GAN-Guided Diffusion Models for Generating Clinically Meaningful Multimodal Neuroimaging Data

Reihaneh Hassanzadeh[1] **Anees Abrol**[1] **Hamid Reza Hassanzadeh**[2] **Vince D. Calhoun**[1]

[1]Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS):
Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, GA, USA
[2]Courtesy Faculty Appointment, College of Pharmacy, University of Florida, Gainesville, FL, USA
`rhassanzadeh6@gatech.edu` `aabrol@gsu.edu` `hassanzadeh@gatech.edu`
`vcalhoun@gatech.edu`

Abstract

Multimodal brain imaging provides complementary insights into brain structure and function, but its capability is often limited by missing modalities. Traditional imputation and subsampling strategies are computationally simple, but have the risk of introducing bias or discarding valuable samples. Recently, generative models have emerged as powerful alternatives for synthesizing missing modalities. In this study, we introduce a GAN-guided diffusion framework for cross-modality translation, designed to generate both T1-weighted MRI and functional network connectivity (FNC) data. The framework integrates conditional diffusion modeling, adversarial learning, and cycle-consistency, enabling training with both paired and unpaired samples. On Alzheimer's disease data, our approach outperformed baseline methods, achieving higher peak signal-to-noise ratio (PSNR) (24.95) and structural similarity index measure (SSIM) (0.86) for T1 synthesis, as well as improved correlation with real FNCs (0.65). Furthermore, our results demonstrate that the model captures variability across clinical groups without supervision from diagnostic labels, producing realistic and clinically meaningful synthetic modalities for downstream analysis and biomarker discovery.

1 Introduction

Multimodal brain imaging offers a comprehensive view and enables more accurate analysis of the brain, by integrating complementary information from different modalities. For example, studies have shown that structural and functional magnetic resonance imaging (MRI), while each captures different aspects of brain anatomy and activity, can together provide stronger performance in disease classification [1, 2], biomarker discovery [3, 4], and understanding underlying pathophysiological mechanisms [5]. Despite these advantages, neuroimaging datasets are often incomplete, with one or more modalities missing, posing significant challenges for robust analysis.

Commonly used approaches for handling missing modalities, such as mean or zero imputation and subsampling (i.e., excluding subjects with missing modalities), are computationally simple and easy to implement, but may introduce bias into the data distribution or result in the loss of informative features. To overcome these limitations, generative models can be used to synthesize missing modalities from the available ones. Generative adversarial networks (GANs) [6], as one of such models, have been extensively explored for cross-modality synthesis in neuroimaging. For example, conditional GANs have been used to translate between structural brain MRI modalities, such as T1 and T2-weighted MRIs, demonstrating realistic modality translation in multimodal MRI

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Foundation Models for the Brain and Body.

reconstruction [7] and improving segmentation performance [8]. Cycle GANs have also been used to synthesize 7T MRI from 3T MRI, resulting in improved contrast and resolution [9].

Although GANs are capable of generating realistic data, they can suffer from mode collapse, instability, and limited diversity in the generated output. In contrast, diffusion models have recently demonstrated more realistic, accurate, higher quality, and more diverse output in synthesis tasks [10, 11]. For example, latent diffusion models have produced high-quality 3D brain magnetic resonance images with volumetric consistency that quantitatively outperform GAN-based methods [12]. Conditional denoising diffusion probabilistic models (DDPMs) have demonstrated high-quality anatomically consistent brain image synthesis that improves downstream tasks such as segmentation [13], and disease classification [14]. In addition, cycle-guided diffusion models have been introduced to enhance cross-modality MRI synthesis by conditioning the paired diffusion models against each other, significantly improving image quality and consistency in state-of-the-art networks [15].

In this study, we used a GAN-augmented diffusion model [16] to generate T1-weighted MRI and FNC data. The model is a conditional diffusion framework that synthesizes a target modality from a conditioning source modality within a paired-unpaired hybrid cross-modality translation setting. When paired samples are available, the model learns directly from supervised source–target mappings; when one modality is missing, a synthetic version is generated via a cycle-consistent network, enabling training with unpaired data. This design allows the framework to leverage both paired and unpaired datasets, making it well-suited for medical imaging scenarios where missing modalities are common. We focused on the challenging task of translating between FNC and structural MRI, two modalities that differ substantially in structure, dimensionality, and underlying signal representation, whereas most prior work has considered modalities of similar nature and spatial dimensions, such as multi-contrast MRI or MRI-to-CT. Furthermore, we applied the model to Alzheimer's disease data to enhance the quality of the generated samples while preserving diagnostic patterns, including disease-specific signatures, in contrast to prior studies that have largely focused on healthy control datasets.

2 Methods

An overview of the model is illustrated in Fig. 1. Our framework is based on a conditional diffusion model, enhanced with adversarial training and a cycle consistency structure. Fig. 1(A) depicts the cycle-consistency path, which includes two generators, G_Φ^A and G_Φ^B, and a discriminator, D_Φ^B. The generator G_Φ^B produces a synthetic sample $\tilde{\mathbf{y}}^B$ in the domain B given an input $\mathbf{x}_0^A$ from the domain A. This generated sample is then translated back to the domain A using G_Φ^A, producing a reconstructed sample $\check{\mathbf{x}}_0^A$. The discriminator D_Φ^B distinguishes between real and generated samples in domain B.

The adversarial objectives used to train the generator and discriminator follow the non-saturating formulation [16]:

$$\mathcal{L}_{G_\Phi} = \mathbb{E}_{p_\Phi(\mathbf{y}|\mathbf{x}_0)}\big[-\log\big(D_\Phi(\tilde{\mathbf{y}})\big)\big] \tag{1}$$

$$\mathcal{L}_{D_\Phi} = \mathbb{E}_{q(\mathbf{y}|\mathbf{x}_0)}\big[-\log\big(D_\Phi(\mathbf{y})\big)\big] + \mathbb{E}_{p_\Phi(\mathbf{y}|\mathbf{x}_0)}\big[-\log\big(1 - D_\Phi(\tilde{\mathbf{y}})\big)\big] \tag{2}$$

where $p_\Phi(\mathbf{y} \mid \mathbf{x}_0)$ denotes the estimated distribution of a target sample conditioned on the source, while $q(\mathbf{y} \mid \mathbf{x}_0)$ denotes the true conditional distribution.

As shown in Fig. 1(B), the conditional diffusion process incorporates adversarial learning to approximate $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$ [17] with a small number of diffusion steps ($T \approx$ 4–10), enabling faster reverse diffusion and sampling. The generator G_θ^A progressively denoises the input to produce an estimate $\hat{\mathbf{x}}_{t-1}$. The discriminator D_θ^A distinguishes between real samples $\mathbf{x}_{t-1} \sim q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$ and generated samples $\hat{\mathbf{x}}_{t-1} \sim p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$.

The corresponding non-saturating loss terms can be expressed as follows:

$$\mathcal{L}_{G_\theta} = \mathbb{E}_{t,\, q(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y}),\, p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y})}\big[-\log\big(D_\theta(\hat{\mathbf{x}}_{t-1})\big)\big] \tag{3}$$

$$\begin{aligned}\mathcal{L}_{D_\theta} = \mathbb{E}_{t,\, q(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y})}\Big[&\mathbb{E}_{q(\mathbf{x}_{t-1}|\mathbf{x}_t,y)}\big[-\log\big(D_\theta(\mathbf{x}_{t-1})\big)\big] \\ &+ \mathbb{E}_{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y})}\big[-\log\big(1 - D_\theta(\hat{\mathbf{x}}_{t-1})\big)\big] + \eta\, \mathbb{E}_{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y})}\big\|\nabla_{\mathbf{x}_{t-1}} D_\theta(\mathbf{x}_{t-1})\big\|_2^2\Big]\end{aligned} \tag{4}$$

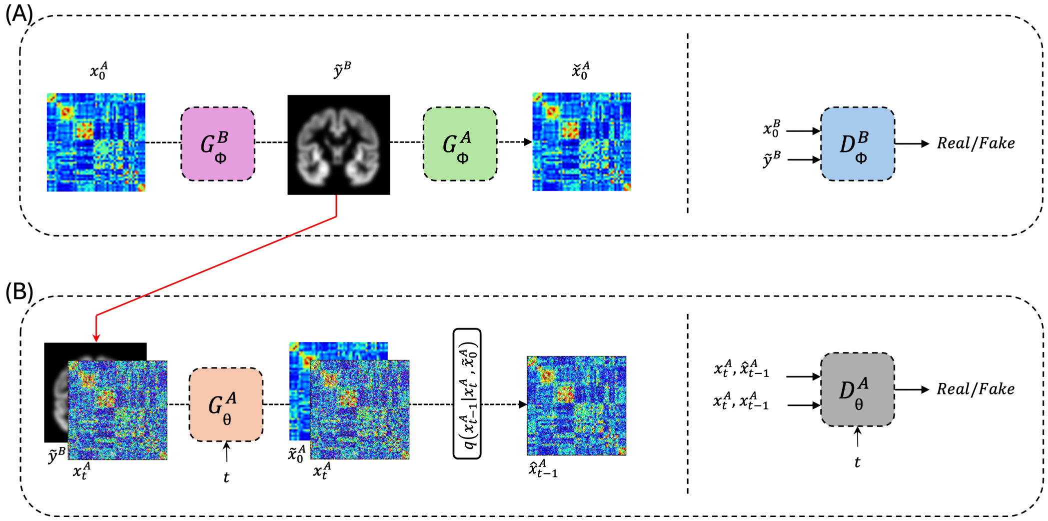

Figure 1: Model overview. (A) **Cycle-consistency path:** G_Φ^B translates an input $\mathbf{x}_0^A$ from modality A (e.g., FNC) into modality B (e.g., T1), producing $\tilde{\mathbf{y}}^B$. The second generator G_Φ^A maps $\tilde{\mathbf{y}}^B$ back to modality A, yielding a reconstruction $\check{\mathbf{x}}_0^A$. The discriminator D_Φ^B evaluates generated $\tilde{\mathbf{y}}^B$ against real samples from modality B. (B) **Conditional diffusion path:** Given a conditioning input (either a real paired sample or a synthetic one from the cycle path), G_θ^A predicts denoised samples at intermediate timesteps, progressively refining toward the target modality output. D_θ^A enforces adversarial realism in the generated outputs.

where $t \sim \mathcal{U}[0, T](\text{step} = 1)$, $T \approx$ 4–10, and η controls the gradient penalty strength. Assuming $\mathbf{x}_t \perp\!\!\!\perp y \mid \mathbf{x}_0$, $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$ in Eq. 4 can be approximated as $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0)$. Following [16], sampling can be performed via:

$$q(\mathbf{x}_{t-k} \mid \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\big(\mathbf{x}_{t-1};\, \bar{\mu}(\mathbf{x}_t, \mathbf{x}_0),\, \bar{\gamma} I\big)$$

The estimated distribution $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$ is obtained via:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y}) := q\big(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \check{\mathbf{x}}_0 = G_\theta(\mathbf{x}_t, \mathbf{y}, t)\big) \tag{5}$$

Furthermore, a cycle consistency loss was applied to the cycle path to ensure that the generated sample $\tilde{\mathbf{y}}^{A,B}$ could be reconstructed back, imposing a minimal difference between the real sample $\mathbf{x}^{A,B}$ and the reconstructed sample $\check{\mathbf{x}}_0^{A,B}$, and to the diffusion path to enforce a minimal difference between the real sample $\mathbf{x}_0^{A,B}$ and the synthetic target sample $\hat{\mathbf{x}}_0^{A,B}$:

$$\mathcal{L}_{\text{cyc}} = \mathbb{E}_{t,\, q(\mathbf{x}_0^{A,B}),\, q(\mathbf{x}_t^{A,B} | \mathbf{x}_0^{A,B})} \sum_{d \in \{A,B\}} \big[\lambda_{\text{cyc}}^{\phi} \|\mathbf{x}_0^d - \check{\mathbf{x}}_0^d\|_1 + \lambda_{\text{cyc}}^{\theta} \|\mathbf{x}_0^d - \hat{\mathbf{x}}_0^d\|_1\big] \tag{6}$$

where $\lambda_{\text{cyc}}^{\phi}$ and $\lambda_{\text{cyc}}^{\theta}$ denote the weights of the cycle path and the diffusion path components, respectively.

Finally, we introduced a subject-aware loss to learn subject-specific mappings by minimizing the L_1 distance between real and generated samples by the generators for paired subjects, as computed below:

$$\mathcal{L}_{\text{subj}} = \mathbb{E}_{t,\, q(\mathbf{x}_0^{A,B}),\, q(\mathbf{x}_t^{A,B} | \mathbf{x}_0^{A,B})} \sum_{d \in \{A,B\}} \big[\lambda_{\text{subj}}^{\phi} \|\mathbf{x}_0^d - \tilde{\mathbf{x}}^d\|_1 + \lambda_{\text{subj}}^{\theta} \|\mathbf{x}_0^d - \tilde{\mathbf{x}}_0^d\|_1\big] \tag{7}$$

where $\lambda_{\text{subj}}^{\phi}$ and $\lambda_{\text{subj}}^{\theta}$ control the contributions of the cycle path and diffusion path generators, respectively, to the subject-aware loss.

The total objective combines the adversarial, cycle-consistency, and subject-aware losses as:

$$\mathcal{L}_G^{\text{total}} = \lambda_\phi \sum_{d \in \{A,B\}} \mathcal{L}_{G_\phi^d} + \lambda_\theta \sum_{d \in \{A,B\}} \mathcal{L}_{G_\theta^d} + \lambda_{\text{cyc}}\, \mathcal{L}_{\text{cyc}} + \lambda_{\text{subj}}\, \mathcal{L}_{\text{subj}} \tag{8}$$

$$\mathcal{L}_D^{\text{total}} = \lambda_\phi \sum_{d\in\{A,B\}} \mathcal{L}_{D_\phi^d} + \lambda_\theta \sum_{d\in\{A,B\}} \mathcal{L}_{D_\theta^d} \tag{9}$$

where λ_ϕ and λ_θ are trade-off weights between the diffusion and cycle-consistency networks, and λ_{cyc} and λ_{subj} control the contributions of the cycle-consistency and subject-aware losses, respectively.

3 Experimental Results

3.1 Dataset

In this study, we focus on two modalities: T1-weighted structural MRI and FNC derived from resting-state fMRI. fMRI scans were processed using the NeuroMark independent component analysis (ICA) framework [18], which extracts 53 reproducible functional brain components. The temporal correlations between all pairs of these components were computed to form a 53 × 53 connectivity matrix, referred to as the FNC map. These components are grouped into seven functional network domains: subcortical (SC), auditory (AU), visual (VI), sensorimotor (SM), cognitive control (CC), default mode (DM), and cerebellar (CB) [19].

We utilized data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [20], which contains 8,372 samples from 1,709 subjects. Table 1 summarizes the distribution of samples and subjects between modalities and diagnostic groups, including control (CN), Alzheimer's disease (AD), and mild cognitive impairment (MCI). As shown in the last column (FNC and T1), only a small fraction of the dataset contains both T1-weighted MRI and FNC for the same subject, with the FNC modality being the least represented.

Table 1: Data distribution across modalities and diagnostic groups.

AD	# of Samples (# of Subjects)		
	FNC	T1	FNC and T1
CN	411 (214)	2246 (553)	356 (170)
AD	168 (85)	1640 (583)	151 (73)
MCI	465 (203)	3227 (883)	412 (165)

3.2 Model Performance

Fig. 2 demonstrates the ability of the proposed model to generate realistic and diagnostically meaningful FNC maps. Panel (A) shows the mean and variance across subjects, where the generated FNCs closely replicate the functional connectivity structures observed in the real data. The cell-wise variance plots further indicate that variability patterns are largely preserved in the generated FNCs, suggesting that the model does not simply memorize the mean structure but also captures subject-level variability. Panel (B) compares group-level FNC differences across diagnostic categories (CN, MCI, AD) between real and generated samples. The results show the model's ability to preserve diagnostic contrasts and encode clinically meaningful differences, despite not being guided by diagnostic labels during training.

Fig. 3 illustrates the quality and diagnostic relevance of the generated T1-weighted MRI images compared with the real ones. Panel (A) shows the mean and variance across subjects, indicating that the generated images closely match the anatomical structures and variability patterns of the real data. Panel (B) presents group-level differences (CN-AD, CN-MCI, AD-MCI), showing that the generated images not only reproduce the spatial distribution of differences but also capture disease-related changes observed in the real data, such as atrophy in medial temporal and cortical regions.

Table 2 compares the proposed model with several baselines, including a conditional DDPM [11] and CycleGAN [21]. We evaluated the generated T1 images against their corresponding real ones using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [22], and the generated FNC maps against their corresponding real ones using Pearson correlation (PCorr). The results demonstrate the superiority of our model over the baselines. Specifically, our approach achieved a PSNR of 24.95 and an SSIM of 0.86 in the T1 generation task, as well as a higher correlation of 0.65 in the FNC generation task. These results indicate improved image fidelity, structural similarity, and preservation of functional connectivity patterns.

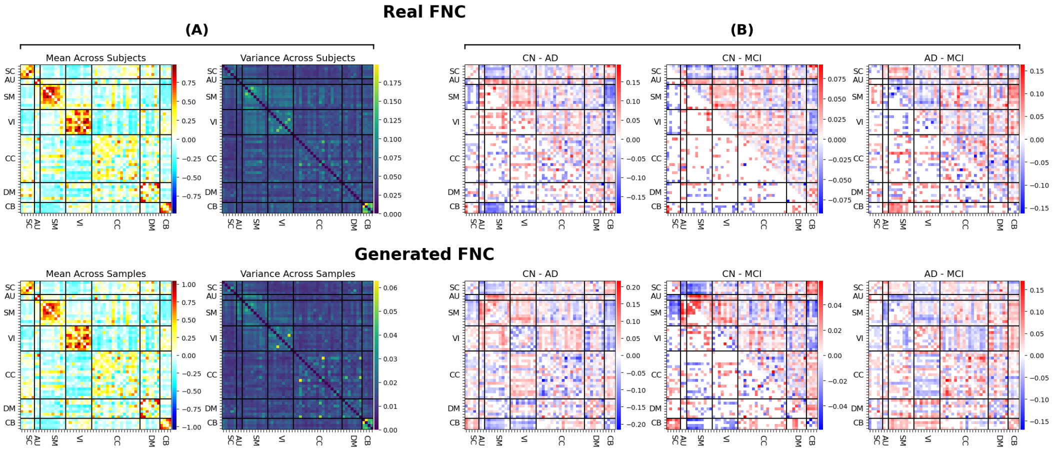

Figure 2: (A) Mean and variance between real and generated FNC maps. (B) Real and generated FNC group differences across diagnostic categories, CN, MCI, and AD. Each matrix shows the difference in mean connectivity values between two groups in the upper triangle, and statistically significant differences ($p < 0.05$, uncorrected) from two-sample t-tests in the lower triangle.

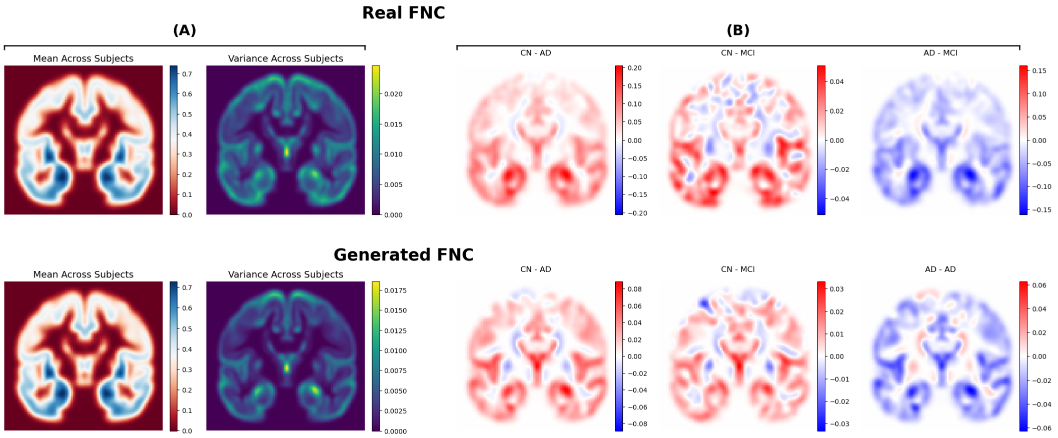

Figure 3: (A) Mean and variance between real and generated T1 images. (B) Real and generated T1 group differences across diagnostic categories, CN, MCI, and AD.

Table 2: Comparison of evaluation metric across models.

	FNC − > T1		**T1 − > FNC**
Model	**PSNR**	**SSIM**	**PCorr**
Our model	**24.95**±1.753	**0.86**±0.025	**0.65**±0.011
DDPM	24.48±2.635	0.83±0.097	0.56±0.021
cycle-GAN	23.63±1.426	0.82±0.027	0.49±0.035

4 Conclusion

In this study, we introduced a GAN-guided diffusion framework for cross-modality translation in neuroimaging, with a focus on synthesizing T1-weighted MRI and functional network connectivity (FNC) data. By combining conditional diffusion modeling, adversarial training, and cycle consistency, our approach is able to leverage both paired and unpaired data, making it well suited to scenarios with missing modalities. Experiments on Alzheimer's disease data sets demonstrated that the proposed model outperforms baseline approaches, achieving superior quantitative performance while preserving anatomical detail, functional connectivity structure, and clinically relevant group differences. In general, these results suggest its potential use in downstream tasks such as disease classification and biomarker discovery.

References

[1] Kai Zhou, Jie Li, Rui Huang, Jiali Yu, Rong Li, Wei Liao, Fengmei Lu, Xiaofei Hu, Huafu Chen, and Qing Gao. Brain structural features with functional priori to classify parkinson's disease and multiple system atrophy using diagnostic mri. *Scientific Reports*, 15(1):21310, 2025.

[2] Junghoe Kim and Jong-Hwan Lee. Integration of structural and functional magnetic resonance imaging improves mild cognitive impairment detection. *Magnetic resonance imaging*, 31(5):718–732, 2013.

[3] Ai-Ling Lin, Angela R Laird, Peter T Fox, and Jia-Hong Gao. Multimodal mri neuroimaging biomarkers for cognitive normal adults, amnestic mild cognitive impairment, and alzheimer s disease. *Neurology research international*, 2012(1):907409, 2012.

[4] Nanyan Zhu, Chen Liu, Xinyang Feng, Dipika Sikka, Sabrina Gjerswold-Selleck, Scott A Small, and Jia Guo. Deep learning identifies neuroimaging signatures of alzheimer's disease using structural and synthesized functional mri data. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 216–220. IEEE, 2021.

[5] Min Ai, Yu Liu, Dan Liu, Chengxi Yan, Xia Wang, and Xun Chen. Research progress in predicting the conversion from mild cognitive impairment to alzheimer's disease via multimodal mri and artificial intelligence. *Frontiers in Neurology*, 16:1596632, 2025.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[7] Daisuke Kawahara and Yasushi Nagata. T1-weighted and t2-weighted mri image synthesis with convolutional generative adversarial networks. *reports of practical Oncology and radiotherapy*, 26(1):35–42, 2021.

[8] Qianye Yang, Nannan Li, Zixu Zhao, Xingyu Fan, Eric I-Chao Chang, and Yan Xu. Mri cross-modality image-to-image translation. *Scientific reports*, 10(1):3753, 2020.

[9] Eduardo Diniz, Tales Santini, Helmet Karim, Howard J Aizenstein, and Tamer S Ibrahim. Cross-modality image translation of 3 tesla magnetic resonance imaging to 7 tesla using generative adversarial networks. *Human Brain Mapping*, 46(9):e70246, 2025.

[10] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[12] Lingting Zhu, Zeyue Xue, Zhenchao Jin, Xian Liu, Jingzhen He, Ziwei Liu, and Lequan Yu. Make-a-volume: Leveraging latent diffusion models for cross-modality 3d brain mri synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 592–601. Springer, 2023.

[13] Zolnamar Dorjsembe, Hsing-Kuo Pao, Sodtavilan Odonchimed, and Furen Xiao. Conditional diffusion models for semantic 3d brain mri synthesis. *IEEE Journal of Biomedical and Health Informatics*, 28(7):4084–4093, 2024.

[14] Alex Ling Yu Hung, Kai Zhao, Haoxin Zheng, Ran Yan, Steven S Raman, Demetri Terzopoulos, and Kyunghyun Sung. Med-cdiff: Conditional medical image generation with diffusion models. *Bioengineering*, 10(11):1258, 2023.

[15] Shaoyan Pan, Zach Eidex, Mojtaba Safari, Richard Qiu, and Xiaofeng Yang. Cycle-guided denoising diffusion probability model for 3d cross-modality mri synthesis. In *Medical Imaging 2025: Clinical and Biomedical Imaging*, volume 13410, pages 515–522. SPIE, 2025.

[16] Muzaffer Özbey, Onat Dalmaz, Salman UH Dar, Hasan A Bedel, Şaban Özturk, Alper Güngör, and Tolga Çukur. Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Medical Imaging*, 42(12):3524–3539, 2023.

[17] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. *arXiv preprint arXiv:2112.07804*, 2021.

[18] Yuhui Du, Zhishui Fu, Vince D. Calhoun, et al. Neuromark: A fully automated ica method to identify effective fmri markers of brain disorders. *NeuroImage: Clinical*, 28:102375, 2020.

[19] Reihaneh Hassanzadeh, Anees Abrol, Godfrey Pearlson, Jessica A Turner, and Vince D Calhoun. A confounder controlled machine learning approach: Group analysis and classification of schizophrenia and alzheimer's disease using resting-state functional network connectivity. *Plos one*, 19(5):e0293053, 2024.

[20] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.

[21] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[22] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[25] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

5 Technical Appendices and Supplementary Material

5.1 Data Preprocessing

The T1 data were segmented into tissue probability maps for gray matter, white matter, and cerebral spinal fluid using SPM12. The gray matter images were then warped to standard space, modulated and smoothed using a Gaussian kernel with an FWHM = 10mm. The preprocessed gray matter volume images had a dimensionality of 121×145×121 in the voxel space, with the voxel size of 1.5×1.5×1.5 mm^3.

For the resting-state fMRI data sets, we removed the first five time points to ensure signal equilibrium and adaptation of subjects to scanner noise. Then we performed slice timing correction and rigid body motion correction using the SPM toolbox, followed by warping the images into the standard Montreal Neurological Institute (MNI) template using an echo-planar imaging (EPI) template and the old SPM12 normalization module. Lastly, the data were resampled to 3x3x3 mm^3 isotropic voxels, resulting in image dimensionality of 53x63x52 in the voxel space, and smoothed using a Gaussian kernel with a full width at half maximum of 6mm. We also implemented thorough quality control (QC) on the preprocessed fMRI images to discard the images that exhibited (1) poor correlation with individual and group data masks, (2) markedly briefer scan lengths, and (3) high head motion parameters ($>3^\circ$ rotations and $>$3mm translations).

5.2 Implementation Details

5.2.1 Network Architectures

The diffusion generators G_θ^A and G_θ^B were implemented using a UNet backbone [23], each consisting of multiple downsampling and upsampling residual blocks with skip connections, group normalization, and Swish activations. Temporal information is injected at each residual block through a learned projection of the sinusoidal timestep embedding, which is added to the feature activations. The input to each generator is formed by concatenating the noisy target modality with the conditioning modality along the channel dimension. The discriminators D_θ^A and D_θ^B are convolutional networks, each consisting of a series of 4×4 convolutional layers with stride 2, doubling the number of feature maps at each stage, followed by LeakyReLU activation. The conditioning modality is concatenated with the generated or real target along the channel dimension.

The cycle-consistency generators G_Φ^A and G_Φ^B were based on ResNet-style architectures [24] with six residual blocks. Each residual block contains two 3×3 convolutional layers, instance normalization, and ReLU activation, with reflection padding to avoid edge artifacts. The discriminators D_Φ and D_θ are PatchGANs [25], consisting of an initial 4×4 convolution with stride 2 and LeakyReLU, followed by a sequence of stride-2 convolutions with doubling feature channels, and ending with a 4×4 convolution that outputs a grid of real/fake scores.

In baselines, in order to have a robust comparison, the DDPM model used similar network architectures as the diffusion path in out network and CycleGAN used the similar network architectures for generators and discriminators.

5.2.2 Symmetry Constraint

For experiments involving the synthesis of FNC matrices, we enforced the inherent symmetry property of functional connectivity. After generating a predicted FNC matrix from the network, we retained only the upper triangular portion (excluding the diagonal) and mirrored it to the lower triangular portion. This step enforces the undirected nature of functional brain networks, reduces the number of independent parameters that must be learned, and improves training stability.

5.2.3 Data Splits

We split the dataset into 60% training, 20% validation, and 20% test sets, while ensuring subject-level separation to prevent data leakage across splits.

5.2.4 Hyperparameter Selection

Hyperparameters were tuned using the validation set to optimize Pearson correlation and structural similarity (SSIM) between generated and real samples. We performed a grid search over learning rates $1\mathrm{e}{-4}, 2\mathrm{e}{-4}, 5\mathrm{e}{-5}$, batch sizes $16, 32$, diffusion timesteps $T \in 4$, $\lambda_{\text{cyc}} \in 0.5, 1, 2$, and $\lambda_{\text{subj}} \in 0.5, 1, 2$, while λ_ϕ and λ_θ were fixed at 1. Training was run for 100 epochs, and the model from the final epoch was selected for evaluation. For the baselines, the DDPM was trained with $T = 1000$ and a cosine scheduler on the paired data, while CycleGAN was trained on all the data. Both baselines were trained for 100 epochs, with a learning rate of $1\mathrm{e}{-4}$ for CycleGAN and $2\mathrm{e}{-4}$ for DDPM.