



Task-Induced Pyramid and Attention GAN for Multimodal Brain Image Imputation and Classification in Alzheimer's Disease

Xingyu Gao, Feng Shi , Dinggang Shen, and Manhua Liu , and the Alzheimer's Disease Neuroimaging Initiative

Abstract—With the advance of medical imaging technologies, multimodal images such as magnetic resonance images (MRI) and positron emission tomography (PET) can capture subtle structural and functional changes of brain, facilitating the diagnosis of brain diseases such as Alzheimer's disease (AD). In practice, multimodal images may be incomplete since PET is often missing due to high financial costs or availability. Most of the existing methods simply excluded subjects with missing data, which unfortunately reduced the sample size. In addition, how to extract and combine multimodal features is still challenging. To address these problems, we propose a deep learning framework to integrate a task-induced pyramid and attention generative adversarial network (TPA-GAN) with a pathwise transfer dense convolution network (PT-DCN) for imputation and classification of multimodal brain images. First, we propose a TPA-GAN to integrate pyramid convolution and attention module as well as disease classification task into GAN for generating the missing PET data with their MRI. Then, with the imputed multimodal images, we build a dense convolution network with pathwise transfer blocks to gradually learn and combine multimodal features

for final disease classification. Experiments are performed on ADNI-1/2 datasets to evaluate our method, achieving superior performance in image imputation and brain disease diagnosis compared to state-of-the-art methods.

Index Terms—Multimodal brain images, generative adversarial network, dense convolution network, image classification, Alzheimer's disease.

I. INTRODUCTION

WITH the advance of neuroimaging technologies, different imaging modalities such as magnetic resonance imaging (MRI) and positron emission tomography (PET) have been used to capture structural and functional information of human brain. Multimodal images can provide powerful and complementary information for more accurate and earlier diagnosis of brain diseases such as Alzheimer's disease (AD) and mild cognitive impairment (MCI) [1]. AD is a progressive and irreversible neurodegenerative dementia, characterized by memory loss and cognitive impairment [2]. MCI is the prodromal stage of AD and can be divided into stable MCI (sMCI) and progressive MCI (pMCI). Currently, no effective treatment has been found to cure AD and existing medicines can merely alleviate symptoms and slow down its progression [3]. Thus, its early diagnosis is important for intervention. Benefitting from rich information in multimodal images, machine learning methods have been investigated to combine MRI and PET for early diagnosis of AD [4].

There are several categories of methods for machine learning based disease diagnosis. One category of methods investigated machine learning technologies on handcrafted features for disease diagnosis. Zhang *et al.* [5] proposed a multi-kernel support vector machine to integrate multimodal features including tissue volumes of 93 regions of interest (ROIs) in MRI scans, intensity values of PET and CSF biomarkers for disease classification. Instead of using multi-kernel learning, Gray *et al.* [6] proposed a random forest model based on a similarity measurement framework for multimodal brain disease classification. Young *et al.* [7] introduced a Gaussian process classifier for multimodal probabilistic prediction of MCI conversion to AD within 3 years. A manifold regularized multi-task feature learning method was

Manuscript received October 2, 2020; revised March 20, 2021, June 16, 2021, and June 19, 2021; accepted July 7, 2021. Date of publication July 19, 2021; date of current version January 5, 2022. This work was supported by the Natural Science Foundation of Shanghai under Grant 20ZR1426300, in part by the National Natural Science Foundation of China (NSFC) under Grants 61773263 and U19B2035, in part by the Shanghai Jiao Tong University Scientific and Technological Innovation Funds under Grant 2019QYB02, in part by an ECNU-SJTU joint grant from the Basic Research Project of Shanghai Science and Technology Commission under Grant 19JC1410102, and in part by the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0102. (Corresponding author: Manhua Liu.)

Xingyu Gao is with the Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xingyu.gao@sjtu.edu.cn).

Feng Shi is with the Department of Research and Development, Shanghai United Imaging Intelligence Company, Ltd., Shanghai 200232, China (e-mail: feng.shi@united-imaging.com).

Dinggang Shen is with the School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China, and with the Department of Research and Development, Shanghai United Imaging Intelligence Company, Ltd., Shanghai 200232, China, and also with the Department of Artificial Intelligence, Korea University, Seoul 02841, Republic of Korea (e-mail: dinggang.shen@gmail.com).

Manhua Liu is with the MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: mhlui@sjtu.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JBHI.2020.3097721>.

Digital Object Identifier 10.1109/JBHI.2021.3097721

proposed to combine the multimodal features for disease classification, which can preserve both intrinsic relatedness and distribution among multimodal data [8]. Tong *et al.* [9] proposed a nonlinear graph fusion network to combine multimodal features for the classification of AD. Shi *et al.* [10] proposed to use the feature-level coupled interaction of multimodal data for the diagnosis of AD and MCI.

Recently, deep learning networks were investigated to extract latent features from ROIs on different image modalities for AD classification. Liu *et al.* [11] extracted a set of latent features from 83 ROIs on MRI and PET scans, and utilized a zero-masking strategy to combine complementary information of multimodal data for AD diagnosis. Suk *et al.* [12] used a stacked autoencoder to learn latent high-level features from multimodal ROIs, followed by a multi-kernel support vector machine to combine these features for AD classification. Lei *et al.* [13] introduced a discriminative sparse learning method with relational regularization to jointly predict the clinical score and judge disease stages using multimodal data. Liu *et al.* [14] designed a view-aligned hypergraph learning network to explicitly model the coherence among multimodal data.

Instead of building deep learning networks on handcrafted features, some methods build deep networks directly on multimodal brain images to learn features for disease diagnosis. A deep neural network with a restricted Boltzmann machine was proposed to learn the hierarchical feature representation of 3D image patches for multimodal AD/MCI diagnosis [15]. A cascaded convolutional neural network was proposed to jointly learn the multi-level and multimodal features of MRI and PET for AD classification [16].

The abovementioned methods could make full use of complementary information of multimodal data. However, it is still challenging to deal with incomplete multimodal data. In practice, patient movement, insufficient image quality, financial consideration or facility availability could result into incomplete multimodal data. To address this problem, Yuan *et al.* [17] proposed a multi-source feature learning method by jointly analyzing the incomplete multiple heterogeneous neuroimaging data for disease classification. Recently, a latent representation learning method was proposed to utilize inter-modality association for AD diagnosis [18]. To utilize incomplete data, another intuitive strategy is to directly generate the missing data with another available modality data. Li *et al.* [19] proposed a deep learning framework through the estimation of missing PET from MRI for multimodal disease diagnosis. Recently, the generative adversarial networks were also investigated to impute missing modality data by combining feature consistency and voxel-wise consistency for multimodal data imputation [20]–[22].

Although many efforts have been made to generate the missing data, it is still challenging to accurately generate the key information to identify abnormalities of diseases. In this paper, we propose a hybrid deep learning framework that contains a task-induced pyramid and attention generative adversarial network (TPA-GAN) and a pathwise transfer dense convolution network (PT-DCN) for multimodal image imputation and classification, respectively. The TPA-GAN is proposed to integrate

TABLE I
DEMOGRAPHIC AND CLINICAL INFORMATION OF THE STUDIED SUBJECTS

Dataset	Category	Gender (F/M)	Age (Years)	CDR	MMSE
ADNI1	AD	97/99	75.7 \pm 7.7	0.8 \pm 0.3	23.3 \pm 2.0
	pMCI	66/102	74.8 \pm 6.9	0.5 \pm 0.0	26.6 \pm 1.7
	sMCI	77/153	75.0 \pm 7.7	0.5 \pm 0.0	27.3 \pm 1.8
	CN	110/117	75.9 \pm 5.0	0.0 \pm 0.0	29.1 \pm 1.0
ADNI2	AD	66/90	75.7 \pm 7.9	0.9 \pm 0.4	21.8 \pm 3.8
	pMCI	31/35	72.9 \pm 7.3	0.6 \pm 0.2	25.7 \pm 2.2
	sMCI	50/62	72.6 \pm 7.9	0.5 \pm 0.1	27.7 \pm 2.2
	CN	106/94	74.1 \pm 6.3	0.0 \pm 0.1	28.9 \pm 1.3

Abbreviations: AD: Alzheimer's disease; pMCI: progressive mild cognitive impairment; sMCI: stable mild cognitive impairment; CN: cognitive normal; M: male; F: female; CDR: Clinical Dementia Rating; MMSE: Mini-Mental State Examination.

the pyramid convolution and attention module as well as a task-induced network into GAN to generate the missing PET with their corresponding MRI. With the imputed multimodal images, a dense convolution network with pathwise transfer blocks is further built to gradually learn and combine multimodal features for final disease classification. Experiments are performed on ADNI dataset to test the efficacy of the proposed method on AD diagnosis and MCI conversion prognosis.

The rest of this paper is organized as follows. In Section II, we introduce the dataset used in this study and the image pre-processing procedures. In Section III, we present the proposed method in detail. The experimental results are provided in Section IV. In Section V, we discuss the advantages and limitations of the proposed method. A conclusion is given in Section VI.

II. MATERIALS

A. Dataset

In this work, the multimodal brain images are from the public database of Alzheimer's Disease Neuroimaging Initiative (ADNI) [23]. For the ADNI study, thousands of participants were recruited across North America to collect the brain images and clinical assessments. The study was approved by the Institutional Review Board (IRB) of participating institutes. The initial five-year study of ADNI-1 was further extended to ADNI-2 in 2009. In this study, we use the T1-weighted MRI and FDG-PET from ADNI-1 and ADNI-2 datasets to train and evaluate our method. Table I shows the demographic and clinic information of the studied subjects from ADNI. In ADNI-1, MRI data are obtained with 1.5T scanners from the baseline visits of 821 subjects including 196 AD, 168 pMCI, 230 sMCI and 227 CN, while PET data are available in 396 of these subjects. In ADNI-2, MRI data are obtained with 3T scanners from the baseline visits of 156 AD, 66 pMCI, 112 sMCI and 200 CN subjects, while PET data are available in 487 of these subjects. More details about the protocol of data acquisition could refer to the official website.

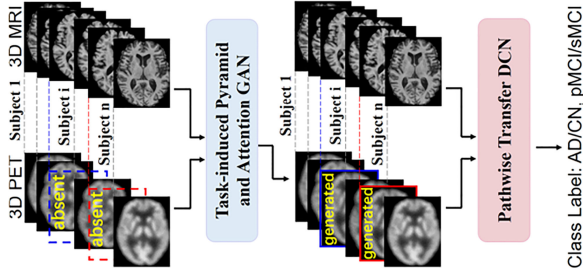


Fig. 1. Overview of our proposed deep learning framework with a task-induced pyramid and attention GAN (TPA-GAN) and a pathwise transfer dense convolution network (PT-DCN) for multimodal brain image imputation and classification, respectively.

B. Image Pre-Processing

All MRI and PET images in this study are pre-processed by FMRIB Software Library (FSL) 6.0.3, which is available at the website <https://fsl.fmrib.ox.ac.uk/>. First, BET algorithm is used for brain extraction [24]. As the result, the skull is stripped from the source image space. Next, the MRI images are aligned to Montreal Neurological Institute T1 standard template space (MNI152_T1_1mm) with the FLIRT linear registration algorithm [25], instead of nonlinear registration [26]–[28], it can save computational time during the application stage. The PET images are aligned to their corresponding T1 images of the specific subjects and brought to MNI space. After linear registration, the MRI and PET images are cropped to the size of $152 \times 188 \times 152$ by removing the voxels of zero values in the periphery of brain. To further reduce the computational complexity, all MRI and PET images are downsampled to the size of $76 \times 94 \times 76$ for the input of our proposed models.

III. METHODS

A. Overview of the Proposed Method

Given the preprocessed MRI and PET brain images, there are two main difficulties for multimodal classification of AD or MCI, i.e., the missing PET data and the multimodal fusion. In this work, we propose a deep learning framework consisting of two deep networks: TPA-GAN for imputation of missing PET images and PT-DCN for multimodal classification, as shown in Fig. 1. The details are presented in the following subsections.

B. Task-Induced Pyramid and Attention Generative Adversarial Network (TPA-GAN) for Image Imputation

To address the problem of modality incompleteness, the generative adversarial network (GAN) is investigated to generate the missing PET with the MRI based on their correlations. GAN consists of generator and discriminator. The generator aims to impute the missing data as close as possible to the real data, while the discriminator aims to learn features for distinguishing the imputed image from its real image. The training process stops when the discriminator could *no longer* recognize that the imputed image is fake.

However, there are some limitations in this method. First, a chain of convolution layers with the same kernel size and pooling layers are used in generator, which may be difficult to capture different levels of details from source images. Second, although the discriminator of general GAN can regularize image distribution and visual quality, it does not preserve the important features related to the task of disease classification. Third, the general GAN does not consider long-range structural correlations, resulting in the blurring of generated images. To address above problems, we propose a TPA-GAN for imputation of missing PET. Fig. 2 shows the network architecture of our TPA-GAN, which consists of three subnetworks: pyramid and attention generator, standard discriminator, and task-induced discriminator.

1) *Pyramid and Attention Generator*: In this work, we build a generator of TPA-GAN based on the U-net structure which contains contracting and expanding parts, as shown in Fig. 2. The contracting part consists of a series of convolution layers followed by a rectified linear unit (ReLU) and a $2 \times 2 \times 2$ max pooling operation with stride as 2 for feature down-sampling. It aims to learn the features from MRI for image imputation. A typical cnn usually uses a chain of convolution layers with small kernel size and pooling layers, to gradually reduce the size of the input and increase the receptive field. Different from that, we apply two pyramid convolution blocks in the contracting part to capture different levels of details from MRI. To process the input image at multiple scales, the first pyramid convolution block contains three different kernel sizes, i.e., $3 \times 3 \times 3$, $5 \times 5 \times 5$ and $7 \times 7 \times 7$. The second pyramid convolution block contains two different kernel sizes, i.e., $3 \times 3 \times 3$ and $5 \times 5 \times 5$. The following convolution kernels are set to $3 \times 3 \times 3$. The pyramid convolution block can not only enlarge the receptive field, but also capture different levels of image details.

The expanding part of generator is to generate high-resolution image details with the lower-resolution feature maps. It contains three up-convolutional blocks, each consisting of an up-sampling layer followed by two $3 \times 3 \times 3$ convolutions with ReLU activation. A self-attention mechanism was introduced into the GAN to help model long-range and multi-level dependencies across image regions [29]. To consider long-range structural correlations, we apply the self-attention module to the last up-convolutional block. In generator, the convolution layers can process the information in local neighborhood while the self-attention module can model long-range correlations, which help generate better image details.

Let $x \in R^{C \times N}$ be the image feature learned from the previous latent layers, where C is the number of channels and N is the number of feature locations. To calculate the attention, according to the network, the feature is transformed into a feature space f as $f(x) = W_f x$. The receptive extent of the model when generating the region can be calculated as:

$$\eta_j = \frac{\exp(f(x_j))}{\sum_{j=1}^N \exp(f(x_j))} \quad (1)$$

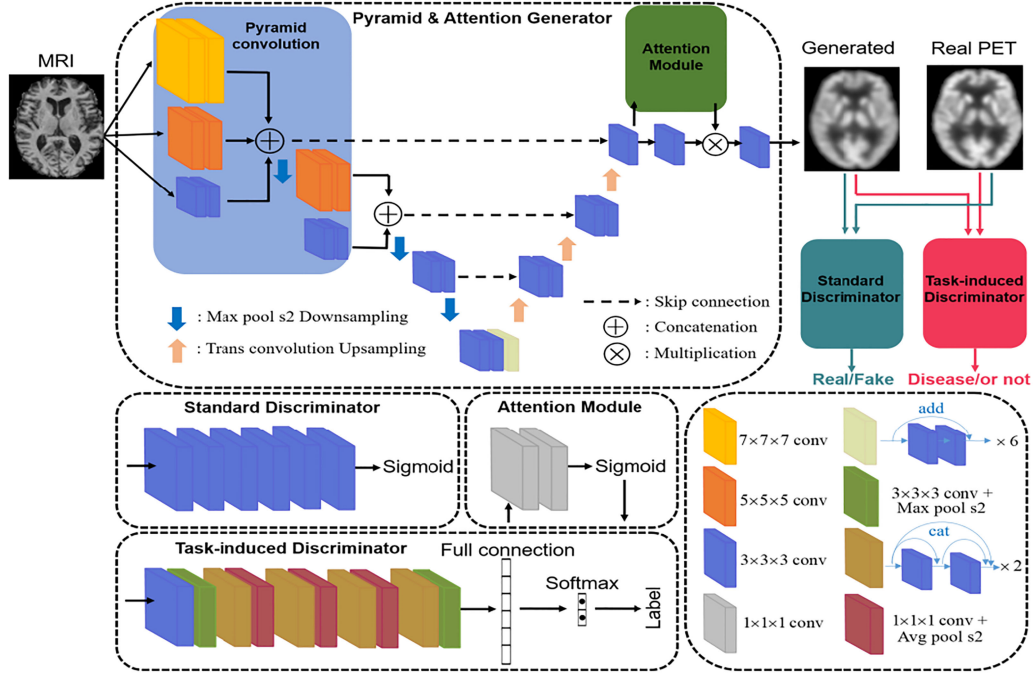


Fig. 2. Network architecture of our proposed a task-induced pyramid and attention GAN (TPA-GAN) for imputation of multimodal brain images.

then, the attention map $a = (a_1, a_2, \dots, a_j, \dots, a_N) \in R^{C \times N}$ is computed as:

$$a_j = v \left(\sum_{j=1}^N \eta_j \cdot \varphi(x_j) \right) \text{ s.t } \varphi(x) = W_\varphi x, v(x) = W_v x \quad (2)$$

where W_f , W_φ and W_v are the matrices of weights leaned from the task of image imputation.

Assume χ_M^i and χ_P^i be MRI and PET for the i^{th} subject, respectively. Since the missing data are always PET images, the generator G is single-oriented and the loss function is formulated as:

$$L_G = E_{x \sim P_{MRI}(x)} [\log(1 - D_{std}(G(\chi_M^i)))] \quad (3)$$

where D_{std} denotes the standard discriminator.

2) Standard Discriminator: The standard discriminator plays an important role in discriminating that the generated data is real or not. It consists of 6 convolution layers with the increasing channels of 32, 64, 128, 256, 512 [20], [21]. The $3 \times 3 \times 3$ convolution with stride as 2 is used for pooling. Lastly, the loss function of standard discriminator is defined as:

$$L_{D_{std}} = E_{x \sim P_{PET}(x)} [\log(1 - D_{std}(\chi_P^i))] + E_{x \sim P_{MRI}(x)} [\log(D_{std}(G(\chi_M^i)))] \quad (4)$$

the loss function optimizes that the output of D_{std} with χ_P^i is real, while the output of D_{std} with $G(\chi_M^i)$ is fake.

3) Task-Induced Discriminator: In recent years, the objective of various GANs is to generate visually graceful images. In medical imaging analysis, the images are from normal subjects and abnormal subjects with diseases. Thus, the medical image generation needs to consider the abnormal changes, which should be included in the generated results if applicable. To

achieve this, we design a task-induced discriminator for jointly learning the image generation and disease classification, so that the features of disease classification can be used to improve the generation of images with abnormal changes. Fig. 2 illustrates the network architecture of the task-induced discriminator which is a classifier for disease diagnosis. To capture more image details, dense blocks are used as the backbone of task-induced discriminator. The network contains one convolutional layer with a max-pooling layer and 4 dense blocks with each one followed by a transition layer, which uses pooling to compress the receptive field of features and the number of channels. Finally, one fully-connected layer with softmax activation is applied to generate the predicted label. The cross-entropy loss is used for task-induced discriminator, which is computed as:

$$L_{D_{task}} = E_{x \sim P_{MRI}(x)} [-y_i * \log(D_{task}(G(\chi_M^i))) - (1 - y_i) * \log(1 - D_{task}(G(\chi_M^i)))] \quad (5)$$

where D_{task} denotes the task-induced discriminator, it outputs the probability of AD. And y_i denotes the label of i^{th} subject.

4) Combined Loss Function: The optimization function of the whole network is from the GAN losses, i.e., $L_{GAN} = L_G + L_{D_{std}}$ and the task-induced discriminator loss $L_{D_{task}}$. In addition, to improve the quality of generated images, we introduce two pixel-wise losses: the L1 loss and the structural similarity index matrix (SSIM) loss. The L1 loss is computed as:

$$L_1 = \|\hat{\chi}_P - \chi_P\|_1 \quad (6)$$

where $\hat{\chi}_P = G(\chi_M)$ and $\|\cdot\|_1$ is the L_1 norm operator. The SSIM loss is one of the most perceptual indicators for assessing

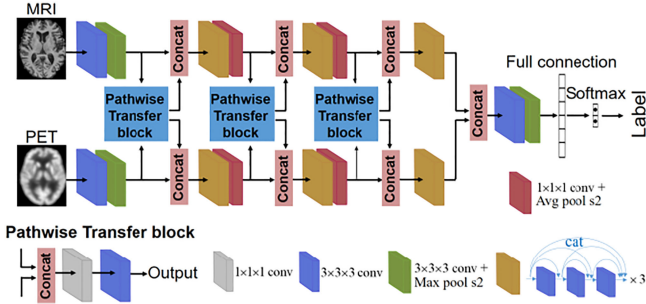


Fig. 3. Architecture of our proposed pathwise transfer dense convolution network (PT-DCN) for multimodal classification of AD/MCI.

the similarity between two images [30], as defined as:

$$L_{SSIM} = \frac{1}{|M|} \sum_{P \in P(\hat{\chi}_P, \chi_P)} SSIM(\hat{\chi}_P, \chi_P) \quad (7)$$

$$SSIM(\hat{\chi}_P, \chi_P) = \frac{(2\mu_{\hat{\chi}_P}\mu_{\chi_P} + C_1)(2\sigma_{\hat{\chi}_P\chi_P} + C_2)}{(\mu_{\hat{\chi}_P}^2 + \mu_{\chi_P}^2 + C_1)(\sigma_{\hat{\chi}_P}^2 + \sigma_{\chi_P}^2 + C_2)} \quad (8)$$

where M is the number of local windows of the image, $\mu_{\hat{\chi}_P}$, $\sigma_{\hat{\chi}_P}$ and μ_{χ_P} , σ_{χ_P} are the mean value and the standard deviation of the generated image $\hat{\chi}_P$ and its corresponding ground truth χ_P , respectively. $C_1 = (k_1 L_1)^2$ and $C_2 = (k_2 L_2)^2$ are used to avoid the zero denominator, where L_1, L_2 are the ranges of voxel intensities and k_1, k_2 are constants far less than 1. And,

$$\sigma_{\hat{\chi}_P\chi_P} = \frac{1}{N-1} \sum_{i=1}^N (\hat{\chi}_P^i - \mu_{\hat{\chi}_P})(\chi_P^i - \mu_{\chi_P}) \quad (9)$$

In summary, the final optimization loss of our proposed TPA-GAN combines the pixel-wise L_1 loss, SSIM loss, GAN loss, and task-induced discriminator loss with weights as:

$$L = \gamma(L_1 + L_{SSIM}) + \lambda L_{GAN} + \varsigma L_{Dtask} \quad (10)$$

where γ, λ and ς are weights for three individual losses, which are optimized to minimize the combined loss function.

C. Pathwise Transfer Deep Convolution Network (PT-DCN) for Multimodal Disease Classification

After the missing PET is imputed by TPA-GAN, another problem is to learn and combine the features of MRI and PET for disease classification. We propose a pathwise transfer deep convolution network to gradually learn and combine the multi-level and multimodal features of MRI and PET for brain disease classification, as shown in Fig. 3. The network contains three pathwise transfer blocks and two paths of dense convolutional networks with each one learning features from single-modality data.

In each path, we use one convolution layer followed by a pooling layer, and three dense blocks followed by transition layers. The networks of two paths have the same structure for parallel single-modality input. In each dense block, every convolution layer receives all feature maps from previous convolution layers through shortcut connection. The size of dense block is

set to 3 in each level, which is different from the task-induced discriminator. Between two dense blocks, a transition layer with one $1 \times 1 \times 1$ convolution layer and one average pooling layer is set to reduce the dimension and size of feature maps.

The pathwise transfer blocks are proposed to make full use of complementary information of different modality images. In this work, pathwise transfer block consists of one concatenation layer, one $1 \times 1 \times 1$ convolution layer, one batch normalization layer with ReLU activation, and one $3 \times 3 \times 3$ convolution layer. The pathwise transfer blocks are used to communicate information across these two paths of PET and MRI, which can improve the performance of classification model.

IV. EXPERIMENTS

In this section, experiments are performed to test the proposed deep learning framework on the imputation of missing data and the multimodal classification of disease status.

A. Implementation Details

Our proposed deep learning method was implemented with Pytorch framework. All experiments were conducted on Ubuntu16.04 and GPU of NVIDIA GeForce GTX1080 Ti. In the experiments, we use ADNI-1 dataset to train the networks and use ADNI-2 dataset to test the performance of our proposed networks. The experiments are conducted for both image imputation and classification. To better show the results, the subjects used for image imputation and classification are slightly different. In the experiments of image imputation by TPA-GAN, only the paired real PET and MRI data are used for training and testing. In the experiments of disease classification by PT-DCN, apart from all real PET and MRI data, the missing PET data are imputed for training and testing. To evaluate our method, additional experiments with ADNI-2 for training and ADNI-1 for testing are provided in *Supplementary Materials*.

In the training process of TPA-GAN, we first train the pyramid and attention GAN and the task-induced discriminator individually. Then, these two pre-trained models are combined to fine-tune the whole TPA-GAN to impute the missing PET data. In the fine-tuning process, the weights of task-induced discriminator are fixed, while its loss is used to update the parameters of generator, which can help the generator reconstruct the pathological changes. Adam optimizer is applied in the training process with the batch size as 2 in both networks. For pyramid and attention GAN, the learning rates of generator and standard discriminator are set to 0.0001 and 0.0004, respectively. For task-induced discriminator, the learning rate is set to 0.0001.

After imputing the missing PET images by our TPA-GAN, we evaluate the PT-DCN on the classification tasks of AD vs CN and pMCI vs sMCI with the complete PET and MRI. Adam optimizer is used in the training process with the batch size set to 2 and learning rate set to 0.0001.

B. Effectiveness of Image Imputation

To test the effectiveness of image imputation, we perform a series of ablation studies of TPA-GAN including the ‘‘Basic

TABLE II

QUANTITATIVE RESULTS OF ABLATION STUDIES FOR GENERATED PET IMAGES ON ADNI-2 DATASET. PSNR, SSIM, MSE AND MMD ARE USED TO EVALUATE THE GENERATION QUALITY OF IMPUTED PET IMAGES (CONSIST OF AD/CN AND PMCI/SMCI), AND AUC IS USED TO EVALUATE THE CLASSIFICATION RESULTS OF IMPUTED PET IMAGES BY TASK-INDUCED DISCRIMINATOR FOR AD VS. CN (AUC*) AND PMCI VS. SMCI (AUC+)

Method	SSIM	PSNR	MSE	MMD	AUC*(%)	AUC+(%)
Cycle GAN [20]	0.880 ± 0.04	27.2 ± 2.67	388 ± 475	0.249 ± 0.25	68.0	62.7
Basic GAN	0.868 ± 0.04	26.1 ± 2.39	392 ± 444	0.224 ± 0.24	61.4	55.6
Pyramid GAN	0.899 ± 0.04	27.3 ± 2.65	360 ± 440	0.221 ± 0.24	71.8	59.7
Attention GAN	0.900 ± 0.03	27.6 ± 1.47	209 ± 348	0.117 ± 0.15	76.1	61.4
PA GAN	0.913 ± 0.03	28.5 ± 1.74	204 ± 347	0.129 ± 0.14	81.0	64.0
TPA-GAN	0.915 ± 0.04	29.0 ± 2.99	184 ± 346	0.107 ± 0.14	88.3	72.4

GAN” and its variants with pyramid convolution block, self-attention and their combination, denoted as “Pyramid GAN”, “Attention GAN” and “PA GAN”, respectively. Specifically, the Basic GAN is implemented with the adversarial loss and U-net structure for the generator. Pyramid GAN is implemented by adding the pyramid convolution layers to the Basic GAN. The Attention GAN is implemented by adding self-attention module to the last block of generator in the Basic GAN. The PA GAN is implemented with the pyramid convolution layers and attention module on the generator of Basic GAN.

We compare the effectiveness of image imputation by different methods using five measures: structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), mean squared error (MSE), maximum mean discrepancy (MMD), and area under the receiver operating characteristic (AUC). The average results on all testing subjects (containing AD/CN and pMCI/sMCI) from ADNI-2 dataset are provided in Table II. For SSIM and PSNR measures, higher values indicate better image quality. For MSE and MMD measures, lower values indicate better image quality. With generated PET images, AUC is used to evaluate the classification ability of AD vs CN (AUC*) and pMCI vs sMCI (AUC+) by the task-induced discriminator. The task-induced discriminator has been pre-trained on real PET from ADNI-1 dataset and tested on the generated PET from ADNI-2 dataset. We also compare our method with the cycle-consistent GAN (Cycle-GAN) [20] in Table II. For fair comparison, we implement the cycle-consistent GAN method with our best efforts on the same training and testing dataset as ours. From the results of SSIM, PSNR, MSE and MMD, we can see that the qualities of the generated PET images are gradually improved by adding the pyramid convolution, attention module and the task-induced discriminator to the Basic GAN. The classification results of AUC (AUC* and AUC+) are shown in Table II. Benefit that the abnormal changes of PET images caused by diseases can be better generated by introducing the task-induced discriminator, our TPA-GAN achieves better performance than Cycle-GAN. Moreover, we visually compare the generated PET images by our TPA-GAN to other ablation studies for two typical subjects from test dataset, which are shown in *Supplementary Materials*.

C. Effectiveness of Disease Classification

In this section, we perform the experiments to test the proposed PT-DCN method on the multimodal classifications of AD vs NC and pMCI vs sMCI. To evaluate proposed method, we compute five performance measures: accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1S) and AUC.

First, we compare the multimodal classification with single-modal classification as shown in Table III. In the experiments, the classification on MRI or PET images is implemented with one path deep dense convolutional network (DCN). We also use simple concatenation to combine the MRI and PET features for comparison. For fair comparison, we use the same subjects of ADNI 2 with the paired real PET and MRI data for testing both the single- and multimodal classifications. From Table III, we can see that PT-DCN performs better than single modality and simple concatenation. The pathwise transfer blocks can make full use of complementary information at multiple levels of MRI and PET images. The additional training data of generated PET can further improve the multimodal classification performances, which demonstrates the effectiveness of our TPA-GAN (w/ TPA-GAN).

Second, we have compared our method with other multimodal methods that are also based on the baseline visits of PET and MRI from ADNI database. Similar to other methods, our method is trained on ADNI-1 dataset and tested on ADNI-2 dataset in the experiment. The missing PET images from ADNI-1 and ADNI-2 dataset are imputed by our TPA-GAN. Then, all subjects with complete multimodal data are used to train and test our PT-DCN. The results are shown in Table IV. Specifically, the compared methods include the hierarchical ensemble classification method [31], the group-constrained sparse learning method [32], the missing data estimation method [19], the nonlinear graph fusion network [9], the landmark-based deep feature learning model [33] and the latent representation learning method [18]. It is not easy to implement these methods, because their source codes and pre-trained models are not released. For comparison, the results reported in their papers are directly used in Table IV. From the results, we can see that our proposed method performs better than other methods in the tasks of AD diagnosis and MCI prognosis.

TABLE III

COMPARISON OF CLASSIFICATION RESULTS BY USING THE SINGLE- AND MULTIMODAL FOR AD VS. CN AND PMCI VS. SMCI, WITH THEIR MODELS TRAINED ON ADNI-1 AND TESTED ON ADNI-2. THE TEST SET CONTAINS ONLY THE PAIRED REAL PET AND MRI DATA, WHILE THE TRAINING SET INCLUDES THE PAIRED REAL PET AND MRI DATA ("W/O TPA-GAN") OR REAL PET AND MRI DATA AS WELL AS THE IMPUTED PET BY TPA-GAN ("W/ TPA-GAN")

Modality	Method	AD vs. CN (%)					pMCI vs. sMCI (%)				
		ACC	SEN	SPE	AUC	F1S	ACC	SEN	SPE	AUC	F1S
MRI	Single DCN	85.4	80.1	89.5	90.2	83.3	63.9	63.1	64.5	66.6	59.0
PET	Single DCN	88.8	86.1	90.8	94.3	87.0	67.7	69.2	66.7	73.7	63.8
MRI+PET	Simple Concatenation	89.1	86.1	91.4	94.8	87.3	64.6	63.1	65.6	71.3	59.4
	Our PT DCN (w/o TPA-GAN)	90.3	88.9	91.4	95.5	88.9	74.7	72.3	76.3	77.5	70.1
	Our PT DCN (w/ TPA-GAN)	92.7	91.7	93.5	96.4	91.7	75.3	70.8	78.4	77.8	70.2

TABLE IV

COMPARISON OF THE MULTIMODAL CLASSIFICATION PERFORMANCES REPORTED IN THE LITERATURE. THE MODELS ARE TRAINED ON ADNI-1 DATASET AND TESTED ON ADNI-2 DATASET

Method	Subjects				AD vs. CN (%)					pMCI vs. sMCI (%)				
	AD	pMCI	sMCI	CN	ACC	SEN	SPE	AUC	F1S	ACC	SEN	SPE	AUC	F1S
Liu et. al., 2014 [31]	198	-	-	229	82.2	77.4	86.1	88.1	79.4	74.1	67.4	75.0	73.8	38.4
Wee et. al., 2014 [32]	370	149	562	440	87.4	84.2	90.0	92.0	85.5	71.0	74.4	70.6	75.0	38.1
Li et al., 2014 [19]	93	-	-	101	-	-	-	89.8	-	-	-	-	-	-
Tong et. al., 2017 [9]	37	-	-	35	88.6	-	-	94.8	-	-	-	-	-	-
Liu et. al., 2018 [33]	370	149	562	440	90.9	87.9	93.3	96.1	89.5	73.5	74.4	73.4	80.9	40.3
Zhou et. al., 2019 [18]	171	157	205	204	-	-	-	-	-	74.3	-	-	75.5	-
Ours	352	234	342	427	92.0	89.1	94.0	95.6	90.5	75.3	77.3	74.1	78.6	69.9

V. DISCUSSION

For multimodal classification of incomplete brain imaging data, we first construct a TPA-GAN network for imputation of the missing modality. The pyramid convolution layers and attention module as well as the task-induced discriminator are integrated into TPA-GAN to generate the missing PET with its corresponding MRI. Specifically, the pyramid convolution layers can capture multi-level features of MRI, while the attention module can effectively eliminate the redundant information and accelerate the convergence of network. The task-induced discriminator can help generate the disease-specific image details. Results and comparisons in Table II have demonstrated improvements of these contributions. TPA-GAN can *not only* generate the image details of PET, *but also* preserve the important information to the task of disease classification.

With the complete multimodal brain images, PT-DCN is built with pathwise transfer blocks to gradually learn and combine multimodal features for final disease classification. The pathwise transfer blocks can make full use of complementary information at multiple levels between MRI and PET images. Through the data augmentation of TPA-GAN, our proposed method can further improve the performance of multimodal AD/MCI diagnosis. Compared to the existing multimodal classification methods, our proposed method has achieved promising performance.

However, there are still some limitations in the proposed method which can be improved in the future. First, in addition to MRI and PET data, other brain imaging data such as DTI

can be included to further improve the performance of disease diagnosis. Second, the models are trained on ADNI-1 and tested on ADNI-2 independently, without considering the differences between these databases. In fact, most of MRI in ADNI-1 are captured by 1.5T scanners while the MRI in ADNI-2 are captured by 3T scanners, which may affect the performance of our proposed method. In the future, data adaptation or transferring techniques can be investigated to alleviate the problem.

VI. CONCLUSION

In this paper, we have proposed two deep learning frameworks: 1) task-induced pyramid and attention GAN (TPA-GAN), and 2) pathwise transfer dense convolutional network (PT-DCN). The TPA-GAN combines the pyramid convolution and attention module with a task-induced discriminator to improve the image imputation. With complete PET and MRI, PT-DCN gradually combines the multi-level and complementary multimodal features for disease classification. Experimental results on ADNI database have demonstrated the effectiveness of our proposed method.

REFERENCES

- [1] I. Campbell-Taylor, B. James, S. Leurgans, and D. Bennett, "Contribution of Alzheimer disease to mortality in the united states," *Neurology*, vol. 83, no. 14, pp. 1302–1302, 2014.

- [2] L. Zhan *et al.*, "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease," *Front. Aging Neurosci.*, vol. 7, 2015, Art. no. 48.
- [3] M. Liu *et al.*, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, 2020, Art. no. 116459.
- [4] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, and S. C. Johnson, "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset," *NeuroImage*, vol. 48, no. 1, pp. 138–149, 2009.
- [5] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [6] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, 2013.
- [7] J. Young, M. Modat, M. J. Cardoso, A. Mendelson, D. Cash, and S. Ourselin, "Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment," *NeuroImage Clin.*, vol. 2, pp. 735–745, 2013.
- [8] B. Jie *et al.*, "Manifold regularized multitask feature learning for multimodality disease classification: Manifold regularized multitask feature learning," *Hum. Brain Mapp.*, vol. 36, no. 2, pp. 489–507, 2015.
- [9] T. Tong, K. Gray, Q. Gao, L. Chen, and D. Rueckert, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognit.*, vol. 63, pp. 171–181, 2017.
- [10] Y. Shi, H.-I. Suk, Y. Gao, S.-W. Lee, and D. Shen, "Leveraging coupled interaction for multimodal Alzheimer's disease diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 1, pp. 186–200, Jan. 2020.
- [11] S. Liu *et al.*, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015.
- [12] Alzheimer's Dis. Neuroimaging Initiative, H.-I. Suk, S.-W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Funct.*, vol. 220, no. 2, pp. 841–859, 2015.
- [13] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational-Regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1102–1113, Apr. 2017.
- [14] M. Liu, J. Zhang, P.-T. Yap, and D. Shen, "View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data," *Med. Image Anal.*, vol. 36, pp. 123–134, 2017.
- [15] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, 2014.
- [16] Alzheimer's Dis. Neuroimaging Initiative, M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, no. 3/4, pp. 295–308, 2018.
- [17] L. Yuan, Y. Wang, P. M. Thompson, V. A. Narayan, and J. Ye, "Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data," *NeuroImage*, vol. 61, no. 3, pp. 622–632, 2012.
- [18] T. Zhou, M. Liu, K.-H. Thung, and D. Shen, "Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2411–2422, Oct. 2019.
- [19] R. Li *et al.*, "Deep learning based imaging data completion for improved brain disease diagnosis," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2014, vol. 8675, pp. 305–312.
- [20] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, and D. Shen, "Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2018, vol. 11072, pp. 455–463.
- [21] Y. Pan, M. Liu, C. Lian, Y. Xia, and D. Shen, "Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2019, vol. 11766, pp. 137–145.
- [22] Y. Pan, M. Liu, C. Lian, Y. Xia, and D. Shen, "Spatially-constrained fisher representation for brain disease identification with incomplete multi-modal neuroimages," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2965–2975, Sep. 2020.
- [23] C. R. Jack *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.
- [24] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp.*, vol. 17, no. 3, pp. 143–155, 2002.
- [25] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, 2002.
- [26] H. Jia, P. T. Yap, and D. Shen, "Iterative multi-atlas-based multi-image segmentation with tree-based registration," *NeuroImage*, vol. 59, no. 1, pp. 422–430, 2012.
- [27] H. Jia, G. Wu, Q. Wang, and D. Shen, "ABSORB: Atlas building by self-organized registration and bundling," *NeuroImage*, vol. 51, no. 3, pp. 1057–1070, 2010.
- [28] J. Fan, X. Cao, P. T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," *Med. Image Anal.*, vol. 54, pp. 193–206, 2019.
- [29] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.
- [30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [31] M. Liu *et al.*, "Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis: AD diagnosis," *Hum. Brain Mapp.*, vol. 35, no. 4, pp. 1305–1319, 2014.
- [32] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, "Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification," *Brain Struct. Funct.*, vol. 219, no. 2, pp. 641–656, 2014.
- [33] M. Liu, J. Zhang, D. Nie, P.-T. Yap, and D. Shen, "Anatomical landmark based deep feature representation for MR images in brain disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1476–1485, 2018.