Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Structure to Function and Back Again: A GAN-Guided Diffusion Framework for Generating Clinically Meaningful Multimodal Neuroimaging Data

Reihaneh Hassanzadeh

rhassanzadeh6@gatech.edu

Georgia Institute of Technology

Anees Abrol

Georgia State University

Hamid Reza Hassanzadeh

University of Florida

Vince D. Calhoun

Georgia Institute of Technology

Article

Keywords:

Posted Date: June 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9684909/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

From Structure to Function and Back Again: A GAN-Guided Diffusion Framework for Generating Clinically Meaningful Multimodal Neuroimaging Data

Reihaneh Hassanzadeh[1,2,*], Anees Abrol[1], Hamid Reza Hassanzadeh[3], and Vince D. Calhoun[1,2]

[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]Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA
[3]Courtesy Faculty Appointment, College of Pharmacy, University of Florida, Gainesville, FL, USA
[*]rhassanzadeh6@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 magnetic resonance imaging (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.

Introduction

Multimodal brain imaging offers a comprehensive view of brain organization by integrating complementary information from different modalities, enabling more accurate analysis of brain structure and function[1]. For example, studies have shown that structural and functional magnetic resonance imaging, while each captures different aspects of brain anatomy and activity, together provide stronger performance in disease classification[2,3], biomarker discovery[4–6], and understanding underlying pathophysiological mechanisms[7].

Despite the advantages of multimodal imaging, fully paired multimodal datasets remain uncommon in practice, even in large-scale studies, due to several factors such as differences in acquisition protocols across sites, cost and scan-time constraints, and subject intolerance to lengthy scanning sessions.

As a result, multimodal studies often handle missing modalities by discarding incomplete subjects (subsampling) or relying on simple imputation strategies such as mean or zero imputation. However, these approaches can substantially reduce sample sizes and statistical power or introduce bias into the data distribution. To overcome these limitations, generative models offer a promising alternative by synthesizing missing modalities from available ones, enabling the use of both paired and unpaired data while preserving modality-specific variability.

This capability is particularly valuable in neuroimaging studies where some modalities, such as structural MRI, are widely available while others, such as functional MRI, may be missing or inconsistently acquired. A framework capable of learning cross-modal relationships and synthesizing missing data can therefore improve statistical power and broaden the applicability of multimodal analysis methods.

In this study, we employ a GAN-augmented diffusion framework[8] to generate gray matter T1-weighted MRI and functional network connectivity (FNC) data. The proposed model synthesizes a target modality from a conditioning source modality within a hybrid paired–unpaired cross-modality translation setting. When paired samples are available, the model learns supervised mappings between modalities; when one modality is missing, a synthetic counterpart is generated through a cycle-consistent network, enabling effective training with unpaired data. This design allows the framework to leverage both paired and unpaired

datasets, making it well suited for medical imaging scenarios with incomplete multimodal acquisition. We evaluate the framework on Alzheimer's disease data to demonstrate that the generated samples preserve clinically meaningful variability and disease-related patterns, in contrast to many prior studies that primarily focused on data from healthy individuals.

The main contributions of this work are summarized as follows:

- We implement a GAN-guided diffusion framework capable of learning cross-modality mappings using both paired and unpaired multimodal neuroimaging data.
- We demonstrate effective translation between highly heterogeneous modalities, namely structural MRI and functional connectivity representations.
- We show that the synthesized modalities preserve clinically meaningful variability and diagnostic group differences without requiring diagnostic supervision during training.
- We provide quantitative and qualitative evidence that the proposed framework improves synthesis quality compared with diffusion-only and GAN-based baselines.

Related Work

Generative models have become a powerful approach for addressing missing modalities in multimodal studies[9,10]. Early work primarily relied on generative adversarial networks (GANs)[11] for cross-modality synthesis[12–16], motivated by their ability to learn complex data distributions and generate visually realistic samples. To better control the generation process, studies adopted conditional GANs (cGANs)[15], where the synthesis is guided by an input modality or auxiliary information. cGAN-based approaches have been widely applied to translation across structural MRI modalities[17,18], CT-MRI[19], CT-PET[12], PET-MRI[20,21], enabling realistic modality reconstruction and improving downstream tasks including segmentation[22], registration[23,24], and medical diagnosis improvement[21]. CycleGAN-based methods[25] further extended cross-modality synthesis to unpaired training settings, allowing translation between modalities without requiring paired samples[26–28]. Such approaches are particularly attractive for medical imaging scenarios where paired multimodal data are scarce or infeasible to acquire and typically.

Despite their success, GAN-based models are known to suffer from several limitations, including training instability, mode collapse, and limited diversity in generated samples[29]. These issues can be especially problematic in clinical neuroimaging applications, where capturing population-level and subject-specific variability is critical. As a result, recent work has increasingly explored diffusion-based generative models as a more stable alternative for medical image synthesis[30–32].

Diffusion models have demonstrated improved stability and synthesis fidelity across image generation tasks[33]. In neuroimaging applications, denoising diffusion probabilistic models (DDPMs)[31] have produced high-quality medical images with strong volumetric consistency, quantitatively outperforming GAN-based approaches[34]. Conditional DDPMs have further demonstrated anatomically consistent brain image synthesis that improves downstream applications such as segmentation and disease classification[35,36]. Compared with GAN-based approaches, diffusion models generally provide more stable training and better coverage of population variability, making them increasingly attractive for medical image synthesis[37].

Several recent studies have explored hybrid or guided diffusion frameworks to improve cross-modality synthesis. For example, adversarially guided diffusion models incorporate discriminators to encourage realism during the denoising process, bridging the strengths of GANs and diffusion models[8]. Additionally, cycle-guided diffusion models have been introduced to enhance cross-modality MRI synthesis by conditioning paired diffusion processes against each other, leading to improved structural consistency and synthesis quality[38]. While these approaches represent important advances, they typically focus on translating between modalities with similar spatial representations, such as multi-contrast MRI or MRI-to-CT mappings, and often assume the availability of paired training data.

In contrast, translation between structurally heterogeneous modalities, such as structural MRI and functional connectivity representations derived from fMRI, remains relatively underexplored. Moreover, most existing diffusion-based approaches do not explicitly address the realistic scenario in which datasets contain a mixture of paired and unpaired samples across modalities. The proposed framework addresses these limitations by integrating adversarial learning, diffusion modeling, and cycle-consistency within a unified architecture that supports both paired and unpaired training.

Methods

An overview of the proposed framework is illustrated in Fig. 1. The model combines a cycle-consistent translation network (Fig. 1(A)) and a conditional diffusion model (Fig. 1(B)) to enable cross-modality synthesis using both paired and unpaired data. When paired samples are available, the diffusion model learns a direct mapping between modalities. When samples are unpaired, a cycle-consistent translation network generates synthetic counterparts that allow the diffusion model to continue

Table 1. Summary of main notations used in the proposed framework.

Symbol	Description
x_0^A, x_0^B	Clean samples in modalities A and B
x_t	Noisy sample at diffusion timestep t
$\hat{x}_{t-1}$	Generated denoised sample at step $t-1$
$\check{x}_0$	Reconstructed sample after cycle translation
$\tilde{x}$	Synthetic modality generated by cycle path
$p_\theta(\cdot)$	Model-estimated diffusion distribution
$q(\cdot)$	True data diffusion distribution
G_ϕ	Cycle-consistency generators
G_θ	Diffusion generators
D_ϕ	Cycle-path discriminator
D_θ	Diffusion-path discriminator
y	Conditioning modality input
t	Diffusion timestep index
T	Total diffusion steps
λ_{cyc}	Cycle-consistency loss weight
λ_{subj}	Subject-aware loss weight

training in the absence of paired observations. In this way, the framework jointly leverages paired and unpaired data while maintaining consistency across modalities.

Fig. 1(A) illustrates the cycle-consistency path, which includes two generators, G_ϕ^A and G_ϕ^B, and a discriminator D_ϕ^B. Generator G_ϕ^B produces a synthetic sample $\tilde{y}_B$ in domain B from an input x_0^A in domain A. This generated sample is then translated back into domain A using G_ϕ^A, producing a reconstructed sample $\check{x}_0^A$. The discriminator D_ϕ^B distinguishes between real samples and generated samples in domain B. This cycle-consistency path is primarily used to generate synthetic modality counterparts when paired samples are unavailable, allowing the diffusion model to be trained on both real paired data and synthetic pairs derived from unpaired samples.

The adversarial objectives used to train the generator and discriminator follow the non-saturating formulation[8]:

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

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

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. Table 1 summarizes the main notations used throughout this section for reference.

Fig. 1(B) illustrates the conditional diffusion path used for modality synthesis. Given a conditioning input (either a real paired sample or a synthetic one produced by the cycle path), the generator G_θ^A progressively denoises intermediate samples to reconstruct the target modality. Starting from noisy representations, the model iteratively predicts cleaner samples until the final modality estimate is obtained.

The conditional diffusion process incorporates adversarial learning to approximate the reverse diffusion transition $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$[39], where $\mathbf{y}$ denotes the conditioning modality. By combining diffusion modeling with adversarial training, the framework encourages realistic intermediate reconstructions while requiring only a small number of diffusion steps ($T \approx 4$–10), enabling faster reverse diffusion and efficient sampling compared with standard diffusion models. Discriminator D_θ^A enforces adversarial realism by distinguishing between real samples $\mathbf{x}_{t-1} \sim q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$ and generated outputs $\hat{\mathbf{x}}_{t-1} \sim p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$ at intermediate diffusion steps.

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

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

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

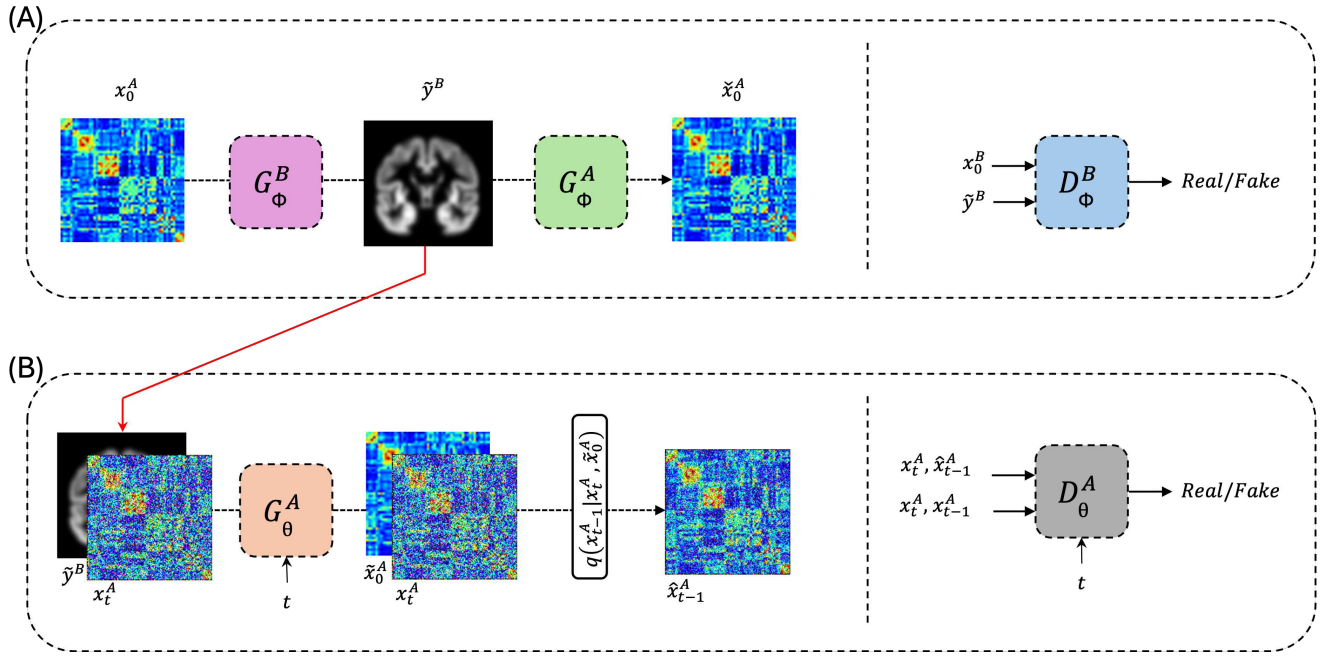


Figure 1. Model overview. (A) **Cycle-consistency path:** G_Φ^B translates an input $\mathbf{x}_0^A$ from modality A (e.g., FNC) into modality B (e.g., T1), producing $\tilde{\mathbf{y}}^B$. The second generator G_Φ^A maps $\tilde{\mathbf{y}}^B$ back to modality A, yielding a reconstruction $\check{\mathbf{x}}_0^A$. 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 $t \sim U[0,T]$, and η controls the gradient penalty strength. At each training iteration, a timestep t is sampled uniformly, a noisy sample x_t is drawn from the forward diffusion process $q(x_t \mid x_0, y)$, and the generator predicts $\hat{x}_{t-1}$ according to $p_\theta(x_{t-1} \mid x_t, y)$; the expectations in Eq. 3 and 4 therefore average discriminator feedback over these stochastic diffusion transitions encountered during training. 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[8], 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 $\hat{\mathbf{x}}_0^{A,B}$:

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

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

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

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 adversarial, cycle-consistency, and subject-aware losses as:

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

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

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.

Experimental Results

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[40], which extracts 53 reproducible functional brain components. Following ICA and band-pass filtering, 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)[41].

We utilized data from the Alzheimer's Disease Neuroimaging Initiative (ADNI)[42], which contains 8,372 samples from 1,709 subjects. Table 2 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 [would update terminology a bit to be more precise: T1-weighted MRI (or structural MRI) and functional MRI/fMRI are modalities, gray matter (GM) and FNC are high dimensional features generated from these modalities] for the same subject, with the FNC modality being the least represented.

Table 2. 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)

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 3 compares the proposed model with several baselines, including a conditional DDPM[31] and CycleGAN[25]. We evaluated the generated T1 images against their corresponding real ones using the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM)[43], and the generated FNC maps against their corresponding real ones using Pearson correlation (PCorr). The results demonstrate improved performance 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 enhanced image fidelity, structural similarity, and preservation of functional connectivity patterns.

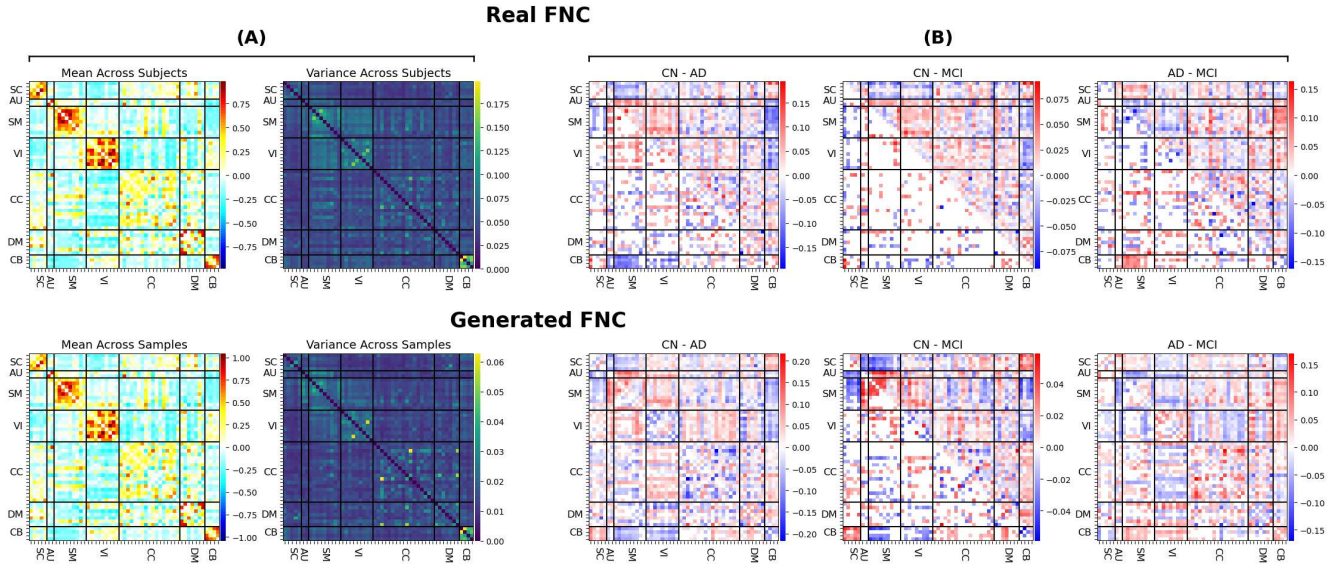


Figure 2. (A) Mean and variance between real and generated FNC maps. (B) Real and generated FNC group differences across diagnostic categories, CN, MCI, and AD. Each matrix shows the difference in mean connectivity values between two groups in the upper triangle, and statistically significant differences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

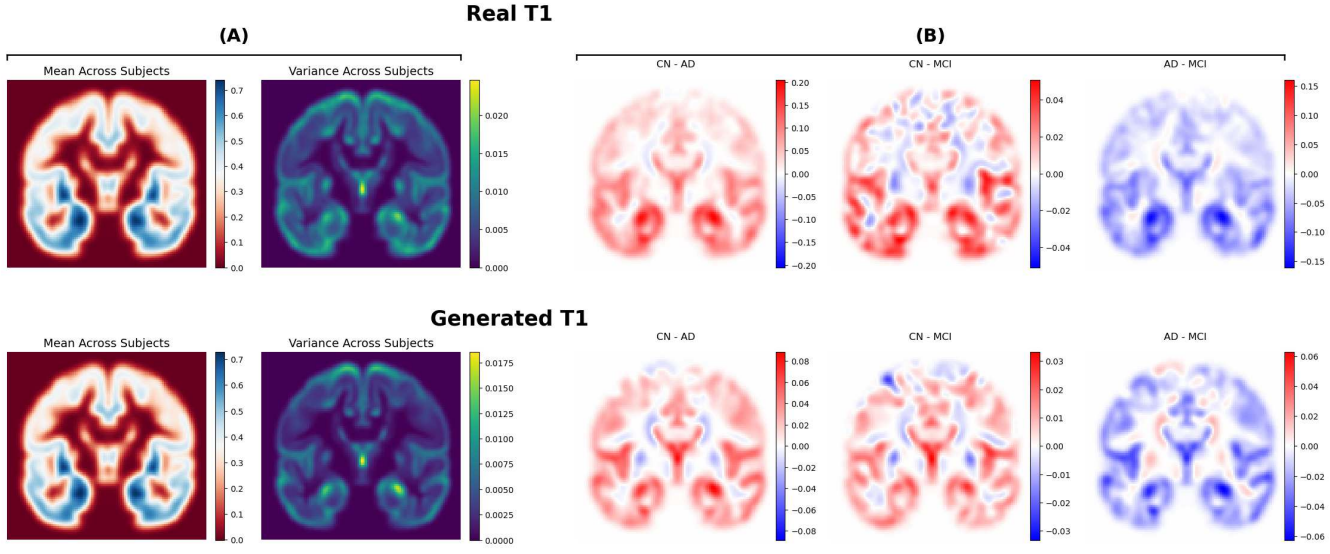


Figure 3. (A) Mean and variance between real and generated T1 images. (B) Real and generated T1 group differences across diagnostic categories, CN, MCI, and AD.

Table 3. 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

Discussion

In this work, we presented a GAN-guided diffusion model for cross-modality translation between T1-weighted MRI and functional network connectivity (FNC). Unlike previous studies that focused primarily on translating between structurally similar modalities (e.g., variations of MRI sequences), our model handled two highly heterogeneous representations, structural MRI and connectivity matrices. Through a hybrid training scheme combining adversarial learning and diffusion modeling, the framework reconstructs high-fidelity output while preserving both anatomical detail and functional network structure. The cycle-consistency and subject-aware losses further help capture individual-level variation rather than collapsing toward population averages. As shown by the similarity of mean and variance maps between real and generated data, the synthesized samples preserved both global structure and subject-specific variability. Moreover, the model reproduced clinically relevant group differences between CN, MCI, and AD subjects despite not being trained on diagnostic labels, suggesting that the cross-modal mapping itself embeds disease-related information, including structural atrophy in T1 MRI and patterns of hypo-/hyper-connectivity in FNC.

Quantitative evaluations also highlight the advantages of the proposed framework, with higher PSNR and SSIM for T1 synthesis and substantially stronger correlations for FNC synthesis compared to DDPM and CycleGAN baselines. These improvements suggest greater structural fidelity, more accurate functional organization, and improved robustness to missing modalities, supporting the potential utility of the model for downstream tasks such as data augmentation, classification, and biomarker discovery.

Beyond quantitative performance, the ability of the model to preserve diagnostic contrasts has important scientific implications. Replicated diagnostic differences in both FNC and T1 images indicate that the model learns clinically meaningful patterns that emerge from multimodal relationships rather than explicit supervision. This suggests the potential for analyses where multimodal completeness is limited, such as disease classification, multimodal biomarker identification, or perhaps when acquisition of one modality is not technically feasible. Importantly, synthetic modalities provide an alternative to zero-imputation or subject exclusion policies that often introduce bias or reduce sample sizes in neuroimaging studies.

This study has several limitations. First, the ADNI dataset contains substantial variability due to multi-site and multi-scanner acquisition differences; although the model performed well, generalization to data from unseen sites or different preprocessing pipelines remains to be tested. Second, FNC construction relies on an ICA-based pipeline that imposes its own assumptions on the data. The hybrid GAN–diffusion framework also requires careful tuning of multiple loss terms and hyperparameters. Third, while T1-weighted scans are inherently 3D, we used a representative 2D slice to make training computationally tractable and aligned with the dimensionality of FNC data; this choice may limit the model's ability to capture certain 3D anatomical relationships.

Future extensions of this work include the incorporation of additional imaging modalities such as PET or EEG to create a unified multimodal diffusion framework capable of synthesizing richer brain representations. Additionally, model interpretability techniques could be applied to understand cross-modal relationships—for example, identifying which structural regions most strongly predict FNC patterns and vice-versa. Another promising direction is to evaluate whether synthesized modalities enhance downstream tasks such as diagnosis, progression prediction, or patient stratification. Finally, scaling the model to full 3D diffusion architectures may enable more detailed anatomical synthesis and better capture disease-specific morphological changes.

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.

Data Availability

The data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu/). Access to the ADNI dataset is available upon registration and compliance with the ADNI data use agreement. The processed data and code used for this study are available from the corresponding author upon reasonable request.

Funding

This research was supported by the National Institutes of Health (NIH) under award number 1R01AG090597.

References

1. Calhoun, V. D. & Sui, J. Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. *Biol. psychiatry: cognitive neuroscience neuroimaging* **1**, 230–244 (2016).
2. Zhou, K. *et al.* Brain structural features with functional priori to classify parkinson's disease and multiple system atrophy using diagnostic mri. *Sci. Reports* **15**, 21310 (2025).
3. Kim, J. & Lee, J.-H. Integration of structural and functional magnetic resonance imaging improves mild cognitive impairment detection. *Magn. resonance imaging* **31**, 718–732 (2013).
4. Lin, A.-L., Laird, A. R., Fox, P. T. & Gao, J.-H. Multimodal mri neuroimaging biomarkers for cognitive normal adults, amnestic mild cognitive impairment, and alzheimer s disease. *Neurol. research international* **2012**, 907409 (2012).
5. Zhu, N. *et al.* 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)*, 216–220 (IEEE, 2021).
6. Calhoun, V. D. & Adali, T. Feature-based fusion of medical imaging data. *IEEE Transactions on Inf. Technol. Biomed.* **13**, 711–720 (2008).
7. Ai, M. *et al.* Research progress in predicting the conversion from mild cognitive impairment to alzheimer's disease via multimodal mri and artificial intelligence. *Front. Neurol.* **16**, 1596632 (2025).
8. Özbey, M. *et al.* Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Med. Imaging* **42**, 3524–3539 (2023).
9. Xie, G. *et al.* Cross-modality neuroimage synthesis: A survey. *ACM computing surveys* **56**, 1–28 (2023).
10. Wu, R., Wang, H., Chen, H.-T. & Carneiro, G. Deep multimodal learning with missing modality: A survey. *arXiv preprint arXiv:2409.07825* (2024).
11. Goodfellow, I. *et al.* Generative adversarial networks. *Commun. ACM* **63**, 139–144 (2020).
12. Ben-Cohen, A. *et al.* Cross-modality synthesis from ct to pet using fcn and gan networks for improved automated lesion detection. *Eng. Appl. Artif. Intell.* **78**, 186–194 (2019).
13. Yu, B. *et al.* Ea-gans: edge-aware generative adversarial networks for cross-modality mr image synthesis. *IEEE transactions on medical imaging* **38**, 1750–1762 (2019).
14. Yu, B. *et al.* Sample-adaptive gans: linking global and local mappings for cross-modality mr image synthesis. *IEEE transactions on medical imaging* **39**, 2339–2350 (2020).
15. Qin, Z., Liu, Z., Zhu, P. & Ling, W. Style transfer in conditional gans for cross-modality synthesis of brain magnetic resonance images. *Comput. Biol. Medicine* **148**, 105928 (2022).
16. Bi, Y., Abrol, A., Jia, S., Sui, J. & Calhoun, V. D. Gray matters: Vit-gan framework for identifying schizophrenia biomarkers linking structural mri and functional network connectivity. *NeuroImage* **297**, 120674 (2024).
17. Zhao, P., Pan, H. & Xia, S. Mri-trans-gan: 3d mri cross-modality translation. In *2021 40th Chinese control conference (CCC)*, 7229–7234 (IEEE, 2021).
18. Yang, Q. *et al.* Mri cross-modality image-to-image translation. *Sci. reports* **10**, 3753 (2020).
19. Liu, M. *et al.* Multi-conditional constraint generative adversarial network-based mr imaging from ct scan data. *Sensors* **22**, 4043 (2022).

20. Peng, Y. & Qi, J. Cm-gans: Cross-modal generative adversarial networks for common representation learning. *ACM Transactions on Multimed. Comput. Commun. Appl. (TOMM)* **15**, 1–24 (2019).

21. Sikka, A., Peri, S., Virk, J. S., Niyaz, U. & Bathula, D. R. Mri-to-pet cross-modality translation using globally & locally aware gan (gla-gan) for multi-modal diagnosis of alzheimer's disease. *The J. Precis. Medicine: Heal. Dis.* **2**, 100004 (2025).

22. Li, J. *et al.* Cross-modality synthesis aiding lung tumor segmentation on multi-modal mri images. *Biomed. Signal Process. Control.* **76**, 103655 (2022).

23. Han, T. *et al.* Review of generative adversarial networks in mono-and cross-modal biomedical image registration. *Front. Neuroinformatics* **16**, 933230 (2022).

24. Zhong, Y. *et al.* Unsupervised fusion of misaligned pat and mri images via mutually reinforcing cross-modality image generation and registration. *IEEE transactions on medical imaging* **43**, 1702–1714 (2023).

25. 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*, 2223–2232 (2017).

26. Diniz, E., Santini, T., Karim, H., Aizenstein, H. J. & Ibrahim, T. S. Cross-modality image translation of 3 tesla magnetic resonance imaging to 7 tesla using generative adversarial networks. *Hum. Brain Mapp.* **46**, e70246 (2025).

27. Hassanzadeh, R., Abrol, A., Hassanzadeh, H. R. & Calhoun, V. D. Cross-modality translation with generative adversarial networks to unveil alzheimer's disease biomarkers. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1–4 (IEEE, 2024).

28. Hassanzadeh, R., Abrol, A., Hassanzadeh, H. R. & Calhoun, V. D. A generative imputation method for multimodal alzheimer's disease diagnosis. *arXiv preprint arXiv:2508.09271* (2025).

29. Sharma, P., Kumar, M., Sharma, H. K. & Biju, S. M. Generative adversarial networks (gans): introduction, taxonomy, variants, limitations, and applications. *Multimed. tools applications* **83**, 88811–88858 (2024).

30. Dhariwal, P. & Nichol, A. Diffusion models beat gans on image synthesis. *Adv. neural information processing systems* **34**, 8780–8794 (2021).

31. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Adv. neural information processing systems* **33**, 6840–6851 (2020).

32. Meng, X., Sun, K., Xu, J., He, X. & Shen, D. Multi-modal modality-masked diffusion network for brain mri synthesis with random modality missing. *IEEE Transactions on Med. Imaging* **43**, 2587–2598 (2024).

33. 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*, 10684–10695 (2022).

34. Zhu, L. *et al.* 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*, 592–601 (Springer, 2023).

35. Dorjsembe, Z., Pao, H.-K., Odonchimed, S. & Xiao, F. Conditional diffusion models for semantic 3d brain mri synthesis. *IEEE J. Biomed. Heal. Informatics* **28**, 4084–4093 (2024).

36. Hung, A. L. Y. *et al.* Med-cdiff: Conditional medical image generation with diffusion models. *Bioengineering* **10**, 1258 (2023).

37. Nichol, A. Q. & Dhariwal, P. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, 8162–8171 (PMLR, 2021).

38. Pan, S., Eidex, Z., Safari, M., Qiu, R. & Yang, X. Cycle-guided denoising diffusion probability model for 3d cross-modality mri synthesis. In *Medical Imaging 2025: Clinical and Biomedical Imaging*, vol. 13410, 515–522 (SPIE, 2025).

39. Xiao, Z., Kreis, K. & Vahdat, A. Tackling the generative learning trilemma with denoising diffusion gans. *arXiv preprint arXiv:2112.07804* (2021).

40. Du, Y., Fu, Z., Calhoun, V. D. *et al.* Neuromark: A fully automated ica method to identify effective fmri markers of brain disorders. *NeuroImage: Clin.* **28**, 102375, DOI: 10.1016/j.nicl.2020.102375 (2020).

41. Hassanzadeh, R., Abrol, A., Pearlson, G., Turner, J. A. & Calhoun, V. D. 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**, e0293053 (2024).

42. Jack Jr, C. R. *et al.* The alzheimer's disease neuroimaging initiative (adni): Mri methods. *J. Magn. Reson. Imaging: An Off. J. Int. Soc. for Magn. Reson. Medicine* **27**, 685–691 (2008).

43. Hore, A. & Ziou, D. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, 2366–2369 (IEEE, 2010).

44. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241 (Springer, 2015).

45. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).

46. 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*, 1125–1134 (2017).

Technical Appendices and Supplementary Material

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 a 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).

Implementation Details

Network Architectures

The diffusion generators G_θ^A and G_θ^B were implemented using a UNet backbone[44], 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[45] 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[46], 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 our network and CycleGAN used similar network architectures for generators and discriminators.

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.

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.

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 $1e{-}4, 2e{-}4, 5e{-}5$, 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 $1e{-}4$ for CycleGAN and $2e{-}4$ for DDPM.