Disease-image-specific Learning for Diagnosis-oriented Neuroimage Synthesis with Incomplete Multi-Modality Data

Yongsheng Pan, Mingxia Liu, Yong Xia, and Dinggang Shen, *Fellow, IEEE*

Abstract—Incomplete data problem is commonly existing in disease diagnosis with multi-modality neuroimages, to track which, some methods have been proposed to utilize all available subjects by imputing missing neuroimages. However, these methods usually treat image synthesis and disease diagnosis as two standalone tasks, thus ignoring the specificity conveyed in different modalities, *i.e.*, different modalities may highlight different disease-relevant regions in the brain. To this end, we propose a disease-image-specific deep learning (DSDL) framework for joint neuroimage synthesis and disease diagnosis using incomplete multi-modality neuroimages. Specifically, with each whole-brain scan as input, we first design a Disease-image-Specific Network (DSNet) with a spatial cosine module to implicitly model the disease-image specificity. We then develop a Feature-consistency Generative Adversarial Network (FGAN) to impute missing neuroimages, where feature maps (generated by DSNet) of a synthetic image and its respective real image are encouraged to be consistent while preserving the disease-image-specific information. Since our FGAN is correlated with DSNet, missing neuroimages can be synthesized in a diagnosis-oriented manner. Experimental results on three datasets suggest that our method can *not only* generate reasonable neuroimages, *but also* achieve state-of-the-art performance in both tasks of Alzheimer's disease identification and mild cognitive impairment conversion prediction.

Index Terms—Multi-modality neuroimaging, generative adversarial network, missing image synthesis, brain disease diagnosis.

1 Introduction

Multi-modality neuroimaging data, such as structural magnetic resonance imaging (MRI) and fluorodeoxyglucose positron emission tomography (PET), have been shown that they can provide complementary information to improve the computer-aided diagnosis performance of Alzheimer's disease (AD) and mild cognitive impairment (MCI) [1], [2], [3], [4]. In practice, the missing data problem has been remaining a common challenge in automated brain disease diagnosis using multi-modality neuroimaging data, since subjects may lack a specific modality due to patient dropout or poor data quality. For example, more than 800 subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) database [5] have baseline MRI scans, but only $\sim$ 400 subjects have baseline PET data.

Conventional methods typically discard those modality-incomplete subjects and use only modality-complete subjects to train diagnosis models [1], [6], [7], [8], [9], [10], [11]. Such strategy significantly reduces the number of training samples and also ignores the useful information provided by data-missing subjects, thus degrading the diagnostic performance. Several data imputation methods [12], [13] have been proposed to estimate the hand-crafted features of missing data subjects using the features of data-complete subjects. Thus, multi-view learning methods [2], [3], [14] can be developed to make use of all subjects. However, these methods rely on hand-crafted imaging features, which may not be discriminated for brain disease diagnosis, thus leading to sub-optimal learning performance.

A more promising alternative is to directly estimate missing data through deep learning [15], [16]. In our previous work, we directly impute missing PET images based on their corresponding MRI scans by the cycle-consistency generative adversarial network (CycGAN) [4]. This model, however, equally treats all voxels in each brain volume, thus ignoring the *disease-image specificity* conveyed in multi-modality neuroimaging data. Herein, such disease-image specificity is two-fold: (1) *not all regions in an MRI/PET scan are relevant with a specific brain disease* [17]; and (2) *disease-relevant brain regions may differ in different modalities* (e.g., MRI and PET) [7], [18]. For the first aspect, existing deep learning methods usually treat all brain regions equally in the image synthesis process, ignoring that several regions (e.g., hippocampus and amygdala) are highly relevant with AD/MCI [17], [19], [20] in comparison to other regions. For the second aspect, existing methods directly synthesize images of one modality (e.g., PET) based on images of another modality (e.g., MRI), without considering the modality gap in terms of disease-relevant regions [7], [11],

- *Y. Pan and Y. Xia were partially supported by the National Natural Science Foundation of China under Grant 61771397, the Science and Technology Innovation Committee of Shenzhen Municipality, China, under Grant JCYJ20180306171334997, and the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grant CX201835. M. Liu and D. Shen were partially supported by NIH grant (No. AG041721). Corresponding authors: Yong Xia, Mingxia Liu, and Dinggang Shen.*

- *Y. Pan and Y. Xia are with School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China. (E-mail: {yspan@mail.,yxia@}nwpu.edu.cn) M. Liu and D. Shen are with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. D. Shen is also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea. (E-mails: mxliu@med.unc.edu, idea.uncch@gmail.com)*

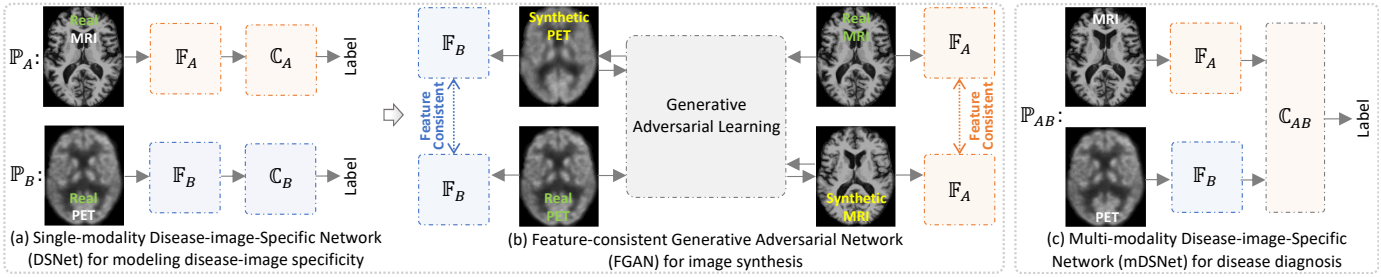


Fig. 1. Illustration of our disease-image-specific deep learning (DSDL) framework. Two major components are included: (a) two single-modality Disease-image-Specific Networks (DSNet) for classification and learning disease-image specificity for MRI (*i.e.*, $\mathbb{P}_A = \mathbb{F}_A + \mathbb{C}_A$) and PET (*i.e.*, $\mathbb{P}_B = \mathbb{F}_B + \mathbb{C}_B$), respectively, and (b) a Feature-consistency Generative Adversarial Network (FGAN) for missing image synthesis, encouraging feature maps (e.g., generated by $\mathbb{F}_A$) of a synthetic image and its real image to be consistent. Note that $\mathbb{F}_A$ and $\mathbb{F}_B$ in (b) are initialized by those learned in (a) and kept frozen in FGAN. Based on complete (after imputation via FGAN) paired MRI and PET scans, we further develop a multi-modality DSNet (*i.e.*, $\mathbb{P}_{AB} = [\mathbb{F}_A, \mathbb{F}_B] + \mathbb{C}_{AB}$) for brain disease identification by concatenating feature maps of MRI and PET (c). In (a) and (c), the backbone feature extractors (e.g., $\mathbb{F}_A$ and $\mathbb{F}_B$) are followed by a spatial cosine module (e.g., $\mathbb{C}_A$, $\mathbb{C}_B$, and $\mathbb{C}_{AB}$) for classification.

[18]. It is worth noting that previous studies have shown that disease diagnosis models can implicitly or explicitly capture the disease-image specificity through regions-of-interest (ROIs) and anatomical landmarks [4], [7], [8], [17], [19], [21]. Therefore, to capture and utilize the disease-image specificity, it is intuitively desirable to integrate disease diagnosis and image synthesis into a unified framework, by imputing missing neuroimages in a diagnosis-oriented manner.

In this paper, we propose a *disease-image-specific deep learning (DSDL) framework* for joint disease diagnosis and image synthesis using incomplete multi-modality neuroimages (see Fig. 1). As shown in Fig. 1 (a)-(b), our method mainly contains two single-modality Disease-image-Specific Network (DSNet) for MRI- and PET-based disease diagnosis and a Feature-consistency Generative Adversarial Network (FGAN) for image synthesis. Herein, DSNet encodes disease-image specificity in MRI- and PET-based feature maps to assist the training of FGAN, while FGAN imputes missing images to improve the diagnostic performance. Since DSNet and FGAN can be trained jointly, missing neuroimages can be synthesized in a diagnosis-oriented manner. Using complete MRI and PET scans (after imputation), we can perform disease diagnosis via the proposed multi-modality DSNet (shown in Fig. 1 (c)). Experimental results on subjects from three public datasets suggest that our method can *not only* synthesize reasonable MR and PET images, *but also* achieve the state-of-the-art results in both AD identification and MCI conversion prediction.

Comparing to our previous works, the contributions of this work are as follows. More detailed information could be seen in Section I of the *Supplementary Materials*.

(1) We proposed a unified framework called DSDL for joint image synthesis and AD diagnosis using incomplete multi-modality neuroimages. The missing images are imputed in a diagnosis-oriented manner, and hence the synthetic neuroimages are more consistent with real neuroimages from a diagnostic point of view.

(2) We designed a spatial cosine module to model the disease-image specificity in whole-brain MRI/PET scans implicitly and automatically.

(3) We proposed a feature-consistency constraint, which can assist the image synthesis model to preserve the disease-relevant information during modality transformation.

2 Related Work

2.1 Synthesis of Missing Neuroimages

By providing complementary information, multi-modality neuroimaging data have shown to be effective in achieving holistic understanding of the brain and improving automated identification performance of brain disorders [22]. Since subjects may lack a specific imaging modality the missing data problem has been remaining a common challenge in multi-modality-based diagnosis systems.

Rather than estimating the hand-crafted features of missing images, many machine learning techniques have been developed to impute the missing images directly. For problems with multi-modality imaging data, a popular solution is to perform cross-modality estimation, *i.e.*, synthesizing an image of a specific modality based on the corresponding image of another modality. For instance, Huynh *et al.* [23] proposed to estimate patches of each computerized tomography (CT) image from its corresponding MR image patches using the structured random forest and auto-context model. Jog *et al.* [24] attempted to transform T_1-weighted MR patches to T_2-weighted MR patches by training a bagged ensemble of regression trees. Bano *et al.* [25] designed a cross-modality convolutional neural network (CNN) with a multi-branch architecture operated on various spatial resolution levels for inference between T_1-weighted and T_2-weighted MRI scans. Li *et al.* [22] proposed a 3-layer CNN for estimating PET images from MRI scans for data completion.

Generative adversarial networks (GANs) have been developed for image transformation from a source domain/modality to a target domain/modality [26], [27], [28], [29], [30], [31], [32], [33]. A typical GAN [26] consists of two neural networks: (a) a generator trained to synthesize an output that approximates the real data distribution, and (2) a discriminator trained to differentiate between the synthetic and real images. Recently, variants of GAN have been used in synthesizing medical images [4], [34], [35], [36], [37]. Ben *et al.* [35] combined a fully CNN with a conditional GAN to predict PET from CT images. Yi *et al.* [36] used GAN to generate missing magnetic resonance angiography images from T_1- and T_2-weighted MR images. Sun *et al.* [11] studied the latent variable representations of different modalities and proposed a flow-based generative model for MRI-to-PET image generation. Pan *et al.* [4] employed Cyc-

GAN to predict PET images from MRI scans and achieved good results in brain disease diagnosis using both real and synthetic multi-modality images. Yan *et al.* [37] added a structure-consistency loss to the original CycGAN and applied it to estimate MRI data from CT data. In general, these GAN-based image synthesis techniques only pose constraints on data distribution, without considering the discriminative capability of synthetic images in a particular task (e.g., neuroimaging-based brain disease diagnosis). Hence, it is desirable to synthesize missing images via GAN in a task-oriented manner.

2.2 Identification of Disease-relevant Brain Regions

Multi-modality neuroimages have been widely used in automated diagnosis of brain disorders, such as AD and MCI. Previous studies have verified that there exits disease-image specificity. For example, AD and MCI are relevant with brain atrophy, especially in specific regions such as hippocampus and amygdala [17], [38], [39]. Zhang *et al.* [7] combined the volumetric features extracted from 93 regions of interest (ROIs) in both MRI and PET scans and reported that the most discriminative MRI- and PET-based features are from different brain regions. Zhang *et al.* [8] further studied the volumetric features of PET and MRI and found that the most discriminative brain regions differ in different tasks. Wachinger *et al.* [19] studied shape asymmetries of neuroanatomical structures across brain regions and found that the subcortical structures in AD is not symmetric, e.g., shape asymmetry in hippocampus, amygdala, caudate and cortex is predictive of disease onset. Cui *et al.* [39] focused on hippocampus regions and used a 3D densely connected CNN to combine global shape and local visual features of hippocampus to enhance the performance of AD classification. However, these ROI-based methods ignore relative changes in multiple regions, while relying solely on rigid partition of ROI may ignore small or subtle changes caused by diseases.

To tackle this limitation, patch-based methods have been proposed to capture diseased-relevant pathology in a more flexible manner, without using pre-defined ROIs. Liu *et al.* [40], [41] partitioned each volume into multiple 3D patches and hierarchically combined patch-based features for AD/MCI identification. Suk *et al.* [9] developed a deep Boltzmann machine to find latent hierarchical feature representation from paired 3D patches of MRI and PET. Based on hand-crafted morphological features, Zhang *et al.* [42], [43] first defined multiple disease-relevant anatomical landmarks via group-wise comparison, and then extracted features from image patches around these landmarks for automated disease diagnosis. Similarly, Li *et al.* [44] detected anatomical landmarks and developed a multi-channel CNN for identifying autism spectrum disorder. Liu *et al.* [21] and Pan *et al.* [4] proposed deep learning models with multiple sub-networks to learn image-level representations from multiple local patches located by anatomical landmarks. Note that these methods generally rely on hand-crafted features to select disease-relevant locations (via ROIs or anatomical landmarks), Hence, they have to treat patch selection and classifier training as two standalone steps, thus leading to that those selected patches may not well coordinated subsequent classifiers.

Without pre-defining disease-relevant regions and patches, Lian *et al.* [17] developed a hierarchical fully convolutional network (FCN) to automatically identify discriminative local patches and regions in whole-brain MR images, upon which task-driven MRI features were then jointly learned and fused to construct hierarchical classification models for disease identification. However, this method cannot explicitly reveal the importance of different brain regions, and also cannot be applied to problems with incomplete multi-modality images.

3 Method

3.1 Problem Formulation

We aim to construct a computer-aided diagnosis system based on multi-modality data, such as MRI (denoted as $\mathcal{A}$) and PET (denoted as $\mathcal{B}$) data. Denote $\mathbf{M} = \{(\mathbf{A}_i, \mathbf{B}_i, \mathbf{y}_i)\}_{i=1}^{N}$ as a dataset consisting of N subjects, where $\mathbf{A}_i \in \mathcal{A}$ and $\mathbf{B}_i \in \mathcal{B}$ represent, respectively, the MRI scan, PET scan of the i^{th} subject. Also, $\mathbf{y}_i \in \{0, 1\}$ denotes the class label of the i^{th} subject, e.g., 1 for AD and 0 for cognitively normal (CN) subjects. An automated diagnosis model with multi-modality data can be formulated as

$$\hat{\mathbf{y}}_i = \mathbb{P}\left(\mathbf{A}_i, \mathbf{B}_i\right), \tag{1}$$

where $\hat{\mathbf{y}}_i$ is the estimated label for the i^{th} subject.

In practice, however, not all subjects have complete data of both modalities. Accordingly, we assume only the first N_c subjects have complete data (*i.e.*, paired MRI and PET scans) and the remaining $N - N_c$ subjects have only one imaging modality (e.g., MRI). The diagnosis model $\mathbb{P}$ can only be learned based on the first N_c subjects with complete multi-modality data as

$$\hat{\mathbb{P}} = \arg\min_{\mathbb{P}} \sum_{i=1}^{N_c} [\mathbb{P}\left(\mathbf{A}_i, \mathbf{B}_i\right) - \mathbf{y}_i], \tag{2}$$

where those $N - N_c$ modality-incomplete subjects cannot be used for model learning. Meanwhile, the model defined in Eq. 2 cannot be used to perform prediction for test subjects with only one imaging modality.

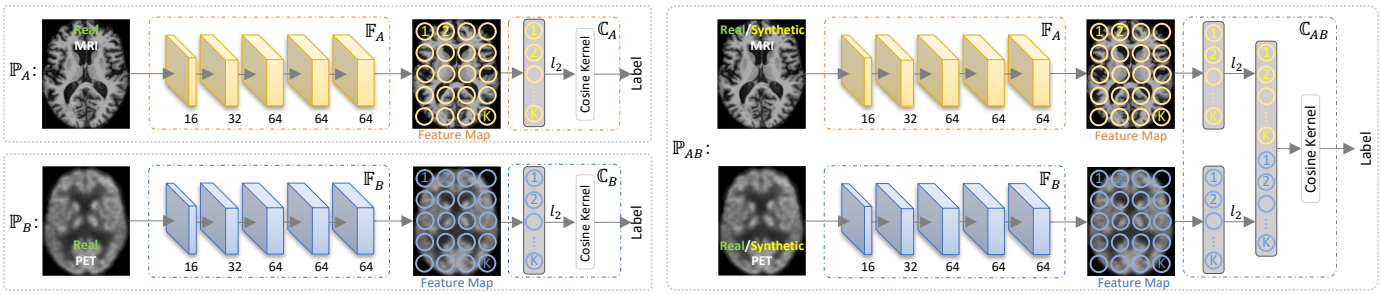
To address this issue, one can impute the missing data (*i.e.*, $\mathbf{B}_i$) for the i^{th} subject, by estimating a virtual $\hat{\mathbf{B}}_i$ based on the available modality (*i.e.*, $\mathbf{A}_i$), considering the underlying relevance between two imaging modalities. Denote $\mathbb{G}_A : \mathcal{A} \rightarrow \mathcal{B}$ as the mapping function from MRI to PET, *i.e.*, $\hat{\mathbf{B}}_i = \mathbb{G}_A\left(\mathbf{A}_i\right)$. Then, the diagnosis model can be executed on modality-incomplete subjects as

$$\hat{\mathbf{y}}^i = \mathbb{P}(\mathbf{A}_i, \mathbf{B}_i) \approx \mathbb{P}(\mathbf{A}_i, \hat{\mathbf{B}}_i) = \mathbb{P}\left(\mathbf{A}_i, \mathbb{G}_A\left(\mathbf{A}_i\right)\right). \tag{3}$$

Based on modality-complete (after imputation) data, $\mathbb{P}$ can be learned by using all subjects as

$$\begin{aligned} \hat{\mathbb{P}} = & \arg\min_{\mathbb{P}} \sum_{i=1}^{N_c} [\mathbb{P}\left(\mathbf{A}_i, \mathbf{B}_i\right) - \mathbf{y}_i] \\ & + \sum_{i=N_c+1}^{N} [\mathbb{P}\left(\mathbf{A}_i, \mathbb{G}_A(\mathbf{A}_i)\right) - \mathbf{y}_i]. \end{aligned} \tag{4}$$

According to Eqs. 3-4, there are two sequential tasks in the automated diagnosis of incomplete multi-modality data, including (**1**) *learning a reliable mapping function for*



(a) Single-modality Disease-image Specific Network (DSNet) (b) Multi-modality Disease-image Specific Network (mDSNet)

Fig. 2. Illustration of the proposed disease-image-specific network (DSNet) for disease classification and modeling disease-image specificity. (a) two single-modality DSNets (*i.e.*, $\mathbb{P}_A$ and $\mathbb{P}_B$) using MRI and PET data, receptively, with each containing a backbone (*i.e.*, $\mathbb{F}_A$ or $\mathbb{F}_B$) for feature extraction and a classifier (*i.e.*, $\mathbb{C}_A$ or $\mathbb{C}_B$) for classification. (b) a multi-modality DSNet (mDSNet) that use paired MRI and PET data as input (*i.e.*, $\mathbb{P}_{AB}$), with two parallel backbones (*i.e.*, $\mathbb{F}_A$ for MRI and $\mathbb{F}_B$ for PET) and a classifier (*i.e.*, $\mathbb{C}_{AB}$ based on concatenation of features maps generated from two modalities). The backbones (*i.e.*, $\mathbb{F}_A$ and $\mathbb{F}_B$) in single-modality DSNet and mDSNet share the same network architecture but have different input modalities, containing 5 convolutional layers (size: $3 \times 3 \times 3$) with instance normalization and "relu" activation. Also, feature maps of the first 4 convolutional layers are max-pooled, while the feature map of the last layer in each backbone is average-pooled with the stride of $2 \times 2 \times 2$. Here, K denote the elements in the feature map generated by $\mathbb{F}_A$ or $\mathbb{F}_B$ and $K = 4 \times 5 \times 4$ in this work.

data imputation (*i.e.*, $\mathbb{G}_A : \mathcal{A} \rightarrow \mathcal{B}$) to synthesize missing data for modality-incomplete subjects, and (**2**) *learning a classification model* (*i.e.*, $\mathbb{P}$) to effectively use multi-modality data for brain disease diagnosis. If these two tasks are performed independently [4], [13], the synthesized data may not be well coordinated with the subsequent diagnosis task. Therefore, we propose the DSDL framework to jointly perform both tasks of image synthesis and disease diagnosis. As illustrated in Fig. 1, this framework include three major components: (1) two single-modality **DSNets**, *i.e.*, $\mathbb{P}_A$ and $\mathbb{P}_B$, for disease diagnosis and learning disease-image specificity; (2) a **FGAN** for missing image synthesis; and (3) a multi-modality DSNet (*i.e.*, $\mathbb{P}_{AB}$) for brain disease identification. By jointly training DSNet and FGAN, we can encourage that the disease-image specificity learned by DSNet can be preserved in the image synthesis process, and also those synthetic images are task-oriented for disease diagnosis by focusing on disease-relevant brain regions in each modality.

3.2 Single-modality DSNet

A specific brain disease is often highly relevant with particular regions [17], [18], [19], [21], and disease-relevant regions may differ in MRI and PET scans [7], [8]. To model such disease-image specificity, we propose two single-modality DSNets (see Fig. 2 (a)) for real MRI and real PET scans, respectively. Using both models, we can directly extract features from each input whole-brain image, and identify disease-relevant regions implicitly in each modality. The identified disease-image specificity will be further employed to aid the image synthesis process conducted by FGAN.

3.2.1 Network Architecture

Each single-modality DSNet contains sequentially a backbone feature extraction module (*i.e.*, $\mathbb{F}_A$ or $\mathbb{F}_B$) and a classification module (*i.e.*, $\mathbb{C}_A$ or $\mathbb{C}_B$). The feature extraction module has 5 Conv layers, with 16, 32, 64, 64, and 64 channels, respectively. The first 4 and the last Conv layers are respectively followed by the max-pooling and average-pooling with the stride of 2 and the kernel size of $3 \times 3 \times 3$. For an input image, the feature extraction module outputs its feature maps at each Conv layer. The classification module first l_2-normalizes the feature vectors in the feature map of the 5^{th} Conv layer, then concatenates them to construct a spatial representation, and finally uses a fully-connected layer with a spatial cosine kernel to compute the probability score of a subject belonging to a particular category.

3.2.2 Classification with Spatial Cosine Kernel

To facilitate analysis, we decompose each MR/PET image $\mathbf{X}$ ($\mathbf{X}$ stand for $\mathbf{A} \in \mathcal{A}$ or $\mathbf{B} \in \mathcal{B}$) into a disease-relevant part and a residual normal part. After feature extraction by $\mathbb{F}_X(*)$, the output feature map $\mathbf{U}$ can be decomposed accordingly into the disease-relevant part $\mathbf{U}_d$ and the residual normal part $\mathbf{U}_r$

$$\mathbf{U} = \mathbb{F}_X(\mathbf{X}) = \alpha \mathbf{U}_d + (1-\alpha)\mathbf{U}_r, \tag{5}$$

where α is a coefficient that weighs the relationship between those two parts of a specific subject. Since the residual normal part $\mathbf{U}_r$ is not relevant to the disease, the diagnosis result should be independent of it. In other words, the response of the classifier $\mathbb{C}_X(*)$ to the entire feature map is only relevant with the disease-relevant part, i.e., $\mathbb{C}_X(\mathbf{U}) = \mathbb{C}_X(\mathbf{U}_d)$.

Since it is difficult to estimate the true value of α for each brain image, we propose a *spatial cosine module* to suppress the effect of α in DSNet, making the disease-relevant features conspicuous and easy to be captured. Denote $\mathbf{U} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_K\}$ as the feature map generated by the final Conv layer in a feature extractor, where the k^{th} ($k = 1, \cdots, K$) element is a vector corresponding to a the k^{th} spatial location in the brain and $K = 4 \times 5 \times 4$ is the number of elements in the feature map when input size is $144 \times 176 \times 144$. We first perform l_2-normalization on each vector in $\mathbf{U}$, and then concatenate them as the spatial representation of an MRI/PET scan:

$$\tilde{\mathbf{U}} = \left(\frac{\boldsymbol{v}_1^{\mathrm{T}}}{\|\boldsymbol{v}_1\|_2}, \frac{\boldsymbol{v}_2^{\mathrm{T}}}{\|\boldsymbol{v}_2\|_2}, \cdots, \frac{\boldsymbol{v}_K^{\mathrm{T}}}{\|\boldsymbol{v}_K\|_2} \right)^{\mathrm{T}}, \tag{6}$$

through which we can avoid estimating the values of α for different images. Instead of using only the first-order

representation in Eq. 6, we further propose the following multiple-order features to represent each input image:

$$\tilde{\tilde{\mathbf{U}}} = \left(\frac{v_1^{\mathrm{T}}}{\|v_1\|_2}, \frac{(v_1^{\mathrm{T}})^2}{\|v_1\|_2^2}, \cdots, \frac{v_K^{\mathrm{T}}}{\|v_K\|_2}, \frac{(v_K^{\mathrm{T}})^2}{\|v_K\|_2^2} \right)^{\mathrm{T}}. \tag{7}$$

Suppose $\mathbb{C}_X(*)$ is a classifier with hyperplane parameters $\mathbf{w}$. With the feature representation $\mathbf{u}$ ($\tilde{\mathbf{U}}$ or $\tilde{\tilde{\mathbf{U}}}$) of an input scan, $\mathbb{C}_X(*)$ is defined as the following spatial cosine kernel

$$\mathbb{C}(\mathbf{u}; \mathbf{w}) = \cos\langle \mathbf{u}, \mathbf{w} \rangle = \frac{\mathbf{u}^{\mathrm{T}}\mathbf{w}}{\|\mathbf{u}\|_2 \|\mathbf{w}\|_2} = \frac{\mathbf{u}^{\mathrm{T}}}{\|\mathbf{u}\|_2} \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|_2}, \tag{8}$$

which is equivalent to the product of l_2-normalized $\mathbf{u}$ and $\mathbf{w}$ (both having the constant unit norm). The constant norm forces $\mathbb{F}(*)$ to focus on the disease-relevant part, since all features have the same norm after l_2-normalization (in Eq. 8), and thus suppresses the influence of the residual normal part. More detailed explain of the theory of spatial cosine kernel could be seen in Section 5.4.

3.3 FGAN

Since a pair of MR and PET images scanned from the same subject have underlying relevance but probably different disease-relevant regions, we develop **FGAN** to synthesize a missing PET image based on its corresponding MRI with a feature-consistency constraint for generating diagnosis-oriented images and an adversarial constraint for generating real-like data. As shown in Fig. 3, our FGAN mainly contains a generative adversarial learning component and two feature-consistency components for MRI and PET modalities, respectively.

3.3.1 Generative Adversarial Learning Component

Using the generative adversarial learning component, we aim to learn an image generator $\mathbb{G}_A : \mathcal{A} \rightarrow \mathcal{B}$ on modality-complete training subjects, and thus can impute the missing PET image for a test subject with only MRI scan via $\mathbb{G}_A(\mathbf{A})$. An inverse mapping $\mathbb{G}_B : \mathcal{B} \rightarrow \mathcal{A}$ ($\mathbb{G}_B = \mathbb{G}_A^{-1}$) is also learned to build the bi-directional mapping between MRI and PET domains. Denote $\mathbb{D}_A$ and $\mathbb{D}_B$ as two discriminators that can tell whether an input image is real or synthetic, corresponding to the MRI and PET domains, respectively. With two generators (*i.e.*, $\mathbb{G}_A$ and $\mathbb{G}_B$) and two discriminators (*i.e.*, $\mathbb{D}_A$ and $\mathbb{D}_B$), the *adversarial loss* is defined as

$$\begin{aligned} &\mathfrak{L}_a(\mathcal{A}, \mathcal{B}; \mathbb{G}_A, \mathbb{G}_B, \mathbb{D}_A, \mathbb{D}_B) \\ &= \mathop{\mathrm{E}}_{(\mathbf{A},\mathbf{B})\in\mathbf{M}} [(\log(\mathbb{D}_B(\mathbf{B})) + \log(1 - \mathbb{D}_B(\mathbb{G}_A(\mathbf{A})))) \\ &\qquad + (\log(\mathbb{D}_A(\mathbf{A})) + \log(1 - \mathbb{D}_A(\mathbb{G}_B(\mathbf{B}))))]. \end{aligned} \tag{9}$$

3.3.2 Feature-consistency Component

To employ the disease-image specificity into FGAN, we design the *feature-consistency constraint* to encourage that the feature maps of a synthetic image should be consistent with feature maps of its corresponding real image. With each feature extractor (e.g., $\mathbb{F}_A$) containing 5 Conv layers, we denote the feature map of its j^{th} layer as $\mathbb{F}_{A,j}$ ($j = 1, \cdots, 5$). As shown in the left and right parts of Fig. 3, to encourage features of a synthetic and its respective real images to be consistent at different abstraction levels, we design the feature-consistency constraint for PET and MRI as

$$\begin{aligned} &\|\mathbb{F}_A(\mathbb{G}_B(\mathbf{B})) - \mathbb{F}_A(\mathbf{A})\| + \|\mathbb{F}_B(\mathbb{G}_A(\mathbf{A})) - \mathbb{F}_B(\mathbf{B})\| \\ &= \sum_{i=1}^{5} (\|\mathbb{F}_{A,j}(\mathbb{G}_B(\mathbf{B})) - \mathbb{F}_{A,j}(\mathbf{B})\| \\ &\qquad + \|\mathbb{F}_{B,j}(\mathbb{G}_A(\mathbf{A})) - \mathbb{F}_{B,j}(\mathbf{B})\|), \end{aligned} \tag{10}$$

through which the disease-image specificity identified by DSNet (based on real MRI/PET data) can be used to constraint FGAN to focus on those modality-specific disease-relevant regions, rather than the whole-brain image. Namely, FGAN is encouraged to generate diagnosis-oriented images by using the feature-consistency constraint components.

Based on the feature-consistency constraint, the proposed *feature-consistency loss* is defined as

$$\begin{aligned} &\mathfrak{L}_f(\mathcal{A}, \mathcal{B}; \mathbb{G}_A, \mathbb{G}_B, \mathbb{F}_A, \mathbb{F}_B) \\ &= \mathop{\mathrm{E}}_{(\mathbf{A},\mathbf{B})\in\mathbf{M}} [\|\mathbb{F}_A(\mathbb{G}_B(\mathbf{B})) - \mathbb{F}_A(\mathbf{A})\| \\ &\qquad + \|\mathbb{F}_B(\mathbb{G}_A(\mathbf{A})) - \mathbb{F}_B(\mathbf{B})\|], \end{aligned} \tag{11}$$

which encourages that a pair of synthetic and real scans from the same modality share the same disease-image specificity.

Finally, the overall loss function of FGAN is defined as

$$\begin{aligned} &\mathfrak{L}(\mathcal{A}, \mathcal{B}; \mathbb{G}_A, \mathbb{G}_B, \mathbb{D}_A, \mathbb{D}_B, \mathbb{F}_A, \mathbb{F}_B) \\ &= \mathfrak{L}_f(\mathcal{A}, \mathcal{B}; \mathbb{G}_A, \mathbb{G}_B, \mathbb{F}_A, \mathbb{F}_B) + \mathfrak{L}_a(\mathcal{A}, \mathcal{B}; \mathbb{G}_A, \mathbb{G}_B, \mathbb{D}_A, \mathbb{D}_B). \end{aligned} \tag{12}$$

3.3.3 Network Architecture

As shown in Fig. 3, each generator (e.g., $\mathbb{G}_A$) in our FGAN consists of 3 Conv layers (with 8, 16, and 32 channels, respectively) to extract the knowledge of images in the original domain (e.g., $\mathcal{A}$), 6 residual network blocks (RNBs) [45] to transfer the knowledge from the original domain to the target domain (e.g., $\mathcal{B}$), and 2 deconvolutional (Deconv) layers (with 32 and 16 channels, respectively) and 1 Conv layer (with 1 channel) to construct the image in the target domain. Each discriminator (e.g., $\mathbb{D}_B$) contains 5 Conv layers, with 16, 32, 64, 128, and 1 channel(s), respectively. It outputs an indicator to tell whether the input pair of real image (e.g., $\mathbf{B}$) and synthetic image (e.g., $\mathbb{G}_A(\mathbf{A})$) are distinguishable (output: 0) or not (output: 1). Besides, each feature-consistency component contains two parallel subnetworks (e.g., $\mathbb{F}_B$) that share the same architecture and parameters with the feature extractor in each modality in our DSNet (e.g., $\mathbb{P}_B$). Note that $\mathbb{F}_A$ and $\mathbb{F}_B$ in Fig. 1 (b) are initialized by those learned in Fig. 1 (a) and kept frozen in FGAN. It inputs a pair of real image (e.g., $\mathbb{B}$) and synthetic image (e.g., $\mathbb{G}_A(\mathbf{A})$), and outputs a differential score to indicate the similarity between the feature maps of the real and its corresponding synthetic image. Hence, the disease-image specificity learned in DSNet can be used to aid the image imputation process in FGAN, *i.e.*, the modality-specific disease-relevant regions will be more effectively synthesized in a diagnosis-oriented manner. In turn, synthetic images could be more relevant to the task of brain disease diagnosis, thus boosting the learning performance.

3.3.4 Model Extension

Besides the loss function defined in Eq. 12, we further extend our FGAN model by using several additional constraints, such as the voxel-wise-consistency constraint [28]

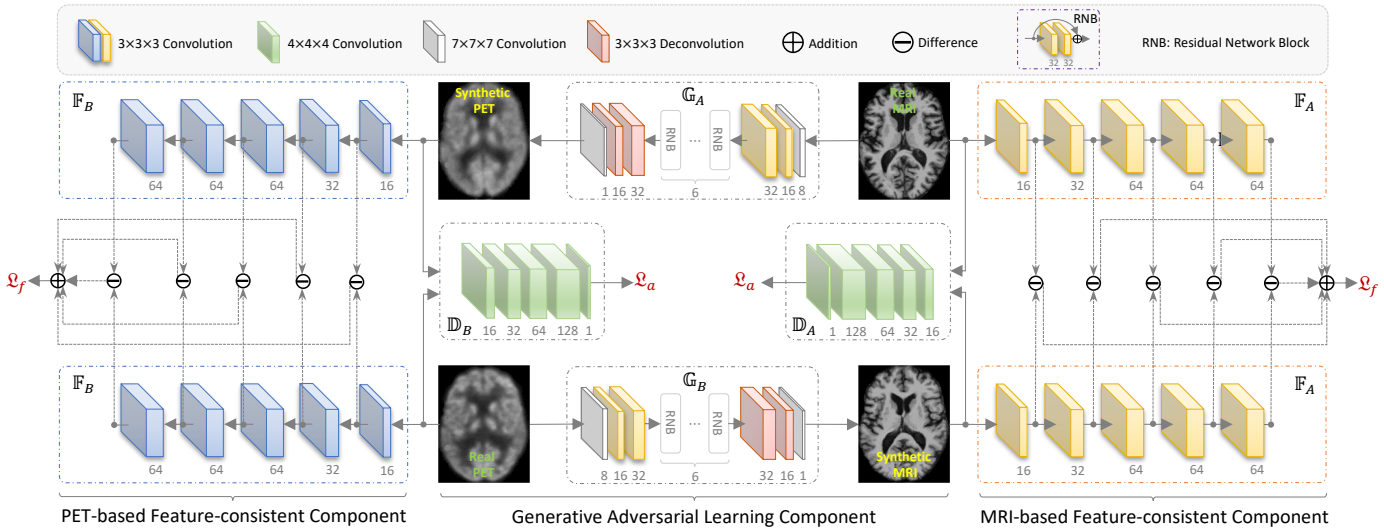


Fig. 3. Illustration of our feature-consistency generative adversarial network (FGAN) for image synthesis. It contains (1) two feature-consistency components (i.e., MRI-based and PET-based components) to encourage feature maps of a synthetic image to be consistent with those of its corresponding real image, and (2) a generative adversarial learning component to synthesize images under the constraints of feature consistency (i.e., $\mathfrak{L}_f$) and distribution consistency (i.e., $\mathfrak{L}_a$). Note that $\mathbb{F}_A$ and $\mathbb{F}_B$ in two feature-consistency components have same architecture but are learned in MRI-based and PET-based DSNet models, respectively, through which the disease-image specificity learned in DSNets will be employed in the image synthesis process, encouraging FGAN to focus on those disease-relevant regions in each modality. Also, the adversarial components, i.e., $\mathbb{D}_A$ and $\mathbb{D}_B$), are used to constrain the synthetic MRI and PET scans follow the same data distribution of those real MRI and PET scans, respectively. Besides, two generators (i.e., $\mathbb{G}_A$ and $\mathbb{G}_B$) are learned to construct bi-directional mappings between two imaging modalities.

and cycle-consistency constraint [4], [32]. In this work, we denote the FGAN with an additional voxel-wise-consistency constraint as **FVoxGAN**, and denote the FGAN with an addition cycle-consistency constraint as **FCycGAN**. Specifically, the corresponding losses of FVoxGAN and FCycGAN are defined, respectively, as

$$\begin{aligned}&\mathfrak{L}_{E1}(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B,\mathbb{D}_A,\mathbb{D}_B,\mathbb{F}_A,\mathbb{F}_B)\\&=\mathfrak{L}_f(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B,\mathbb{F}_A,\mathbb{F}_B)+\mathfrak{L}_a(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B,\mathbb{D}_A,\mathbb{D}_B)\\&\quad+\mathfrak{L}_v(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B),\end{aligned}\tag{13}$$

and

$$\begin{aligned}&\mathfrak{L}_{E2}(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B,\mathbb{D}_A,\mathbb{D}_B,\mathbb{F}_A,\mathbb{F}_B)\\&=\mathfrak{L}_f(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B,\mathbb{F}_A,\mathbb{F}_B)+\mathfrak{L}_a(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B,\mathbb{D}_A,\mathbb{D}_B)\\&\quad+\mathfrak{L}_c(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B),\end{aligned}\tag{14}$$

where the voxel-wise-consistency loss in Eq. 13 and cycle-consistency loss in Eq. 14 are defined, respectively, as

$$\begin{aligned}&\mathfrak{L}_v(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B)\\&=\mathop{\mathbb{E}}_{(\mathbf{A},\mathbf{B})\in\mathbf{M}}[\|\mathbb{G}_B(\mathbf{B})-\mathbf{A}\|+\|\mathbb{G}_A(\mathbf{A})-\mathbf{B}\|],\end{aligned}\tag{15}$$

and

$$\begin{aligned}&\mathfrak{L}_c(\mathcal{A},\mathcal{B};\mathbb{G}_A,\mathbb{G}_B)\\&=\mathop{\mathbb{E}}_{(\mathbf{A},\mathbf{B})\in\mathbf{M}}[\|\mathbb{G}_B(\mathbb{G}_A(\mathbf{A}))-\mathbf{A}\|+\|\mathbb{G}_A(\mathbb{G}_B(\mathbf{B}))-\mathbf{B}\|].\end{aligned}\tag{16}$$

Different from the adversarial loss and feature-consistency loss that rely on specific sub-networks, the voxel-wise-consistency loss in FVoxGAN and cycle-consistency loss in FCycGAN can be directly apply to paired real and synthetic images. Therefore, our FGAN and its two extensions share the same network architecture (see Fig. 3) but have different loss functions.

3.4 Multi-modality DSNet

Using the proposed FGAN, one can obtain synthetic MRI/PET scans for modality-incomplete subjects, and thus each subject will be represented by complete multi-modality data (*i.e.*, a pair of MRI and PET scans). To handle classification problems with multi-modality data, we extend our single-modality DSNet to a multi-modality version, called **mDSNet**.

As shown in Fig. 2 (b), our mDSNet $\mathbb{P}_{AB}$ consists of sequentially two parts: (1) two parallel feature extraction modules (*i.e.*, $\mathbb{F}_A$ and $\mathbb{F}_B$) followed by a feature map concatenation operation to fuse futures of MRI and PET and (2) a classifier (*i.e.*, $\mathbb{C}_{AB}$) with the multi-order feature representation $\tilde{\tilde{\mathbf{U}}}$ (see Eq. 7) as its input. The feature extraction modules in mDSNet have the same network architecture as DSNet but different parameters. The proposed mDSNet is trained from sketch using both the real and synthetic images. Since the classifier $\mathbb{C}_{AB}$ can also input the first-order feature representation $\tilde{\mathbf{U}}$ (see Eq. 6), we denote mDSNet with first-order features as **mDSNet-1st** in this work.

3.5 Implementation Details

In the *training* stage, the proposed DSDL framework is executed via the following three stages: (1) using only the real MRI (i.e., $\mathbb{A}$) and real PET (i.e., $\mathbb{B}$) data to train two single-modality DSNets (*i.e.*, $\mathbb{P}_A$ and $\mathbb{P}_B$), respectively, for Disease-image Specificity Identification; (2) using modality-complete subjects with real MRI and PET data to train FGAN for image synthesis; and (3) using the complete (after imputation) paired MRI and PET scans of all training subjects to train mDSNet for disease diagnosis. Note that, to augment the number of samples, The combination $(\mathbf{A}_i, \mathbb{G}_A(\mathbf{A}_i))$ will also be used as a training sample to train mDSNet even $\mathbf{B}_i$ exists.

In the *test* stage, for an unseen test subject with complete multi-modality data, we directly feed its MRI and PET scans to mDSNet for classification. Given an unseen test subject with missing MRI/PET scan, we first impute the missing scan using our trained FGAN model and then feed the

complete (after imputation) paired MRI and PET scans into mDSNet for classification.

Our proposed models are implemented by Python with Tensorflow on a platform with GTX 1080 Ti and Intel Core i7-8700. We first train two DSNets $\mathbb{P}_A$ and $\mathbb{P}_B$ for 40 epochs using complete subjects (i.e., those with paired PET and MR images). We then train $\mathbb{D}_A$ and $\mathbb{D}_B$ by minimizing $-\mathfrak{L}_a(*)$ with fixed $\mathbb{G}_A$ and $\mathbb{G}_B$, and train $\mathbb{G}_A$ and $\mathbb{G}_B$ by minimizing $\mathfrak{L}(*)$ with fixed $\mathbb{D}_A$ and $\mathbb{D}_B$, iteratively, for 100 epochs. After that, we train mDSNet for 40 epochs using both real and synthetic data. We use SGD solver with a learning rate 1×10^{-3} to train DSNets and mDSNet, and use the Adam solver [46] with a learning rate of 2×10^{-3} to train FGAN. The batch size is set to 1 due to the limitation of GPU memory. More details on hyper-parameter settings could be seen in the *Supplementary Materials*.

4 Experiments

4.1 Materials and Image Pre-processing

We first evaluated the proposed method on two subsets of ADNI [5], including ADNI-1 and ADNI-2. Subjects in these two datasets were divided into four categories: (1) AD, (2) CN, (3) progressive MCI (pMCI) that would progress to AD within 36 months after baseline, and (4) static MCI (sMCI) that would not progress to AD. After removing these subjects appearing in both ADNI-1 and ADNI-2 from ADNI-2, there are 205 AD, 231 CN, 165 pMCI and 147 sMCI subjects in ADNI-1, while there are 162 AD, 209 CN, 89 pMCI and 256 sMCI subjects in ADNI-2. All the above subjects in ADNI-1 and ADNI-2 have baseline MRI data, while only 356 and 581 of these subjects have PET images in ADNI-1 and ADNI-2, respectively. We also used the AIBL [47] dataset for performance evaluation, containing 71 AD and 447 CN subjects. Similar to ADNI-1 and ADNI-2, all subjects in AIBL have MRI scans, and only part of them have PET scans. The demographic and clinical information of three datasets are listed in Table 1. More details of these data are described in Section 2 of the *Supplementary Materials*.

For data pre-processing, we first performed skull-stripping on MRI scans using FreeSurfer [48], and then linearly aligned each PET scan to its corresponding MRI scan. Next, we affine each MRI scan to the commonly-used MNI template using SPM [49], while its corresponding PET scan (if existed) is also affined using the same affination parameter. In this way, each pair of MRI and PET scans of a same subject have the spatial correspondence.

We performed two groups of experiments on the ADNI-1 and ADNI-2 datasets: synthesizing MR and PET images and diagnosing brain diseases, including AD identification (AD vs. CN classification) and MCI conversion prediction (pMCI vs. sMCI classification). To evaluate the generalization capability of our proposed models for image synthesis and disease classification, we further performed an extra group of experiments on the AIBL dataset.

TABLE 1
The demographic and clinical information of studied subjects with four categories (Cat.) from three datasets. The education (Edu.) years and the mini-mental state examination (MMSE) values are reported in terms of mean ± standard deviation. M: Male; F: Female.

Dataset	Cat.	MRI/PET	M/F	Age	Edu.	MMSE
ADNI-1	AD	205/95	106/99	76 ± 8	14 ± 4	23 ± 2
	CN	231/102	119/112	76 ± 5	16 ± 3	29 ± 1
	pMCI	165/76	100/65	75 ± 7	16 ± 3	27 ± 2
	sMCI	147/83	101/46	75 ± 8	16 ± 4	27 ± 2
ADNI-2	AD	162/142	95/70	75 ± 8	16 ± 3	23 ± 3
	CN	209/186	99/110	73 ± 6	16 ± 3	27 ± 2
	pMCI	89/80	52/37	73 ± 7	16 ± 3	28 ± 2
	sMCI	256/173	146/110	71 ± 8	16 ± 3	27 ± 2
AIBL	AD	71/62	30/41	73 ± 8	–	27 ± 4
	CN	447/407	192/254	72 ± 7	–	28 ± 4

4.2 Evaluation of Synthetic Neuroimages

4.2.1 Competing Methods

We first evaluate the quality of synthetic images generated by three typical GANs, including (1) conventional **GAN** with only an adversarial loss, (2) cycle-consistency GAN (**CycGAN**) [4], [32] with an additional cycle-consistency constraint, and (3) GAN with a voxel-wise-consistency constraint (**VoxGAN**) [28], our **FGAN**, and its two variants, i.e., **FVoxGAN** and **FCycGAN**. For the fair comparison, FGAN and its variants share the same network architecture as shown in Fig. 3 but have different losses. The other three GANs have the same architecture as the generative adversarial learning component of FGAN (the middle part of Fig. 3) but different losses.

4.2.2 Experimental Setup

We trained six GANs using the subjects with real MRI and PET scans in ADNI-1, and tested the trained models on complete subjects (with both real MRI and real PET scans) in ADNI-2. We used four metrics to measure the quality of synthetic images, including (1) the mean absolute error (MAE), (2) the mean square error (MSE), (3) peak signal-to-noise ratio (PSNR), and (4) structural similarity index measure (SSIM) [50].

To evaluate the reliability of synthetic MR and PET images in disease diagnosis, we further reported the values of the area under receiver operating characteristic (AUC) achieved by our single-modality DSNet model on both AD identification (denoted as AUC*) and MCI conversion prediction (denoted as AUC†). We first trained MRI- and PET-based DSNet models on complete subjects (i.e., with both real MRI and real PET scans) in ADNI-1, respectively, and then applied these two DSNets to subjects in ADNI-2 represented by synthetic MR and PET images, respectively, for classification.

4.2.3 Results of Image Synthesis

In Table 2, we report the results achieved by six different methods in synthesizing MRI and PET scans. Four interesting observations can be found from Table 2. *First*, five advanced GAN methods (i.e., CycGAN, VoxGAN, FCycGAN, FVoxGAN, and FGAN) generally yield better results than the baseline GAN. This implies that the cycle-consistency loss, voxel-wise-consistency loss, and feature-consistency loss are positive constraints to help synthesize images with higher quality, in terms of both image similarity and discrimination in subsequent diagnosis tasks. It verifies that spatial structure information is useful for computer-aided AD diagnosis. *Second*, The AUC values obtained by using synthetic images generated by our FGAN-based methods

TABLE 2
Results(% except PSNR) of image synthesis achieved by six different methods for MRI and PET scans of subjects in ADNI-2, with the models trained on ADNI-1.

Method	Synthetic MRI						Synthetic PET					
	MAE	MSE	SSIM	PSNR	AUC*	AUC†	MAE	MSE	SSIM	PSNR	AUC*	AUC†
GAN	10.66 ±0.77	3.68 ±0.54	59.34 ±2.78	26.43 ±0.59	66.69	52.86	10.79 ±1.00	3.05 ±0.63	57.41 ±2.82	27.27 ±0.77	52.92	51.54
CycGAN	10.16 ±0.78	3.33 ±0.51	60.55 ±3.06	26.86 ±0.65	82.35	65.68	10.36 ±1.25	2.53 ±0.69	57.41 ±3.54	27.56 ±0.96	57.50	52.55
VoxGAN	**8.10** ±0.81	**2.17** ±0.46	**69.74** ±3.95	**28.75** ±0.79	88.12	66.85	**7.61** ±1.42	**1.55** ±0.62	**70.23** ±5.48	**30.40** ±1.48	82.66	68.65
FGAN (Ours)	8.64 ±0.83	2.27 ±0.47	68.20 ±3.65	28.56 ±0.78	**94.06**	78.05	8.24 ±1.37	1.80 ±0.63	67.34 ±4.97	29.70 ±1.29	86.05	70.63
FCycGAN (Ours)	8.69 ±0.87	2.39 ±0.51	67.20 ±3.52	28.34 ±0.81	93.22	**78.92**	8.54 ±1.51	1.94 ±0.69	65.85 ±5.21	29.38 ±1.36	86.10	69.24
FVoxGAN (Ours)	8.38 ±0.87	2.26 ±0.52	68.25 ±3.63	28.59 ±0.81	93.23	78.45	8.42 ±1.37	1.89 ±0.64	66.16 ±4.92	29.48 ±1.27	**86.10**	**72.22**

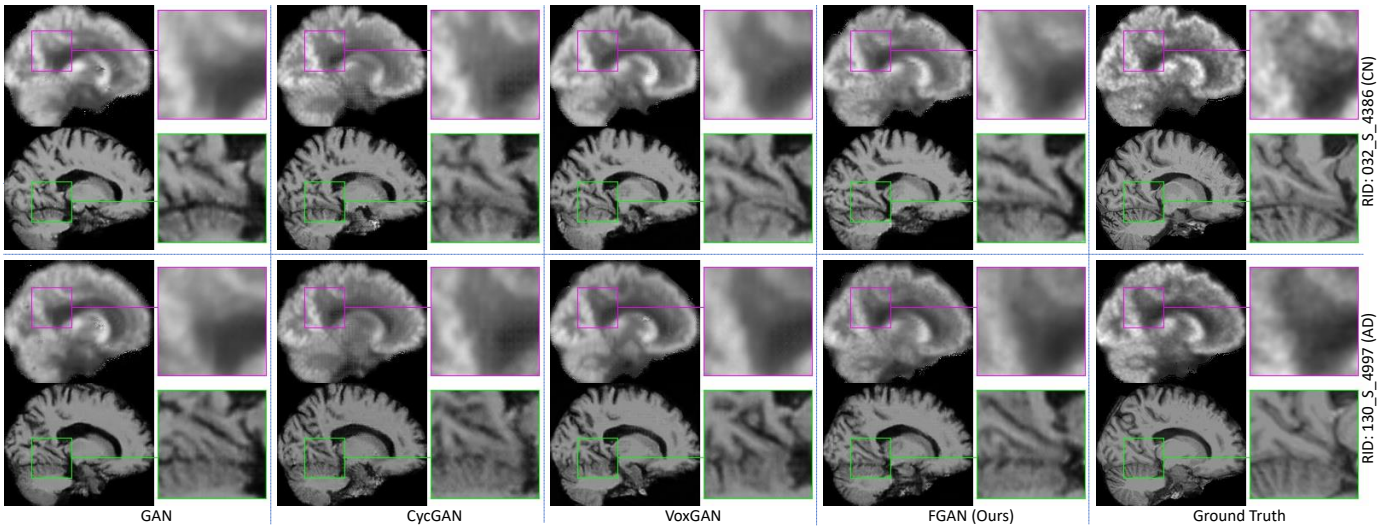


Fig. 4. PET and MRI scans synthesized by four methods for two typical subjects (Roster IDs: 4386, 4997) in ADNI-2, along with their corresponding ground-truth images. All six image synthesis models are trained on ADNI-1.

TABLE 3
Diagnosis results (%) achieved by six different methods, with classification models trained on ADNI-1 and tested on ADNI2. Methods marked as "-M" denote that only subjects with real MRI scans are used for model training, methods marked as "-C" denote that modality-complete subjects (with real paired MRI and PET scans) are used for model training, while the remaining methods employ all subjects (with both real images and synthetic PET images generated by FGAN) in two classification tasks.

Method	AD *vs.* CN Classification						pMCI *vs.* sMCI Classification					
	AUC	ACC	SPE	SEN	F1S	MCC	AUC	ACC	SPE	SEN	F1S	MCC
ROI-M	86.22	79.41	83.64	76.08	78.19	59.30	70.48	66.38	62.92	67.58	49.12	27.21
PBM-M	88.11	82.22	77.36	86.07	79.35	63.83	73.21	68.70	58.43	72.27	49.06	28.04
LDMIL-M	95.77	90.37	88.48	**91.87**	89.02	80.46	80.35	73.33	73.03	73.44	58.56	41.77
CNN-M	93.87	88.24	87.27	89.00	86.75	76.18	76.59	72.46	66.29	74.61	55.40	37.29
mDSNet-1st-M (Ours)	95.68	90.11	**89.70**	90.43	88.89	79.99	**82.36**	75.65	73.03	76.56	60.75	45.14
mDSNet-M (Ours)	**96.31**	**90.64**	**89.70**	91.39	**89.42**	**81.03**	81.84	**76.23**	**74.16**	**76.95**	**61.68**	**46.52**
ROI-C	88.95	81.49	82.99	80.32	79.74	62.92	69.17	66.92	42.50	77.78	44.16	20.74
PBM-C	87.87	82.39	80.27	84.04	80.00	64.27	68.04	68.85	60.00	72.78	54.24	31.28
LDMIL-C	95.64	90.75	88.44	**92.55**	89.35	81.18	82.62	76.92	71.25	**79.44**	65.52	48.67
CNN-C	94.40	89.25	88.44	89.89	87.84	78.22	78.70	75.00	67.50	78.33	62.43	44.13
mDSNet-1st-C (Ours)	95.78	90.45	89.80	90.96	89.19	80.64	82.74	76.15	**76.25**	76.11	66.30	49.33
mDSNet-C (Ours)	**96.26**	**91.46**	**90.85**	91.94	**90.21**	**82.65**	**83.06**	**77.47**	75.00	78.61	**67.80**	**51.27**
ROI	90.51	83.42	84.24	82.78	81.76	66.69	72.31	71.30	52.81	77.73	48.70	29.12
PBM	91.46	84.49	81.21	87.08	82.21	68.48	72.79	71.88	64.04	74.61	54.03	35.37
LDMIL	96.76	91.71	88.48	94.26	90.40	83.17	83.89	79.71	71.91	82.42	64.65	51.13
CNN	94.72	89.57	85.45	92.82	87.85	78.82	79.98	76.52	69.66	78.91	60.49	44.98
mDSNet-1st (Ours)	97.02	92.25	**90.91**	93.30	91.18	84.27	83.94	**80.00**	70.79	**83.20**	64.61	51.20
mDSNet (Ours)	**97.23**	**93.05**	**90.91**	**94.74**	**92.02**	**85.88**	**84.44**	79.71	**75.28**	81.25	**65.69**	**52.47**

(i.e., FGAN, FCycGAN, and FVoxGAN) are significantly higher than those using synthetic images generated by three competing methods (i.e., GAN, CycGAN, and VoxGAN) in both classification tasks. The possible reason is that the proposed feature-consistency constraint is effective to encourage GAN models to generate diagnosis-oriented images (rather than focusing on whole-brain regions), thus helping boost the performance of brain disease diagnosis. *Besides*, regarding four metrics for image quality (i.e., MAE, MSE, SSIM and PSNR), our FGAN models consistently outper-

forms CycGAN and GAN in synthesizing both MRI and PET scans, but only achieves comparable results compared with VoxGAN. This implies that the proposed feature-consistency loss is a strong constraint to encourage good quality of synthetic images, but not as strong as the voxel-wise-consistency loss used in VoxGAN. However, using the voxel-wise-consistency loss and feature-consistency loss simultaneously (as we do in FVoxGAN) does not significantly improve the quality of synthetic images. It implies that the feature-consistency loss and voxel-wise-consistency loss may have potential competitive relationships. *Furthermore*, the AUC values (*i.e.* AUC* and AUC†) achieved by six methods based on synthetic PET data are observably lower than based on synthetic MRI, which indicates it may be more hard to synthesize classifiable PET from MRI than synthesize MRI from PET. The possible reason is that PET scans contain comparable structure information with MRI while MRI contain insufficient functional information to be passed to synthetic PET.

Fig. 4 visualizes the ground truth MR and PET images of a CN subject (Roster ID: 4386) and an AD subject (Roster ID: 4997) in ADNI-2 and the synthetic images generated by GAN, cycGAN (used in our previous work), VoxGAN, and our FGAN, respectively. To show sufficient details, the region in the pink / green rectangular on each image was enlarged and displayed to the right of the image. It reveals that the images synthesized by our FGAN (4^{th} column) are more consistent with the ground truth (5^{th} column) than those synthesized by other GANs (1^{st}-3^{rd} columns), particularly in terms of the ventricle size and sulcus width. It can be attributed to the fact that, comparing to the voxel-consistency constraint and cycle-consistency constraint, the feature-consistency constraint used in FGAN is a high-level constraint, which can encourage pattern similarities rather than only voxel-level similarities. More views and more examples and supplied in Figs. S3-S5 of the *Supplementary Materials*.

4.3 Evaluation of Automated Diseases Diagnosis

4.3.1 Competing Methods

We further evaluated our **mDSNet** on both tasks of AD identification and MCI conversion prediction against two conventional methods using concatenated MRI and PET features, i.e., (1) **ROI** method [7], [51], and (2) patch-based morphology (**PBM**) [9], [52] and two deep learning models, i.e., (3) landmark-based deep multi-instance learning (**LDMIL**) method [21], and (4) a conventional CNN method. For ROI, PBM and LDMIL methods, we used default parameter settings in their original papers. For the fair comparison, the CNN method shares the similar network architecture with our mDSNet (see Fig. 2 (b)) but a different classification module. That is, CNN globally averages the feature map of the last Conv layer and uses a fully connected layer for classification (instead of using the spatial cosine module in mDSNet). To investigate the effect of multi-order representation in Eq. 7, we compare the **mDSNet-1st** that use first-order features defined in Eq. 6.

4.3.2 Experimental Setup

These classification methods utilize all subjects with both real multi-modality scans and synthetic PET images generated by our FGAN. We also performed experiments on complete subjects (with real paired MRI and PET scans), and denoted the corresponding methods as "-C". Since all subjects have MRI scans in ADNI-1 and ADNI-2 datasets, we also reported the results of different methods using only MRI modality, and denoted the corresponding methods as "-M". For all methods, classifiers are trained on ADNI-1, and tested on the independent ADNI-2 and AIBL datasets, respectively. We employ six metrics for performance evaluation in disease diagnosis, including (1) AUC, (2) accuracy (ACC), (3) sensitivity (SEN), (4) specificity (SPE), (5) F1-Score (F1S), and (6) Matthews correlation coefficient (MCC) [53].

4.3.3 Disease Identification Results on ADNI-2

With models trained on ADNI-1, the disease classification results achieved by different methods on ADNI-2 are reported in Table 3. From Table 3, we can see that our mDSNet generally achieves the best performance in most cases. For instance, using both real and synthetic images, our mDSNet method achieves the highest AUC values (97.23%, 84.44%) in the tasks of AD vs. CN and pMCI vs. sMCI classification. This suggests that our mDSNet is reliable in automated AD diagnosis and progression prediction of MCI patients, which is potentially very useful in practice. *Besides*, our mDSNet yields slightly better results compared to LDMIL that pre-defines disease-relevant regions in brain images via anatomical landmarks [54]. This implies that the proposed spatial cosine kernel provides an efficient strategy to capture the disease-image specificity embedded in neuroimages. *On the other hand*, methods (e.g., mDSNet) using all subjects with complete multi-modality data (after imputation via FGAN) consistently outperform their counterparts (e.g., mDSNet-C) that utilize modality-complete subjects with real MRI and PET scans, and are superior to their counterparts (e.g., mDSNet-M) using all subjects with real MRI scans. For example, mDSNet achieves an MCC value of 52.47 in MCI conversion prediction, which is higher than the results of mDSNet-C (46.52) and mDSNet-M (51.27). The possible reason could be that, compared with mDSNet-C, more subjects are used for model training in mDSNet. Even though mDSet-M (using only real MRI data) and mDSNet used the same number of training subjects, mDSNet takes advantage of data from an additional imaging modality (i.e., real and synthetic PET images). These results demonstrate that neuroimages generated by our FGAN model are useful in promoting the diagnostic performance.

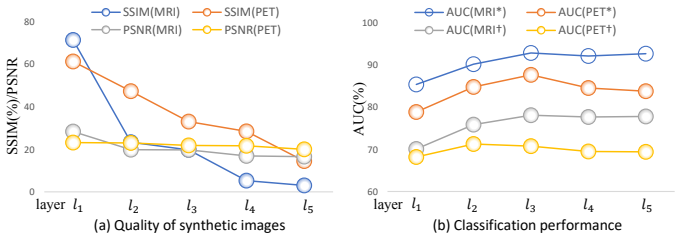
4.3.4 Disease Identification Results on AIBL

We further use AIBL as the testing set for classification performance evaluation. Different from the ADNI, we only use synthesized PET in AIBL for testing to simulate the case that all subjects refused PET scanning. For the fair comparison, six methods utilize all subjects with real MRI scans and synthetic PET scans (generated by FGAN). Results of different methods in AD vs. CN classification are reported in Table 4, considering that AIBL contains a limited number of MCI subjects.

From Table 4, we can see that four deep learning methods (i.e., LDMIL, CNN, mDSNet-1st, and mDSNet) generally outperform two conventional approaches (i.e., ROI

TABLE 4
Diagnosis results (%) achieved by six different methods using all subjects with only real MRI scans (denoted as "-M") and with both real MRI images and synthetic PET images (generated by FGAN) in AC vs. CN. classification. Classification models are trained on ADNI-1 and tested on AIBL.

Method	AD *vs.* CN Classification					
	AUC	ACC	SPE	SEN	F1S	MCC
ROI-M	83.49	73.69	76.06	73.32	44.26	36.02
PBM-M	86.34	78.34	80.28	78.03	50.44	43.80
LDMIL-M	93.81	87.64	**87.32**	87.70	65.96	61.70
CNN-M	91.23	83.78	83.09	83.89	58.41	53.00
DSNet-1st-M (Ours)	93.98	88.80	85.92	89.26	67.78	63.43
DSNet-M (Ours)	**94.39**	**89.77**	83.10	**90.83**	**69.01**	**64.41**
ROI	87.25	79.69	85.92	78.70	53.74	48.45
PBM	90.63	80.66	85.92	79.82	54.95	49.76
LDMIL	93.64	88.03	**87.32**	88.14	66.67	62.45
GANN	92.77	88.42	84.51	89.04	66.67	62.05
mDSNet-1st (Ours)	94.37	89.77	**87.32**	90.16	70.06	66.05
mDSNet (Ours)	**94.92**	**90.35**	**87.32**	**90.83**	**71.26**	**67.34**



(a) Quality of synthetic images (b) Classification performance

Fig. 5. Quality (a) and classification performance (b) of synthetic images generated by five FGAN variants using the feature-consistency constraint on feature maps at only one single Conv layer (e.g., the i-th layer l_i) in $\mathbb{F}_A$ and $\mathbb{F}_B$.

TABLE 5
Comparison of our mDSNet on AIBL while using only ADNI-1 or using both ADNI-1 and ADNI-2 for training in stage (3).

Method	AD *vs.* CN Classification					
	AUC	ACC	SPE	SEN	F1S	MCC
mDSNet	94.92	90.35	87.32	90.83	71.26	67.34
mDSNet-2	**95.31**	**90.73**	**88.73**	**91.05**	**72.41**	**68.74**

and PBM) that use hand-crafted imaging features. This suggests that integrating feature extraction and classifier construction into a unified framework can boost the diagnosis performance. *Besides*, mDSNet based on multi-order representation usually outperforms mDSNet-1st (using first-order features) and CNN (using average-pooling-based features). This implies that using both first-order and second-order features in our mDSNet is more efficient in capturing the disease-image specificity embedded in neuroimages, compared with using only first-order representation and average pooling based features.

5 Discussion

5.1 Effect of Feature Maps

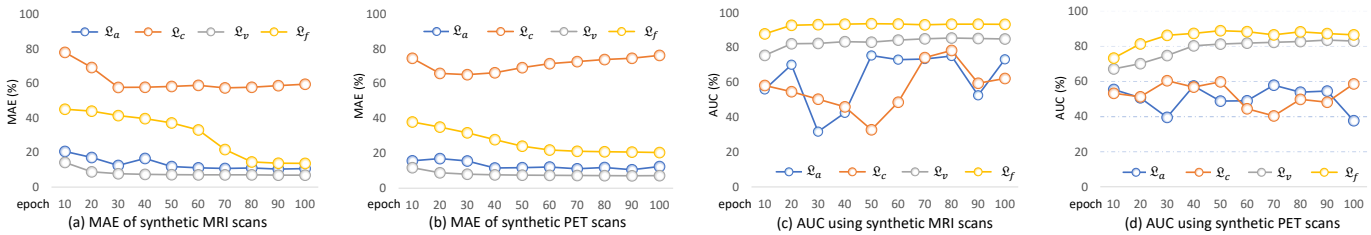
In the feature-consistency component (Fig. 3), we add the feature-consistency constraint at each of five Conv layers. To investigate the effect of our feature-consistency constraint at different layers, we perform an experiment by adding such a constraint at only one single Conv layer, and denote the generated FGAN variants as $\{l_i\}_{i=1}^{5}$. Based on the resulting FGAN models, we can obtain synthetic MRI and PET images. The image quality of these images (regarding SSIM and PSNR) and their performance in AD vs. CN classification (with AUC values marked as *) and pMCI vs. sMCI classification (with AUC values marked as †) are shown in Fig. 5. Fig. 5 reveals that adding the feature-consistency constraint on early layers (e.g., l_1) generally results in better synthesis results (in term of SSIM and PSNR) but lower classification results (in term of AUCs). Also, models using such a constraint at the late layers (e.g. l_5) usually generate the reverse results (i.e., better classification performance but low image quality). Therefore, we add the feature-consistency constraint to feature maps of all five Conv layers to balance the synthesis results and classification results.

5.2 Comparison of Different Losses

To evaluate the influence of different losses used in our image synthesis models (i.e., FGAN, FCycGAN and FVox-GAN), we further train the generative model (with only $\mathbb{G}_A$ and $\mathbb{G}_B$ in Fig. 3) with only one of the following losses: (1) adversarial loss ($\mathfrak{L}_a$), (2) cycle-consistency loss ($\mathfrak{L}_c$), (3) voxel-wise-consistency loss ($\mathfrak{L}_v$), and (4) feature-consistency loss ($\mathfrak{L}_f$). Using each of four different losses, the corresponding generative model can generate synthetic MRI and PET scans. In Fig. 6 (a)-(b), we show the resulting MAE values of these synthetic image along the training epochs. Based on these generated images, we further report the AUC values of our mDSNet in AD vs. CN classification in Fig. 6 (c)-(d), with models trained on ADNI-1 and tested on ADNI-2. Fig. 6 (a)-(b) suggests, when using four different losses, the model that uses only the voxel-wise-consistency loss can produce the most visually realistic image (with respect to MAE values). From Fig. 6 (c)-(d), one can observe that the model using the feature-consistency loss consistently achieves the best classification performance (regarding AUC values), compared with those using the other three losses. Meanwhile, models with the voxel-wise-consistency loss and our feature-consistency loss are more stable compared to those with cycle-consistency loss and adversarial loss. This could be due to the fact that the calculation of the latter two losses relies on updating modules, i.e., another generator and discriminator.

5.3 Enhancing Diagnosis Performance by More Subjects

In the previous experiments, we only use these subjects in ADNI-1 to train our diagnosis model. Since more data may result in better performance, it is possible to use more subjects (e.g., those from ADNI-2) to further improve the performance of the diagnosis performance. Accordingly, we perform another experiment by using both the images in ADNI-1 and ADNI-2 to train the diagnosis model (i.e., mDSNet) but using only complete subjects in ADNI-1 to train the FGAN model. Then we apply the obtained models on the AIBL dataset, with results (denoted as mDSNet-2) reported in Table 5 as well as the original results (denoted by mDSNet). It seems this strategy slightly improves the diagnosis performance (e.g. the AUC score increases from 94.92% to 95.31%). The possible reason could be that more training data make the learned model more general among different sites, thus improving the diagnosis performance.



(a) MAE of synthetic MRI scans (b) MAE of synthetic PET scans (c) AUC using synthetic MRI scans (d) AUC using synthetic PET scans

Fig. 6. Performance of the generative component of FGAN in image synthesis versus the numbers of training epochs. The model was trained on ADNI-1 with only single loss and tested on ADNI-2. Four loss functions were evaluated, including the adversarial loss ($\mathfrak{L}_a$), cycle-consistency loss ($\mathfrak{L}_c$), voxel-wise-consistency loss($\mathfrak{L}_v$), and feature-consistency loss ($\mathfrak{L}_f$).

5.4 More Details of Spatial Cosine Kernel

Following the description of spatial cosine kernel in section 3.2.2, the input of our DSNet is a neuroimage of size $144 \times 176 \times 144$ and the l_2-normalized feature map (spatial representation) is denoted as $\mathbf{u} = (u_1, u_2, \cdots, u_K)$ where $u_k = \frac{v_k}{\|v_k\|_2}$. The spatial cosine kernel (defined in Eq. 8) is

$$\mathbb{C}(\mathbf{u}, \mathbf{w}) = \cos\langle \mathbf{u}, \mathbf{w} \rangle = \frac{\mathbf{u}^{\mathrm{T}}\mathbf{w}}{\|\mathbf{u}\|_2 \|\mathbf{w}\|_2} \tag{17}$$

where $\mathbf{w}$ is the ensemble of hyper-parameters. Due to having the same dimension as $\mathbf{u}$, $\mathbf{w}$ can be partitioned into K elements, i.e., $\mathbf{w} = (w_1, w_2, \cdots, w_K)$, where w_k has same dimension as u_k. Let $\beta_k = \frac{\sqrt{K}\|w_k\|_2}{\|\mathbf{w}\|_2}$, ($\mathsf{E}_{\beta \in \{\beta_1, \beta_2, \cdots, \beta_K\}} \beta^2 = 1$), then the spatial cosine kernel can be rewritten as accumulating the similarity between w_k and the k-th element in a feature map as

$$\mathbb{C}(\mathbf{u}; \mathbf{w}) = \frac{1}{K} \sum_{k=1}^{K} \frac{v_k^{\mathrm{T}}}{\|v_k\|_2} \frac{w_k}{\|w_k\|_2} \beta_k = \frac{1}{K} \sum_{k=1}^{K} \beta_k C(v_k, w_k) \tag{18}$$

Herein, $\mathbb{C}(v_k, w_k) \in [-1, 1]$ is a cosine kernel that indicates the group difference corresponding to the k-th elements. Meanwhile, β_k, directly proportional to the norm of w_k, indicates the contribution coefficient of the k-th cosine kernel to the classification result. Since β_k is learned automatically and implicitly while training DSNet, it is very convenient to capture the disease-relevant patterns by finding the disease-relevant part of the feature map ($\mathbf{U}_d$), in which each u_k corresponds to a large β_k.

Furthermore, it should be noted that the disease-relevant part ($\mathbf{U}_d$), the residual normal part ($\mathbf{U}_r$), and α are mutually dependent but only one constraint in Eq. 5 should be satisfied. Hence, the decomposition of disease-relevant part and residual normal part is not unique. For example, a possible decomposition of a feature map $\mathbf{U}$ could be

$$\begin{cases} \alpha \mathbf{U}_d = \left(\frac{w_1 v_1 w_1}{\|w_1\|_2^2}, \frac{w_2 v_2 w_2}{\|w_2\|_2^2}, \cdots, \frac{w_K v_K w_K}{\|w_K\|_2^2} \right) \\ (1-\alpha)\mathbf{U}_r = \left(v_1 - \frac{w_1 v_1 w_1}{\|w_1\|_2^2}, v_2 - \frac{w_2 v_2 w_2}{\|w_2\|_2^2}, \cdots, v_K - \frac{w_K v_K w_K}{\|w_K\|_2^2} \right) \end{cases} \tag{19}$$

where the disease-relevant component of each pattern is placed to the disease-relevant part, and the disease-irrelevant component of each pattern is placed to the residual normal part. In this case, $\mathbb{C}(\mathbf{U}, \mathbf{w}) = \mathbb{C}(\alpha \mathbf{U}_d, \mathbf{w})$.

Alternatively, we can use a threshold τ to binarize β_k as follows

$$\alpha_k = \begin{cases} 1, \text{if } \ \beta_k > \tau \\ 0, \text{otherwise} \end{cases} \tag{20}$$

and thus decompose $\mathbf{U}$ as

$$\begin{cases} \alpha \mathbf{U}_d = (\alpha_1 v_1, \alpha_2 v_2, \cdots, \alpha_K v_K) \\ (1-\alpha)\mathbf{U}_r = (1-\alpha_1 v_1, (1-\alpha_2) v_2, \cdots, (1-\alpha_K) v_K) \end{cases} \tag{21}$$

It means that we directly place the disease-relevant patterns to disease-relevant parts and place the disease-irrelevant patterns to residual normal part. In this case, $\mathbb{C}(\mathbf{U}, \mathbf{w}) \approx \mathbb{C}(\alpha \mathbf{U}_d, \mathbf{w})$. When considering only the prediction labels, we have

$$\mathbb{C}(\mathbf{U}, \mathbf{w}) \approx \mathbb{C}(\alpha \mathbf{U}_d, \mathbf{w}) \approx \mathbb{C}(\frac{\alpha}{2} \mathbf{U}_d, \mathbf{w}) \tag{22}$$

Namely, if the disease-relevant parts and residual normal parts can be separated, the coefficient α will not affect the predicted labels.

5.5 Statistical Significance Analysis

The major hypothesis in this work is that a generative model with the feature-consistency constraint can generate the synthetic images which are diagnostically similar to real images. It means that the synthetic images and corresponding real images can deliver similar diagnosis of medical conditions. To verify this, we calculated the inter-class averaged dissimilarity (ICAD) of 80 brain regions in synthetic MRI and PET images and displayed them in Fig. 7. The definition of ICAD is as follows.

Suppose the i-th ($i \in \{1, 2, \cdots, N\}$) scan has a feature map $\mathbf{u}_i = (u_{1,i}, u_{2,i}, \cdots, u_{K,i})$ ($\|u_{k,i}\|_2 = 1$) and a class label $y_i \in \{0, 1\}$. The similarity of two feature maps $\mathbf{u}_i$ and $\mathbf{u}_j$ is measured by the spatial cosine kernel,

$$\mathbb{C}(u_i, u_j) = \frac{\mathbf{u}_i^{\mathrm{T}} \mathbf{u}_j}{\|\mathbf{u}_i\|_2 \|\mathbf{u}_j\|_2} = \frac{1}{K} \sum_{k=1}^{K} u_{k,i}^{\mathrm{T}} u_{k,j} \tag{23}$$

Then, the ICAD, denoted by S, can be calculated as

$$S = 1 - \mathop{\mathsf{E}}_{y_i \neq y_j} \mathbb{C}(\mathbf{u}_i, \mathbf{u}_j) = 1 - \frac{1}{K} \sum_{k=1}^{K} \mathop{\mathsf{E}}_{y_i \neq y_j} u_{k,i}^{\mathrm{T}} u_{k,j} \tag{24}$$

A large ICAD value indicates that two classes are easy to be distinguished. Similarly, we can use

$$s_k = 1 - \mathop{\mathsf{E}}_{y_i \neq y_j} u_{k,i}^{\mathrm{T}} u_{k,j} \tag{25}$$

to measure the distinguishability regarding the k-th location.

For each modality in Fig. 7, the top six rows (denoted as $S_1, S_2, \cdots, S_6$ from top to bottom) are the ICAD of the synthetic images generated by GAN, cycGAN, voxGAN, FGAN, FcycGAN, and FvoxGAN, respectively, and the 7^{th} row (denoted as S_7) is the ICAD of real images. The bottom row for each modality depicts the values of $\{\beta_k\}$, which

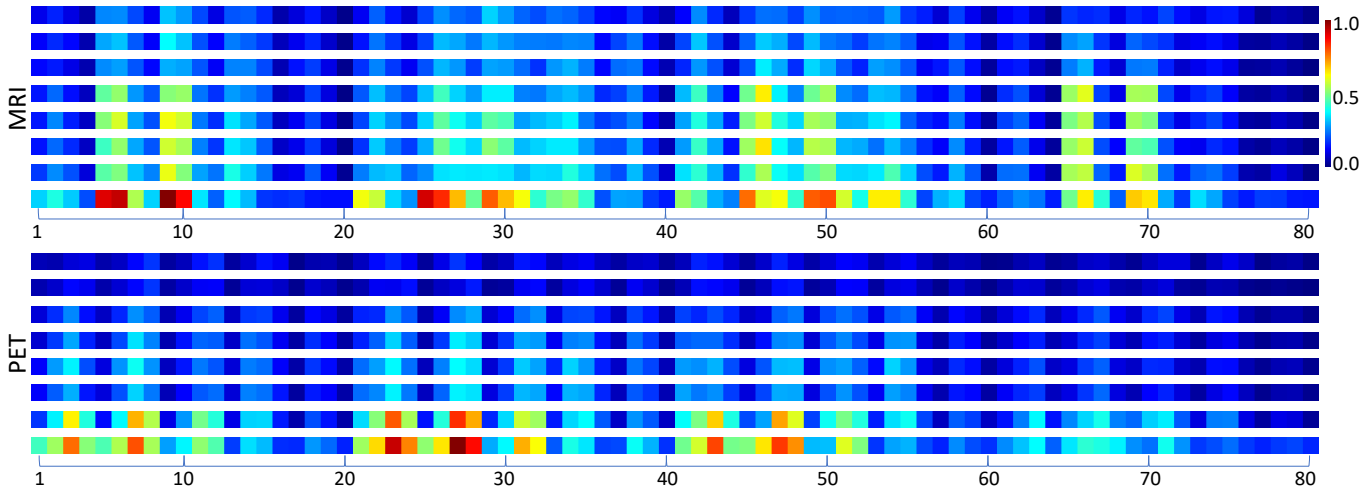


Fig. 7. The inter-class averaged dissimilarities (ICAD) and rescaled contribution coefficients of 80 brain region in synthetic MRI and PET images. For each modality, the top six rows are the ICAD of the synthesized images generated by GAN, cycGAN, voxGAN, FGAN, FcycGAN, and FvoxGAN, respectively, the 7^{th} row is the ICAD of real images, and the 8^{th} row is the rescaled value of contribution coefficient. It shows that the disease-relevant regions are different across two modalities.

have been rescaled to $[0, 1]$, indicating the contribution coefficient of each region to the classification task. Where, a larger value means that the region is more relevant to the classification task. Comparing the 7^{th} row and the bottom row, it suggests that most of the regions with higher inter-class dissimilarities are recognized as more relevant to disease diagnosis. Meanwhile, it reveals that most disease-relevant regions are different across two modalities, which renders the cross-modality image synthesis a challenging task and also makes the feature-consistency constraint a must for any solutions to this task.

We now conduct the statistical significance analysis for our major hypothesis. The null hypotheses (H_0), alternative hypotheses (H_1), and the p-values of H_0 for MRI and for PET are listed in Table 6. It shows that all obtained p-values are smaller than 0.05. Therefore, H_1 is accepted,, which means that a generative model with the feature-consistency constraint can generate images more diagnostically similar to real images.

5.6 "Hallucination" in Generative Unseen Patterns

Currently, it is arguable to first train a generative model on available data for absent data generation and then to use the generated data, together with the available data, to train another model for classification. Especially, the generated data may be affected by the dataset used to train the generative model. For example, if each training image contains a tumor, the data generated by the trained model may also contain a tumor, although it is expected to generate the scan of a normal control. On the other hand, if each training data was acquired from a normal control, it is less impossible for the trained generative model to generate a scan with tumors.

In this study, the abovementioned "hallucination" issue must be addressed when we attempt to use cross-modality image generation to impute missing PET data for the multi-modality based AD diagnosis. The reason lies in the fact that, at an early stage of AD progression, subtle functional changes can be detected by PET long before any structural changes become evident on MRI scans. Thus, when using the MRI scan of a subject with early AD to generate a PET scan, the generative model must be able to produce the disease-related patterns, which do not exist on the input MRI scan. Otherwise, the generated PET scan may not be indeed helpful for disease diagnosis. The major contribution of this work is to address this issue using several tricks, listed below.

- The dataset used to train the generative model contains the cases from all categories (e.g., AD, CN, and MCI). Hence, there will be no "unseen" case in the inference stage.
- We perform image generation in a supervised way. It means that we know the class label of each input MRI scan and accordingly know, statistically, the disease-image-specifics of the PET scans of that class.
- We introduce a feature-consistency constraint to the generative model, encouraging the generated PET scan to preserve the disease-image-specifics for the subsequent diagnosis task. Specifically, the feature-consistency constraint encourages the multi-layer feature maps of a synthetic PET scan (produced by DSNet) and the features of real PET scans of the same class to be consistent. In this way, our FGAN correlates with DSNet, and hence the generated PET scans become consistent with real PET scans from the perspective of classification/diagnosis.

The proposed solution was evaluated in two perspectives: (1) visual similarity and (2) clinical usefulness. The results in Table 2 and Fig. 4 show that the synthetic PET scans generated by our FGAN are similar to real PET scans. Also, the results in Table 3 show that, with our imputed PET scans, the proposed multi-modality based classification model achieves substantially improved performance for AD diagnosis.

Now let us further elaborate the advantage of the proposed solutions over existing generative models like GAN, cycGAN, voxGAN, and condition GAN. The target of this study is to impute missing PET scans and thus to improve the multi-modality diagnosis of AD. However, the distri-

TABLE 6
Hypotheses and obtained p-values in our statistical significance analysis.

Index	Hypotheses H_0	Hypotheses H_1	p-values (MRI)	**p**-values (PET)
1	$(S_1 - S_7)^2 <= (S_4 - S_7)^2$	$(S_1 - S_7)^2 > (S_4 - S_7)^2$	1.64×10^{-4}	7.63×10^{-11}
2	$(S_2 - S_7)^2 <= (S_5 - S_7)^2$	$(S_2 - S_7)^2 > (S_5 - S_7)^2$	4.95×10^{-6}	9.92×10^{-11}
3	$(S_3 - S_7)^2 <= (S_6 - S_7)^2$	$(S_3 - S_7)^2 > (S_6 - S_7)^2$	9.10×10^{-7}	1.00×10^{-6}

bution matching constraints used in existing generative models may not be able to preserve the discriminative information in the generated images. Although the l_1 (MAE) loss encourages the pixel/voxel-wise consistency, it remains incapable of preserving sufficient discriminative information, as evidenced in our supporting experiments and the experiments in [55]. The potential reason that these generative models are not suitable for this study is that both distribution matching constraints and pixel-wise-consistency are class-independent. Take the pixel-wise-consistency for example, if the input is an MRI scan of an AD subject, the output PET scan is forced to be pixel-wise-consistent with training PET scans, which are from both AD subjects and normal controls. Although the generated PET scan looks like a real PET scan, it may not contain enough disease-image-specifics, i.e., the pathological patterns of AD that can be observed using PET. In contrast, the proposed FGAN generates images in a supervised way, in which the feature-consistency constraint encourages the multi-layer feature maps of a generated PET scan (produced by DSNet) to be consistent with the features of real PET scans of the same class. Hence, the PET scan generated by our FGAN contains rich discriminative information and can help the subsequent diagnosis task.

5.7 Potential Applications in other Scenarios

In practice, applications of image-image translation (besides cross-modality image translation) have been increasingly used in many vision applications, such as security surveillance and autonomous driving. In these applications, sometimes we may encounter a similar issue of requiring a specific modality of data. Although the proposed DSDL framework was developed for neuroimage synthesis and AD diagnosis, the ideas of supervised image generation and feature-consistency constraint are generic and can be extended to these applications of task-specific image-image translation.

Various image-image translation tasks, including "edges to photo", "aerial to map", "day to night", and "BW to color", have been summarized in [28]. However, using only the distribution-match constraint or pixel-wise-consistency constraint may not handle it well when we want to use the generated images for a classification purpose. For instance, on the task of "BW to color" for flower classification, the color information, which plays a pivotal role in distinguishing a flower species from others, should be generated during the translation from a grayscale image to its color version. In this scenario, we can apply our proposed feature-consistency constraint to encourage the generated color images to preserve the discriminative information for flower classification (as Section 7 in the *Supplemental Materials*).

5.8 Limitations and Future Work

The proposed DSDL framework has three major limitations.

First, the spatial cosine kernel used in this model has a fixed stride (i.e., 32 voxels along each axis). Hence, it can only roughly capture disease-specific regions with a fixed size of $32 \times 32 \times 32$ voxels. To capture more precise disease-specific regions, hierarchical structures or multi-scale features will be investigated in our future work.

Second, PET protocols used for building these three databases are very different, particularly those for AIBL. The proposed method has no mechanism to handle this issue. As a result, it is hard to use the PET scans in AIBL, and we have to use synthetic PET scans for AIBL. Hence, data harmonization / adaptation techniques [56], [57] will be studied in our future work to capture unified features regardless of scanning protocols.

Third, the pre-processing pipeline used for this study is purely hand-crafted. It relies heavily on the experience of operators and can hardly be optimized for unseen datasets. In our further work, we plan to embed the pre-processing steps into the target task to avoid the devastating effect caused by inappropriate pre-processing.

6 Conclusion

We proposed a disease-image-specific deep learning framework for task-oriented neuroimage synthesis based on incomplete multi-modality data, where a diagnosis network is employed to provide disease-image specificity to an image synthesis network. Specifically, we designed a single-modality disease-image-specific network (DSNet) trained on whole-brain images to implicitly capture the disease-relevant information conveyed in MRI and PET. We then developed a feature-consistency generative adversarial network (FGAN) to synthesize missing neuroimages, by encouraging that feature maps (generated by DSNet) of each synthetic image and its respective real image to be consistent. We further proposed a multi-modality DSNet (mDSNet) for disease diagnosis using complete (after imputation) MRI and PET scans. Experiments on three public datasets demonstrate that our method can generate reasonable neuroimages and achieve the state-of-the-art performance in AD identification and MCI conversion prediction.

References

[1] J. Zhou, L. Yuan, J. Liu, and J. Ye, "A multi-task learning formulation for predicting disease progression," in *ACM SIGKDD*. ACM, 2011, pp. 814–822.

[2] 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.

[3] S. Xiang, L. Yuan, W. Fan, Y. Wang, P. M. Thompson, and J. Ye, "Bi-level multi-source learning for heterogeneous block-wise missing data," *NeuroImage*, vol. 102, pp. 192–206, 2014.

[4] 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. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 455–463.

[5] C. R. J. Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.

[6] V. D. Calhoun and J. Sui, "Multimodal fusion of brain imaging data: A key to finding the missing link(s) in complex mental illness," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 1, no. 3, pp. 230–244, 2016.

[7] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, and ADNI, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.

[8] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.

[9] 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.

[10] M. Wang, D. Zhang, D. Shen, and M. Liu, "Multi-task exclusive relationship learning for Alzheimer's disease progression prediction with longitudinal data," *Med. Image Anal.*, vol. 53, pp. 111–122, 2019.

[11] H. Sun, R. Mehta, H. H. Zhou, Z. Huang, S. C. Johnson, V. Prabhakaran, and V. Singh, "Dual-glow: Conditional flow-based generative model for modality transfer," *arXiv preprint arXiv:1908.08074*, 2019.

[12] R. Parker, *Missing Data Problems in Machine Learning*. Saarbrücken, Germany, Germany: VDM Verlag, 2010.

[13] K.-H. Thung, C.-Y. Wee, P.-T. Yap, and D. Shen, "Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion," *NeuroImage*, vol. 91, pp. 386–400, 2014.

[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] J. M. Wolterink, K. Kamnitsas, C. Ledig, and I. Išgum, "Generative adversarial networks and adversarial methods in biomedical image analysis," *arXiv preprint arXiv:1810.10352*, 2018.

[16] S. Kazeminia, C. Baur, A. Kuijper, B. Van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "GANs for medical image analysis," *arXiv preprint arXiv:1809.06222*, 2018.

[17] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, April 2020.

[18] B. Cheng, M. Liu, D. Zhang, B. C. Munsell, and D. Shen, "Domain transfer learning for MCI conversion prediction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1805–1817, 2015.

[19] C. Wachinger, D. H. Salat, M. Weiner, and M. Reuter, "Whole-brain analysis reveals increased neuroanatomical asymmetries in dementia for hippocampus and amygdala," *Brain*, vol. 139, no. 12, pp. 3253–3266, 2016.

[20] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, 2011.

[21] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, 2018.

[22] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, and S. Ji, "Deep learning based imaging data completion for improved brain disease diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2014, pp. 305–312.

[23] T. Huynh, Y. Gao, J. Kang, L. Wang, P. Zhang, J. Lian, and D. Shen, "Estimating CT image from MRI data using structured random forest and auto-context model," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 174–183, 2015.

[24] A. Jog, S. Roy, A. Carass, and J. L. Prince, "Magnetic resonance image synthesis through patch regression," in *Proc. Int. Symposium on Biomedical Imaging*. IEEE, 2013, pp. 350–353.

[25] S. Bano, M. Asad, A. E. Fetit, and I. Rekik, "XmoNet: A fully convolutional network for cross-modality MR image inference," in *International Workshop on Predictive Intelligence In Medicine*. Springer, 2018, pp. 129–137.

[26] I. Goodfellow, J. Pougetabadie, M. Mirza, B. Xu, D. Wardefarley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[27] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2172–2180.

[28] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1125–1134.

[29] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Machine Learning*, 2017, pp. 214–223.

[30] L. Yu, W. Zhang, J. Wang, and Y. Yu, "SeqGAN: Sequence generative adversarial nets with policy gradient," in *Proc. AAAI Conf. Artificial Intelligence*, 2017, pp. 2852–2858.

[31] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 469–477.

[32] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc.IEEE Int. Conf. Computer Vision*, 2017, pp. 2223–2232.

[33] T. Qiao, J. Zhang, D. Xu, and D. Tao, "Mirrorgan: Learning text-to-image generation by redescription," *arXiv preprint arXiv:1903.05854*, 2019.

[34] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *arXiv preprint arXiv:1809.07294*, 2018.

[35] A. Ben-Cohen, E. Klang, S. P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M. M. Amitai, and H. Greenspan, "Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection," *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 186–194, 2019.

[36] S. Olut, Y. H. Sahin, U. Demir, and G. Unal, "Generative adversarial training for MRA image synthesis using multi-contrast MRI," in *International Workshop on PRedictive Intelligence In MEdicine*. Springer, 2018, pp. 147–154.

[37] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, Z. Xu, and J. Prince, "Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 174–182.

[38] J. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V. De La Sayette, and F. Eustache, "In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease," *NeuroImage*, vol. 14, no. 2, pp. 298–309, 2001.

[39] R. Cui and M. Liu, "Hippocampus analysis by combination of 3D densenet and shapes for Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Informat.*, 2018.

[40] M. Liu, D. Zhang, and D. Shen, "Ensemble sparse classification of Alzheimer's disease," *NeuroImage*, vol. 60, no. 2, pp. 1106–1116, 2012.

[41] ——, "Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis," *Human Brain Mapping*, vol. 35, no. 4, pp. 1305–1319, 2014.

[42] J. Zhang, M. Liu, L. An, Y. Gao, and D. Shen, "Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 6, pp. 1607–1616, 2017.

[43] J. Zhang, Y. Gao, Y. Gao, B. C. Munsell, and D. Shen, "Detecting anatomical landmarks for fast Alzheimer's disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2524–2533, 2016.

[44] G. Li, M. Liu, Q. Sun, D. Shen, and L. Wang, "Early diagnosis of autism disease by multi-channel CNNs," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Springer, 2018, pp. 303–309.

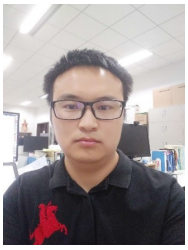
[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[47] K. A. Ellis, A. I. Bush, D. Darby, D. De Fazio, J. Foster, P. Hudson, N. T. Lautenschlager, N. Lenzo, R. N. Martins, P. Maruff *et al.*, "The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease," *International Psychogeriatrics*, vol. 21, no. 4, pp. 672–687, 2009.

[48] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012.

[49] F. Kurth, C. Gaser, and E. Luders, "A 12-step user guide for analyzing voxel-wise gray matter asymmetries in statistical parametric mapping (SPM)," *Nature Protocols*, vol. 10, no. 2, p. 293, 2015.
[50] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *ICPR*, 2010, pp. 2366–2369.
[51] H. Rusinek, M. J. De Leon, A. E. George, L. A. Stylopoulos, R. Chandra, G. Smith, T. Rand, M. Mourino, and H. Kowalski, "Alzheimer disease: Measuring loss of cerebral gray matter with MR imaging," *Radiology*, vol. 178, no. 1, pp. 109–114, 1991.
[52] P. Coupé, S. F. Eskildsen, J. V. Manjón, V. S. Fonov, and D. L. Collins, "Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease," *NeuroImage*, vol. 59, no. 4, pp. 3736–3747, 2012.
[53] O. O. Koyejo, N. Natarajan, P. K. Ravikumar, and I. S. Dhillon, "Consistent binary classification with generalized performance metrics," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2744–2752.
[54] 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.
[55] J. P. Cohen, M. Luck, and S. Honari, "Distribution matching losses can hallucinate features in medical image translation," in *Proc. Int. Conf. Med. Image Comput. Computer Assisted Intervention*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds. Cham: Springer International Publishing, 2018, pp. 529–536.
[56] M. Wang, D. Zhang, J. Huang, P.-T. Yap, D. Shen, and M. Liu, "Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 644–655, March 2020.
[57] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1640–1650.



Yongsheng Pan (S'18) received his B.E. degree in computer science and technology, in 2015, from Northwestern Polytechnical University (NPU), Xi'an, China. He is currently working toward the Ph.D. degree at the School of Computer Science and Engineering, Northwestern Polytechnical University (NPU). His research areas include neuroimage analysis, computer vision and machine learning.

Mingxia Liu received her B.S. and M.S. degrees from Shandong Normal University, Shandong, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2015. She is a Senior Member of IEEE. Her current research interests include machine learning, pattern recognition, and medical image analysis.

Yong Xia (S'05-M'08) received his B.E., M.E., and Ph.D. degrees in computer science and technology from Northwestern Polytechnical University (NPU), Xi'an, China, in 2001, 2004, and 2007, respectively. He is currently a Professor at the School of Computer Science and Engineering, NPU. His research interests include medical image analysis, computer-aided diagnosis, pattern recognition, machine learning, and data mining.

Dinggang Shen (SM'07-F'18) is Jeffrey Houpt Distinguished Investigator, and a Professor of Radiology, Biomedical Research Imaging Center (BRIC), Computer Science, and Biomedical Engineering in the University of North Carolina at Chapel Hill (UNC-CH). He is currently directing the Center for Image Analysis and Informatics, the Image Display, Enhancement, and Analysis (IDEA) Lab in the Department of Radiology, and also the medical image analysis core in the BRIC. He was a tenure-track assistant professor in the University of Pennsylvanian (UPenn) and a faculty member in the Johns Hopkins University. His research interests include medical image analysis, computer vision, and pattern recognition. He has published more than 1000 papers in the international journals and conference proceedings. He serves as an editorial board member for eight international journals. He has also served in the Board of Directors, the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society, in 2012-2015, and will be General Chair for MICCAI 2019. He is Fellow of IEEE, Fellow of The American Institute for Medical and Biological Engineering (AIMBE), and Fellow of The International Association for Pattern Recognition (IAPR).

Disease-image-specific Learning for Diagnosis-oriented Neuroimage Synthesis with Incomplete Multi-Modality Data – *Supplementary Materials*

Yongsheng Pan, Mingxia Liu, Yong Xia, and Dinggang Shen, *Fellow, IEEE*

In what follows, we first show the advantages of the proposed method comparing to our previous works in Section 1 and more details about the datasets we used in Section 2. Then, we present more discussion on hyper-parameter settings in Section 3, including the batch sizes, training epochs, and backbones. We also explain more details about the advantages of our DSDL framework (Sections 4 and 5) and more samples of synthetic neuroimages (Section 6). In addition, we discussed the potential applications in other scenarios (Section 7).

1 Technical Novelty

Due to its ability to provide complementary structural and functional information, multi-modal neuroimaging (e.g., MRI and PET) has been commonly used for the diagnosis of neurodegenerative disorders such as the Alzheimer's disease (AD). However, the missing data problem is almost inevitable in clinical practice due to various reasons, e.g., PET scanning may be rejected by some subjects due to high cost or concern of radioactive exposure. Many studies on multi-modal neuroimages simply discard the subjects with missing PET, leading to a significant decrease of the number of training subjects. However, deep learning-based diagnostic models, which have become the de facto standard in medical image analysis, are prone to over-fit the training dataset and hence exhibit unreliable performance, if the training dataset is small.

To solve the missing data problem, we attempted, in our previous works (e.g., [1]), to impute the missing PET data using the available MRI data based on the observation that there probably exists underlying relevance between the images acquired from the same subject but using different modalities. We first resorted 3D Cycle-consistency Generative Adversarial Networks (3D-cGAN) to learn the bi-directional mapping between relevant image domains (i.e., across PET and MRI), where the cycle-consistency loss is used to capture their probable underlying relationship. Then, we developed a landmark-based multi-modal multi-instance learning (LM3IL) model to use the complete MRI and PET data (i.e., real or synthetic PET + real MRI) for AD diagnosis and mild cognitive impairment (MCI) conversion prediction. This previous work achieve good performance because (1) we directly imputed missing PET scans to almost double the number of training subjects, leading to a more reliably-learned diagnostic model; (2) we performed the classification only on the patches around disease-related landmarks, which are pre-defined manually.

However, this previous work is an early attempt to learn data imputation and classification under a two-stage framework in a data-driven manner and has three limitations. First, the image synthesis and disease diagnosis are treated as two standalone tasks, and hence the difference of specificities conveyed by two modalities is ignored. Second, the cycle consistency we used is a weak constraint to preserve the disease information, since it only encourages pixel/voxel consistency after two transformations (i.e., transformed through two synthesis models), not encouraging the consistency of disease-relevant information. Third, the classification performance relies highly on

- *Y. Pan and Y. Xia were partially supported by the National Natural Science Foundation of China under Grant 61771397, the Science and Technology Innovation Committee of Shenzhen Municipality, China, under Grant JCYJ20180306171334997, and the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grant CX201835. M. Liu and D. Shen were partially supported by NIH grant (No. AG041721). Corresponding authors: Yong Xia, Mingxia Liu, and Dinggang Shen.*

- *Y. Pan and Y. Xia are with School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China. Y. Pan is also with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. (E-mail: {yspan@mail.,yxia@}nwpu.edu.cn) M. Liu and D. Shen are with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. D. Shen is also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea. (E-mails: mxliu@med.unc.edu, Dinggang.Shen@gmail.com)*

This work was finished when Y. Pan was visiting the University of North Carolina at Chapel Hill. Part of the data used in this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The investigators within the ADNI contributed to the design and implementation of ADNI and provided data but did not participate in analysis or writing of this article.

the precision of landmarks. However, no landmark set is universally recognized as precise and comprehensive, since the pathological changes can be subtle in the early course of the disease and there can be some overlap with other neurodegenerative types.

To address these issues, we proposed a disease-image-specific deep learning (DSDL) framework for *joint neuroimage synthesis and disease diagnosis* using incomplete multi-modality neuroimages. Specifically, we first designed a Disease-image-Specific Network (DSNet) with a spatial cosine module to implicitly model the disease-image specificity, and then developed a Feature-consistency Generative Adversarial Network (FGAN) to impute missing neuroimages. During the image synthesis, we aim to preserve the disease-image-specific information via using the feature-consistency constraint, which encourages the multi-layer feature maps (generated by DSNet) of a synthetic image and its corresponding real image to be consistent. For instance, at an early stage of the AD process, subtle physiological changes can be detected by PET long before any changes are apparent on MRI. In this case, when imputing the missing PET scan based on the available MRI scan, our previous method cannot generate the physiological changes since they are not apparent on the MRI scan. However, with the help of the proposed the feature-consistency constraint, our FGAN can generate the synthetic PET scan that contains the physiological changes via implicitly learning from same-class PET scans in the training dataset. In this way, our FGAN is correlated with DSNet, leading the missing neuroimages to be imputed in a diagnosis-oriented manner. Hence the synthetic neuroimages are more consistent with real neuroimages from a diagnostic point of view.

2 Dataset Introduction

ADNI: The Alzheimer's Disease Neuroimaging Initiative (ADNI) study [2] is the most widely open-access study for Alzheimer's Disease (AD) and is jointly funded by the National Institutes of Health (NIH) and industry via the Foundation for the NIH. It is a longitudinal multi-site observational study of elderly individuals with normal cognition, mild cognitive impairment (MCI), or AD. Healthy elderly controls are sampled at 0, 6, 12, 24, and 36 months. Subjects with MCI are sampled at 0, 6, 12, 18, 24, and 36 months. AD subjects are sampled at 0, 6, 12, and 24 months. The follow-up study assesses how well the information (alone or in combination) obtained from MRI, 18F- FDG PET, etc., can measure the disease progression in three groups of elderly subjects mentioned above.

- The ADNI-1 phase was launched in October 2004 and has lasted for 5 years, during which more than 800 subjects have been collected. All subjects are from multiple participating sites in North America (United States and Canada). All subjects were scanned with 1.5 T MRI at each time point, and half subjects were scanned with FDG PET.
- The ADNI-2 phase was launched in September 2011 and has lasted for 5 years. Except for the subjects in ADNI-1, more than 800 additional subjects have been collected during this phase. In ADNI-2, all subjects were scanned with 3T MRI using similar T1-weighted imaging parameters to those used in ADNI-1. About half subjects were scanned with FDG PET using similar parameters to those for ADNI-1.

AIBL: The Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) [3] is a study to discover which biomarkers, cognitive characteristics, and health and lifestyle factors determine subsequent development of symptomatic AD. It is a 4.5-year prospective longitudinal study of cognition, launched in November 2006, which is the largest study of its kind in Australia. This study collected more than 1000 subjects with AD, MCI, and healthy volunteers from two sites (i.e., Perth and Melbourne). In AIBL, MRI was performed at 1.5T/3T MRI with similar T1-weighted imaging parameters to those used in ADNI, but PET was performed with the parameters different from those used in ADNI.

For more details, we listed the statistics of the protocols used for acquiring the baseline MRI and PET scans in Table SII and Table SIII, respectively in the Supplementary Materials. The ADNI organization requires a uniform protocol for data acquirement. As for the AIBL database, the MRI scanning protocols are similar to the one used for ADNI (ADNI-1 and ADNI-2), but the PET scanning protocols are slightly different. For example, the slice thickness of most PET scans in AIBL is either 2.0 mm or 3.0 mm, which is different from that of ADNI PET scans, and the radioisotope for most PET scans in ADNI is F-18 but for 43% PET scans in AIBL is C-11. Meanwhile, most ADNI contributors provide data for both ADNI1 and ADNI2 cohorts. Since similar scanning protocols and the same imaging site may lead to less-diverse imaging quality, the model learned on ADNI-1 can adapt to the data in ADNI-2. To cope with the quality diversity of PET scans in AIBL, we directly used the synthetic PET for this study.

3 Hyper-parameter Discussion

3.1 Influence of Batch Size

Due to the limitation of GPU memory, we cannot use a large batch size. To assess the impact of this hyperparameter on the model's performance, we set different batch sizes and performed the AD *vs.* CN classification task again using different batch sizes. It shows that, when setting the batch size to 1, 2, 3, and 4, the proposed model achieves an AUC of 0.9631, 0.9622, 0.9588, and 0.9596, respectively. The results indicate that the performance of our model is not sensitive to the batch size. Hence, considering the performance and complexity, we empirically set the batch size to 1.

3.2 Influence of Maximum Number of Epochs

In Fig. 6, we reported associated performance of image synthesis obtained after training FGAN different epochs. It shows that the performance has a broad dynamic range and has tolerable changes when the number of epochs approaches to 100. In Fig. S1, we further plotted the training loss and test loss of DSNet during the first 100 epochs. It reveals that the test loss of DSNet become relatively stable after 40 epochs. Therefore, we empirically set the maximum number of epochs to 100 and 40 for FGAN and DSNet, respectively.

TABLE S1
Protocols of baseline MRI scans in ADNI-1, ADNI-2, and AIBL.

Protocol	Dataset	Parameter (Number of Subjects)
Acquisition Plane	ADNI-1	SAGITTAL (844)
	ADNI-2	SAGITTAL (846)
	AIBL	SAGITTAL (665)
Slice Thickness	ADNI-1	1.2 (844)
	ADNI-2	1.2 (846)
	AIBL	1.0 (137), 1.2 (528)
Matrix Z	ADNI-1	146.0 (1), 160.0 (332), 162.0 (1), 166.0 (284), 170.0 (90), 180.0 (124), 184.0 (12)
	ADNI-2	170.0 (166), 176.0 (466), 196.0 (214)
	AIBL	160.0 (527), 170.0 (137), 159.0 (1), 153.0 (1)
Acquisition Type	ADNI-1	3D (844)
	ADNI-2	3D (846)
	AIBL	3D (665)
Manufacturer	ADNI-1	GE Medical Systems (410), Philips Medical Systems (103), SIEMENS (331)
	ADNI-2	GE Medical Systems (214), Philips Medical Systems (165), SIEMENS (466), Philips Healthcare (1)
	AIBL	SIEMENS (665)
Mfg Model	ADNI-1	Achieva (13), Intera (69), Intera Achieva (6), Avanto (64), GENESIS_SIGNA (78), Gyroscan Intera (12), Intera (3), SIGNA EXCITE (307), SIGNA HDx (25), Sonata (99), SonataVision (7), Symphony (161)
	ADNI-2	Achieva (108), Discovery MR750 (90), Discovery MR750w (9), GEMINI (14), Intera (28), SIGNA HDx (9), SIGNA HDxt (9), Skyra (52), TrioTim (279), Verio (135)
	AIBL	Avanto (116), TrioTim (426), Verio (123)
Field Strength	ADNI-1	1.5 (844)
	ADNI-2	3.0 (846)
	AIBL	1.5 (116), 3.0 (549)
Weighting	ADNI-1	T1 (844)
	ADNI-2	T1 (846)
	AIBL	T1 (665)

TABLE S2
Protocols of baseline PET scans in ADNI-1, ADNI-2, and AIBL.

Protocol	Dataset	Parameter (Number of Subjects)
Slice Thickness	ADNI-1	1.2 (51), 2.0 (71), 2.4 (110), 3.3 (21), 3.4 (53), 4.3 (93)
	ADNI-2	1.2 (41), 2.0 (254), 2.4 (157), 3.3 (156), 3.4 (20), 4.3 (45)
	AIBL	2.0 (355), 3.0 (142), 3.3 (109)
Manufacturer	ADNI-1	CPS (59), GE MEDICAL SYSTMS (67), GEMS (47), Philips Medical Systems (39), Siemens ECAT (51) Siemens/CTI (136)
	ADNI-2	CPS (53), GE MEDICAL SYSTMS (184), GEMS (17), Philips Medical Systems (91), Siemens (123), Siemens ECAT (41), Siemens/CTI (164)
	AIBL	GE MEDICAL SYSTMS (109), SIEMENS (142), Philips Medical Systems (355)
Mfg Model	ADNI-1	ACCEL (17), Advance (47), Allegro Body (C) (10), Discovery HR (5), Discovery LS (46), Discovery RX (3), Discovery ST (13), EXACT (ACS 1) (3), EXACT (ACS 2) (6), G-PET Brain (C), Gemini TF(C) (4), Guardian Body(C) (17), HR+ (110), HRRT (51), LSO PET/CT (5), LSO PET/CT (Pico electronics) (22), LSO PET/CT HI-REZ (32)
	ADNI-2	1093 (14), 1094 (51), ACCEL (7), Advance (17), Allegro Body (C) (3), Biograph64 (22), Discovery 600 (6), Discovery LS (28), Discovery RX (11), Discovery ST (45), Discovery STE (94), GEMINI TF Big Bore (14) , GEMINI TF TOF 16 (30) , GEMINI TF TOF 64 (20), Guardian Body(C) (12), HR+ (157), HRRT (41), LSO PET/CT (Pico electronics) (13), LSO PET/CT HI-REZ (72), SOMATOM Definition AS_mCT (4)
	AIBL	Allegro Body (C) (353), Biograph128 (123), Biograph128_mCT (19), Discovery 710 (109), GEMINI TF TOF 64 (2)
Radioisotope	ADNI-1	C-11 (8), F-18 (391)
	ADNI-2	F-18 (673)
	AIBL	C-11 (262), F-18 (344)
Radio Pharmaceutical	ADNI-1	11C-PIB (8), 18F-FDG (391)
	ADNI-2	18F-AV45 (284), 18F-FDG (389)
	AIBL	11C-PIB (61), 18F-Flutemetamol (48), Flutemetamol (142), Other (355)
Frames	ADNI-1	1.0 (70), 4.0 (2), 5.0 (1), 6.0 (252), 7.0 (40), 12.0 (1), 15.0 (2), 17.0 (1), 27.0 (2), 28.0 (1), 30.0 (1), 33.0 (19), 38.0 (1), 39.0 (2)
	ADNI-2	1.0 (83), 2.0 (2), 4.0 (247), 6.0 (339), 16.0 (2)
	AIBL	1.0 (498), 4.0 (49), 6.0 (59)

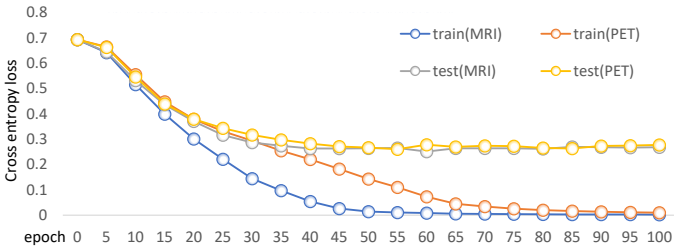


Fig. S1. Training and testing losses for image classification versus different numbers of training epochs.

3.3 Influence of Generative Model Backbone

In general, there are two alternative backbone structures, i.e., Decoder-Encoder (DE) backbone and UNet backbone [4], available for the generative components. The structures of DE and UNet were displayed in Fig. S2 (left), where the major differences are the skip connections and feature concatenation used in UNet. Fig. S2 (right) shows the performance metrics of AD $vs.$ CN classification obtained by using either the MRI scan or PET scan synthesized by either DE backbone or UNet backbone. It reveals that these two structures achieve similar results, e.g., the AUC values of

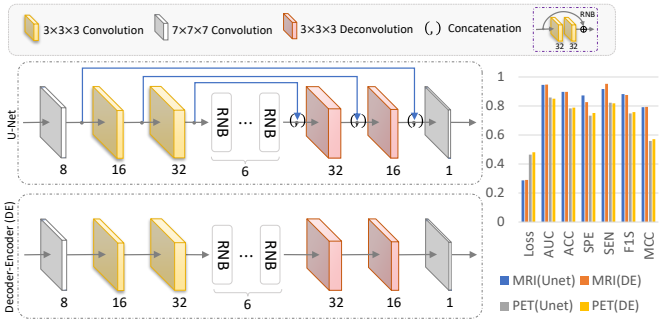


Fig. S2. The structure (left) of Decoder-Encoder (DE) and UNet backbone are displayed in the left while the metric values on image classification of scans synthesized by DE and UNet for MRI and PET modality, respectively.

MRI scans generated by UNet backbone and DE backbone are 0.9459 and0.9475, respectively. The reason may lie in the fact that the spatial consistency between a pair of MRI and PET scans are not strictly held. Following the principle of Occam's razor, we use the DE backbone, which is simpler than the UNet backbone, for this study.

4 Advantages of Synthetic Data

Compared to "unseen-modality classification methods", namely, using different networks for feature extraction and a shared network for classification, our model has the following three major advantages.

- For each incomplete case, the information of the missing modality is transformed not only from the other modality, but also from the complete cases in the training set via training the image synthesize model. For instance, the missing PET image of an incomplete AD subject is synthesized based on both the MRI image of that subject and the PET images of AD subjects in the training set, since FGAN used for image synthesis was trained on both MRI and PET images in the training set. Therefore, even if a disease specificity has not become evidenced on the structural MRI image, our model can still "guess" it according to the PET images of AD cases in the training set, given that the image synthesis and classification are performed in a unified framework in our solution.
- Suppose our model, in the worst case, cannot benefit from the complete cases in the training set at all. From the information point of view, our model uses the fake "multi-modality" information, which is only the information of the available modality. In this case, our model becomes a "real unseen-modality classification method".
- The proposed multi-modality classification model is much simpler than the "real unseen-modality classification" model, in which one or two of the feature extraction branches is active in each training epoch. Alternatively, if training independently two models for feature extraction and another model for classification, the system cannot benefit from the unique advantage of "learning image representation and classification in a unified framework for simultaneous optimization".

5 Advantages of Cascade Model

In our experiments, we followed the cascade strategy to train DSNet and FGAN components rather than training then together in an end-to-end strategy. There are four major reasons for using the cascade model, instead of an end-to-end one.

- *First*, the feature-consistency constraint is defined in the feature extraction part, and hence is varying during training DSNet. If jointly training DSNet and FGAN in an end-to-end model, the less-optimal feature-s obtained in the early training stage will undermine the convergence of FGAN.
- *Second*, we design the feature-consistency constraint to capture the disease-specific information in real data and then use the information to guide FGAN to synthesize real-like scans. Thus, it derives the information only from existing real data and is independent to training FGAN. Therefore, jointly or step-wise iteratively training DSNet and FGAN will not capture more disease-specific information, but requires more GPU memory and computational resources.
- *Third*, we need a feature extraction module to measure the effect of feature-consistency constraint on FGAN. Thus, we utilize the well-trained DSNet with freezing weights while training FGAN.
- *Finally*, as a classification model, DSNet can hardly converge synchronously with FGAN. Hence, if we jointly train DSNet and FGAN in an end-to-end way, the asynchronous convergence of both components will lead one component to under-fitting or the other component to over-fitting.

6 More disease-related Visual Examples

Besides samples in Fig. 4 in the main text, we supplied more views of high-resolution examples in Figs. S3-S5, where four typical subjects (Roster IDs: 4386, 4765, 4997, and 4417) in ADNI-2 are shown. Basically, it can be seen from Fig. S3 (sagittal MRI views) that the sizes of ventricle in the 4^{th}, 5^{th}, 6^{th} columns are more like the ground truth (7^{th} column) than the 1^{st}, 2^{nd}, and 3^{rd} columns. It suggests the feature-consistency constraint can coexist with other consistency constraint, with minimal impact of voxel-wise-consistency and cycle-consistency constraints.

Taking into account the results listed in Table 2 (main text), the conclusion can be supported that the feature-consistency constraint can help preserve more diagnosis information during the transformation between two modalities without dropping the visual quality. However, there may be no metric that can cover all concerned aspects. Therefore, we still suggest considering the suitable choice of constraints for a specific task, e.g., using the adversarial loss to keep the distribution (texture, structure) similarity, using the pixel-wise-consistency constraint to keep intensity consistency, and using the feature-consistency constraint to keep diagnosis consistency.

7 Application in Another Scenario

We applied the proposed DSDL framework to a natural image classification scenario: using grayscale images to

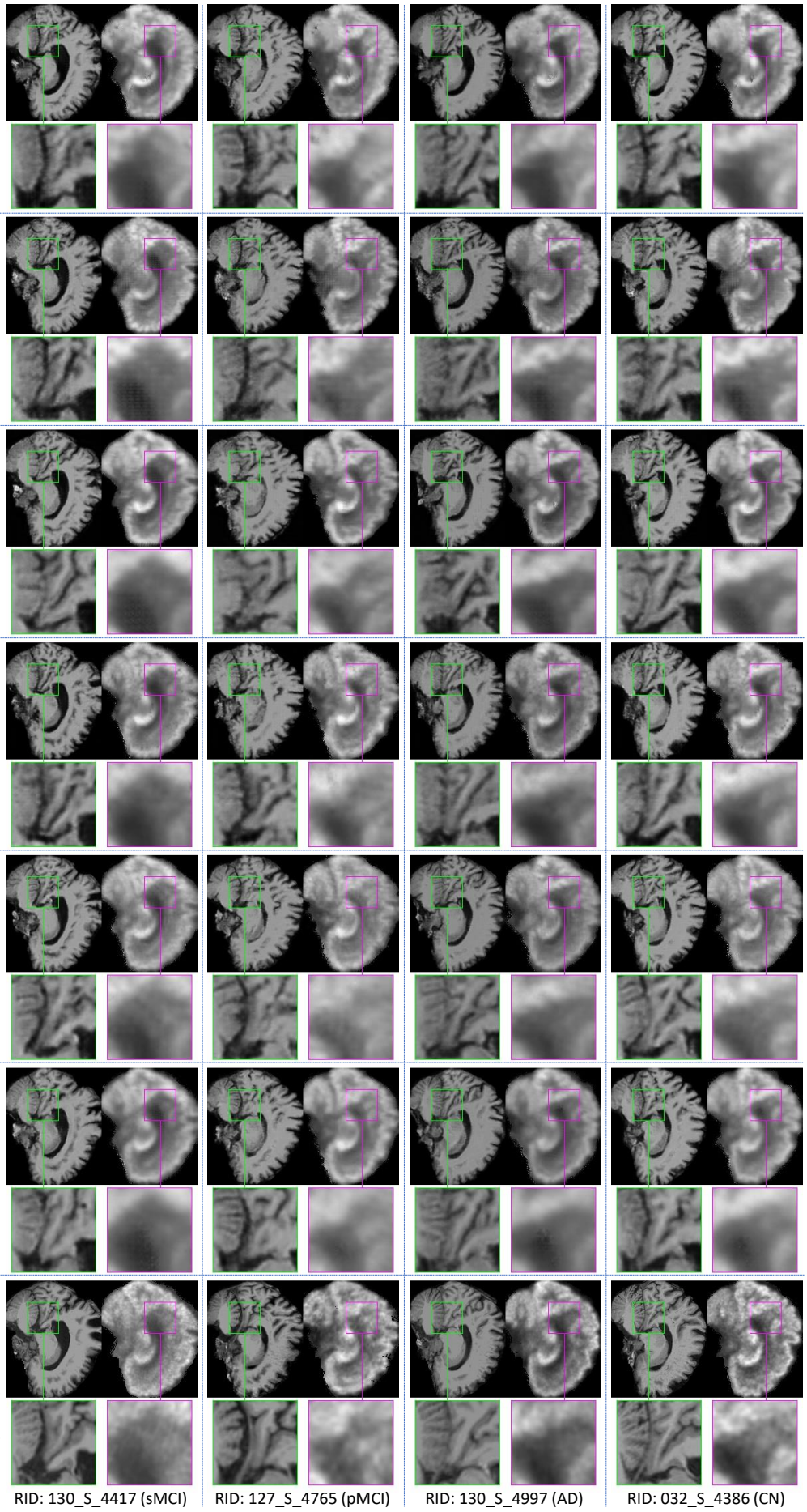


Fig. S3. Sagittal views of PET and MRI scans synthesized by six methods for four typical subjects (Roster ID: 4386, 4765, 4997, and 4417) in ADNI-2, along with their corresponding ground-truth images. All six image synthesis models are trained on ADNI-1.

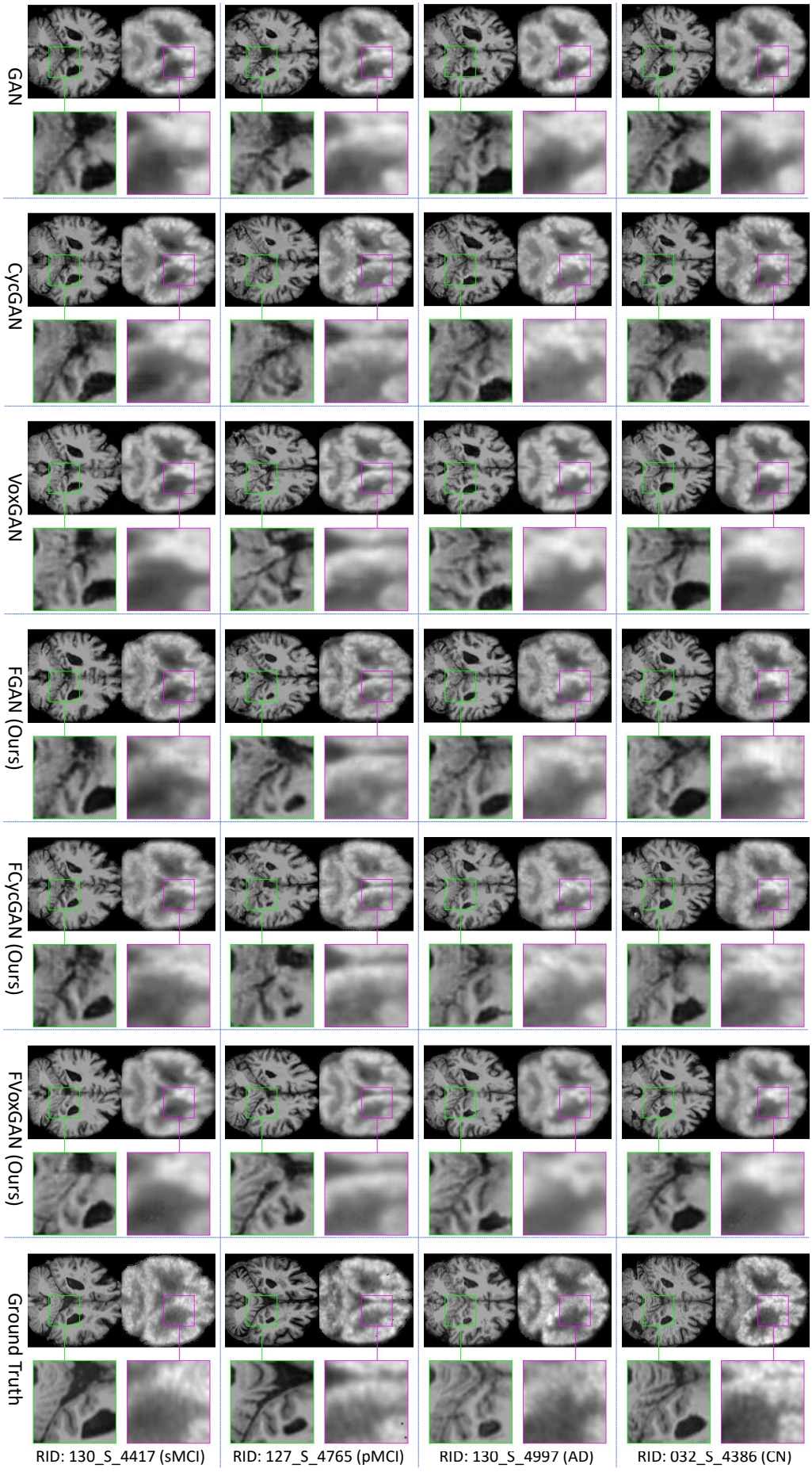


Fig. S4. Coronal views of PET and MRI scans synthesized by six methods for four typical subjects (Roster ID: 4386, 4765, 4997, and 4417) in ADNI-2, along with their corresponding ground-truth images. All six image synthesis models are trained on ADNI-1.

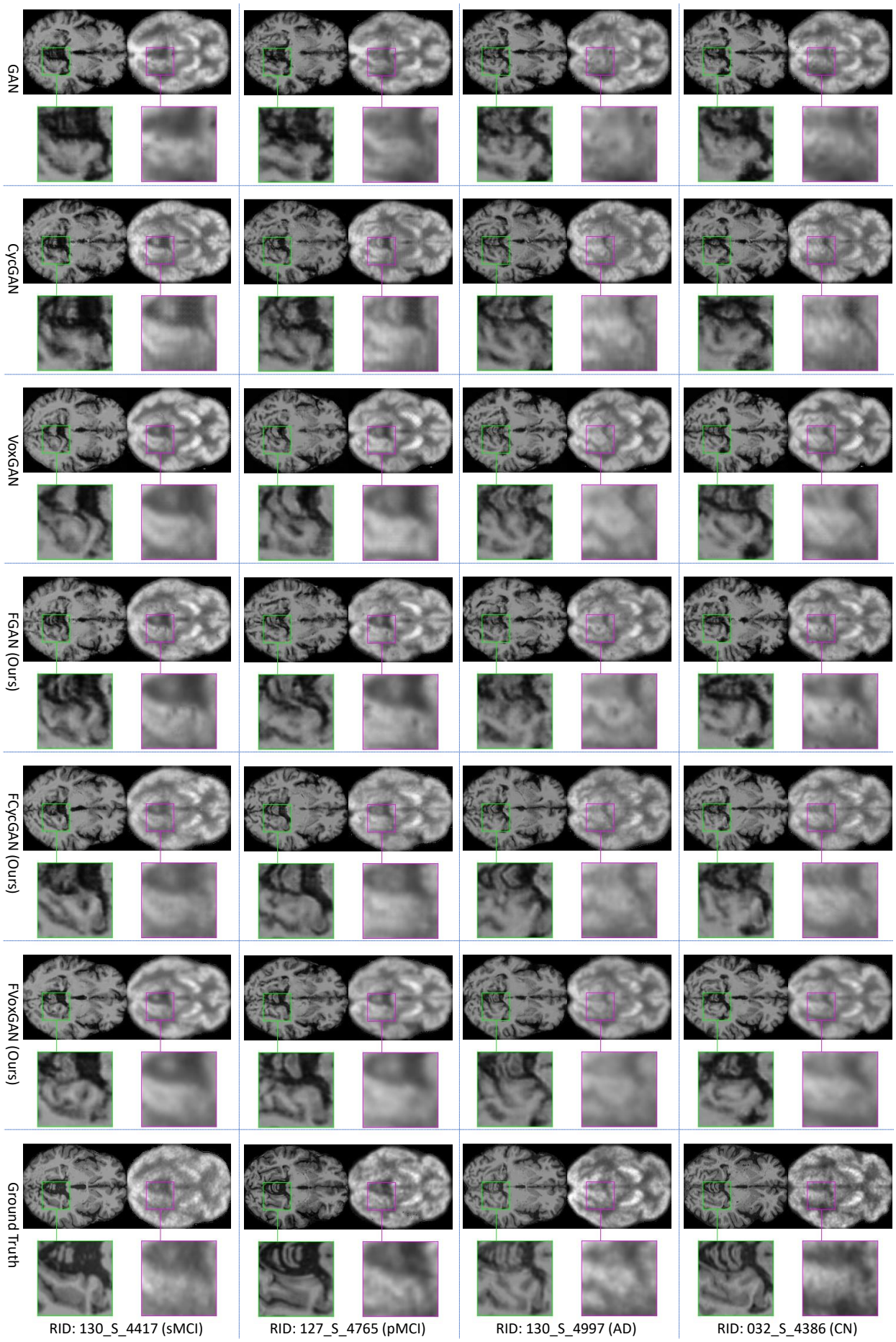


Fig. S5. Axial views of PET and MRI scans synthesized by six methods for four typical subjects (Roster ID: 4386, 4765, 4997, and 4417) in ADNI-2, along with their corresponding ground-truth images. All six image synthesis models are trained on ADNI-1.

generate the unknown color images [4] and then jointly using both grayscale and color images for classification. Two benchmark datasets for fine-grained classification were considered. The Oxford Flower-17 (F17) [5] dataset contains 17 flower species with 80 color images per species. The Oxford Pet-37 (P37) [6] dataset contains $7,349$ color images of 12 kinds of cats and 25 kinds of dogs with roughly 200 images per class. In both datasets, the images suffer from large variations in scale, pose, viewpoint angle and illumination. We randomly selected 75% images per category for training and used the rest for test.

To adapt our method to this problem, we replace 3D (de)convolutional kernels to 2D kernels and set the input size to 224×224. Since these images have no rigid structure consistency, the effect of spatial cosine kernel is suppressed. Noted that, although our DSNet may not be the best choice for 2D image classification, it is useful for verifying the effectiveness of the proposed feature-consistency constraint. The quality of synthetic color images was measured by the mean absolute error (MAE), mean square error (MSE), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR), and the performance of image classification was measure by the area under receiver operating characteristic (AUC), accuracy (ACC), average precision score (APS), and F1-Score (F1S).

The experiments include four stages. In the *first* stage, we trained two DSNets on color images and grayscale images, respectively, and reported the classification performance of each DSNet on the test set in Table S3. In the *second* stage, we trained different image generative models to transfer grayscale images to color images and reported the quality of synthetic color images in Table S4. In the *third* stage, the synthetic color images generated in the second stage were fed to the DSNet trained on real color images in the first stage, and the classification performance was reported in Table S5. In the *fourth* stage, the predicted scores achieved by DSNets on grayscale images and synthetic color images were simply averaged to mimic the multi-modality data. The corresponding classification performance was reported in Table S6. Besides, we show several samples from the F17 dataset and P37 dataset and the corresponding synthetic color images obtained by PixGAN, FGAN, and FPixGAN in Fig. S6 and Fig. S7. The following four conclusions can be drawn from these results.

$First$, the classification performance achieved by using grayscale images is obviously lower than that achieved by using color images on both datasets (see Table S3). Thus, it is possible to boost the performance of grayscale image classification as long as we can use grayscale images to generate the missing color images reasonably.

$Second$, the experimental evidence provided in [7] demonstrates that (1) the distribution matching constraints used in GANs may not be able to preserve discriminative information for either unpaired or paired data translation, leading to mis-diagnosis of medical conditions, and (2) using the l_1 (MAE) loss, equivalent to a pixel-wise-consistency constraint, seems to be helpful when the image quality metric is MAE, which matches the l_1 loss rather than measuring the classification performance. To evaluate how much discriminative information is preserved by each generative model, we gave the performance of using only the discriminative loss (GAN-d), feature-consistency loss (GAN-f), and pixel-wise consistency loss (GAN-p) in the 1^{rd} - 3^{th} rows of Table S4 and Table S5, respectively. It shows that GAN-f achieves the best image classification performance, GAN-p achieves best quality of synthetic images, and GAN-d, which may be good at distribution matching, achieves lower performance than GAN-f and GAN-p in color image generation and classification. This conclusion is consistent with the conclusion that distribution matching can hardly preserve discriminative information [7]. Therefore, we suggest considering a suitable constraint for each specific task, e.g., using the adversarial loss to keep the distribution (texture / structure) similarity, using the pixel-wise-consistency constraint to keep intensity consistency, and using the proposed feature-consistency constraint to keep classification consistency.

$Third$, jointly using different constraints may lead to balanced performance. PixGAN jointly uses the adversarial loss pixel-wise consistency loss, FGAN jointly uses the adversarial loss and feature-consistency loss, and FPixGAN jointly uses all three losses. The performance of PixGAN, FGAN, and FPixGAN was displayed in the 4^{rd} - 6^{th} rows of Table S4 and Table S5, respectively. Some samples from the F17 dataset and P37 dataset and the corresponding synthetic color images obtained by PixGAN, FGAN, and FPixGAN were illustrated in Fig. S6 and Fig. S7. It reveals that FGAN outperforms PixGAN in terms of image classification, but underperforms it in terms of image synthesis. FpixGAN achieves similar quality of synthetic images to PixGAN and similar classification performance to FGAN. It demonstrates again that the feature-consistency constraint is effective to preserve discriminative information. Moreover, multiple constraints can be jointly used to balanced performance in terms of both image generation and classification.

$Fourth$, combining the grayscale images with the synthetic color images to form pseudo multi-modality images and using them to perform the classification task may lead to improved performance (see Table S6). Especially, the performance of classifying pseudo multi-modality data, in which the missing color images were generated by the GANs with the feature-consistency constraint (e.g., FGAN and FPixGAN), is even compatible to that of classifying real color images on the F17 dataset. It suggests that the proposed feature-consistency constraint can be successfully applied to the transform of grayscale images to color images for classification purpose.

REFERENCES

[1] 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. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 455–463.

[2] C. R. J. Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.

[3] K. A. Ellis, A. I. Bush, D. Darby, D. De Fazio, J. Foster, P. Hudson, N. T. Lautenschlager, N. Lenzo, R. N. Martins, P. Maruff *et al.*, "The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease," *International Psychogeriatrics*, vol. 21, no. 4, pp. 672–687, 2009.

Fig. S6. Some samples from the Oxford Flower-17 dataset. The top and bottom rows are the paired grayscale and color images, while the 2^{nd}-4^{th} rows are the synthesized images generated by PixGAN, FGAN, and FPixGAN, respectively.

Fig. S7. Some samples from and Oxford Pet-37 dataset. The top and bottom rows are the paired grayscale and color images, while the 2^{nd}-4^{th} rows are the synthesized images generated by PixGAN, FGAN, and FPixGAN, respectively.

TABLE S3
Classification results achieved by DSNet on grayscale images and color images

Data Modality	Oxford Flower-17				Oxford Pet-37			
	AUC	APS	ACC	F1S	AUC	APS	ACC	F1S
Color	99.39	93.62	87.65	87.65	96.04	63.05	58.87	58.86
Gray	98.81	89.50	83.24	83.35	96.15	59.29	53.86	53.98

TABLE S4
Image quality of synthesized color images generated by six different GANs from grayscale images

Synthesis Model	Oxford Flower-17				Oxford Pet-37			
	MAE	MSE	SSIM	PSNR	MAE	MSE	SSIM	PSNR
GAN-d	74.69	92.42	2.38	6.73	60.25	73.18	9.99	8.77
GAN-f	21.69	25.59	63.83	18.59	20.68	23.89	61.51	19.34
GAN-p	10.01	16.09	82.20	21.89	11.47	15.30	74.85	23.08
PixGAN	11.48	17.89	77.18	20.97	11.10	15.02	75.75	23.27
FGAN	20.31	26.33	57.36	17.61	20.10	23.29	62.41	19.64
FPixGAN	12.95	18.48	77.34	20.69	11.71	15.35	74.86	23.05

TABLE S5
Classification results of DSNet on those synthesized color images generated by six different GANs from grayscale images

Synthesis Model	Oxford Flower-17				Oxford Pet-37			
	AUC	APS	ACC	F1S	AUC	APS	ACC	F1S
GAN-d	49.33	7.56	5.59	1.68	56.84	3.99	3.65	1.22
GAN-f	98.26	87.93	81.76	81.75	93.80	50.85	48.42	48.44
GAN-p	96.54	80.57	72.65	72.34	91.13	37.30	35.47	33.71
PixGAN	95.26	75.92	68.53	68.40	90.99	35.56	32.05	28.49
FGAN	98.14	89.06	82.94	82.97	94.05	52.14	48.48	48.38
FPixGAN	97.19	84.93	77.94	77.81	93.74	48.34	46.25	45.88

TABLE S6
Classification results while averaging the predicted scores of each synthesized color image and its corresponding grayscale image

Synthesis Model	Oxford Flower-17				Oxford Pet-37			
	AUC	APS	ACC	F1S	AUC	APS	ACC	F1S
GAN-d	97.72	87.23	64.41	69.31	93.96	54.86	42.00	46.80
GAN-f	99.24	92.29	87.94	87.88	96.25	61.75	55.82	55.91
GAN-p	98.91	90.83	83.82	83.68	95.60	56.83	50.92	50.84
FGAN	99.28	93.02	88.53	88.48	96.30	62.32	56.93	55.89
PixGAN	98.65	88.71	81.47	81.28	95.57	56.47	49.62	49.41
FPixGAN	98.93	90.96	86.18	86.02	96.17	61.04	56.10	56.06

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1125–1134.
[5] M. . Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, 2006, pp. 1447–1454.
[6] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. Jawahar, "Cats and dogs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2012, pp. 3498–3505.
[7] J. P. Cohen, M. Luck, and S. Honari, "Distribution matching losses can hallucinate features in medical image translation," in *Proc. Int. Conf. Med. Image Comput. Computer Assisted Intervention*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds. Cham: Springer International Publishing, 2018, pp. 529–536.