcher-refiner scheme [32], GANDALF for MRI to PET generation in AD diagnosis [27], bidirectional GAN for 3D brain MRI-to-PET synthesis [10], and a GAN-based residual vision Transformers for multimodal medical image synthesis [4]. Diffusion models outperform GANs in capturing complex distributions [28, 8, 5] and are emerging in medical imaging for unconditional generation and cross-contrast MRI translation [35, 22, 21]. Their application to MRI to PET translation, unexplored thus far, offers a promising avenue.

## 2   Proposed Method

Given two datasets $\mathcal{X}_A$ and $\mathcal{X}_B$ from modalities A and B, cross-modal image translation aims to learn a mapping from A to B in a paired manner. In medical imaging, this process is constrained on generating images that both match the ground truth (GT) and preserve pathology evidence. However, current DM-based translation methods focus on style transfer and structural maintenance [31, 14], which are insufficient for medical translation, as shown in Fig. 1.

Hence, we propose establishing a strong interaction with the input modality on the conditional denoising diffusion probabilistic models (DDPM) by integrating input features at multiple scales, giving rise to a symmetric dual-arm architecture. PASTA consists of a conditioner arm, a denoiser arm, and adaptive conditional modules, as shown in Fig. 2. Both arms adopt symmetric U-Net [24]. The interaction between the two arms and the fusion of additional conditions are achieved through the adaptive group normalization layers (AdaGN) [5]. This symmetric design ensures that matching blocks across the two arms share the same spatial resolution, enabling multi-scale feature map interactions.

**Conditioner Arm:** We define our training data as $\mathcal{D}_\mathrm{T} = (\mathcal{M}_\mathrm{T}^i, \mathcal{P}_\mathrm{T}^i)_{i=1}^N$, which comprises $N$ pairs of MRI $\mathcal{M} \in \mathbb{R}^{H \times W \times D}$ and its corresponding PET

$\mathcal{P} \in \mathbb{R}^{H \times W \times D}$. The objective is to learn a model $\boldsymbol{G}(\cdot)$ on $\mathcal{D}_\mathrm{T}$, such that given any unseen MRI input $\mathcal{M}_\mathrm{s} \notin \mathcal{D}_\mathrm{T}$, its PET counterpart is inferred as $\mathcal{P}_\mathrm{s} = \boldsymbol{G}(\mathcal{M}_\mathrm{s})$. The MRI input $\mathcal{M}$ first goes through the conditioner arm to derive multi-scale task-specific representations $\boldsymbol{h}_m$, given by $\hat{\mathcal{M}} = \phi_\omega(\mathcal{M}; \boldsymbol{h}_m)$, where $\phi_\omega(\cdot)$ is the conditioner model parameterized by $\omega$. The task-specific representations $\boldsymbol{h}_m = \{\boldsymbol{h}_m^1, \ldots, \boldsymbol{h}_m^n\}$ includes the intermediate feature maps from $\phi_\omega(\cdot)$ at multiple scales, with $n$ the number of residual blocks in the conditioner arm. These representations will facilitate the PET synthesis in the other arm. The conditioner arm executes predefined tasks to convert MRI into $\boldsymbol{h}_m$, including MRI reconstruction and MRI-to-PET translation, with each task dictating its own training objective. Taking MRI-to-PET translation as the predefined task, the arm is trained to minimize the pixel-level distance between the original PET and the conditioner output, using distance metric $dist(\cdot)$ like $L_1$ or $L_2$:

$$\mathcal{L}_{task}(\omega) = \mathbb{E}_{\mathcal{M},\mathcal{P}} \ dist(\phi_\omega(\mathcal{M}), \mathcal{P}). \tag{1}$$

We employ MRI-to-PET translation as the predefined task, as it gives the best empirical performance. The efficacy of alternative tasks is investigated in Sec. 4.

**Adaptive Conditional Module:** We propose to use the adaptive group normalization layers (AdaGN) [5] to apply multi-modal conditions to the feature maps $\boldsymbol{h} = \{\boldsymbol{h}^1, \ldots, \boldsymbol{h}^n\}$ in each residual block from the denoiser arm. Our AdaGN layer adapts conditions of: 1) Timestep $t$ in the diffusion process; 2) Task-specific representations $\boldsymbol{h}_m$ from the conditioner arm at corresponding scales; 3) Clinical data $\boldsymbol{c} \in \mathbb{R}^{c \times n}$ of an individual subject. Our AdaGN is defined as:

$$\mathrm{AdaGN}(\boldsymbol{h}, t, \boldsymbol{c}, \boldsymbol{h_m}) = \boldsymbol{c_s}(\boldsymbol{h_m}(\boldsymbol{t_s} \ \mathrm{GroupNorm}(\boldsymbol{h}) + \boldsymbol{t_b}), \tag{2}$$

where $(\boldsymbol{t_s}, \boldsymbol{t_b}) \in \mathbb{R}^{2 \times c} = \mathrm{MLP}(pos(t))$ is the output of a multilayer perceptron (MLP) with a sinusoidal encoding function $pos(\cdot)$, and $\boldsymbol{c_s} = \mathrm{MLP}(\boldsymbol{c})$. These modules are used throughout the dual-arm architecture. To enhance the accuracy and pathology preservation in generated PET scans, we incorporate available clinical data into our model, including demographic information (age, gender, education), cognitive scores (MMSE [6], ADAS-Cog-13 [20]), and AD biomarker ApoE4 [30]. Overall, the AdaGN layer fuses the multi-modal conditions including both structural and pathological evidence.

**Denoiser Arm** performs the reverse process of DDPM by restoring the clean PET $\mathcal{P}_0$ from the noise. It produces $\mathcal{P}_t$ at each diffusion timestep $t$ starting from the Gaussian noise $\epsilon$ at $t = T$, conditioned on multi-modal variables via AdaGN:

$$\mathcal{P}_{0:T} = \mathcal{P}_T \prod_{t=1}^{T} \mathbf{x}_\theta \left( \mathcal{P}_{t-1} \mid \boldsymbol{h}_m, \boldsymbol{c} \right), \tag{3}$$

where $\mathbf{x}_\theta(\cdot)$ denotes the denoising model parameterized by $\theta$. The symmetric layout of PASTA allows the feature maps in $\mathbf{x}_\theta$ at each scale to be conditioned by an equivalent scale of the task representation from the conditioner arm, which augments the impact of the conditional modality. To further enhance pathology awareness, we integrate MetaROIs [17] as pathology priors to guide the model
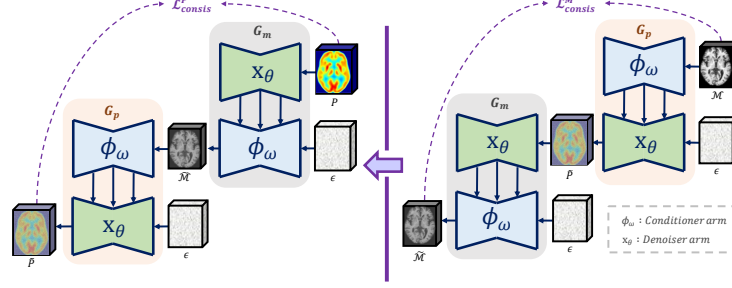
Fig. 3: Cycle exchange consistency (CycleEx) strategy of PASTA.

to the important hypometabolic regions of abnormal metabolic changes in AD patients. MetaROIs are a set of pre-defined regions of interest based on coordinates cited frequently in other PET studies comparing AD and normal subjects. We convert MetaROIs into a loss weighting map $\boldsymbol{\lambda}_R \in \mathbb{R}^{H \times W \times D}$ and add to the denoiser arm. During training, the MetaROIs of the denoised PET images will be more penalized when they deviate from the GT PET, given by:

$$\mathcal{L}_{diff}^t(\theta) = \boldsymbol{\lambda}_R \cdot \mathbb{E}_{\mathcal{P}_0, \epsilon}[\|\mathcal{P}_0 - \mathbf{x}_\theta(\alpha_t \mathcal{P}_0 + \sigma_t \epsilon, \boldsymbol{c}, \boldsymbol{h}_m)\|_2^2]. \quad (4)$$

**Cycle Exchange Consistency (CycleEx):** PASTA goes through a CycleEx strategy shown in Fig. 3. CycleEx stems from the idea in [34], where the learned translation mapping follows cycle-consistency: for each image $\mathcal{M}^i$ from the MRI domain, given two mappings $\boldsymbol{G}_p$: $\mathcal{M} \to \mathcal{P}$ and $\boldsymbol{G}_m$: $\mathcal{P} \to \mathcal{M}$, the image translation cycle should bring $\mathcal{M}^i$ back to its original image, i.e. $\mathcal{M}^i \to \boldsymbol{G}_p(\mathcal{M}^i) \to \boldsymbol{G}_m(\boldsymbol{G}_p(\mathcal{M}^i)) \approx \mathcal{M}^i$, as forward cycle consistency. Similarly, for each image $\mathcal{P}^i$ from the PET domain, $\boldsymbol{G}_p$ and $\boldsymbol{G}_m$ should also satisfy backward cycle consistency: $\mathcal{P}^i \to \boldsymbol{G}_m(\mathcal{P}^i) \to \boldsymbol{G}_p(\boldsymbol{G}_m(\mathcal{P}^i)) \approx \mathcal{P}^i$. In the CycleEx, the mapping $\boldsymbol{G}_m$ shares the same network as $\boldsymbol{G}_p$, but only with the exchanged conditioner and denoiser arm: in the forward cycle, the conditioner arm $\phi_\omega$ used to process MRI to $\boldsymbol{h}_m$ during mapping $\boldsymbol{G}_p$, will be reused for denoising to synthesize MRI in $\boldsymbol{G}_m$; the denoiser arm $\mathbf{x}_\theta$ used to synthesize PET in $\boldsymbol{G}_p$ will be reused to process PET to representations $\boldsymbol{h}_p$ in $\boldsymbol{G}_m$. The backward cycle acts similarly. Due to the symmetric nature of the two arms, this exchange can be achieved conveniently. This cycling approach requires only two trainable networks $\phi_\omega$ and $\mathbf{x}_\theta$. CycleEx introduces three more conditional diffusion processes without adding additional learnable parameters. Both cycles introduce the cycle-consistency loss:

$$\mathcal{L}_{cycle}(\omega, \theta) = \mathbb{E}_\mathcal{M} \; dist(\boldsymbol{G}_m(\boldsymbol{G}_p(\mathcal{M})), \mathcal{M}) + \mathbb{E}_\mathcal{P} \; dist(\boldsymbol{G}_p(\boldsymbol{G}_m(\mathcal{P}), \mathcal{P}), \quad (5)$$

where $dist(\cdot)$ can be $L_1$ or $L_2$. Such a setup enforces the information sharing between the two arms, adding additional supervision and regularization to the image translation. Finally, the combined training objective of PASTA is:

$$\mathcal{L} = \lambda_{task} * \mathcal{L}_{task} + \lambda_{diff} * \mathcal{L}_{diff} + \lambda_{cycle} * \mathcal{L}_{cycle}, \quad (6)$$

where $\lambda_{task}$, $\lambda_{diff}$, and $\lambda_{cycle}$ are constant factors.

**Volumetric Generation:** PASTA introduces a 2.5D strategy for volumetric synthesis. A full 3D network offers inherent inter-slice consistency but is limited by high computational demands. Moreover, the shortage of paired, multimodal medical data hinders its proper training. Thus, we adopt 2D convolutional layers in the network with 2D slices as input, but cater to the input channel with $N$ consecutive neighboring slices of each input slice along the same axis. After training, the network produces the target slice with its $N$ neighbors. These neighbors are weighted by the distance to the target slice, with closer ones higher. Summing all the weighted slices and averaging overlaps for each slice position yields a consistent 3D scan. This strategy enables the model to efficiently mitigate the slice inconsistencies introduced by the 2D network.

## 3    Experiments

**Models and Hyperparameters:** We adopt U-Net for diffusion models as in [5] and DDIM sampling strategy [29] with timestep $T = 1000$, input neighboring slices number $N = 15$, and $\lambda_{task} = 0.1$, $\lambda_{diff} = 1$, $\lambda_{cycle} = 1$ for the training objective after an exhaustive search. Sec. A.1 reports more model parameters.

**Datasets and Preprocessing:** We use 1,248 paired T1-weighted MRI and PET from the Alzheimer's disease neuroimaging initiative (ADNI) database [12]. We include the data of cognitively normal (CN, n=379), subjects with mild cognitive impairment (MCI, n=611), and Alzheimer's disease (AD, n=257). Both modalities are co-registered with the size of $96 \times 112 \times 96$. Data are split into train/validation/test sets using only baseline visits, ensuring that diagnosis, age, and sex are balanced across sets. Sec. A.2 details on data preprocessing.

**Baselines:** Our baseline methods include Pix2Pix [11], CycleGAN [34], ResVit [4], BBDM, and BBDM-LDM [18]. ResVit [4] is a GAN-based method integrating ResNet and ViT for medical image translation. Unfortunately, other GAN-based MRI to PET translation approaches do not provide open-source codes [26, 27, 32]. To ensure the inclusion of representative GAN-based methods in our comparison, we implemented the widely-used image translation techniques Pix2Pix [11] and CycleGAN [34]. While none of the SOTA DM-based translation models has been used for MRI to PET translation, BBDM [18] stands out for its superior replicability and performance, making it our choice for adaptation and comparison. We also include its variation BBDM-LDM based on latent diffusion models (LDM) [23]. The same training and evaluation data are used as in PASTA.

**Evaluation:** A comprehensive evaluation of all methods is performed both qualitatively and quantitatively. We compute mean absolute error (MAE), mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structure similarity index (SSIM) between the real and synthesized PET. We implement a downstream AD classification task with 5-fold cross-validation to further validate the pathology preservation. For qualitative assessment, we present our generative results with additional evaluation on 3D-SSP maps and fairness in Sec. A.4.
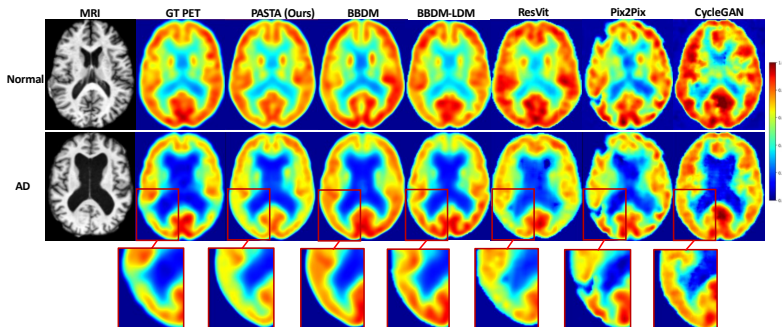
Fig. 4: Qualitative results for a normal subject (top) and an AD patient (middle) with magnified specific pathology (hypometabolism in temporoparietal lobes).

## 4    Results

**Qualitative Results and Clinical Assessment:** Fig. 4 shows qualitative results for PASTA and baseline methods. We analyzed these results together with our clinical experts. It clearly shows that PET generated by PASTA have higher similarity and fidelity to the real ones than other methods. In Pix2Pix and CycleGAN, synthetic scans deviate severely from the ground truth (GT) with obvious artifacts. For AD patients, PASTA's generation accurately grasps the dropped metabolism in the temporoparietal lobe, a region highly associated with AD, as in the GT. Other DM-based models, BBDM and BBDM-LDM, recover the structure well in the generated PET but lack in transferring pathology correctly. ResVit, developed for medical translation, improves pathology awareness but generates less accurate anatomical structures. Overall, PASTA synthesizes scans with superior consistency to the GT both pathologically and structurally.

Based on feedback from our clinical collaborators, PASTA's PET images are deemed realistic and closely comparable to actual PET. Although generated images tend to be smoother, this is not considered a drawback in clinical settings. Nuclear physicians typically apply filters on PET, and AD diagnosis does not require high-resolution edge details [9]. Further, pathology is less pronounced in generated PET, which is to be expected as the synthesis relies on MRI, a modality less sensitive to functional alterations. Yet, synthesized PETs still offer higher pathological sensitivity for AD diagnosis than corresponding MRIs.

**Quantitative Results:** Tab. 1 reports quantitative metrics between GT and synthesized PET for all methods. Consistent with the qualitative results, PASTA generates PETs with the highest quality, owning the lowest MAE (3.45%), MSE (0.43%), and highest PSNR (24.59), SSIM (86.29%). DM-based method BBDM consistently has the second-best results, with a similar performance for BBDM-LDM. However, the other GAN-based baselines, especially CycleGAN, cannot reach on-par performance. These results confirm the potential of DMs.

**Classification Results for AD Diagnosis:** To assess the synthesized 3D PET in AD diagnosis, we train AD classifiers on MRI, MRI with clinical data

Table 1: Quantitative comparison between PASTA and the baselines.

| Method | MAE(%) ↓ | MSE(%) ↓ | PSNR ↑ | SSIM(%) ↑ |
|---|---|---|---|---|
| CycleGAN [34] | 9.83 | 2.61 | 16.38 | 47.48 |
| Pix2Pix [11] | 5.56 | 1.15 | 19.90 | 73.23 |
| ResVit [4] | 6.72 | 1.74 | 19.40 | 69.83 |
| BBDM-LDM [18] | 3.96 | 0.54 | 23.75 | 84.25 |
| BBDM [18] | 3.88 | 0.56 | 23.37 | 84.55 |
| PASTA | **3.45** | **0.43** | **24.59** | **86.29** |

Table 2: Results for AD classification with different input modalities.

| Input | BACC ↑ | F1-Score ↑ | AUC ↑ |
|---|---|---|---|
| MRI | $79.23 \pm 4.30$ | $74.97 \pm 5.95$ | $85.88 \pm 3.79$ |
| MRI+**c** | $81.51 \pm 3.16$ | $77.83 \pm 3.83$ | $\underline{89.19} \pm 1.41$ |
| GT PET | $\mathbf{87.02} \pm 2.35$ | $\mathbf{80.77} \pm 2.62$ | $89.04 \pm 1.88$ |
| Syn PET (ResVit) | $78.54 \pm 4.15$ | $74.41 \pm 5.28$ | $81.80 \pm 4.95$ |
| Syn PET (BBDM) | $72.93 \pm 8.48$ | $70.76 \pm 6.02$ | $83.29 \pm 5.13$ |
| Syn PET (PASTA) | $\underline{83.41} \pm 2.67$ | $\underline{79.98} \pm 3.51$ | $\mathbf{91.63} \pm 2.21$ |

(**c**), GT PET, and synthesized (Syn) PET, respectively, using a 3D ResNet. We use ResVit, BBDM, and PASTA for the PET generation. To mitigate potential domain shift issues and ensure fair comparison, we train and test classifiers on images from the same source. Tab. 2 reports the results, and, as expected, GT PET has a higher performance than MRI. The results for PASTA also surpass MRI in all metrics with an increase of over 4%. Clinical data only boosted MRI's BACC by 2%, suggesting that PASTA's high pathology awareness is likely due to its more effective learning of the interaction between MRI and clinical data. While the results for PASTA in BACC are between those for MRI and GT PET, it is almost on par with GT PET in F1-Score and achieves the highest AUC. The results for BBDM are worse than MRI, confirming its issues with pathology transfer, consistent with Fig. 4. ResVit achieves better accuracy than BBDM, stemming from its better pathology awareness, as shown in the qualitative results, yet is still lower than for MRI. It demonstrates the high potential of PASTA for AD diagnosis and the necessity for pathology-aware transfer.

**Ablation Study:** We perform ablations to verify several designs in PASTA, including CycleEx, the inclusion of pathology priors ($\boldsymbol{\lambda}_R$) and clinical data ($\boldsymbol{c}$), different predefined task in the conditioner arm (MRI reconstruction (M2M)) and its loss weight ($\lambda_{task}$). Sec. A.3 reports further ablations about condition integration and positions. Tab. 3 indicates that CycleEx significantly elevates generation quality, followed by the choice of predefined tasks and its loss weight.

## 5    Conclusion

We introduced PASTA for translating brain MRI to PET with conditional diffusion models. Compared to current DM-based methods, PASTA excels in preserving both structural and pathological details in the target modality, achieved via a

Table 3: Ablation studies on important designs in PASTA.

| Ablation | MAE(%) ↓ | MSE(%) ↓ | PSNR ↑ | SSIM(%) ↑ |
|---|---|---|---|---|
| w/o CycleEx | 3.99 | 0.54 | 23.64 | 85.14 |
| w/o $\boldsymbol{\lambda}_R$ | 3.67 | 0.48 | 24.19 | 85.95 |
| w/o $\boldsymbol{c}$ | 3.57 | 0.46 | 24.39 | 86.12 |
| PASTA (M2M) | 3.78 | 0.50 | 23.78 | 84.77 |
| PASTA ($\lambda_{task} = 1$) | 3.70 | 0.49 | 24.11 | 85.71 |
| PASTA ($\lambda_{task} = 10$) | 3.81 | 0.48 | 24.04 | 83.24 |
| PASTA | **3.45** | **0.43** | **24.59** | **86.29** |

highly interactive dual-arm architecture. CycleEx and the volumetric generation strategy elevated its ability to produce high-quality 3D PET. In AD diagnosis, PASTA reached an improved AUC over the real PET and higher accuracy than MRI. Its unique pathology awareness is likely due to the effective learning from multi-modal conditions and pathology priors, yet further research is required to explore the deeper explanation. Overall, PASTA demonstrated high potential in bridging the gap between structural and functional brain degradation processes.

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

# References

1. Aisen, P.S., Cummings, J., Jack, C.R., Morris, J.C., Sperling, R., Frölich, L., Jones, R.W., Dowsett, S.A., Matthews, B.R., Raskin, J., et al.: On the path to 2025: understanding the alzheimer's disease continuum. Alzheimer's research & therapy **9**, 1–10 (2017)

2. Bau, D., Zhu, J.Y., Wulff, J., Peebles, W., Strobelt, H., Zhou, B., Torralba, A.: Seeing what a gan cannot generate. In: ICCV (2019)

3. Bloudek, L.M., Spackman, D.E., Blankenburg, M., Sullivan, S.D.: Review and meta-analysis of biomarkers and diagnostic imaging in alzheimer's disease. Journal of Alzheimer's Disease **26**(4), 627–645 (2011)

4. Dalmaz, O., Yurt, M., Çukur, T.: Resvit: Residual vision transformers for multi-modal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)

5. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. In: NeurIPS (2021)

6. Folstein, M.F., Folstein, S.E., McHugh, P.R.: "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research **12**(3), 189–198 (1975)

7. Frisoni, G.B., Bocchetta, M., Chételat, G., Rabinovici, G.D., et al.: Imaging markers for alzheimer disease: which vs how. Neurology **81**(5), 487–500 (2013)

8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)

9. Hoffman, J.M., Welsh-Bohmer, K.A., Hanson, M., Crain, B., Hulette, C., Earl, N., Coleman, R.E.: Fdg pet imaging in patients with pathologically verified dementia. Journal of Nuclear Medicine **41**(11), 1920–1928 (2000)

10. 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)

11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR (2017)

12. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging **27**(4), 685–691 (2008)

13. Keppler, J.S., Conti, P.S.: A cost analysis of positron emission tomography. American Journal of Roentgenology **177**(1), 31–40 (2001)

14. Kwon, G., Ye, J.C.: Diffusion-based image translation using disentangled style and content representation. In: ICLR (2023)

15. Kyrtata, N., Emsley, H.C., Sparasci, O., Parkes, L.M., Dickie, B.R.: A systematic review of glucose transport alterations in alzheimer's disease. Frontiers in Neuroscience **15**, 626636 (2021)

16. Lan, H., the Alzheimer Disease Neuroimaging Initiative, Toga, A.W., Sepehrband, F.: Three-dimensional self-attention conditional gan with spectral normalization for multimodal neuroimaging synthesis. Magnetic Resonance in Medicine **86**(3), 1718–1733 (2021)

17. Landau, S.M., Harvey, D., Madison, C.M., Koeppe, R.A., Reiman, E.M., Foster, N.L., Weiner, M.W., Jagust, W.J.: Associations between cognitive, functional, and fdg-pet measures of decline in ad and mci. Neurobiology of Aging **32**(7), 1207–1218 (2011)

18. Li, B., Xue, K., Liu, B., Lai, Y.: Bbdm: Image-to-image translation with brownian bridge diffusion models. In: CVPR (2023)
19. Marcus, C., Mena, E., Subramaniam, R.M.: Brain pet in the diagnosis of alzheimer's disease. Clinical nuclear medicine **39**(10), e413 (2014)
20. Mohs, R.C., Knopman, D., Petersen, R.C., Ferris, S.H., Ernesto, C., Grundman, M., Sano, M., Bieliauskas, L., Geldmacher, D., Clark, C., et al.: Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the alzheimer's disease assessment scale that broaden its scope. Alzheimer Disease & Associated Disorders **11**, 13–21 (1997)
21. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging (2023)
22. Peng, W., Adeli, E., Bosschieter, T., Park, S.H., Zhao, Q., Pohl, K.M.: Generating realistic brain mris via a conditional diffusion probabilistic model. In: MICCAI (2023)
23. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
25. Saharia, C., Chan, W., Chang, H., Lee, C.A., Ho, J., Salimans, T., Fleet, D.J., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH (2022)
26. Shin, H.C., Ihsani, A., Mandava, S., Sreenivas, S.T., Forster, C., Cha, J., Initiative, A.D.N.: Ganbert: Generative adversarial networks with bidirectional encoder representations from transformers for mri to pet synthesis. arXiv:2008.04393 (2020)
27. Shin, H.C., Ihsani, A., Xu, Z., Mandava, S., Sreenivas, S.T., Forster, C., Cha, J.: Gandalf: Generative adversarial networks with discriminator-adaptive loss fine-tuning for alzheimer's disease diagnosis from mri. In: MICCAI (2020)
28. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: ICML (2015)
29. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: ICLR (2021)
30. Strittmatter, W.J., Roses, A.D.: Apolipoprotein e and alzheimer's disease. Annual review of neuroscience **19**(1), 53–77 (1996)
31. Su, X., Song, J., Meng, C., Ermon, S.: Dual diffusion implicit bridges for image-to-image translation. In: ICLR (2023)
32. Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., Colliot, O.: Learning myelin content in multiple sclerosis from multimodal mri through adversarial training. In: MICCAI (2018)
33. 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)
34. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networkss. In: ICCV (2017)
35. Zhu, L., Xue, Z., Jin, Z., Liu, X., He, J., Liu, Z., Yu, L.: Make-a-volume: Leveraging latent diffusion models for cross-modality 3d brain mri synthesis. In: MICCAI (2023)