Morphological feature visualization of Alzheimer's disease via Multidirectional Perception GAN

Wen Yu, Baiying Lei, Yong Liu, Zhiguang Feng, Yong Hu, Yanyan Shen, Shuqiang Wang, Michael K. Ng

***Abstract*—The diagnosis of early stages of Alzheimer's disease (AD) is essential for timely treatment to slow further deterioration. Visualizing the morphological features for the early stages of AD is of great clinical value for early diagnosis. In this work, a novel Multidirectional Perception Generative Adversarial Network (MP-GAN) is proposed to visualize the morphological features indicating the severity of AD for patients of different stages. Specifically, by introducing a novel multidirectional mapping mechanism into the model, the proposed MP-GAN can capture the salient global features efficiently. Thus, by utilizing the class-discriminative map from the generator, the proposed model can clearly delineate the subtle lesions via MR image transformations between the source domain and the pre-defined target domain. Besides, by integrating the adversarial loss, classification loss, cycle consistency loss and $L1$ penalty, a single generator in MP-GAN can learn the class-discriminative maps for multiple-classes. Extensive experimental results on Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrate that MP-GAN achieves superior performance compared with the existing methods. The lesions visualized by MP-GAN are also consistent with what clinicians observe.**

***Index Terms*—Alzheimer's Disease, Lesion visualization, Generative Adversarial Networks, MR images.**

I. Introduction

ALZHEIMER'S DISEASE (AD) is an irreversible and chronic neurodegenerative disease with progressive impairment of memory and other mental functions. It is estimated to be the fifth leading cause of death in elderly people [1]. AD is caused by abnormal cell death in the brain, long before amnestic symptoms are observable [2]. The resulting brain atrophy is visible in structural magnetic resonance (MR) images. To date, AD is incurable but preventable. It is crucial to diagnose the early stages of AD by MR images for timely treatment [3]. Significant memory concern (SMC) and mild cognitive impairment (MCI) are the transitional stages between normal controls (NC) and AD [4]. SMC and MCI present mild symptoms, and the disease-related regions are very subtle in MR images. Currently, the clinical diagnosis procedure is time-consuming and requires extensive clinical training and experience for clinicians. Thus, developing automatic methods by utilizing deep learning to visualize the brain changes for the early stages of AD is highly desirable. It can assist clinicians for early diagnosis and may provide meaningful information on the pathogenesis of cognitive decline. However, this is a challenging task due to several reasons, such as the low-intensity contrast between the lesion and other neighboring regions, the indistinct boundary of the lesion, and the irregular lesion shape.

To visualize features of different Alzheimer's stages in MR images, there already exist several feature visualization methods based on classification. These methods can be classified into two categories. (1) The Regions Of Interest (ROI)-based classification approaches [1], [5]–[7] and patch-based classification approaches [8]. The performance of these methods is limited since the brain ROIs or patches need to be selected based on anatomical brain atlases or biological prior knowledge beforehand. Multiple steps are required to exact features from ROIs or patches for classification and subsequent visualization. Therefore, they tend to be sensitive to parameters and time-consuming; (2) Three strategies to visualize features for a convolutional neural network (CNN) classifier. (i)By editing an input image and observing its effect on the prediction results, the occluded regions which have a significant impact on prediction can be visualized; (ii)By analyzing the gradients of the prediction for an input image, a heatmap can be produced for visualization; (iii)By analyzing the activations of the feature maps for the image, the regions which are responsible for making the specific prediction can be visualized. These classification-based feature visualization methods make their predictions based on local regions most relevant to the particular prediction rather than the whole image, and it may ignore features with low discriminative power if stronger features for the prediction are available. As a result, if there is evidence for a category at multiple locations in the image (such as multiple AD lesions in MR images), some lesions with low discriminative power may be ignored. Moreover, visual features strongly depend on the performance of the classifier, and a large number of labeled samples are required to train a robust model.

To alleviate these issues, a novel Multidirectional Perception Generative Adversarial Network (MP-GAN) is proposed to visualize morphological features in whole-brain MR images. Generative Adversarial Network (GAN) [9], [10] has attracted lots of attention as it is capable of generating realistic data without explicitly modeling the probability density function. In this paper, MP-GAN with a novel multidirectional mapping

W. Yu and B. Lei contributed equally to this work.
W. Yu, Y. Shen and SQ. Wang are with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518060, China. E-mail: sq.wang@siat.ac.cn.
B. Lei is with School of Biomedical Engineering, Health Science Center,Shenzhen University, Shenzhen 518055, China. E-mail:leiby@szu.edu.cn
Yong Liu is with Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China
Yong Hu is with Department of Orthopaedics and Traumatology, University of Hong Kong
Zhiguang Feng is with College of Intelligent Systems Science and Engineering, Harbin Engineering University, 150001 Harbin, China
Michael K. Ng is with Department of Mathematics, University of Hong Kong, Pokfulam, Hong Kong.

mechanism is designed to capture the salient global features efficiently. Specifically, the generator of MP-GAN takes inputs as both MR images and its target domain. Then it flexibly learns a class-discriminative map for the target domain. By adding the class-discriminative map and the input MR image of the source domain, a synthetic MR image of the target domain can be produced. Thus the learned class-discriminative map can capture all brain changes by transforming the MR image between the source domain and the target domain. By visualizing the class-discriminative maps, the subtle and complex lesions that may not be found within one region can be identified. Besides, by designing the hybrid loss, a single generator in MP-GAN can learn the class-discriminative maps for multiple-classes. In this manner, the common features unrelated to the specific domain can be reused during training, therefore the visualization performance is further improved. With this global lesion visualization, clinicians can better exclude undesirable biases and potentially even identify previously unknown characteristics of AD. To the best of our knowledge, the proposed MP-GAN is the first work to visualize the morphological features for different Alzheimer's stages by a single generator. The contributions of this paper are summarized as follows:

1) A novel MP-GAN with a multidirectional mapping mechanism is proposed to capture the salient global features efficiently. By utilizing the class-discriminative map from the generator, the proposed model can clearly delineate the subtle lesions via MR image transformations between the source domain and the target domain.
2) By integrating the adversarial loss, classification loss, cycle consistency loss and $L1$ penalty, a single generator in MP-GAN can learn the class-discriminative maps for multiple-classes. The morphological features indicating different Alzheimer's stages can be visualized by a single MP-GAN model.

The rest of this paper is organized as follows. The related work is reviewed in Section II. The proposed MP-GAN is described in detail in Section III. In Section IV, MP-GAN is tested and compared with existing feature visualization methods to demonstrate its advantage. Finally, concluding remarks and future work are discussed in Section V.

II. Related work

The current feature visualization methods for AD generally fall into two categories: (1)The ROI-based classification approaches and patch-based classification approaches; (2)The CNN-based classification approaches.

For the first category, the brain ROIs or patches were selected based on anatomical brain atlases or biological prior knowledge beforehand, then multiple steps were required to extract features from ROIs or patches for classification. According to classification performance, the most frequently selected ROIs or patches would be visualized. For instance, Lian et al. [8] proposed a hierarchical fully convolutional network (H-FCN) to automatically identify discriminative local patches and regions in MR images for AD computer-aided diagnosis. The hierarchical discriminative locations of brain atrophy at both the patch-level and region-level were visualized. Wang et al. [5] proposed a multi-task exclusive relationship learning (MTERL) approach to predict the cognitive status. The most important ten ROIs for estimating clinical scores were visualized. Jie et al. [6] proposed a classification approach to extract features by integrating both temporal and spatial variabilities from the constructed dynamic connectivity networks (DCNs). Then a multi-kernel SVM model was employed for AD computer-aided diagnosis. The brain ROIs with significant spatial variability for EMCI vs. LMCI and EMCI vs. NC were visualized respectively.

For the second category, there were three strategies to visualize features for CNN.

1) By editing an input image and observing its effect on the prediction results, the occluded regions which had a significant impact on prediction can be visualized. For instance, Zeiler and Fergus [11] proposed an occlusion-based method to visualize the activity within CNN. Different portions of the input image were occluded systematically with a grey square, and the output of the classifier was observed. The occluded regions which cause the probability of the correct class drop significantly would be visualized. Korolev et al. [12] utilized 3D-ResNet for AD classification, and the important regions of the MR image most affected by AD were visualized by the occlusion-based method [11]. Nigri et al. [13] proposed a Swap Test to visualize the areas of the brain image most indicative of AD;
2) By analyzing the gradients of the prediction for an input image, the heatmap can be produced for visualization [14]–[16]. For example, Selvaraju et al. [17] proposed Gradient-weighted class activation mapping (Grad-CAM) for making any CNN-based models more transparent by producing heatmaps. Ancona et al. [18] analyzed four gradient-based feature visualization methods from theoretical and practical perspectives. Springenberg et al. [19] proposed a new variant of the "deconvolution approach" guided backpropagation for visualizing features learned by CNNs. Guided backpropagation can be applied to a broader range of network structures. Bohle et al. [20] utilized layer-wise relevance propagation (LRP) to visualize CNN decisions for AD based on MR images. Sundararajan et al. [21] proposed Integrated Gradients by utilizing an axiomatic framework for feature visualization;
3) By analyzing the activations of the feature maps for the image, the regions which were responsible for making the specific prediction can be visualized. For instance, Zhou et al. [22] proposed Class Activation Mapping (CAM) to visualize the discriminative object parts detected by CNN in a single forward pass. Khan et al. [23] utilized VGG with transfer learning for AD computer-aided diagnosis. CAM was utilized to visualize the discriminative regions in the MR image for model interpretation. Lian et al. [24] proposed a multi-task weakly-supervised attention network (MWAN) by leveraging a fully-trainable dementia attention block for

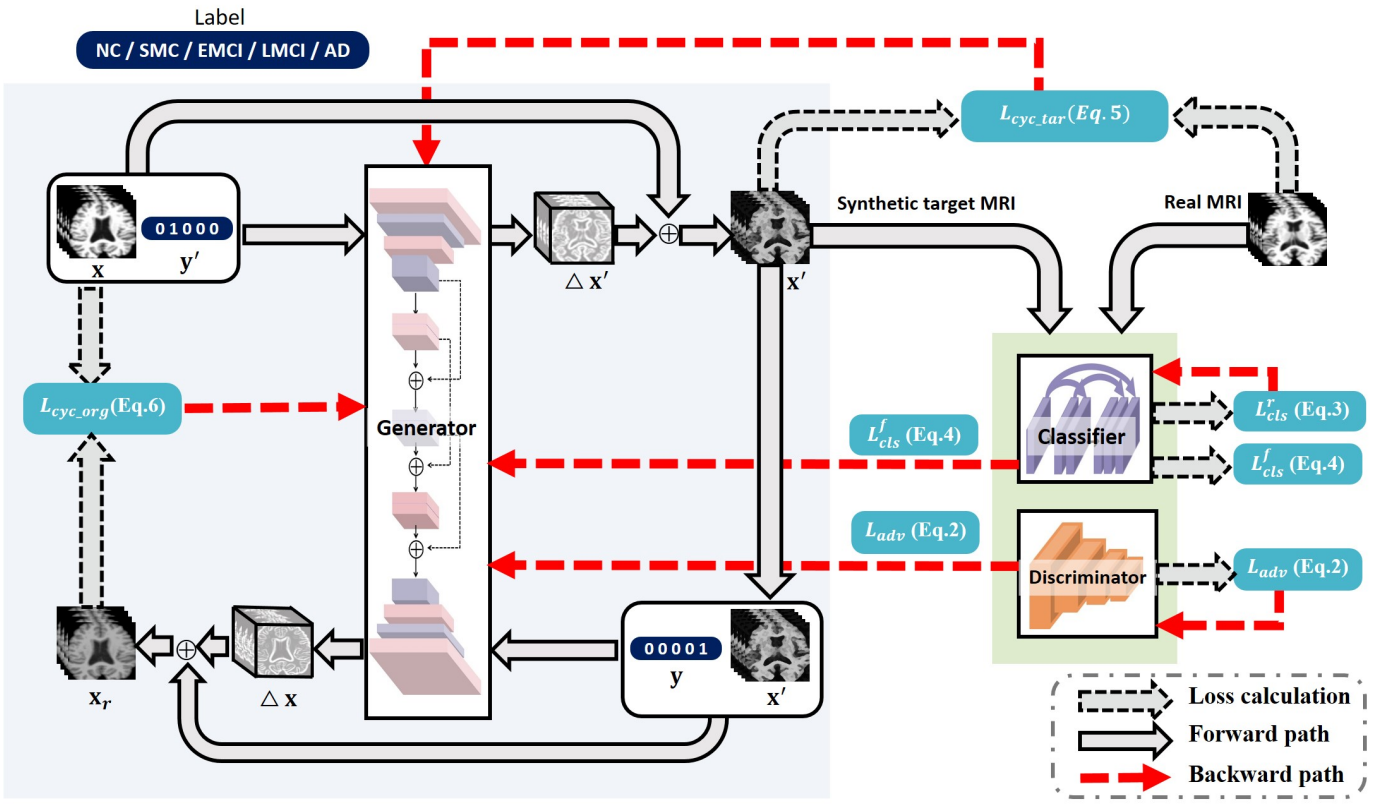

Fig. 1: The flowchart of MP-GAN. It consists of three components: a generator, a classifier, and a discriminator.

regression. The attention maps were visualized by CAM for AD subjects. Sarraf et al. [25] utilized LeNet-5 to classify structural MR images for AD vs. NC. The filters and the features were visualized for interpretation. Furthermore, Baumgartner et al. [26] proposed VA-GAN to visualize attributions distinguishing between MCI and AD. Note that CAM and gradient-based methods were not model-agnostic. They were most limited to neural networks (except [15]) and many required architectural modifications [11], [14], [22] or accessed to intermediate layers [17], [22]

III. The Proposed MP-GAN

A. Overview

The flowchart of MP-GAN is shown in Fig. 1. After data preprocessing (see Section IV-A), the normalized T1-MR images of all classes are fed into MP-GAN. The proposed model learns the class-discriminative maps between all class-pairs for visualizing morphological features. More specifically, the generator aims to capture salient global features in class-discriminative maps. Then the class-discriminative maps are used to transform MR images between the source domain and the target domain. To control semantic information, an auxiliary classifier is introduced based on the generator and discriminator to form the MP-GAN architecture. While the generator produces the class-discriminative maps distinguishing between the source domain and the target domain, the classifier predicts the domain indicating Alzheimer's stage, and the discriminator identifies whether the transformed MR images are real or fake. In this manner, the class-discriminative maps learned by MP-GAN can highlight exactly which regions of the MR image are significant for discrimination between the source domain and the target domain at the voxel-level. The subtle and complex lesions that may not be found within one region can be identified. Furthermore, since the input MR images are high-order with complicated brain structure, MP-GAN is further designed with the following two improvements: (1) 3D Residual Blocks are exploited in the conditional generator so that the features from the low-level can be reused, and the vanishing-gradient problem can be prevented; (2) 3D-DenseNet is utilized in classifier to capture more discriminative features.

B. The Architecture

The proposed MP-GAN is designed to visualize morphological features for multiple-classes. To achieve this, the generator G is designed to produce class-discriminative map Δx which can transform an input MR image x to an output MR image x' conditioned on the target class y', $[G(x, y') + x] \rightarrow x'$. During training, the target class y' is randomly selected so that G learns to produce class-discriminative maps for all class-pairs. By doing so, the target class y' can be predefined, and global features that distinguish between the source domain y and the desired target domain y' can be visualized at the testing stage.

As illustrated in Fig. 1, input MR image x is labeled and y represents the corresponding class. The conditional

generator aims to capture all salient global features in class-discriminative maps Δx. Then Δx is utilized to transform input MR image from the source domain y to the target domain y' in a bidirectional manner. The classifier predicts label y_c given real MR image x by the conditional distribution $p_c(y|x)$, and the discriminator is trained to identify whether the MR image is real or fake. Formally, given an MR image x of source class y and a conditional variable y', the generator can produce a synthetic MR image x' of target class y' by adding the generated class-discriminative map Δx and input MR image x.

$$x' = \Delta x + x = G(x, y') + x, \tag{1}$$

which is indistinguishable from the real MR image of the target domain y'. Thereby, class-discriminative map Δx contains all salient global features which distinguish between two domains y and y'. The change of salient voxels between the source domain y and the target domain y' on the MR image can be visualized by the class-discriminative map.

Adversarial Loss. To make the synthetic target MR images indistinguishable from real MR images, an adversarial loss is defined as

$$\begin{aligned}\mathcal{L}_{adv} =& \mathbb{E}_x \left[\log D(x)\right] + \\ & \mathbb{E}_{x,y'} \left[\log\left(1 - D(G(x, y') + x)\right)\right],\end{aligned} \tag{2}$$

where generator G generates an MR image $[G(x, y') + x]$ conditioned on both the input MR image x and the target class y', while discriminator D attempts to distinguish between real and fake MR images. The G tries to minimize this adversarial loss, while the D tries to maximize it.

Classification Loss. Given an input MR image x and a target class y', the goal of MP-GAN is to produce a class-discriminative map that can transform x into an output MR image x'. x' aims to be classified as the target class y'. To achieve this condition, an independent classifier is introduced and the classification loss is imposed when optimizing generator G. Specifically, the loss function is decomposed into two terms: a classification loss of real images to optimize classifier C, and a classification loss of fake images to optimize generator G. In detail, the former is defined as

$$\mathcal{L}_{cls}^r = \mathbb{E}_{(x,y)\sim p_{\text{real}}(x,y)} \left[-\log p_c(y|x)\right]. \tag{3}$$

By minimizing this classification loss, classifier C learns to classify a real MR image x to its corresponding class y. On the other hand, the loss function for the classification of fake images is defined as

$$\mathcal{L}_{cls}^f = \mathbb{E}_{(x',y')\sim p_g(x,y)} \left[-\log p_c(y'|x')\right]. \tag{4}$$

Generator G tries to minimize the loss $\mathcal{L}_{cls}^f$ to produce the class-discriminative maps for generating MR images x' that can be classified as the target class y'.

Cycle consistency loss. By minimizing the adversarial and classification losses, generator G is trained to generate MR images that are realistic and classified as target class. However, minimizing the losses (Eqs. (2) and Eqs. (4)) does not guarantee that the final transformed images preserve the content of input MR images while changing only the disease-related regions of the input. To alleviate this problem, a forward cycle consistency loss and backward cycle consistency loss [27], [28] are applied to the generator. They are defined as

$$\mathcal{L}_{cyc\_tar} = \mathbb{E}_{x,y',y} \left[\|x'_{\text{real}} - (G(x, y') + x)\|_1\right], \tag{5}$$

$$\begin{aligned}&\mathcal{L}_{cyc\_org} \\ =&\mathbb{E}_{x,y',y} \left[\|x_{\text{real}} - (G(x', y) + x')\|_1\right] \\ =&\mathbb{E}_{x,y',y} \left[\|x_{\text{real}} - (G((G(x, y') + x), y) + x')\|_1\right],\end{aligned} \tag{6}$$

where generator G takes in the transformed MR image x' and the source class y as input and tries to reconstruct the MR image $X_r = G(x', y) + x'$ of the source domain y. The $L1$ norm is adopted as the reconstruction loss. Note that a single generator is reused twice. The generator is first utilized to transform MR images of the source domain y to MR images of the target domain y'. Then it is used to reconstruct the MR image of the source domain y from the synthetic MR images of the target domain y'. For the first utilization, forward cycle consistency loss $\mathcal{L}_{cyc\_tar}$ is adopted. For the second one, backward cycle consistency loss $\mathcal{L}_{cyc\_org}$ is adopted.

***L*1 penalty.** The smallest class-discriminative map Δx that leads to a real MR image of the target domain y' is encouraged. Thus $L1$ penalty is defined as

$$\mathcal{L}_1(\Delta x) = \|\Delta x\|_1, \tag{7}$$

where $\|\cdot\|_1$ is the $L1$ norm.

Total Loss. The total loss functions to optimize D, C, and G are defined respectively as

$$\mathcal{L}_D = -\mathcal{L}_{adv}, \tag{8}$$

$$\mathcal{L}_C = \mathcal{L}_{cls}^r, \tag{9}$$

$$\begin{aligned}\mathcal{L}_G =& \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^f + \lambda_1\mathcal{L}_1(\Delta x) \\ &+ \lambda_{cyc\_org}\mathcal{L}_{cyc\_org} + \lambda_{cyc\_tar}\mathcal{L}_{cyc\_tar},\end{aligned} \tag{10}$$

where λ_{cls}, λ_1 , λ_{cyc_org} and λ_{cyc_tar} are hyperparameters that control the relative importance of classification loss, $L1$ penalty, and cycle consistency loss respectively, compared to the adversarial loss. $\lambda_{cls} = 0.1$, $\lambda_1 = 10$, $\lambda_{cyc\_org} = 10$ and $\lambda_{cyc\_tar} = 1$ are fixed throughout the paper. By optimising Eq.10, MR images of the source domain are transformed to MR images of the target domain by Δx , thus the class-discriminative map Δx can capture all morphological features between the source domain and the target domain.

For AD computer-aided diagnosis, the following class settings are defined to train so that all features between any two domains y and y' can be visualized completely. More specifically, y denotes the source class of the input MR image x, such as NC. y' denotes the target label, such as EMCI. At the training stage, y' is selected randomly according to the following rule.

(1) y={NC}, y'={SMC,EMCI,LMCI,AD},
(2) y={SMC}, y'={NC,EMCI,LMCI,AD},
(3) y={EMCI}, y'={SMC,NC,LMCI,AD},
(4) y={LMCI}, y'={SMC,EMCI,NC,AD},

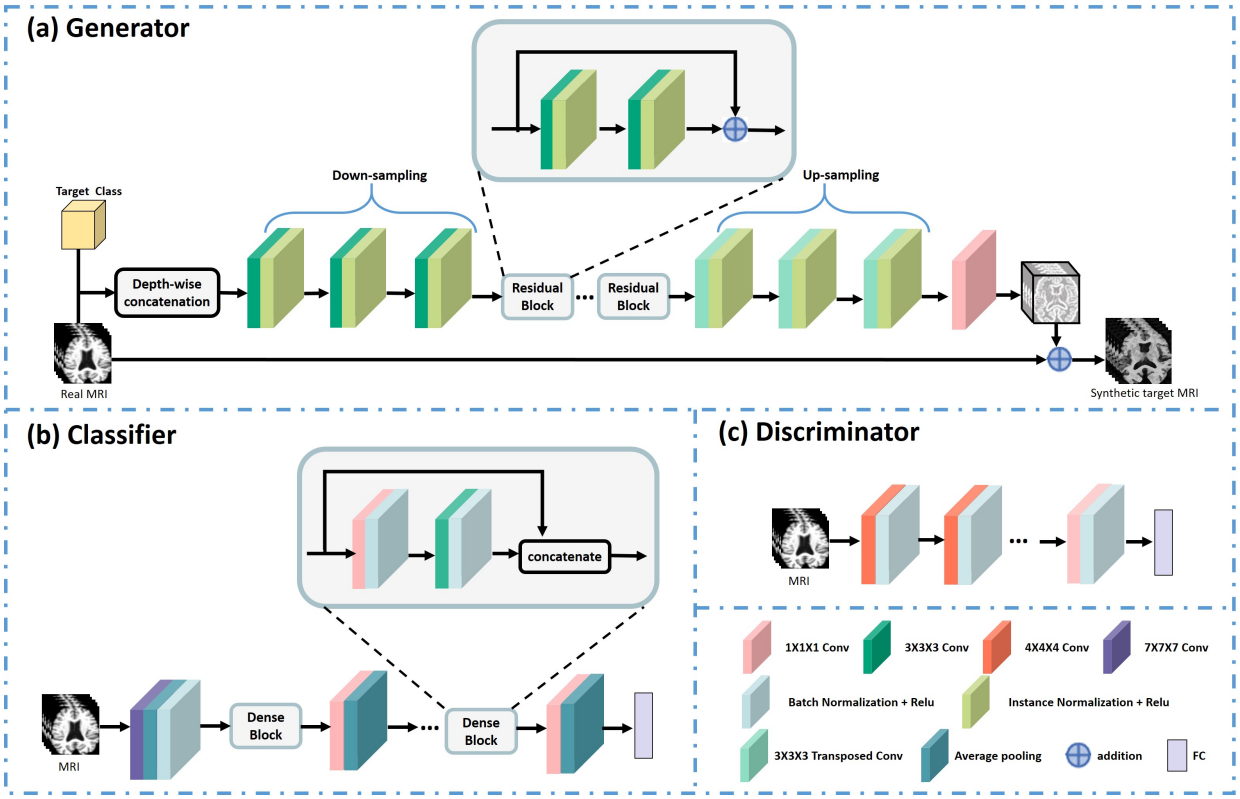

Fig. 2: The network structure of (a) Generator, (b) Classifier, and (c) Discriminator.

(5) y={AD}, y'={SMC,EMCI,LMCI,NC}.

In this manner, the generator is able to learn the class-discriminative maps between any two classes. At the testing stage, y' is pre-defined according to the requirement of user. In this paper, at the testing stage, the morphological features of NC versus all Alzheimer's stages including SMC, EMCI, and LMCI are visualized. MCI is characterized by a slight decline in cognitive abilities. Note that patients with MCI are at increased risk of developing to AD, but do not always do. Thus MCI is significant for morphological feature visualization and further diagnosis.

The network structure of generator, classifier, and discriminator is shown respectively in Fig. 2. The network utilized in the generator is ResNet. 3D-ResNet is expanded by adding a spatial dimension to all convolutional and pooling layers in ResNet for the MR image. By utilizing the shortcut connection, ResNet explicitly reformulates the layers as learning residual functions regarding the input layer, and it transfers feature representations from low layers to the high layers. More specifically, the generator network is composed of two convolutional layers with a stride size of 2 for downsampling, three residual blocks [29], and two transposed convolutional layers with the stride size of two for upsampling. Instance normalization [30] is used in all layers except the last output layer for the generator. $3 \times 3 \times 3$ and $1 \times 1 \times 1$ convolutional filters are employed in generator. The network utilized in the classifier is DenseNet [31]. 3D-DenseNet is expanded by adding a spatial dimension to all convolutional and pooling layers in DenseNet for the MR image. Feature-maps learned by all preceding layers are concatenating along the last dimension for the subsequent layers. Through such dense connectivity, feature-maps are reused and the vanishing-gradient problem is alleviated. Meanwhile, 3D-DenseNet can extract discriminative features related to Alzheimer's stage from the whole MR images efficiently. The details of 3D-denseNet can be found in [31], [32]. In this paper, the depth is set to 30, the growth rate is set to 12, the number of the Dense-BC block is set to 3, and the reduction is set to 0.5. A standard CNN architecture with 7 convolutional layers with $4 \times 4 \times 4$ and $1 \times 1 \times 1$ convolutional filters is adopted in the discriminator. Each convolutional layer is followed by batch normalization [33] and ReLU.

IV. Experiments and Results

A. Dataset and Preprocessing

T1-weighted MR images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public dataset are used for the evaluation purpose. 5316 MR images in ADNI-1, ADNI-go, and ADNI-2 are utilized. It includes 1736 NC subjects, 288 SMC subjects, 1582 EMCI subjects, 616 LMCI subjects, and 1094 AD subjects. Both 1.5T and 3T field strength MR images are used. Table I lists the demographic characteristics of the subjects. Note that the ADNI-2 assessed participants from the ADNI-1 phases in addition to new participant groups in 2011[1]. Different from the ADNI-1 dataset, MCI is divided into two subtypes, including early mild cognitive impairment(EMCI),

[1]http://adni.loni.usc.edu/about/

TABLE I: Demographic characteristics of the subjects in ADNI dataset.

Magnet strength	3T								1.5T			
Source	ADNI-1		ADNI-GO	ADNI-2					ADNI-1		ADNI-GO	ADNI-2
Subject	NC	AD	EMCI	NC	SMC	EMCI	LMCI	AD	NC	AD	NC	NC
Number	42	29	142	190	121	309	177	159	171	175	16	81
Gender(F/M)	27 / 15	20 / 9	67 / 75	95 / 95	64 / 47	139 / 169	80 / 97	68 / 91	89 / 82	85 / 90	8 / 8	45 / 36
Age	76.1±5.1	75.6±7.9	71.7±7.7	74.9±6.8	72.9±5.6	72.3±7.3	72.9±7.6	75.4±7.9	77.7±5.4	76.6±7.5	80.2±4.8	82.5±4.5
Education	16±2.8	14.7±2.9	15.8±2.7	16.4±2.7	16.8±2.5	16±2.7	16.5±2.6	15.8±2.7	16.0±2.9	14.6±3.2	15.5±2.5	15.9±2.9
MMSE	29.3±1.0	20.03±4.8	28.2±1.8	28.7±1.5	28.6±1.7	28.0±2.1	26±3.5	20.8±4.4	29.1±1.2	21.5±4.4	29.4±1.0	28.5±2.6
CDR	0±0.14	1.07±0.4	0.45±0.19	0.07±0.19	0.13±0.23	0.46±0.22	0.58±0.37	0.99±0.46	0±0.19	0.93±0.49	0.07±0.18	0.2±0.35
Samples	149	73	471	723	288	1111	616	501	587	520	72	205

and late mild cognitive impairment(LMCI) in the ADNI-2 dataset. SMC is the transitional stage between NC and MCI. The diagnostic criteria are described in the ADNI procedures manual[2].

All MR images are in the neuroimaging informatics technology initiative (NIfTI) format. They are processed using standard operations in the FSL[3] toolbox [34]–[36] for registering the MR images to MNI space. The preprocessing pipelines contain three parts: (1) removal of redundant tissues; (2) brain area extraction by BET; (3) linear registration by FLIRT [37], [38]. Lastly, the T1-MR image is normalized into the range [-1,1], and is fed into the MP-GAN model as a tensor directly without compressing or downsizing. For evaluation, 80% of the data are allocated for training. The remaining 20% of the data are equally partitioned and used as validation and test data sets respectively. A single MP-GAN model is trained on a training dataset of all categories as mentioned above, then the morphological features between the source domain and the predefined target domain are visualized on the test set. The validation set is used for hyperparameter optimization.

B. Experiment Settings

The proposed MP-GAN is trained on the ADNI dataset from scratch in an end-to-end manner. All methods are implemented in TensorFlow[4]. All experiments are conducted on four NVIDIA GeForce GTX 2080 Ti GPUs. 'Adam' is utilized as the optimizer for stochastic gradient descent (SGD). The batch size is set to 8. The learning rate of both generator and classifier is set to 0.001. The learning rate of the discriminator is set to 10^{-4}.

C. Qualitative Analysis

In this section, comprehensive experiments are conducted to show the effectiveness of MP-GAN. First, the proposed model is compared with 4 methods: (1)Guided Backpropagation [19]; (2)Integrated Gradients [21]; (3)Class Activation Mapping(CAM) [22]; and (4)GAN [26]. For Guided Backpropagation, Integrated Gradients, and CAM, a conventional CNN architecture is used for these networks. Besides, for the CAM method, the last layer is designed as described in [22] and the last two max-pooling layers are omitted. This allows more accurate heatmaps due to the higher resolution of the last feature maps. The proposed method is also compared with the conventional GAN [26] to demonstrate our advantages. For a fair comparison, the network structure of the generator and discriminator in GAN is the same as the proposed MP-GAN, and the loss function of GAN is the conventional adversarial loss. The GAN is trained to visualize important regions in MR images between two predefined demains. Furthermore, the following 4 evaluation groups are set up when compared with the 4 existing methods: (1)NC vs. SMC, (2)NC vs. EMCI, (3)NC vs. LMCI, and (4)NC vs. AD. The main reason for this setup is that more meaningful pathological features can be found by comparing with healthy people. It is worth noting that MR images of all 5 classes are trained using only one MP-GAN model, and the class-discriminative map for each evaluation group is visualized at the test stage. But for the 4 compared methods, one independent binary model is trained for each evaluation group respectively.

To visually show the quality of heatmaps produced by the proposed model and the 4 existing methods, one MR image is taken from each evaluation group for qualitative analysis. From Fig. 3 to Fig. 6, the heatmaps from the sagittal, coronal, and axial views are illustrated for each evaluation group respectively. The figures are shown by progression from SMC to AD in order. From Fig. 3 to Fig. 6, it can be seen that the proposed MP-GAN can visualize subtle lesions with contour edge at a finer scale (i.e., voxel-level). More detailed discriminative regions can be depicted, such as the hippocampus, and the corners and boundaries of the ventricle. The highlighted subtle lesions predicted by MP-GAN are relatively more precise than those generated by the other 4 methods. For example, from Fig. 4, it can be observed that the lesions that have much more blurred boundary and are difficult to recognize can be delineated by MP-GAN. More specifically, the corpus callosum with irregular sulcus is depicted accurately by MP-GAN from the sagittal-view and coronal-view in Fig. 4. Atrophy of the corpus callosum may lead to functional disability because of reduced interhemispheric integration. It is a region that has been examined intensively for indications of EMCI [39]. On the other hand, Integrated Gradients and Guided Backprop tend to focus on some small parts of the lesions rather than the whole lesions. Because some subtle voxels of the lesion might be more salient than the other voxels of the whole lesion. This proves that the feature visualization methods based on classification only focus on the most discriminative features and ignore the rest. It is difficult to interpret the results produced by CAM due to the low-resolution. Moreover, the regions visualized by GAN seem to cover parts of ground-truth affected by the AD for NC vs. AD as shown in Fig. 6. However, they are not close to ground-truth, this is because the training of GAN

[2]http://www.adni-info.org
[3]www.fmrib.ox.ac.uk/fsl
[4]http://www.tensorflow.org/

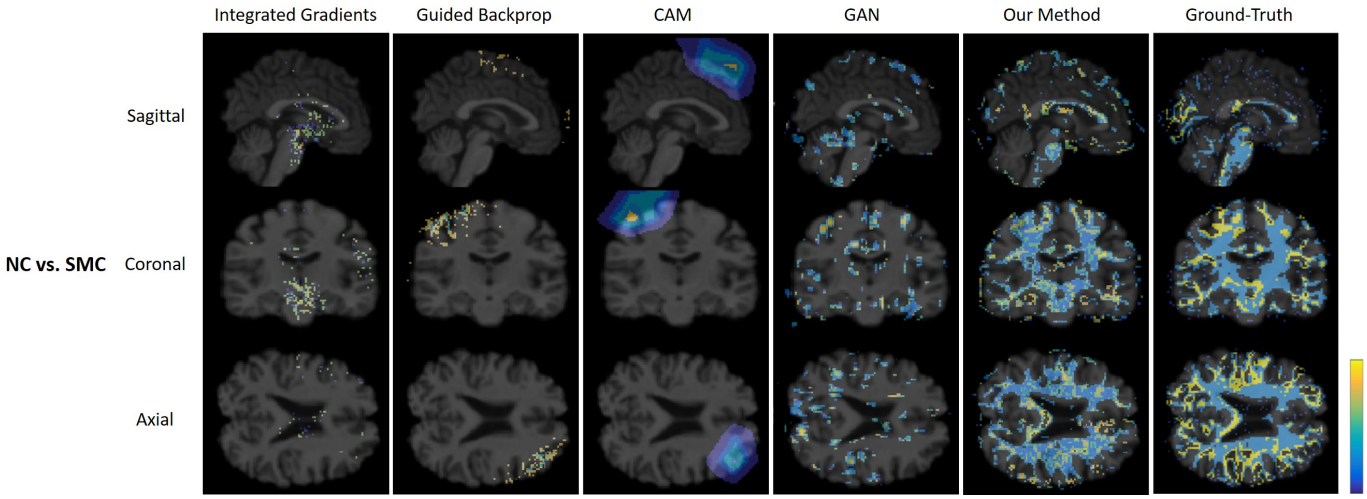

Fig. 3: Heatmaps predicted by Integrated Gradients, Guided Backprop, CAM, GAN, and our method are shown in sagittal,coronal, and axial views for NC vs. SMC respectively.

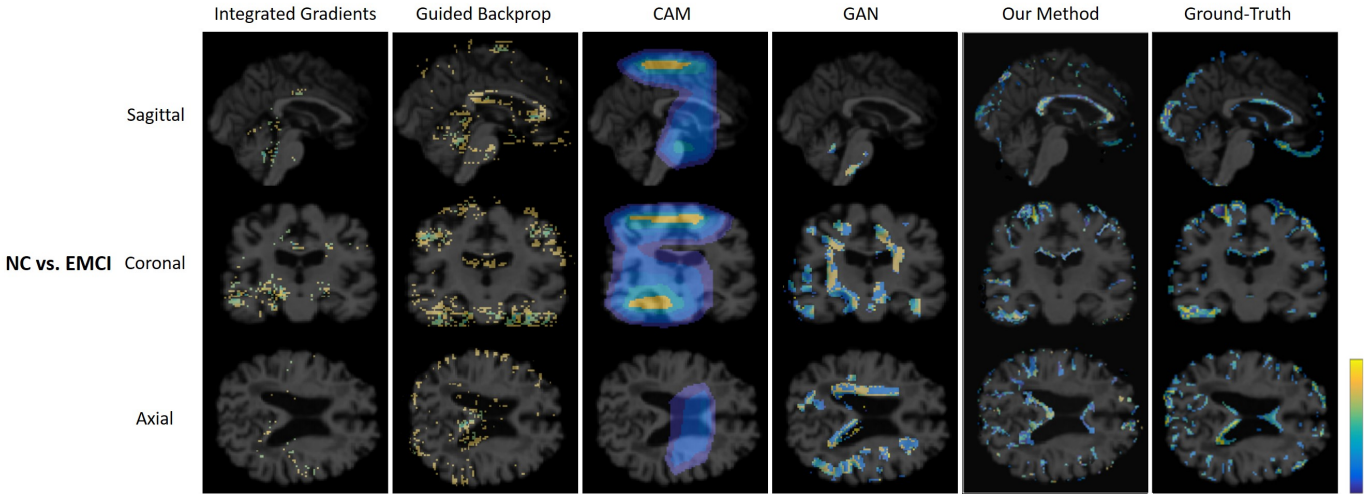

Fig. 4: Heatmaps predicted by Integrated Gradients, Guided Backprop, CAM, GAN, and our method are shown in sagittal,coronal, and axial views for NC vs. EMCI respectively.

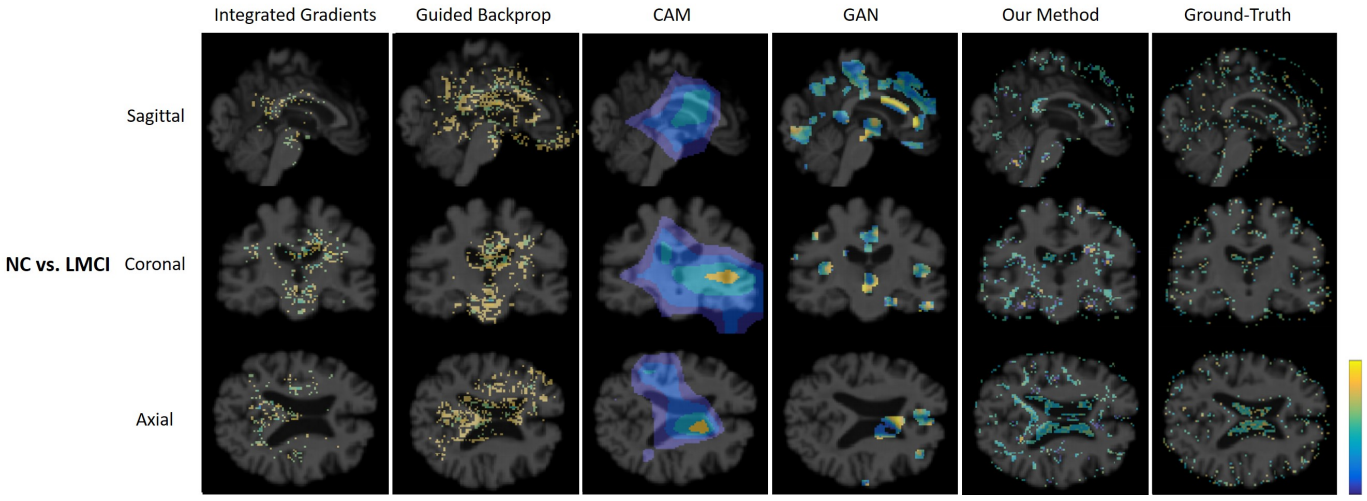

Fig. 5: Heatmaps predicted by Integrated Gradients, Guided Backprop, CAM, GAN, and our method are shown in sagittal,coronal, and axial views for NC vs. LMCI respectively.

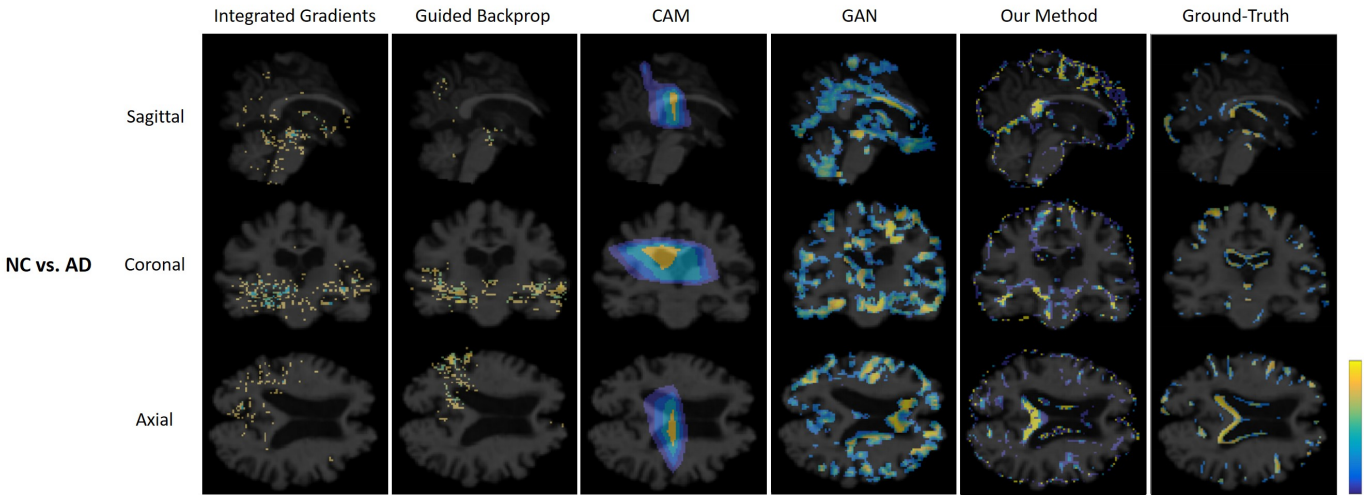

Fig. 6: Heatmaps predicted by Integrated Gradients, Guided Backprop, CAM, GAN, and our method are shown in sagittal,coronal, and axial views for NC vs. AD respectively.

is unstable. In summary, the results of the proposed MP-GAN are closer to the ground-truth compared with the other 4 existing methods. This implies that MP-GAN can benefit from the multidirectional mapping mechanism and the hybrid loss function. MP-GAN is more sensitive to subtle structural changes in MR images caused by cognitive decline.

The ADNI diagnostic criteria for each Alzheimer's stage are briefly described as following. (1) NC participants have no subjective or informant-based complaints of memory decline, and they have a normal cognitive performance; (2) SMC participants have subjective memory concerns assessed by the Cognitive Change Index(CCI). They have no informant-based complaint of memory impairment or decline, and they have a normal cognitive performance on the Wechsler Logical Memory Delayed Recall (LM-delayed) and the Mini-Mental State Examination (MMSE) [40]; (3) EMCI participants have a subtle cognitive decline. Their abnormal memory function is approximately 1 standard deviation below normative performance, and their MMSE total score is greater than 24; (4) LMCI participants have a memory concern. Clinical Dementia Rating (CDR) of LMCI participants is 0.5, and the Memory Box (MB) score must be at least 0.5; (5) AD participants have a significant memory concern. The MMSE score of AD participants is between 20 and 26, and CDR is 0.5 or 1.0.

To further analyze the visualization results of the proposed MP-GAN from a clinical perspective, the two-view slices in another coordinate of (33,55,39) are shown in Fig. 7. Note that the three-view slices shown from Fig. 3 to Fig. 6 are in the coordinate of (44,55,47). From Fig. 7, the following observations can be made. (1) For all four evaluation groups, the proposed MP-GAN can delineate the discriminative lesions clearly. More specifically, lesions visualized by MP-GAN are hippocampus, thalamus, putamen, pallidum, caudate nucleus, amygdala, and insula [12], [41], [42]. It is worth noting that the discriminative capability of these brain regions in clinical diagnosis has already been validated by previous studies [1], [43]–[45]. This implies the feasibility of the proposed MP-GAN; (2) The morphological changes including global atrophy (e. g. smaller volumes of hippocampus or amygdala) and shape changes are visualized by class-discriminative map (indicated by color). These morphological changes are related to AD disease progression and cognitive decline severity; (3) For 4 evaluation groups, identified multiple regions are overlapped or localized at similar brain regions. For instance, the regions of NC vs. LMCI and NC vs. AD are similar because LMCI might develop to AD. Meanwhile, since the features between LMCI and AD are very subtle, thus some visualized regions of NC vs. LMCI and NC vs. AD are overlapped, but the atrophy severity of each lesion is different (indicated by color). The lesions visualized for EMCI vs. AD and NC vs. AD also have some common regions, such as the hippocampus and pallidum. Furthermore, it is reasonable that the overlap regions between NC vs. EMCI and NC vs. AD might not be identified for EMCI vs. AD, and some regions such as the amygdala which are specific to EMCI vs. AD can be identified; (4) Along with the progression from EMCI to AD, from Fig. 7(a) to Fig. 7(c), it can be observed that the intensity values (i.e., light salmon color) in the heatmaps are gradually increased (i.e., change to crimson) at various brain locations, and some of them are accumulated at the annotated regions. These results suggest that the class-discriminative maps generated by the proposed MP-GAN have the potential to provide some extra information regarding the AD progression, and it may reveal the gradual atrophic process of the human brain due to cognitive decline. Furthermore, the severity of cognitive decline is also reflected in ADNI diagnostic criteria for each Alzheimer's stage as described above. In summary, the above observations imply the robustness of MP-GAN in visualizing morphological features for different Alzheimer's stages.

For further visualization analysis, 5 evaluation groups are investigated respectively in Fig. 8. The results show that the important brain regions visualized by the proposed method are consistent with regions in Fig. 7. More specifically, by aligning the automatic anatomical labeling (AAL) map with the class-discriminative maps visualized in Fig. 8, each region in the class-discriminative map will be matched to the specific

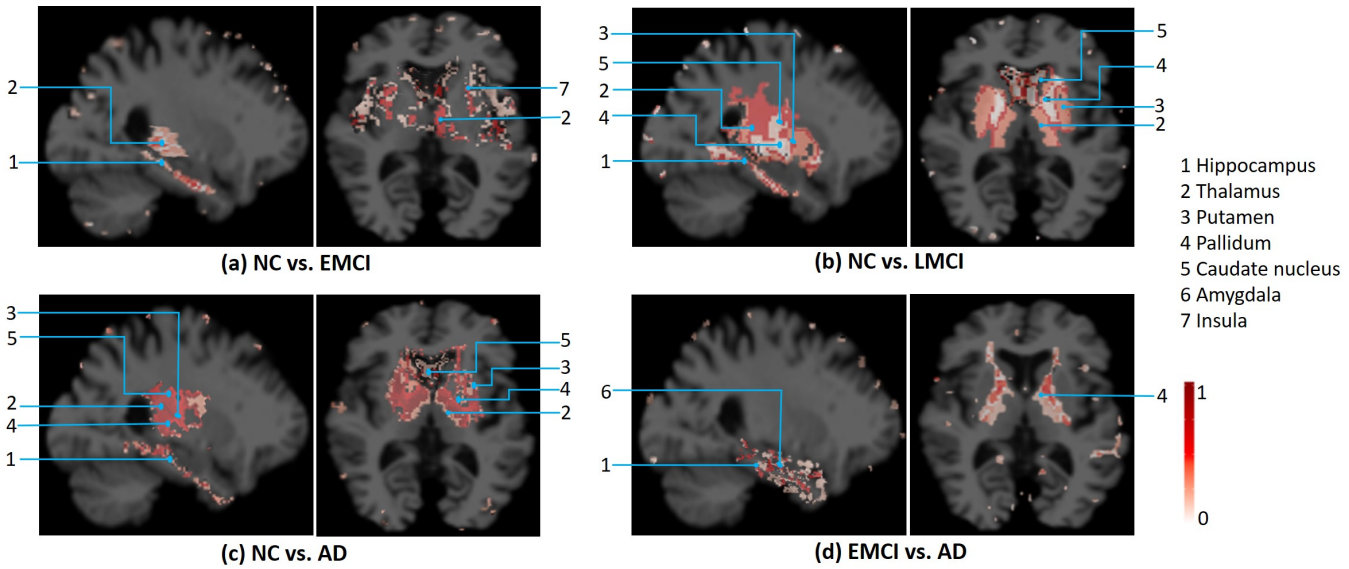

Fig. 7: Class-discriminative maps generated by MP-GAN are shown as a colored overlay over the MR images. The regions affected by the progression of AD are reliably captured by MP-GAN for four evaluation groups respectively.

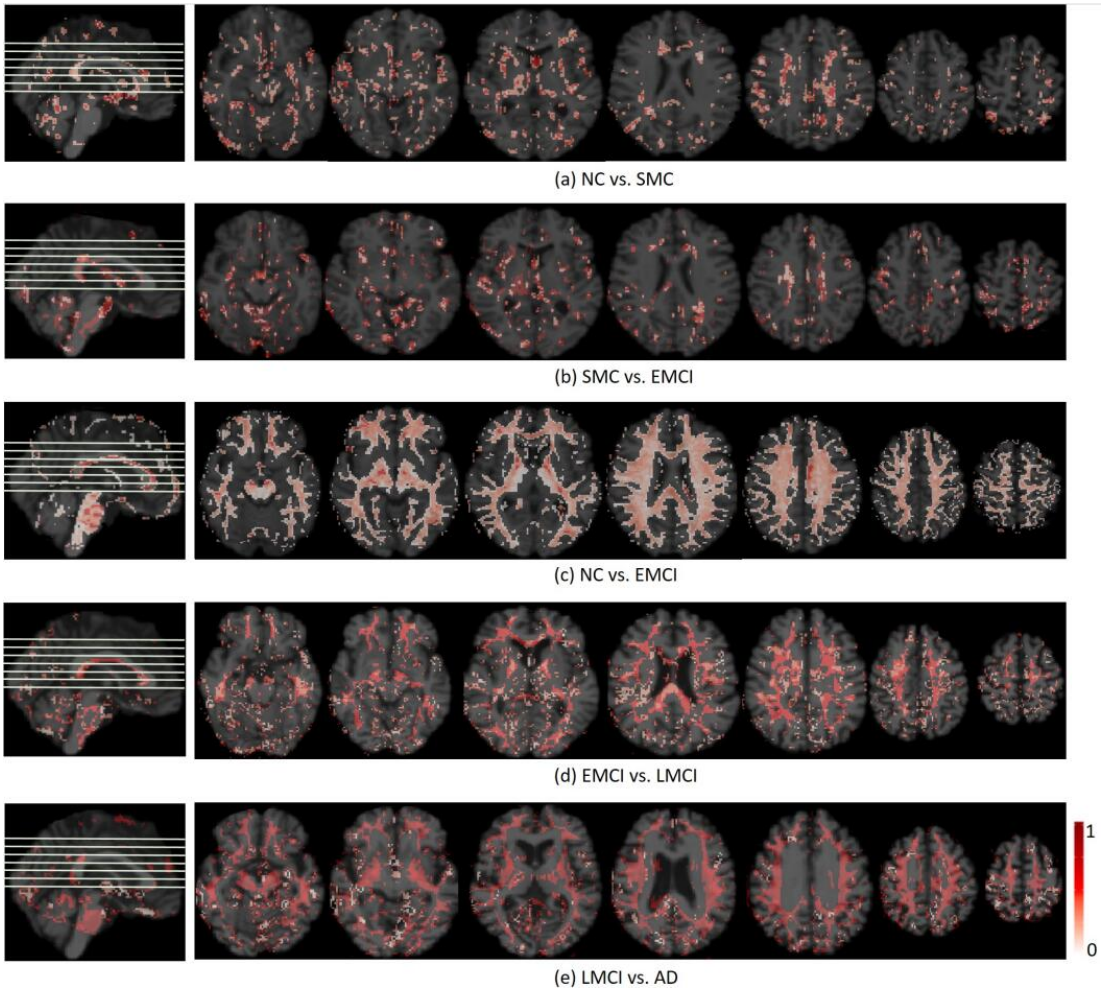

Fig. 8: Distribution of class-discriminative maps visualized by MP-GAN for five evaluation groups respectively.

TABLE II: Indices and names of regions visualized by MP-GAN using the AAL template.

NC vs. SMC		SMC vs. EMCI		NC vs. EMCI		EMCI vs. LMCI		LMCI vs. AD	
ROI index	ROI name	ROI index	ROI name	ROI index	ROI name	ROI index	ROI name	ROI index	ROI name
1	Precentral_L	17	Rolandic_Oper_L	7	Frontal_Mid_L	5	Frontal_Sup_Orb_L	9	Frontal_Mid_Orb_L
8	Frontal_Mid_R	29	Insula_L	36	Cingulum_Post_R	12	Frontal_Inf_Oper_R	14	Frontal_Inf_Tri_R
10	Frontal_Mid_Orb_R	39	ParaHippocampal_L	43	Calcarine_L	37	Hippocampus_L	37	Hippocampus_L
12	Frontal_Inf_Oper_R	42	Amygdala_R	45	Cuneus_L	39	ParaHippocampal_L	38	Hippocampus_R
34	Cingulum_Mid_R	49	Occipital_Sup_L	46	Cuneus_R,	40	ParaHippocampal_R	43	Calcarine_L
39	ParaHippocampal_L	51	Occipital_Mid_L	50	Occipital_Sup_R	43	Calcarine_L	48	Lingual_R
40	Parahippocampal_R	52	Occipital_Mid_R	56	Fusiform_R	47	Lingual_L	52	Occipital_Mid_R
50	Occipital_Sup_R	58	Postcentral_R	57	Postcentral_L	50	Occipital_Sup_R	67	Precuneus_L
57	Postcentral_L	60	Parietal_sup_R	74	Putamen_R	54	Occipital_Inf_R	68	Precuneus_R
78	Thalamus_R	67	Precuneus_L	90	Temporal_Inf_R	55	Fusiform_L	74	Putamen_R

ROI index and name in AAL. The disease-related regions visualized by MP-GAN are listed in Table II. Note that the suffix 'L' denotes the left brain, and the suffix 'R' denotes the right brain. The following observations can be made from Fig. 8 and Table II. (1) The brain regions visualized by the proposed method for NC vs. SMC are precental gyrus, middle frontal gyrus, inferior frontal gyrus, median cingulate, paracingulate gyri, parahippocampal gyrus, superior occipital gyrus, postcentral gyrus and thalamus; (2) The brain regions visualized by the proposed method for SMC vs. EMCI are rolandic operculum, insula, parahippocampal gyrus, amygdala, superior occipital gyrus, middle occipital gyrus, postcentral gyrus, superior parietal gyrus and precuneus; (3) The brain regions visualized by the proposed method for NC vs. EMCI are the middle frontal gyrus, posterior cingulate gyrus, calcarine fissure and surrounding cortex, cuneus, superior occipital gyrus, fusiform gyrus, postcentral gyrus, lenticular nucleus, putamen and inferior temporal gyrus; (4) The brain regions visualized by the proposed method for EMCI vs. LMCI are the superior frontal gyrus, orbital part, inferior frontal gyrus, opercular part, hippocampus, parahippocampal gyrus, calcarine fissure and surrounding cortex, lingual gyrus, inferior occipital gyrus and fusiform gyrus; (5) The brain regions visualized by the proposed method for LMCI vs. AD are the middle frontal gyrus, orbital part, inferior frontal gyrus, triangular part, hippocampus, calcarine fissure and surrounding cortex, lingual gyrus, middle occipital gyrus, precuneus, lenticular nucleus and putamen. These regions also agree with the existing research findings. To sum up, the lesions visualized by the proposed model are highly suggestive and effective for tracking the progression of AD.

The performance of MP-GAN to visualize the subtle lesions in the hippocampus is further investigated. The class-discriminative maps of the hippocampus in the sagittal view are visualized in Fig. 9. Specifically, the following 4 neighborhood evaluation groups are further explored: (a) NC vs. SMC; (b) SMC vs. EMCI; (c) EMCI vs. LMCI; (d) LMCI vs. AD. From Fig. 9, it can be observed that the zoomed regions preserve more details in the hippocampus. In particular, in the earlier stages of AD such as (a) NC vs. SMC and (b) SMC vs. EMCI, the visualized lesions are extremely subtle and scattered around the boundary of the hippocampus. In the later stages of AD such as (c) EMCI vs. LMCI and (d) LMCI vs. AD, the visualized lesions are accumulated at the core region of the hippocampus. Furthermore, Fig. 9(a) to Fig. 9(d) reflect the shape change and atrophy of the hippocampus qualitatively as the progressive deterioration from SMC to AD.

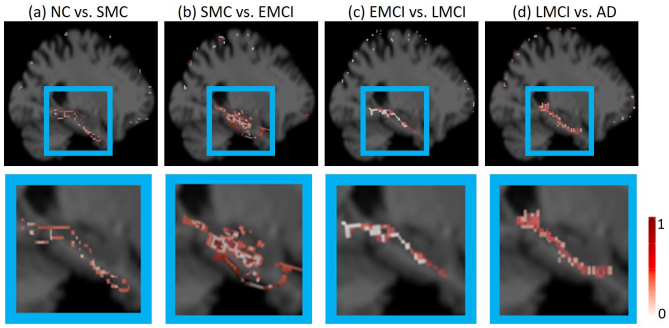

Fig. 9: Visualization results of hippocampus by MP-GAN in sagittal view and corresponding zoomed regions. The subfigures at the bottom are the zoom of the original subfigures for better observation.

It has already been validated by the previous studies [46] that the hippocampus is significant for identifying biomarkers in clinical practice. Although the volume loss and shape change of the hippocampus can not be quantitatively measured in this work, the visualized lesions of the hippocampus are beneficial for identifying the biomarkers in future work. Based on these visualized lesions in Fig. 9, the existing biomarkers such as Brain boundary shift integral (BBSI) [47], Scoring by Non-local Image Patch Estimator(SNIPE) [48], and other grading biomarkers [49] can be computed. Furthermore, new potential biomarkers reflecting the shape change and brain atrophy might be discovered based on these visualized lesions in the hippocampus in future work.

D. Quantitative Analysis

In this section, the following 4 metrics are computed to assess visual quality. (1) Normalized Cross-Correlation (NCC). NCC is calculated between the ground-truth maps and the predicted class-discriminative maps. The higher the NCC, the more correlation between ground-truth maps and the predicted class-discriminative maps. For Integrated Gradients, Guided Backprop, and CAM, the visualized heatmaps for predicting positive class are utilized to calculate the NCC; (2) Peak Signal-to-Noise Ratio (PSNR). PSNR is also calculated between the ground-truth maps and the predicted class-discriminative maps on the test data set. Similar to NCC, the higher the PSNR, the closer between ground-truth maps and the predicted class-discriminative maps; (3) Structural Similarity Index Measure (SSIM) [50]. Different from NCC and PSNR, SSIM in each iteration is calculated between synthetic

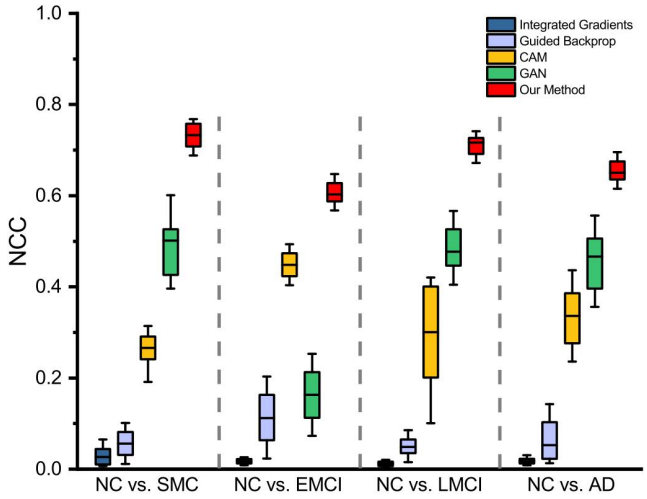

Fig. 10: Box-plots of NCC for different models.

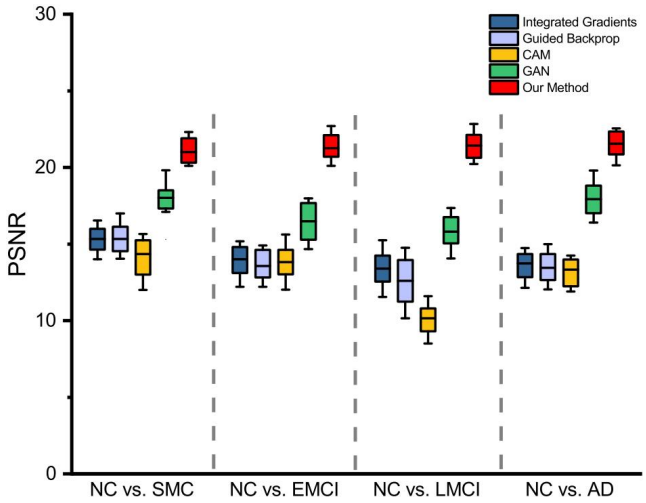

Fig. 11: Box-plots of PSNR for different models.

images and real images on the validation dataset. Higher SSIM indicates better reconstructed MR image quality. By computing SSIM in each iteration, the convergency of the model is further validated; (4) Classification metrics such as AUC, ACC, Sensitivity, and Specificity for data augmentation. Note that the purpose of SSIM and the classification metrics is to demonstrate that the proposed MP-GAN can generate images close to real distribution, thus it validates that MP-GAN can capture salient global features in class-discriminative maps. For NCC and PSNR, the 4 existing methods are compared. For SSIM, only GAN is compared since the other 3 methods are based on classification. Similarly, for the classification metrics, only GAN is compared since the classification performance is based on synthetic data augmentation by the proposed MP-GAN and GAN.

The NCC results shown in Fig. 10 are mostly consistent with the qualitative results shown from Fig. 3 to Fig. 6. The proposed MP-GAN achieves significantly higher NCC than the other 4 existing methods. It indicates that the distribution of class-discriminative maps generated by MP-GAN is the closest to ground-truth maps. The three methods based on classification (Integrated Gradients, Guided Backprop, and CAM) achieve a low NCC score due to its exclusive focus on local features. GAN performs better than 3 classification-based feature visualization methods for NC vs. SMC, NC vs. LMCI, and NC vs. AD. This implies that the GAN architecture can capture global features, which alleviate the limitations of feature visualization methods based on classification. Above all, the proposed MP-GAN achieves the highest correlation scores compared with the other 4 existing methods in all 4 evaluation groups.

From Fig. 11, it can be seen that the proposed MP-GAN achieves the best PSNR compared with the other 4 existing methods. This is also consistent with NCC results in Fig. 10 and the qualitative results shown from Fig. 3 to Fig. 6. The class-discriminative maps visualized by MP-GAN are closer to ground-truth. This is because MP-GAN benefits from the multidirectional mapping mechanism and the hybrid loss function. Meanwhile, MP-GAN can be trained on MR images

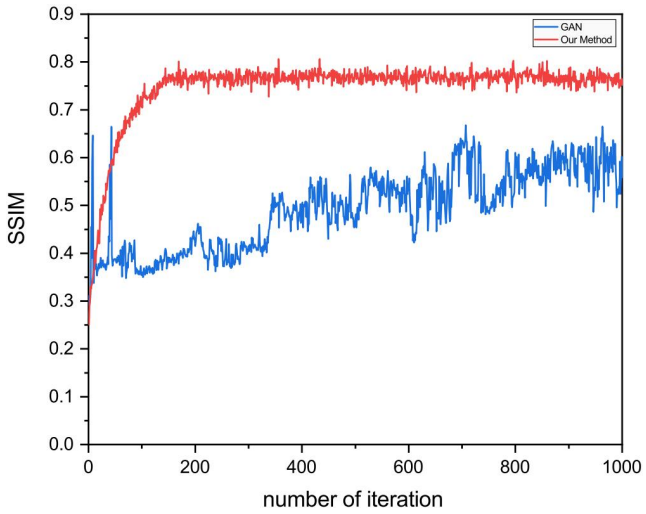

Fig. 12: Convergence curves for NC vs. SMC.

of all classes with only one model. In this manner, the common features unrelated to the disease can be reused, thus all salient global features can be captured in class-discriminative maps for different Alzheimer's stages.

Furthermore, the generation diversity with SSIM is evaluated in each iteration on the validation dataset. The convergence curves of the proposed MP-GAN and GAN are given for 4 evaluation groups: (1)NC vs. SMC, (2)NC vs. EMCI, (3)NC vs. LMCI, and (4)NC vs. AD respectively. From Fig. 12 to Fig. 15, it can be observed that the proposed MP-GAN converges faster than GAN. Meanwhile, MP-GAN performs stably in all 4 evaluation groups. On the other hand, the training of GAN is extremely unstable for NC vs. LMCI, and it can not converge for NC vs. EMCI and NC vs. AD. Again, these results are consistent with the NCC score in Fig. 10. The NCC score of GAN is low for NC vs. EMCI in Fig. 10, because GAN can't converge for NC vs. EMCI as shown in Fig. 13. These results also indicate that the proposed MP-GAN can generate diverse MR images close to the real distribution.

To evaluate the reliability of synthetic MR images generated

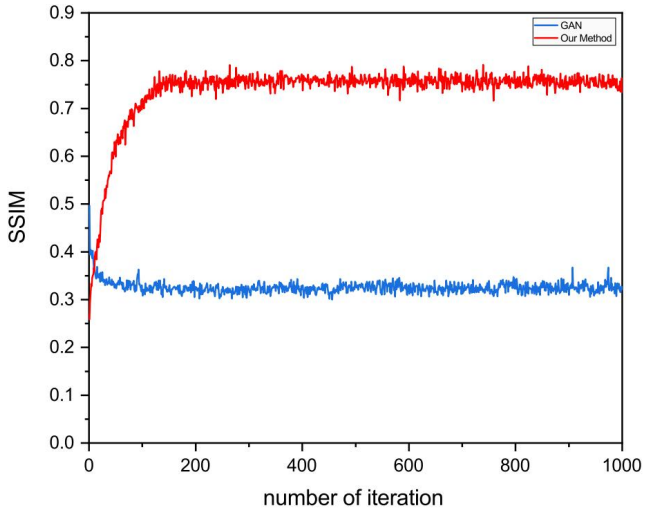

Fig. 13: Convergence curves for NC vs. EMCI.

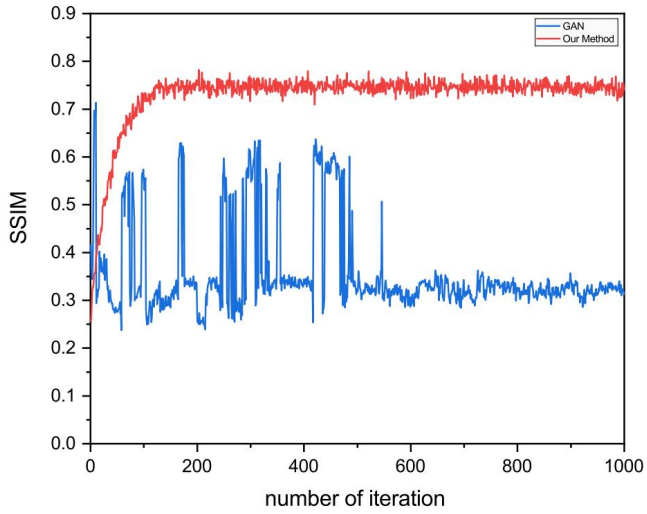

Fig. 14: Convergence curves for NC vs. LMCI.

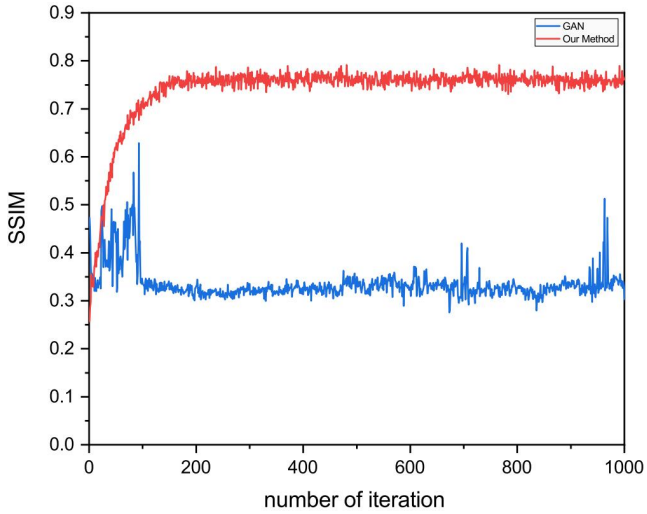

Fig. 15: Convergence curves for NC vs. AD.

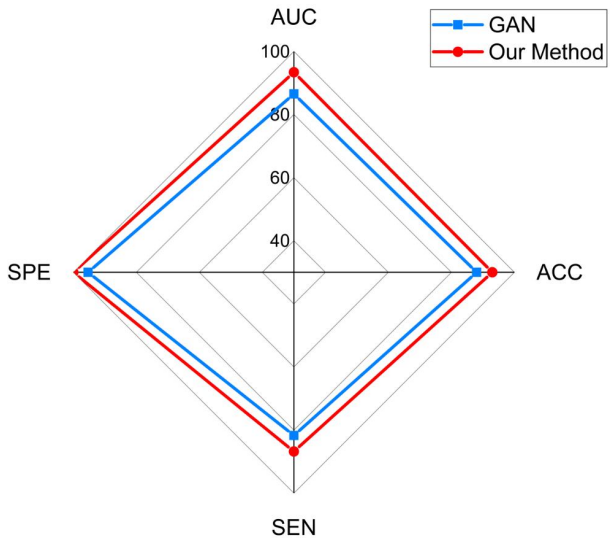

Fig. 16: The classification results of synthetic data augmentation for NC vs. SMC.

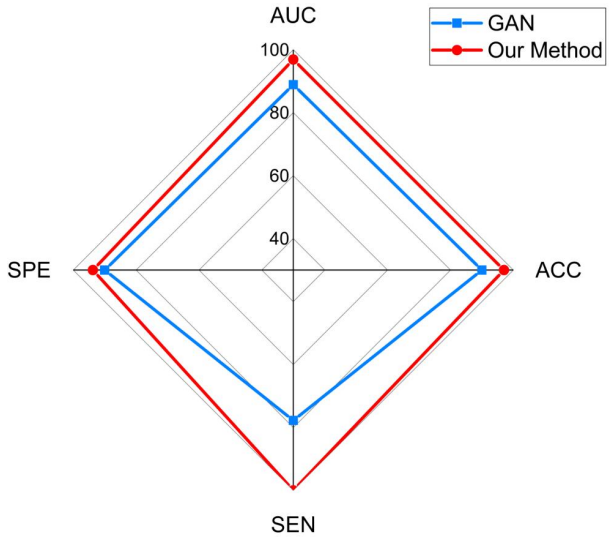

Fig. 17: The classification results of synthetic data augmentation for NC vs. EMCI.

by MP-GAN, the CNN classifier is further trained using synthetic data augmentation. More specifically, the 100 synthesized MR images of each class by MP-GAN and GAN are added to the original training set to form two new augmented training sets separately. Then the CNN model is trained on the two new augmented training sets separately for each evaluation group. During the test stage, the same test set of real MR images are used. From Fig. 16 to Fig. 19, it can be seen that adding synthesized samples by the proposed MP-GAN achieves better classification performance in terms of AUC, accuracy, specificity, and sensitivity. Overall, the synthetic data samples generated by MP-GAN can add additional variability to the original training set, which in turn leads to better performance. This implies that the synthesized MR images generated by MP-GAN not only provide meaningful visualizations but also capture the discriminative features for AD computer-aided

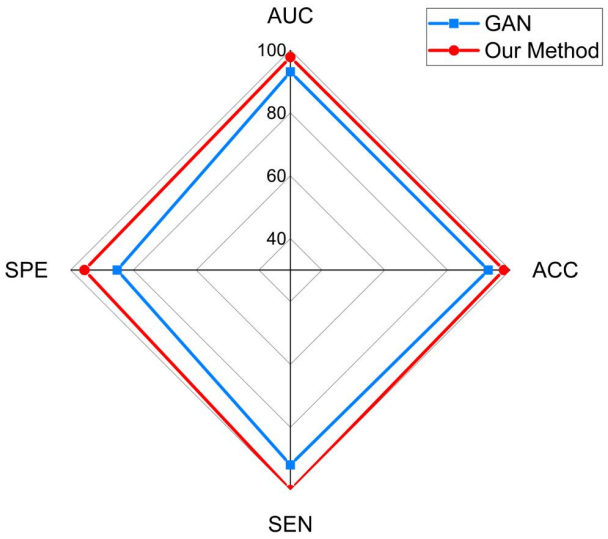

Fig. 18: The classification results of synthetic data augmentation for NC vs. LMCI.

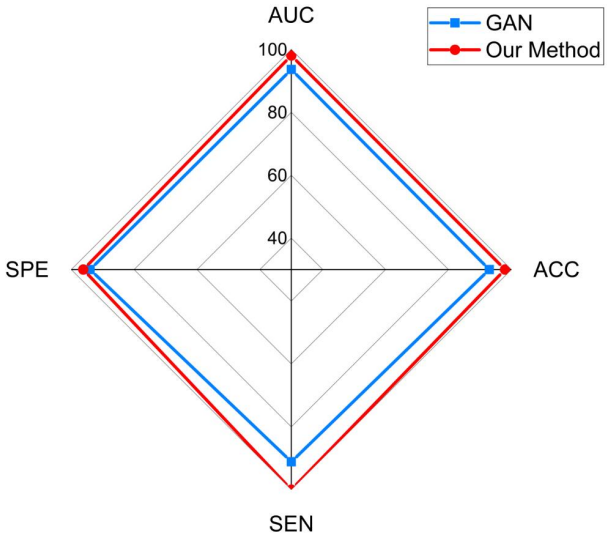

Fig. 19: The classification results of synthetic data augmentation for NC vs. AD.

diagnosis. The proposed MP-GAN can be used as an effective data augmentation method.

V. Conclusion

In this paper, a novel MP-GAN is proposed to visualize the morphological features indicating the severity of AD in whole-brain MR images. By introducing a novel multidirectional mapping mechanism into the model, MP-GAN can capture the salient global features efficiently. Thus, by utilizing the class-discriminative map from the generator, the proposed model can clearly delineate the subtle lesions via MR image transformations between the source domain and the target domain. Besides, by integrating the adversarial loss, classification loss, cycle consistency loss, and $L1$ penalty, a single generator in MP-GAN can learn the class-discriminative maps for multiple-classes. Experimental results on the public ADNI dataset has demonstrated that MP-GAN can visualize multiple lesions affected by the progression of AD accurately. Furthermore, MP-GAN may visualize some new disease-related regions that have not been investigated yet. This can be studied further to discover potential new AD biomarkers in future work.

Acknowledgment

This research is supported by National Natural Science Foundations of China under Grant No. 61872351, Shenzhen Key Basic Research Project under Grant No.JCYJ20180507182506416, ADNI(National Institutes of Health Grant U01 AG024904), DOD ADNI (Department of Defense award number W81XWH-12-2-0012) and HKRGC Grant Numbers: GRF 12200317, 12300218, 12300519 and 17201020.

References

[1] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang, S. L. Risacher, A. J. Saykin, X. Yao, and L. Shen, "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of alzheimer's disease," *Medical Image Analysis*, vol. 60, p. 101625, 2020.

[2] M. W. Bondi, E. C. Edmonds, and D. P. Salmon, "Alzheimer's disease: Past, present, and future," *Journal of the International Neuropsychological Society : JINS*, vol. 23, no. 9-10, p. 818831, October 2017.

[3] Y. Shi, H. I. Suk, Y. Gao, S. W. Lee, and D. Shen, "Leveraging coupled interaction for multimodal alzheimers disease diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 186–200, 2020.

[4] N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti, and F. C. Morabito, "Permutation jaccard distance-based hierarchical clustering to estimate eeg network density modifications in mci subjects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 10, pp. 5122–5135, 2018.

[5] M. Wang, D. Zhang, D. Shen, and M. Liu, "Multi-task exclusive relationship learning for alzheimers disease progression prediction with longitudinal data," *Medical Image Analysis*, vol. 53, pp. 111 – 122, 2019.

[6] B. Jie, M. Liu, and D. Shen, "Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease," *Medical Image Analysis*, vol. 47, pp. 81 – 94, 2018.

[7] G. S. Babu and S. Suresh, "Sequential projection-based metacognitive learning in a radial basis function network for classification problems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 2, pp. 194–206, 2013.

[8] 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 Transactions on Pattern Analysis and Machine Intelligence*, pp. 50–60, 2018.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[10] X. Yi, E. Walia, and P. Babyn, "Generative Adversarial Network in Medical Imaging: A Review," *arXiv e-prints*, Sep. 2018.

[11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision-ECCV 2014*, 2014, pp. 818–833.

[12] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain mri classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 835–838.

[13] E. Nigri, N. Ziviani, F. Cappabianco, A. Antunes, and A. Veloso, "Explainable deep cnns for mri-based diagnosis of alzheimer's disease," *arXiv e-prints*, 2020.

[14] A. Mahendran and A. Vedaldi, "Salient deconvolutional networks," in *Computer Vision – ECCV 2016*, 2016, pp. 120–135.

[15] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: visualising image classification models and saliency maps." ICLR, 2014, pp. 1–8.

[16] J. Yosinski, J. Clune, A. M. Nguyen, T. J. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," *arXiv e-prints*, vol. abs/1506.06579, 2015.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[18] M. Ancona, E. Ceolini, A. C. Öztireli, and M. H. Gross, "A unified view of gradient-based attribution methods for deep neural networks," *arXiv e-prints*, vol. abs/1711.06104, 2017.

[19] J. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," in *ICLR*, 2015.

[20] M. Bohle, F. Eitel, M. Weygandt, and K. Ritter, "Layer-wise relevance propagation for explaining deep neural network decisions in mri-based alzheimer's disease classification," *Frontiers in Aging Neuroscience*, vol. 11, p. 194, 2019.

[21] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," *arXiv e-prints*, vol. abs/1703.01365, 2017.

[22] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[23] N. M. Khan, N. Abraham, and M. Hon, "Transfer learning with intelligent training data selection for prediction of alzheimers disease," *IEEE Access*, vol. 7, pp. 72 726–72 735, 2019.

[24] C. Lian, M. Liu, L. Wang, and D. Shen, "End-to-end dementia status prediction from brain mri using multi-task weakly-supervised attention network," in *Medical Image Computing and Computer Assisted Intervention(MICCAI)*, Jan. 2019, pp. 158–167.

[25] S. Sarraf and G. Tofighi, "Classification of Alzheimer's Disease Structural MRI Data by Deep Learning Convolutional Neural Networks," *arXiv e-prints*, p. arXiv:1607.06583, Jul. 2016.

[26] C. F. Baumgartner, L. M. Koch, K. C. Tezcan, J. X. Ang, and E. Konukoglu, "Visual feature attribution using wasserstein gans," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8309–8319, 2017.

[27] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," *arXiv e-prints*, vol. abs/1703.05192, 2017.

[28] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[30] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv e-prints*, vol. abs/1607.08022, 2016.

[31] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[32] D. Gu, "3d densely connected convolutional network for the recognition of human shopping actions," in *http://dx.doi.org/10.20381/ruor-21013*, 2017.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv e-prints*, vol. abs/1502.03167, 2015.

[34] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, "Bayesian analysis of neuroimaging data in fsl," *NeuroImage*, vol. 45, pp. S173 – S186, 2009.

[35] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. D. Luca], I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano], J. M. Brady, and P. M. Matthews, "Advances in functional and structural mr image analysis and implementation as fsl," *NeuroImage*, vol. 23, pp. S208 – S219, 2004.

[36] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, p. 143155, November 2002.

[37] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, no. 2, pp. 143 – 156, 2001.

[38] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825 – 841, 2002.

[39] F. K. S. J. Ardekani BA, Bachman AH, "Corpus callosum shape changes in early alzheimer's disease: an mri study using the oasis brain database," *Brain Struct Funct*, vol. 219(1), pp. 343–352, 2014.

[40] L. Nie, L. Zhang, L. Meng, X. Song, X. Chang, and X. Li, "Modeling disease progression via multisource multitask learners: A case study with alzheimers disease," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 7, pp. 1508–1519, 2017.

[41] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T.-F. Yuan, "Detection of subjects and brain regions related to alzheimer's disease using 3d mri scans based on eigenbrain and machine learning," *Frontiers in Computational Neuroscience*, vol. 9, p. 66, 2015.

[42] J. M. Rondina, L. K. Ferreira, F. L. de Souza Duran, R. Kubo, C. R. Ono, C. C. Leite, J. Smid, R. Nitrini, C. A. Buchpiguel, and G. F. Busatto, "Selecting the most relevant brain regions to discriminate alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases," *NeuroImage: Clinical*, vol. 17, pp. 628 – 641, 2018.

[43] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, "Deep learning framework for alzheimers disease diagnosis via 3d-cnn and fsbi-lstm," *IEEE Access*, vol. 7, pp. 63 605–63 618, 2019.

[44] H. Braak and E. Braak, "Neuropathological stageing of alzheimer-related changes," *Acta neuropathologica*, vol. 82, no. 4, p. 239259, 1991.

[45] B. C. Dickerson, A. Bakkour, D. H. Salat, E. Feczko, J. Pacheco, D. N. Greve, F. Grodstein, C. I. Wright, D. Blacker, H. D. Rosas, R. A. Sperling, A. Atri, J. H. Growdon, B. T. Hyman, J. C. Morris, B. Fischl, and R. L. Buckner, "The Cortical Signature of Alzheimer's Disease: Regionally Specific Cortical Thinning Relates to Symptom Severity in Very Mild to Mild AD Dementia and is Detectable in Asymptomatic Amyloid-Positive Individuals," *Cerebral Cortex*, vol. 19, no. 3, pp. 497–510, 07 2008.

[46] F. Mrquez and M. A. Yassa, "Neuroimaging biomarkers for alzheimer's disease," *Molecular neurodegeneration*, vol. 14, no. 1, p. 21, June 2019.

[47] N. C. Fox and P. A. Freeborough, "Brain atrophy progression measured from registered serial mri: validation and application to alzheimer's disease," *JMRI*, vol. 7, no. 6, pp. 1069–1075, 1997.

[48] P. Coup, S. F. Eskildsen, J. V. Manjn, 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.

[49] T. Tong, Q. Gao, R. Guerrero, C. Ledig, L. Chen, D. Rueckert, and A. D. N. Initiative, "A novel grading biomarker for the prediction of conversion from mild cognitive impairment to alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 155–165, 2017.

[50] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers*, vol. 2, 2003, pp. 1398–1402.