Alzheimer's Disease Prediction via Brain Structural-Functional Deep Fusing Network

Qiankun Zuo, Yanyan Shen, *Member, IEEE*, Ning Zhong, C. L. Philip Chen,
Baiying Lei, *Senior Member, IEEE*, and Shuqiang Wang, *Senior Member, IEEE*

Abstract—**Fusing structural-functional images of the brain has shown great potential to analyze the deterioration of Alzheimer's disease (AD). However, it is a big challenge to effectively fuse the correlated and complementary information from multimodal neuroimages. In this work, a novel model termed cross-modal transformer generative adversarial network (CT-GAN) is proposed to effectively fuse the functional and structural information contained in functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI). The CT-GAN can learn topological features and generate multimodal connectivity from multimodal imaging data in an efficient end-to-end manner. Moreover, the swapping bi-attention mechanism is designed to gradually align common features and effectively enhance the complementary features between modalities. By analyzing the generated connectivity features, the proposed model can identify AD-related brain connections. Evaluations on the public ADNI dataset show that the proposed CT-GAN can dramatically improve prediction performance and detect AD-related brain regions effectively. The proposed model also provides new insights into detecting AD-related abnormal neural circuits.**

Index Terms—**Multimodal fusion, brain network computing, swapping bi-attention mechanism, generative adversarial strategy.**

I. INTRODUCTION

ALZHEIMER'S disease (AD) is a progressive and irreversible neurodegenerative disease that has become the primary cause of dementia among the elderly [1]. According to statistics [2], there are more than fifty million AD sufferers all around the world. As the patients progress towards AD, they will lose cognitive abilities, including the ability to remember or think, and finally the ability to care for themselves [3], [4]. The high prevalence of AD creates a heavy financial strain on governments as well as the patients' families. Thanks to the development of artificial intelligence, researchers are studying and analyzing AD by using machine learning-based tools [5], [6], [7]. Yet, the exact cause of AD is still unknown. One of the primary causes of the aforementioned challenges is that the brain is a highly complex system, and carrying out cognitive tasks requires topological communications between regions of interest (ROIs). Therefore, the study of brain network computation is beneficial to the diagnosis and analysis of cognitive brain diseases, as well as to exploring potential biomarkers.

The graph is usually used to define the brain network. The ROIs of the whole brain network represent the graph's nodes. The relationships between the ROIs in a brain network represent their edges [8], [9]. In the brain network, there are two main types of connectivities, including functional connectivity (FC) and structural connectivity (SC). FC refers to the statistical correlation between the functional characteristics of two brain regions and can be measured using various neuroimaging techniques. It describes the activity correlation between one brain region and another, providing key information for communication between brain areas. Based on functional magnetic resonance imaging (fMRI) technology, we can determine functional connectivity by calculating the temporal correlation of blood-oxygen-level-dependent (BOLD) signals between different brain regions. The pearson correlation coefficient (PCC) is the most commonly used method to define functional connectivity. The neural fiber connection strength between brain regions is defined as SC. It makes use of diffusion tensor imaging (DTI) to measure water molecular dispersion motion.

The structural and functional connectivity can describe AD patients' pathological features from different perspectives. AD patients exhibit damage to their structural connections [10], which affects information transmission and

processing and results in cognitive dysfunction. Besides, early-stage AD patients show weakened and enhanced changes in the functional connectivity strength [11]. Many studies have employed either FC or SC to discover certain AD-related characteristics that are not detectable using traditional imaging techniques [12], [13]. They demonstrate that brain network approaches in AD studies have more benefits than the traditional imaging approach. Colclough et al. [14] presented a hierarchical inverse covariance algorithm to simultaneously infer connectivity strength at both subject and population levels. To preserve the brain's local geometry or manifold structure, Yu et al. [15] applied a weighted graph regularized sparse representation (WGraphSR) method to obtain brain connectivity. It can not only boost mild cognitive impairment (MCI) prediction performance but also reveal more valuable connections associated with MCI. To explore the causal relations between ROIs and lower individual differences, Li et al. [16] proposed a sparse constrained connectivity inference model to construct a functional network from functional time series and then built a multilayer perceptron classifier for MCI detection. Besides, the high-order features can be represented by a hypergraph, where multiple ROIs share the same edge. They [17] proposed a hypergraph learning-based method to construct brain functional connectivity for a better understanding of the brain's overall structure. To fully make use of the dynamic interactions among brain regions, [18] first derived multiple low-order functional connectivity networks (FCNs) from a series of sliding windows and then constructed a high-order FCN by measuring the topographical similarity between FCNs. The disease-related biomarkers can be successfully recognized. However, most existing investigations of brain networks concentrate on single-modal imaging, which makes it impossible to focus on the integration of structural and functional connectivity information.

Neuroscientific research indicates that AD patients exhibit damage to both structural and functional brain connections, which can lead to cognitive impairment [19]. Single-modal imaging would disregard the opportunity to use complementary information to more deeply understand AD since it might only partially contain AD-related information. Integrating structural and functional connectivity to analyze brain disorders is not only beneficial to model the general relationship between brain structure and function, but also to provide complementary information for exploring and identifying potential abnormalities in brain cognitive disorders [20]. Compared with analyzing AD using single modality (structural connectivity or functional connectivity), fusing structural and functional connectivity to analyze AD can explore pathogenesis and provide potential biomarkers for early AD diagnosis. Structure-function fusion has become a hot topic in the current AD studies [9], [21], [22]. As multimodal brain networks are heterogeneous and concealed in various types of neuroimaging data, knowing how to properly exploit complementary information between modalities is crucial for structural-functional deep fusion. The majority of current methods only employ linear interactions to fuse structural and functional information [23]. For example, [24] proposed a novel deep neural network-based model to effectively fuse structural MRI and

functional MRI by finding linkages between bimodal images. It reveals a significant correlation between the impairments in schizophrenia and the function/structure alignment score. The work in [25] developed a computer-aided detection system to combine structural and functional abnormalities for autism prediction, which discovered autism-related areas affected by impairment loss. Similarly, they [26] utilized multimodal magnetic resonance imaging to study the abnormal structure-function patterns in catatonia. The co-altered interactions in the brain are founded to facilitate visuospatial functions and motor behavior. Due to the fact that changes in brain structure and functional connectivity cannot be entirely explored by linear correlations [20], we applied a graph convolutional network (GCN)-based network to deeply fuse structural and functional connectivity information for mild cognitive impairment diagnosis. The topological properties are fully explored by iteratively updating the fused deep connectome. However, previous studies [27], [28], [29] showed that strong FC typically follows strong SC but rarely the other way around. Clinical research [30], [31], [32] demonstrates that certain regions can compensate for a lowered SC when it happens by increasing functional activity between ROIs.

Deep learning has achieved remarkable success in the field of medical image analysis [33], [34], [35]. As a branch of deep learning, generative adversarial networks (GANs) are widely used to learn complex distributions for medical image computing [36], [37], [38], [39]. Besides, transformers [40] have demonstrated their strong capacity for sequential analysis and successful applications in natural language processing (NLP). Transformers have recently been used for image tasks as a result of their successful NLP applications [41], [42]. Inspired by the above observations, in this paper, the CT-GAN is proposed to combine structural and functional connectivities for AD analysis. Using a swapping bi-attention mechanism, the proposed cross-modal transformer-based network generates multimodal connectivity (MC). A dual-channel separator and a generative adversarial strategy are used to optimize the training of the CT-GAN to maintain the learned MC's robustness. The main contributions to this work are as follows:

- The proposed CT-GAN is proposed to transform the fMRI and DTI into multimodal connectivity for AD analysis by combining the generative adversarial strategy. It not only learns the topological characteristics of non-Euclidean space but also deeply fuses complementary information in an efficient end-to-end manner.
- The swapping bi-attention mechanism (SBM) is developed to effectively align functional information with microstructural information and enhance the complementary information between bimodal images.
- The dual-channel separator with cross-weighting scheme is devised to decompose multimodal connectivity into functional and structural connectivities, which preserve global topological information and ensure the high quality and diversity of the generated connectivities.

The remaining sections of this paper are divided into the following sections: Section II presents the overall design of the proposed CT-GAN. The experimental results, including generation evaluation and classification performance, are presented
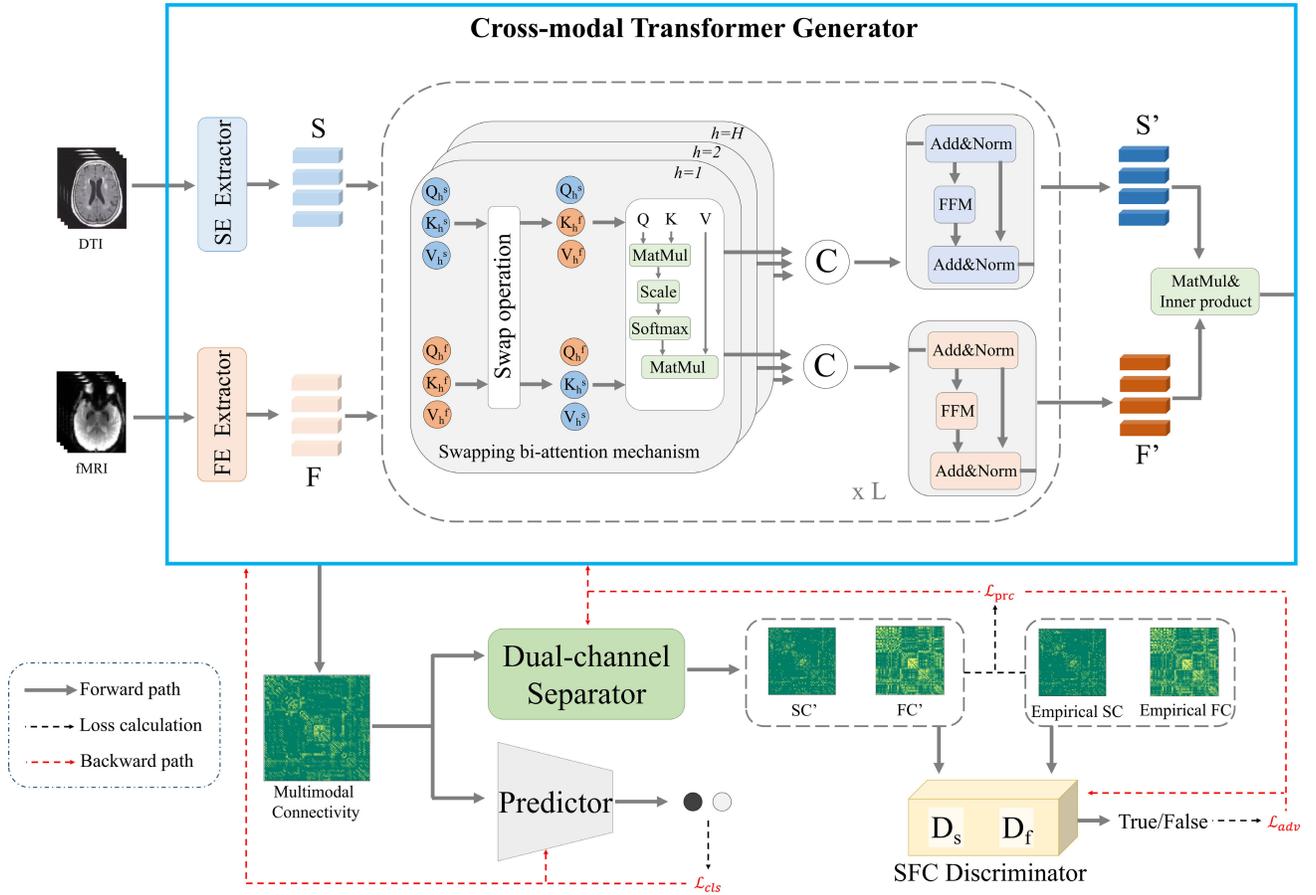
Fig. 1. The framework of the proposed CT-GAN, including four parts: the cross-modal transformer generator, the dual-channel separator, the SFC discriminator, and the predictor. S represents the structural embedding, and F represents the functional embedding. The framework aims to generate multimodal connectivity from DTI and fMRI.

in Section III. The effectiveness of our model is discussed in Section IV. The primary remarks of this study are presented in Section V.

II. METHOD

The architecture of the proposed CT-GAN is shown in Figure 1. Given bimodal images (i.e., fMRI and DTI), the proposed model can learn a non-linear learning network that transforms imaging space into topological connectivity space. To obtain multimodal connectivity, we designed the CT-GAN with four components: 1) the cross-modal transformer generator (G) that is used for inference and generates multimodal connectivity; 2) the dual-channel separator (DS) that decomposes the multimodal connectivity into SC and FC; 3) the structural-functional consistency (SFC) discriminator, which contains two sub-discriminators (i.e., D_s and D_f). Each of them discriminates whether an SC or FC comes from the proposed generator or the software toolboxes; 4) the predictor that assigns AD stages based on the generated multimodal connectivity.

A. Cross-Modal Transformer Generator

1) Embedding Extractor: To embed ROI-based features into the transformer-based network, a routine convolutional neural network is adopted to extract ROI-based features from brain imaging. We design two extractors, including a structural embedding (SE) extractor and a functional embedding (FE)

extractor. As shown in the upper left of Figure 1, the embedding S is computed by successive convolutional filters on the DTI. Specifically, we first design four down-sampling convolutional operations with a kernel size of $3 \times 3 \times 3$ and a stride of 2 to extract local feature maps. The extracted feature maps are then passed through $1 \times 1 \times 1$ filters to fix the channel at N. Finally, each channel map is combined with the brain anatomical information (x, y, z, v) to align the features for each brain region [43]. Similar operations are conducted on the fMRI. The output embeddings $\mathbf{S}$ and $\mathbf{F}$ are given below:

$$\mathbf{S} = SE(DTI, x, y, z, v), \quad \mathbf{F} = FE(fMRI, x, y, z, v) \quad (1)$$

where $\mathbf{S} \in \mathbb{R}^{N \times q}$, $\mathbf{F} \in \mathbb{R}^{N \times q}$.

2) Swapping Bi-Attention Mechanism: The proposed model aims to leverage the transformer's bi-attention mechanism to explore complementary information between structural and functional images. Traditional transformers haven't been thoroughly studied in the context of brain network computing, and they just model relationships between brain regions within a single modality, which fails to effectively explore the complementary information between modalities. To mine the complementary information between fMRI and DTI, we devise the swapping bi-attention mechanism (SBM) to proficiently align functional features with microstructural information. It can facilitate the synergistic exchange of information between bimodal images. In this section, we first introduce

the traditional transformer and then detail the proposed SBM module. The traditional transformer gradually projects an input embedding as $\mathbf{S} \in \mathbb{R}^{N \times q}$ to a target feature embedding as $\mathbf{F} \in \mathbb{R}^{N \times q}$, where N represents the overall number of ROIs. The following is a description of the traditional attention learning process: (1) computing query matrices Q, key matrices K, and value matrices V through a linear projection.

$$Q = XW^q, \quad K = XW^k, \quad V = XW^v \quad (2)$$

where, X represents the $\mathbf{S}$ or $\mathbf{F}$. $W^q \in \mathbb{R}^{q \times q}$, $W^k \in \mathbb{R}^{q \times q}$, and $W^v \in \mathbb{R}^{q \times q}$ are weight parameters. (2) computing the attention of X by the softmax function:

$$\text{Att}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{q}}\right)V. \quad (3)$$

In the SBM module, we first design H heads for each modality to focus on different parts of the learned embedding. The tokens can be computed by

$$Q_h^S = \mathbf{S}W_h^{qs}, \quad K_h^S = \mathbf{S}W_h^{ks}, \quad V_h^S = \mathbf{S}W_h^{vs} \quad (4)$$

$$Q_h^F = \mathbf{F}W_h^{qf}, \quad K_h^F = \mathbf{F}W_h^{kf}, \quad V_h^S = \mathbf{F}W_h^{vf} \quad (5)$$

where, h is the index of all H heads. Each head has the dimension q/H. Each token (i.e., Q_h^S, K_h^S, V_h^S) has the same size, $N \times q/H$. Then, we exchange the tokens between the two modalities and fuse the intermediate features adaptively. For structural modality, the token Q_h^S is combined with the other two tokens (K_h^F and V_h^F) to adaptively bring additional functional information into the structural features. And vice versa for functional modality. The structural and functional swapping bi-attention can be defined as

$$S_h^{Att} = \text{F2SAtt}(Q_h^S, K_h^F, V_h^F) = \text{softmax}\left(\frac{Q_h^S(K_h^F)^T}{\sqrt{q/H}}\right)V_h^F. \quad (6)$$

$$F_h^{Att} = \text{S2FAtt}(Q_h^F, K_h^S, V_h^S) = \text{softmax}\left(\frac{Q_h^F(K_h^S)^T}{\sqrt{q/H}}\right)V_h^S. \quad (7)$$

Finally, we can obtain the cross-modal bi-attention features S^{Att} and F^{Att} with the following formula:

$$S^{Att} = [S_1^{Att}, S_2^{Att}, \ldots, S_H^{Att}] \quad (8)$$

$$F^{Att} = [F_1^{Att}, F_2^{Att}, \ldots, F_H^{Att}] \quad (9)$$

where, [,] denotes the concatenation along the ROI feature direction. The cross-modal attention mechanism allows for more transparent feature alignment and fusion, making it easier to understand how different modalities contribute to feature extraction and ultimately impact multimodal fusion effects.

3) Feed Forward Mapping: After the bi-attention mechanism, the cross-modal bi-attention features F^{Att} and S^{Att} represent the mixed functional feature sequence and mixed structural feature sequence, respectively. We apply linear mapping to them for effective information adjustment. Where, we denote the updated ROI-based features as $\mathbf{F}' \in \mathbb{R}^{N \times q}$ and $\mathbf{S}' \in \mathbb{R}^{N \times q}$, respectively. N is the number of all the ROIs.

The computing processes of the feed-forward mapping (FFM) layer are defined as

$$\mathbf{S}^{norm} = Norm(\mathbf{S} + S^{Att}) \quad (10)$$

$$\mathbf{F}^{norm} = Norm(\mathbf{F} + F^{Att}) \quad (11)$$

$$\mathbf{S}' = Norm(\mathbf{S}^{norm} + FFM(\mathbf{S}^{norm})) \quad (12)$$

$$\mathbf{F}' = Norm(\mathbf{F}^{norm} + FFM(\mathbf{F}^{norm})) \quad (13)$$

4) Connectivity Computation: After L layers of transformer, we obtained the updated mixed structural and functional features $\mathbf{S}'$ and $\mathbf{F}'$. These mixed features contain common and unique information for both modalities. Where, we utilize the PCC to define multimodal connectivity. We first project one modal feature onto the other modal feature, then compute the relationship of paired ROIs with the following formula:

$$\text{MC} = \mathbf{S}'\mathbf{F}'\mathbf{F}'^T\mathbf{S}'^T. \quad (14)$$

where, MC means the final multimodal connectivity (MC) with the size $N \times N$. In the experiment, we follow the fourth method of FC definition in [44] to recompute the multimodal connectivity for AD analysis. The elements in the MC matrix below the threshold are represented by 0, and those above the threshold remain unchanged.

B. Dual-Channel Separator

The MC contains both structural and functional connectivity information. To stabilize the learning process, we design the dual-channel separator to recover the SC and FC from the MC. As shown in Figure 2, the dual-channel separator projects the MC back to two modality-specific connectivities. Considering the topological properties of the human brain, we adopt the cross-weighting scheme to extract global connectivity information for better detachment between structural and functional connectivity. It consists of two branches, which share the first layer and have different weighting parameters in the second and third layers, respectively. The filter is a cross-shape parameter with step size 1. The input and the output for each layer have the same size, except for different channels. Finally, the third layer outputs the reconstructed SC and FC.

C. Structural-Functional Consistency Discriminator

To make the reconstructed SC and FC have the same distribution as the empirical SC and FC, the structural-functional consistency (SFC) discriminator is designed to distinguish empirical SC and FC from the reconstructed SC and FC and provide feedback loss on the generator. It not only makes the dual-channel separator robust but also improves the generator's learning abilities. The SFC discriminator contains two sub-discriminators (i.e., D_s and D_f). Both discriminators share the same network structure. For the sake of narration, we take D_s as an example to describe the detailed computing process. As shown in Figure 3, the input SC is passed through two branches, including local convolution (top) and global convolution (bottom). The top branch contains four 3×3 convolution operations, one average pooling operation, and one flattened layer. The bottom branch contains two kinds of convolution filters (1×90 and 90×1) and one flattened layer. The final output is a value ranging from $0 \sim 1$, which indicates whether the input is a reconstructed or empirical SC.
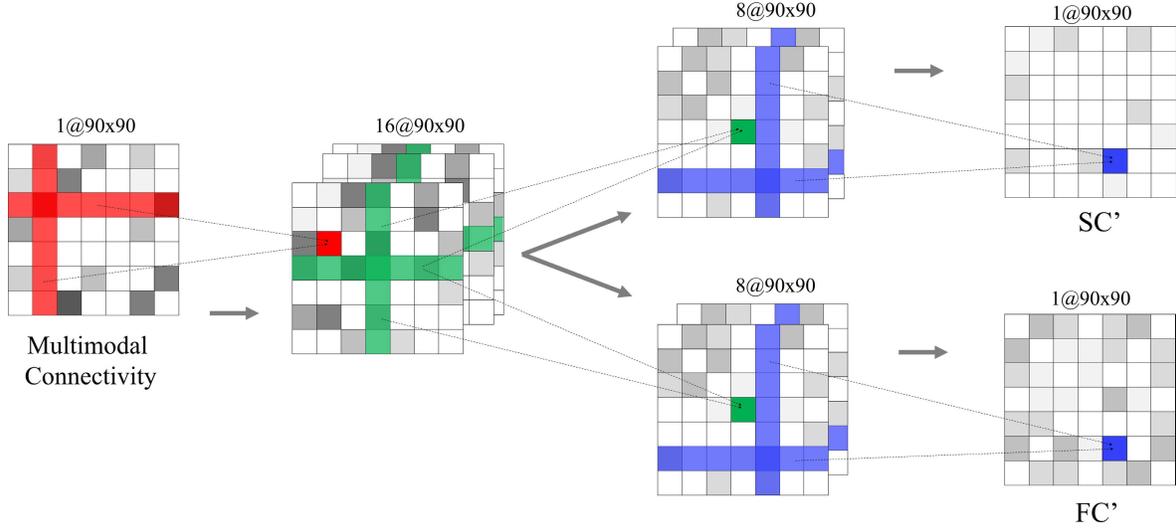
Fig. 2. The network architecture of the dual-channel separator. Given the multimodal connectivity, it outputs the structural connectivity and functional connectivity.
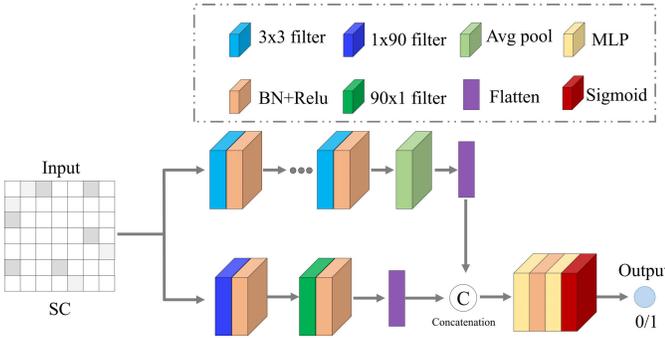

Fig. 3. The network architecture of one sub-discriminator in the SFC discriminator.

D. Loss Function

To effectively fuse the functional and structural information contained in fMR and DTI, the proposed CT-GAN model is optimized using a hybrid loss function that incorporates three types of objective losses: adversarial loss, classification loss, and pair-wise connectivity reconstruction loss. These three loss functions can ensure the quality of multimodal connectivity generation. The loss functions are as follows:

Adversarial Loss. To make the FC' matrix and SC' matrix decoded from the multimodal connectivity matrix as similar to empirical FC matrices and SC matrices as possible, the adversarial losses are defined as follows:

$$\mathcal{L}_{adv}^{SC} = \mathbb{E}_{x \sim P_{SC}}[\log D_s(x)] + \mathbb{E}_{y \sim P_{SC'}}[\log(1 - D_s(y))], \tag{15}$$

$$\mathcal{L}_{adv}^{FC} = \mathbb{E}_{x \sim P_{FC}}[\log D_f(x)] + \mathbb{E}_{y \sim P_{FC'}}[\log(1 - D_f(y))], \tag{16}$$

$$SC', FC' = DS(G(DTI, fMRI)) \tag{17}$$

where G is the cross-modal transformer generator. The distribution of empirical SC matrix is represented by P_{SC} and the distribution of empirical FC matrix is represented by P_{FC}.

Classification Loss. Since MC matrices are usually used to predict the stages of AD, generating discriminative MC matrices can be an indicator of cross-modal fusion quality. The classification loss can guide the generator to learn AD-related MC matrices. They are defined by the following formula:

$$\mathcal{L}_{cls} = \mathbb{E}_{x \sim P_{DTI}, y \sim P_{fMRI}}[-\log p(Y|C(G(x, y)))], \tag{18}$$

where $G(x, y)$ is the output MC of the generator. C means the predictor. Y is the disease stage, including the normal control (NC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and AD. $p(Y|C(G(x, y)))$ is the probability that the subject is predicted to be stage Y.

Pair-wise Connectivity Reconstruction Loss. To impose an additional topological constraint on the cross-modal transformer generator, we add the $L1$ pair-wise connectivity reconstruction loss in the model's optimization process. The overall pair-wise connection gap between empirical FC/SC matrices and FC/SC matrices are minimized by the following formula:

$$\mathcal{L}_{pcr}^{FC} = \|FC - FC'\|_1, \tag{19}$$

$$\mathcal{L}_{pcr}^{SC} = \|SC - SC'\|_1. \tag{20}$$

III. Experiments

A. Preprocessing and Settings

The ADNI (Alzheimer's Disease Neuroimaging Initiative) public dataset is used to test our CT-GAN model. Table I contains full information about the 268 patients whose data we used in this study. Each patient was scanned with both DTI and fMRI. The preprocessing procedure makes use of the AAL 90 atlas. Using the DPARSF toolkit, the top 20 volumes are eliminated, followed by head motion correction, band-pass filtering, Gaussian smoothing, and extracting the time series of all voxels. By following fiber bundles between ROIs,

TABLE I
DATA INFORMATION IN THIS STUDY

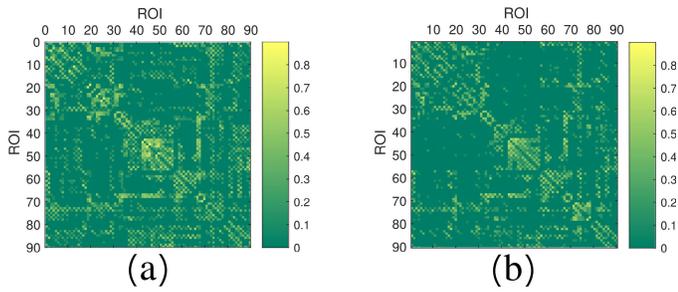| Group	NC(84)	EMCI(80)	LMCI(41)	AD(63)
Male/Female	38M/46F	48M/32F	20M/21F	39M/24F
Age(mean $\pm$ SD)	74.0 $\pm$ 5.9	75.8 $\pm$ 6.1	74.9 $\pm$ 5.3	75.3 $\pm$ 5.5

Fig. 4.   Examples of two multimodal connectivity matrices at different stages of cognitive disease (a) NC; (b) AD.

Fig. 5.   The ten most important brain regions between NC and EMCI groups.

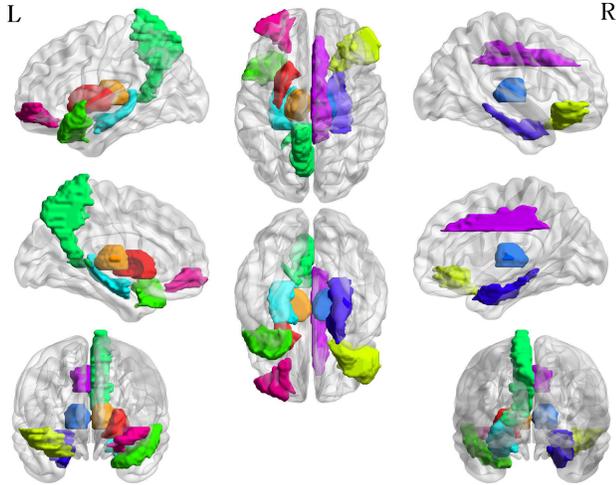Fig. 6.   The ten most important brain regions between EMCI and LMCI groups.
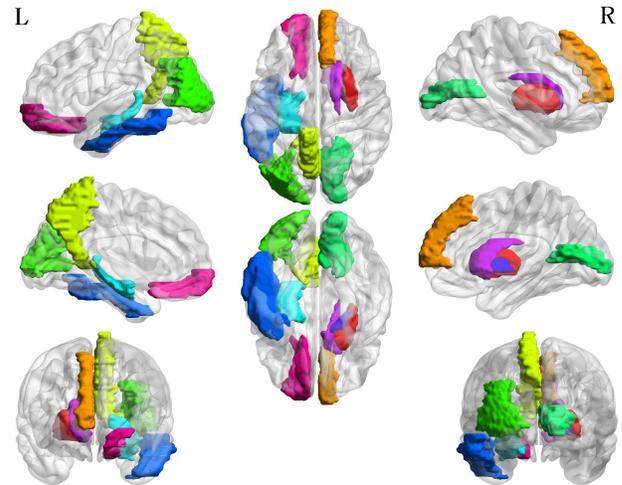
Fig. 7.   The ten most important brain regions between LMCI and AD.

the structural connection is computed. The requirements are configured in PANDA as the fiber tracking halting conditions: a crossing angle of greater than 45 degrees between two traveling directions.

The predictor is implemented by the row-based filters in the work [47]. The embedding dimension in the generator G is set at 128. $L = 5$ layers of transformer are utilized to fuse structural and functional embeddings. The heads in the transformer block are 8. The model's parameters will be updated during the training process using the Adam algorithm. The learning rate is set to 0.001. The weight decay is set to 0.01. The four widely used metrics − accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC) − make up the evaluation criteria.

B. Prediction Results

To demonstrate the generation effect of the proposed model, Figure 4 qualitatively depicts four examples of the generated MC at different stages. Even though the four MCs show the same global connectivity patterns, different connectivity characteristics can be seen in the local area. The MC at the AD stage has the sparsest connectivity features.

To conduct a quantitative analysis of the proposed model's classification, we conducted three binary classification tasks (i.e., NC vs. EMCI, EMCI vs. LMCI, and LMCI vs. AD). Each classification task is operated with a five-fold cross-validation strategy. To evaluate how well various fMRI-DTI

fusion models can capture characteristics associated with AD, two competing models and two classifiers are introduced in our experiments for comparison. Specifically, the multi-modal enhanced graph convolutional network (MMEGCN) [45] and the graph-based deep model (GBDM) [20] output a combined brain network by inputting fMRI and DTI. After the combined brain network has been generated, we adopt two classifiers (GCN [46] and Brainnetcnn [47]) to evaluate the classification performance of the generated brain networks. Table II shows the detailed prediction performance between different competing models. Under different classification tasks, our model achieves superior results to others in terms of different classifiers. Both classifiers have similar classification performances. The best classification results for NC vs. EMCI are ACC value of 90.24%, SEN value of 90.00%, SPE value of 90.48%, and AUC value of 93.26%; the EMCI vs. LMCI task achieves the best ACC value of 93.39%, SEN value of 92.68%, SPE value of 93.75%, and AUC value of 94.51%; the best results for LMCI vs. AD are ACC value of 95.19%, SEN value of 95.24%, SPE value of 95.12%, and AUC value of 94.27%. Overall, the results of the experiments demonstrate that the
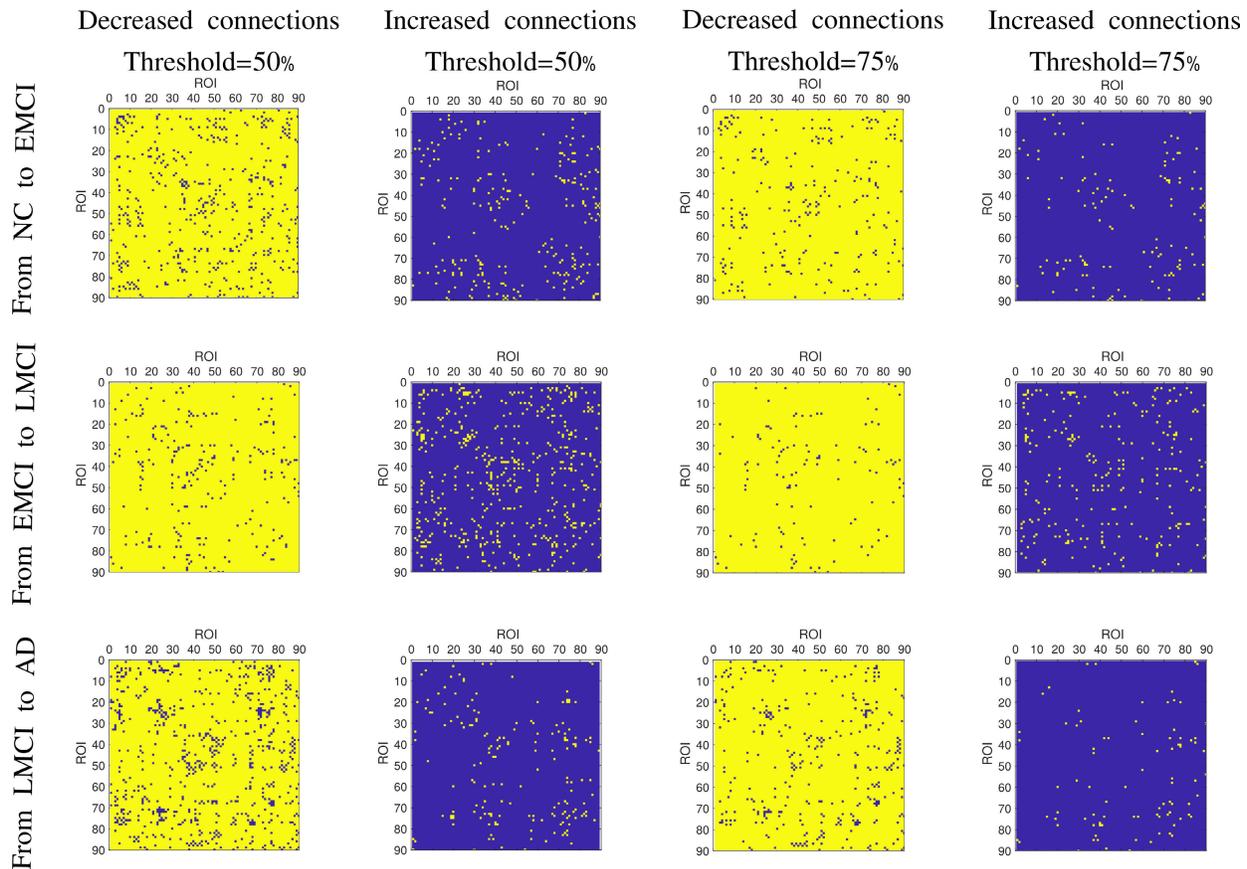
Fig. 8. The altered connectivity between two MC groups. The first and third columns are the decreased connectivity matrices, with the threshold values set at 50% and 75% respectively. The second and fourth columns are the increased connectivity matrices with the threshold values at 50% and 75% respectively.

TABLE II
PREDICTION OF PERFORMANCE UNDER DIFFERENT MODELS AND CLASSIFIERS BY FUSING FMRI AND DTI(%)

| Models	Classifiers	NC vs. EMCI				EMCI vs. LMCI				LMCI vs. AD			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
MMEGCN		87.20	91.25	83.33	89.35	89.26	87.80	90.00	90.70	92.31	92.06	92.68	93.46
GBDM	GCN	87.80	85.00	90.48	90.95	88.43	87.80	88.75	89.45	93.27	93.65	92.68	94.39
Ours		**89.63**	90.00	89.29	**93.44**	**93.39**	**92.68**	**93.75**	**94.51**	**94.23**	93.65	**95.12**	**95.20**
MMEGCN		88.41	87.50	89.29	90.71	88.43	90.24	87.50	88.17	91.35	90.48	92.68	89.47
GBDM	Brainnetcnn	88.41	90.00	86.90	90.31	91.74	90.24	92.50	89.42	93.27	92.06	95.12	91.99
Ours		**90.24**	90.00	**90.48**	**93.26**	**92.56**	90.24	**93.75**	**93.93**	**95.19**	**95.24**	95.12	**94.27**

proposed CT-GAN has the benefit of being more accurate than previous multimodal fusion models in predicting the phases of AD.

To evaluate the AD-related ROIs in the classification tasks, we utilized the LOOCV method [6] to compute the important score for each ROI. To calculate the importance score for each ROI, we first began to remove one row and one column corresponding to one particular ROI in the generated multimodal connectivity matrix. We then computed the mean classification accuracy of the removed connectivity matrices. Subsequently, we derived the ROI's importance score by subtracting this mean accuracy from one. This computation process was repeated for all the ROIs. The scores were then arranged in descending order, and the top 10 ROIs represent the important brain regions. The important ROIs are displayed by the BrainNetviewer tool [48]. As shown in Figure 5, the ten important ROIs for NC vs. EMCI are the left lenticular nucleus putamen, left thalamus, right inferior frontal gyrus orbital part, left temporal pole superior temporal gyrus, left precuneus, left hippocampus, right thalamus, right parahippocampal gyrus, right median cingulate and paracingulate gyri, and left middle frontal gyrus orbital part. For EMCI vs. LMCI, the identified 10 important ROIs in Figure 6 are the right lenticular nucleus putamen, the right calcarine fissure and surrounding cortex, the left precuneus, the left gyrus rectus, the left thalamus, the left superior frontal gyrus orbital part, the right superior temporal gyrus, the right hippocampus, the left superior occipital gyrus, and the left calcarine fissure and surrounding cortex. In Figure 7, the important ROIs between the LMCI and AD groups are the following: right lenticular nucleus putamen,
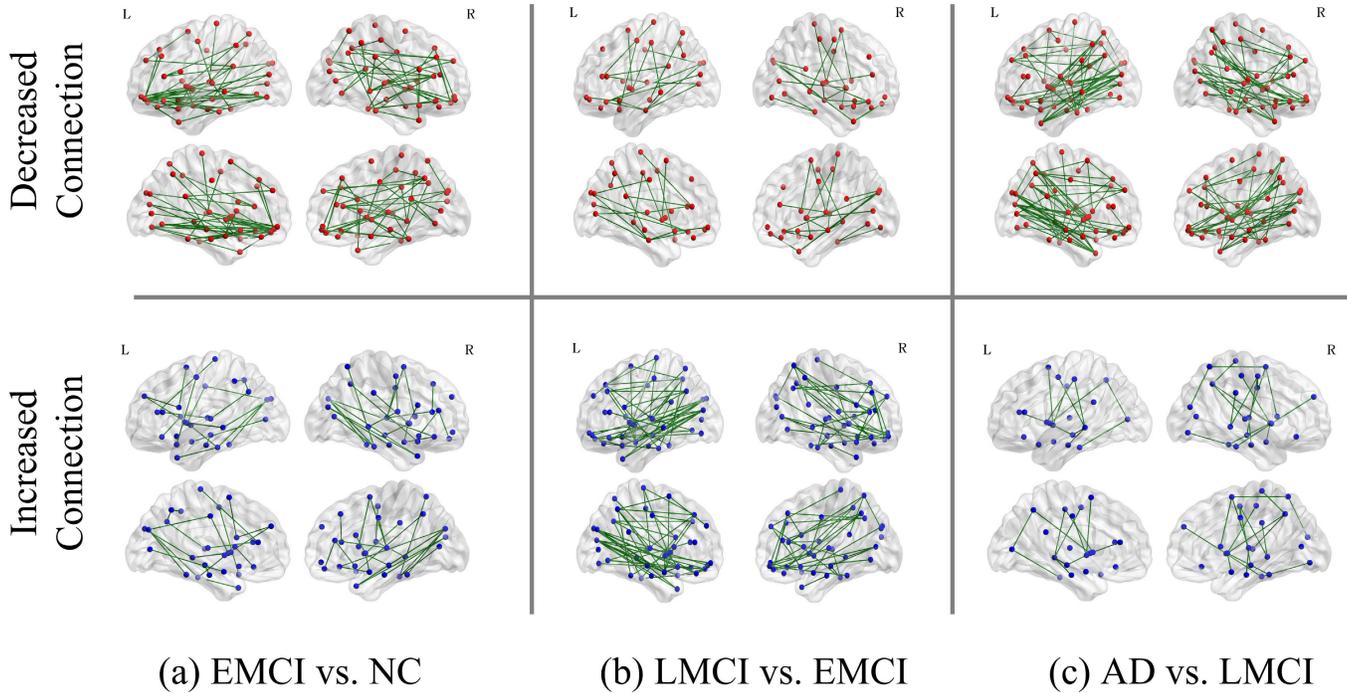
Fig. 9.   The altered connectivities with the threshold value at 75% quantile for the three scenarios.

right superior frontal gyrus medial, left precuneus, left middle occipital gyrus, right calcarine fissure and surrounding cortex, left hippocampus, left inferior temporal gyrus, right lenticular nucleus pallidum, right caudate nucleus, left superior frontal gyrus orbital part. The identified important ROIs are found to be related to Alzheimer's disease and are partly overlapped with previous studies [49], [50], [51].

C. Connectivity Analysis

To analyze the important connections associated with AD, we first compute the averaged MC matrix for each disease group (i.e., NC, EMCI, LMCI, and AD) and then evaluate the difference matrix between adjacent groups. The positive value in the difference matrix means increased connections, and the negative value represents decreased connections. The averaged multimodal connectivity matrices at different stages of AD disease can be obtained by applying the trained generator to DTI and fMRI. The visualization of averaged multimodal connectivity matrices and the change in connectivity with various thresholds are shown in Figure 8. The three rows correspond to the altered connections from NC to EMCI, from EMCI to LMCI, and from LMCI to AD, respectively. The values between $-0.1 \sim 0.1$ are ignored during the analysis. Two threshold values are set for viewing the important connections. The first threshold is 50% quantile values, which are estimated from the positive and negative connectivities. The same operation is implemented on the second 75% threshold value. The more important connections with the 75% threshold value are shown in Figure 9. It can be seen that the decreased connections are greater than the increased connections at the stages of EMCI and AD, while the phenomenon is reversed at the LMCI stage.

To evaluate the most important connections for different stages of AD, we sort the altered connection strength and
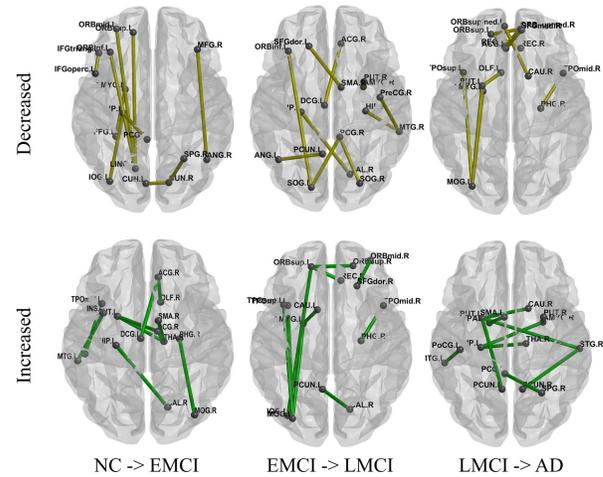
Fig. 10.   The spatial view of the top 10 decreased and increased connections for the three scenarios.

find the top 10 connections for both increased and decreased situations. The results are shown in Figure 10 and Figure 11. For the NC vs. EMCI, the increased connections are SMA.R - THA.R, TPOmid.L - ITG.L, DCG.R - THA.R, OLF.R - ACG.R, INS.L - MTG.L, DCG.R - PUT.L, ACG.R - DCG.L, PUT.L - THA.R, PHG.R - MOG.R, HIP.L - CAL.R; the decreased connections are PCG.L - HIP.L, IFGoperc.L - IFG-triang.L, ORBsup.L - LING.L, CUN.R - SPG.R, ORBinf.L - LING.L, AMYG.L - FFG.L, MFG.R - ANG.R, ORBmid. L - LING.L, HIP.L - IOG.L, CUN.L - CUN.R. As the stage changes from EMCI to LMCI, the increased connectivities are the following: AMYG.L - MOG.L, ORBsup.L - REC.R, CAL.R - PCUN.L, MOG.L - TPOsup.L, MOG.L - TPOmid.L, PHG.R - TPOmid.R, ORBsup.L - ORBsup.R, ORBsup.L - IOG.L, SFGdor.R - ORBmid.R, AMYG.L - CAU.L; and the decreased connectivities are the following: PCG.R - SOG.L,

Fig. 11. A view of the circos plot for the corresponding top 10 decreased and increased connections.

PreCG.R - MTG.R, AMYG.R - PUT.R, SFGdor.L - SMA.R, PCG.R - SOG.R, ACG.R - DCG.L, ORBinf.L - SOG.L, HIP.L - CAL.R, HIP.R - MTG.R, ANG.L - PCUN.L. The top 10 connections from LMCI to AD are increased at PCG.L - SPG.R, HIP.L - THA.R, CAU.R - PUT.L, PoCG.L - MTG.L, PCUN.L - PUT.L, SMA.L - PAL.L, AMYG.R - PUT.R, PUT.L - STG.R, PCUN.R - STG.R, HIP.L - AMYG.R; and decreased at ORBsupmed. R - REC.L, ORBsup.L - REC.R, ORBsup.L - ORBsupmed.R, PHG.R - TPOmid.R, AMYG.L - MOG.L, ORBsupmed.L - CAU.R, OLF.L - PUT.L, SFGmed. R - ACG.L, MOG.L - PUT.L, MOG.L - TPOsup.L.

To show the important ROI-related connections, we counted the overlapping brain regions of the three top 10 ROIs in the prediction results and found six frequent ROIs. These ROIs include the following: left superior frontal gyrus orbital part, left hippocampus, right calcarine fissure and surrounding cortex, left precuneus, right lenticular nucleus putamen, left thalamus Based on the increased and decreased connectivities, the important ROI-related connections are shown in Figure 12. Even though these connections are not listed in the top 10 altered connections, they reflect the changing characteristics at the different stages of AD. For example, the connection ORBsup.L-HIP.L first drops in strength at EMCI, then gains some strength at LMCI, and finally loses strength again at the AD stage. These findings can be explained by the compensation mechanism [11], [52]. The six identified important ROIs have AD-related biological meaning. For example, the

PCUN.L, situated within the parietal lobes on the medial surface of the left cerebral hemisphere, is involved in functions such as visuospatial imagery, episodic memory retrieval, and self-processing operations. It achieves these functions by integrating sensory information from various sources. Damage to the precuneal region could potentially result in disorders related to consciousness, impairment in self-processing, visuospatial challenges, as well as deficits in both episodic and semantic memory. The CAL.R region is primarily responsible for visual processing tasks, encompassing the interpretation of color and motion in the visual field. Injury to CAL.R can lead to specific visual field defects. The patient may experience various forms of visual hallucinations and color agnosias.

D. Ablation Studies

The proposed CT-GAN can generate discriminative MCs for early AD analysis. The prediction precision and altered connectivity evaluation proved the effectiveness of the model. To study the impact of different modules in the model on the prediction results, we (1) replace the SBM module with a traditional multi-head self-attention mechanism; and (2) remove the SFC discriminator. As shown in Figure 13, the removal of the SBM mechanism significantly lowers the prediction performance. As well, the SFC discriminator affects the classification accuracy to some extent.

In order to quantitatively evaluate the effect of structural-functional fusion, we conducted multiple sets of classification
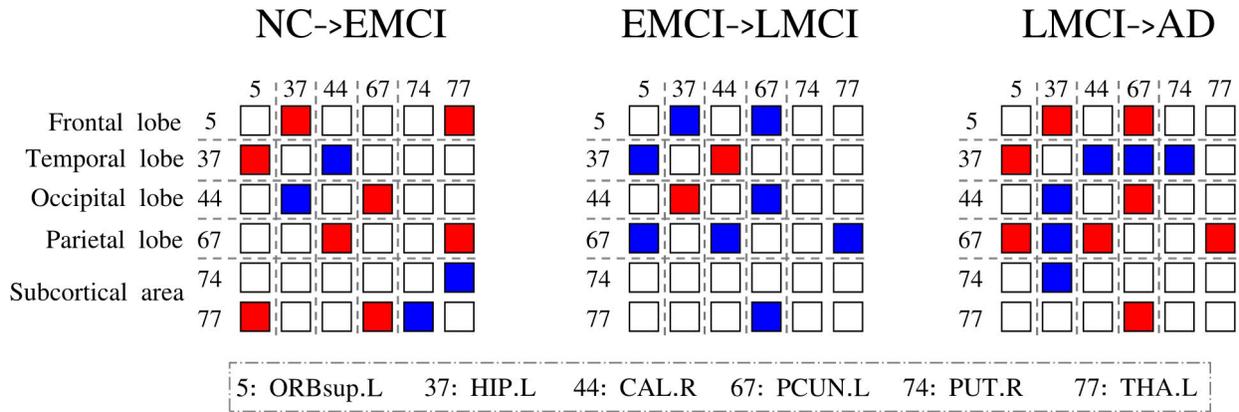
Fig. 12. The altered multimodal connectivities associated with the overlapping six ROIs in the prediction results. The index indicates the corresponding ROI in the AAL90 atlas. The red color represents decreased connections; the blue color represents increased connections. The gray dotted lines divide the six ROIs into five brain lobes.
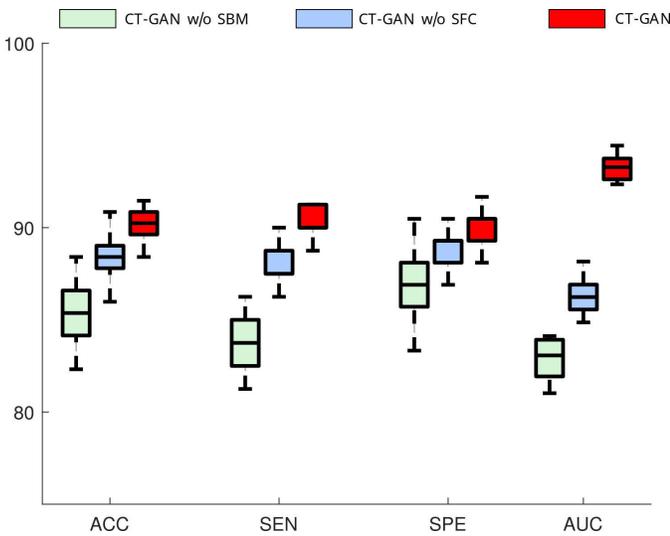

Fig. 13. Influence of different modules in CT-GAN on the prediction performance.

experiments. First, we individually computed the classification performance using functional brain imaging (fMRI) (as shown by the red color in Figure 14). Then, we individually computed the classification performance using structural brain imaging (DTI) (as shown by the blue color in Figure 14). Finally, we fused functional and structural brain imaging and presented the classification results in green color in Figure 14. The experimental results indicates that adding structural brain imaging to functional brain imaging can improve classification accuracy by 7% ∼ 9%. The results also indicate that integrating functional and structural brain imaging to analyze AD can better improve the AD prediction performance than using single modality (either structural or functional brain imaging).

IV. DISCUSSION

Fusing structural and functional brain images to analyze AD can establish complex nonlinear relationships and fully mine complementary information between structure and functional connectivity. It can improve the performance of AD diagnosis and helps explore the pathogenesis of AD. The

results in Table I shows that our model achieves a maximum improvement of 4.96% in accuracy compared to other models for EMCI vs. LMCI. At each stage of AD, our model has high resolution in detecting patients. The highest classification accuracy in NC vs. EMCI, EMCI vs. LMCI, and LMCI vs. AD are 90.24%, 93.39%, and 95.19%, respectively. Compared with other models, our model demonstrates noticeable improvements in classification performance as follows: from NC to EMCI, the classification accuracy has increased by up to 2.43%; from EMCI to LMCI, the classification accuracy has increased by up to 4.96%; and from LMCI to AD, the classification accuracy increased by up to 3.84%.

It should be stressed that an improvement of 2 percentage points is good work in the field of AD diagnosis [53], [54], [55]. Furthermore, our model improves the classification accuracy by at least 8 percentage points compared with single-modality-based methods [56]. For example, in the NC vs. EMCI classification, the classification accuracy of the fMRI-based method is 75%, and the classification accuracy of the DTI-based method is 81.81%. Our model achieved an accuracy of 90.24%, which improves the classification accuracy by 15.24 and 8.43 percentage points over the fMRI-based and DTI-based method, respectively. The 4-5 percentage point improvement in classification performance by our model has important clinical significance in AD diagnosis. AD is a multi-stage progressive neurodegenerative disorder, and accurately identifying different stages of AD enables clinicians to provide more personalized treatment plans for patients. Our model exhibits high discriminative accuracy (above 90%) in four stages of AD. The patients can be accurately identified and receive specific treatment plans according to the characteristics of the identified AD stage, which is more effective to slow down the AD progression.

More importantly, the improvement in classification accuracy is just one of our contributions, and another more important contribution of our work is that we design a unified framework to fuse structural modality and functional modality for AD analysis. The current methods of fusing structural and functional brain images are based on two steps: the first step is to preprocess the brain structural and functional images
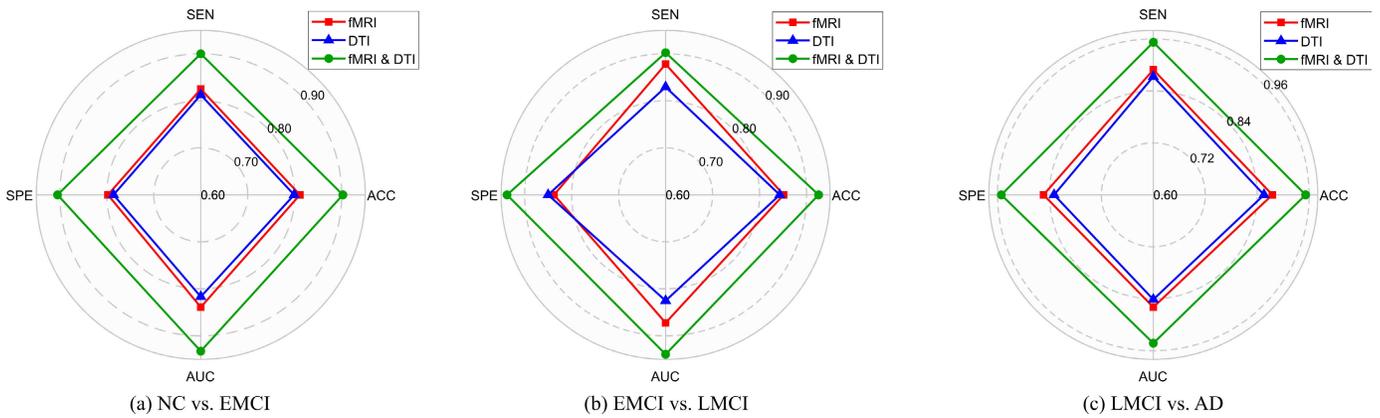
Fig. 14. Comparison of classification performance using single-modal and bimodal images.

to obtain structural and functional features by the software toolbox; the second step is to use the preprocessed structural and functional features to build deep learning models for fusion. The novelty of our model is constructing an end-to-end framework to fuse structural brain imaging (DTI) and functional brain imaging (fMRI) for AD analysis. By gradually aligning functional and structural information, the proposed framework can mine complementary information between modalities and identify AD-related abnormal connection features, which is beneficial to reveal the AD pathogenesis and provide potential biomarkers for early AD diagnosis. The proposed model can perform end-to-end automatic analysis of multi-modal brain images to assist the clinicians in AD diagnosis. It has potential clinical applications in terms of efficiency.

AD is a chronic neurodegenerative disease, and data collection is time-consuming and economically costly. Among the publicly available datasets, the representative dataset containing brain imaging data for all stages of AD is the Alzheimer's Disease Neuroimaging Initiative (ADNI). Since the sample size of the ADNI dataset is not large, we have applied for the UK biobank dataset (https://www.ukbiobank.ac.uk/) to study AD. This dataset is a large biomedical research repository containing magnetic resonance imaging (MRI) brain scan data from more than 500,000 UK participants. This large sample data can be suitable to study neurodegenerative diseases (e.g., Alzheimer's disease). Due to the huge size of the dataset, downloading and preprocessing of brain imaging data is very time-consuming, and we are still in the collection stage of the brain imaging data set. We will validate our model on the UK biobank data in future work.

V. CONCLUSION

In this paper, we propose a novel CT-GAN model to fuse fMRI and DTI and generate multimodal connectivity from fMRI and DTI in an efficient end-to-end manner. The key idea of this work is that mutual conversion between structural and functional information is accomplished using a cross-modal swapping bi-attention mechanism. Therefore, the proposed model can gradually and effectively extract complementary information between modalities. The results of the experiments demonstrated that the multimodal connectivity generated by

our model is more accurate than other multimodal fusion models in terms of classification performance. Furthermore, some AD-related connectomes and brain regions are identified by analyzing the generated multimodal connectivity matrices. These connectomes partially agree with the clinical investigations on AD, which indicates that the proposed model can provide new insights for detecting AD-related abnormal connectivities. In the future, we will extend the CT-GAN to other neurodegenerative diseases for evaluation and analysis.

REFERENCES

[1] M. Dadar et al., "Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease," IEEE Trans. Med. Imag., vol. 36, no. 8, pp. 1758–1768, Aug. 2017.
[2] A. Association, "2019 Alzheimer's disease facts and figures," Alzheimer's Dementia, vol. 15, no. 3, pp. 321–387, Mar. 2019.
[3] P. Scheltens et al., "Alzheimer's disease," Lancet, vol. 388, pp. 505–517, Aug. 2016.
[4] M. Colom-Cadena et al., "The clinical promise of biomarkers of synapse damage or loss in Alzheimer's disease," Alzheimer's Res. Therapy, vol. 12, pp. 1–12, Dec. 2020.
[5] S. Wang, H. Wang, A. C. Cheung, Y. Shen, and M. Gan, "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," in Deep Learning Applications, 2020, pp. 53–73.
[6] S. Yu et al., "Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Lima, Peru, Oct. 2020, pp. 228–237.
[7] Y. Zong, C. Jing, and Q. Zuo, "Multiscale autoencoder with structural-functional attention network for Alzheimer's disease prediction," in Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV), Shenzhen, China, Nov. 2022, pp. 286–297.
[8] Y. Zhang et al., "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," Pattern Recognit., vol. 88, pp. 421–430, Apr. 2019.
[9] X. Song et al., "Multicenter and multichannel pooling GCN for early AD diagnosis based on dual-modality fused brain network," IEEE Trans. Med. Imag., vol. 42, no. 2, pp. 354–367, Feb. 2023.
[10] N. Shu, X. Wang, Q. Bi, T. Zhao, and Y. Han, "Disrupted topologic efficiency of white matter structural connectome in individuals with subjective cognitive decline," Radiology, vol. 286, no. 1, pp. 229–238, Jan. 2018.
[11] D. Berron, D. van Westen, R. Ossenkoppele, O. Strandberg, and O. Hansson, "Medial temporal lobe connectivity and its associations with cognition in early Alzheimer's disease," Brain, vol. 143, no. 4, pp. 1233–1248, Apr. 2020.
[12] E. Jeon, E. Kang, J. Lee, J. Lee, T.-E. Kam, and H.-I. Suk, "Enriched representation learning in resting-state fMRI for early MCI diagnosis," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Lima, Peru, Oct. 2020, pp. 397–406.

[13] S. Wang, Y. Shen, W. Chen, T. Xiao, and J. Hu, "Automatic recognition of mild cognitive impairment from MRI images using expedited convolutional neural networks," in *Proc. Int. Conf. Artif. Neural Netw.*, Oct. 2017, pp. 373–380.

[14] G. L. Colclough, M. W. Woolrich, S. J. Harrison, P. A. Rojas López, P. A. Valdes-Sosa, and S. M. Smith, "Multi-subject hierarchical inverse covariance modelling improves estimation of functional brain networks," *NeuroImage*, vol. 178, pp. 370–384, Sep. 2018.

[15] R. Yu, L. Qiao, M. Chen, S.-W. Lee, X. Fei, and D. Shen, "Weighted graph regularized sparse brain network construction for MCI identification," *Pattern Recognit.*, vol. 90, pp. 220–231, Jun. 2019.

[16] Y. Li, H. Yang, B. Lei, J. Liu, and C.-Y. Wee, "Novel effective connectivity inference using ultra-group constrained orthogonal forward regression and elastic multilayer perceptron classifier for MCI identification," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1227–1239, May 2019.

[17] L. Xiao et al., "Multi-hypergraph learning-based brain functional connectivity analysis in fMRI data," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1746–1758, May 2020.

[18] B. Lei et al., "Diagnosis of early Alzheimer's disease based on dynamic high order networks," *Brain Imag. Behav.*, vol. 15, no. 1, pp. 276–287, 2021.

[19] M. Yu, O. Sporns, and A. J. Saykin, "The human connectome in Alzheimer disease—Relationship to biomarkers and genetics," *Nature Rev. Neurol.*, vol. 17, no. 9, pp. 545–563, 2021.

[20] L. Zhang et al., "Deep fusion of brain structure-function in mild cognitive impairment," *Med. Image Anal.*, vol. 72, Aug. 2021, Art. no. 102082.

[21] X. Hao et al., "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101625.

[22] S. Zhang et al., "Edge-centric effective connection network based on muti-modal MRI for the diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 552, Oct. 2023, Art. no. 126512.

[23] A. Abrol, Z. Fu, Y. Du, and V. D. Calhoun, "Multimodal data fusion of deep learning and dynamic functional connectivity features to predict Alzheimer's disease progression," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 4409–4413.

[24] S. M. Plis et al., "Reading the (functional) writing on the (structural) wall: Multimodal fusion of brain structure and function via a deep neural network based translation approach reveals novel impairments in schizophrenia," *NeuroImage*, vol. 181, pp. 734–747, Nov. 2018.

[25] O. Dekhil et al., "A personalized autism diagnosis CAD system using a fusion of structural MRI and resting-state functional MRI data," *Frontiers Psychiatry*, vol. 10, p. 392, Jul. 2019.

[26] D. Hirjak et al., "Multimodal magnetic resonance imaging data fusion reveals distinct patterns of abnormal brain structure and function in catatonia," *Schizophrenia Bull.*, vol. 46, no. 1, pp. 202–210, Jan. 2020.

[27] C. J. Honey et al., "Predicting human resting-state functional connectivity from structural connectivity," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 6, pp. 2035–2040, Feb. 2009.

[28] K. Li, L. Guo, D. Zhu, X. Hu, J. Han, and T. Liu, "Individual functional ROI optimization via maximization of group-wise consistency of structural and functional profiles," *Neuroinformatics*, vol. 10, no. 3, pp. 225–242, Jul. 2012.

[29] Q. Zuo, B. Lei, Y. Shen, Y. Liu, Z. Feng, and S. Wang, "Multimodal representations learning and adversarial hypergraph fusion for early Alzheimer's disease prediction," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, Beijing, China, Nov. 2021, pp. 479–490.

[30] S. M. Daselaar, V. Iyengar, S. W. Davis, K. Eklund, S. M. Hayes, and R. E. Cabeza, "Less wiring, more firing: Low-performing older adults compensate for impaired white matter with greater neural activity," *Cerebral Cortex*, vol. 25, no. 4, pp. 983–990, Apr. 2015.

[31] B. Lei et al., "Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101652.

[32] P. Cao et al., "Generalized fused group lasso regularized multi-task feature learning for predicting cognitive outcomes in Alzheimers disease," *Comput. Methods Programs Biomed.*, vol. 162, pp. 19–45, Aug. 2018.

[33] S. Wang, Z. Chen, S. You, B. Wang, Y. Shen, and B. Lei, "Brain stroke lesion segmentation using consistent perception generative adversarial network," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8657–8669, Jun. 2022.

[34] H. Wu, C. Lin, J. Liu, Y. Song, Z. Wen, and J. Qin, "Feature masking on non-overlapping regions for detecting dense cells in blood smear image," *IEEE Trans. Med. Imag.*, vol. 42, no. 6, pp. 1668–1680, Jun. 2023.

[35] H. Wu, X. Huang, X. Guo, Z. Wen, and J. Qin, "Cross-image dependency modeling for breast ultrasound segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 6, pp. 1619–1631, Jun. 2023.

[36] W. Yu, B. Lei, M. K. Ng, A. C. Cheung, Y. Shen, and S. Wang, "Tensorizing GAN with high-order pooling for Alzheimer's disease assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4945–4959, Sep. 2022.

[37] S. Hu, B. Lei, S. Wang, Y. Wang, Z. Feng, and Y. Shen, "Bidirectional mapping generative adversarial networks for brain MR to PET synthesis," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 145–157, Jan. 2022.

[38] W. Yu et al., "Morphological feature visualization of Alzheimer's disease via multidirectional perception GAN," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4401–4415, Aug. 2023.

[39] C. Gong et al., "Generative AI for brain image computing and brain network computing: A review," *Frontiers Neurosci.*, vol. 17, Jun. 2023, Art. no. 1203104.

[40] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[41] D. A. Hudson and L. Zitnick, "Generative adversarial transformers," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 4487–4499.

[42] H. Wu, J. Pan, Z. Li, Z. Wen, and J. Qin, "Automated skin lesion segmentation via an adaptive dual attention module," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 357–370, Jan. 2021.

[43] Q. Zuo, L. Lu, L. Wang, J. Zuo, and T. Ouyang, "Constructing brain functional network by adversarial temporal-spatial aligned transformer for early AD analysis," *Frontiers Neurosci.*, vol. 16, Nov. 2022, Art. no. 1087176.

[44] L. Zhang, L. Wang, and D. Zhu, "Predicting brain structural network using functional connectivity," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102463.

[45] L. Liu, Y.-P. Wang, Y. Wang, P. Zhang, and S. Xiong, "An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders," *Med. Image Anal.*, vol. 81, Oct. 2022, Art. no. 102550.

[46] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[47] J. Kawahara et al., "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.

[48] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS ONE*, vol. 8, no. 7, Jul. 2013, Art. no. e68910.

[49] Y. Li et al., "Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification," *Med. Image Anal.*, vol. 52, pp. 80–96, Feb. 2019.

[50] B. Lei et al., "Auto-weighted centralised multi-task learning via integrating functional and structural connectivity for subjective cognitive decline diagnosis," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102248.

[51] B. Lei et al., "Multi-scale enhanced graph convolutional network for mild cognitive impairment detection," *Pattern Recognit.*, vol. 134, Feb. 2023, Art. no. 109106.

[52] M. Montembeault, I. Rouleau, J.-S. Provost, and S. M. Brambati, "Altered gray matter structural covariance networks in early stages of Alzheimer's disease," *Cerebral Cortex*, vol. 26, no. 6, pp. 2650–2662, Jun. 2016.

[53] Y. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, and T. Ma, "Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning," *IEEE Trans. Med. Imag.*, early access, Jul. 13, 2023, doi: 10.1109/TMI.2023.3294967.

[54] L. Meng and Q. Zhang, "Research on early diagnosis of Alzheimer's disease based on dual fusion cluster graph convolutional network," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105212.

[55] J. Liu, Y. Pan, F.-X. Wu, and J. Wang, "Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification," *Neurocomputing*, vol. 400, pp. 322–332, Aug. 2020.

[56] X. Song et al., "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101947.