



BPGAN: Brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis

Jin Zhang^a, Xiaohai He^{a,*}, Linbo Qing^a, Feng Gao^{b,c}, Bin Wang^a

^a College of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan, 610064, China

^b National Interdisciplinary Institute on Aging (NIA), Southwest Jiaotong University, Chengdu, Sichuan, 611756, China

^c External cooperation and liaison office, Southwest Jiaotong University, Chengdu, Sichuan, 611756, China

ARTICLE INFO

Article history:

Received 5 October 2021

Revised 30 January 2022

Accepted 30 January 2022

Keywords:

Alzheimer's disease

Medical imaging synthesis

MRI

PET

Generative adversarial networks

ABSTRACT

Background and Objective: Multi-modal medical images, such as magnetic resonance imaging (MRI) and positron emission tomography (PET), have been widely used for the diagnosis of brain disorder diseases like Alzheimer's disease (AD) since they can provide various information. PET scans can detect cellular changes in organs and tissues earlier than MRI. Unlike MRI, PET data is difficult to acquire due to cost, radiation, or other limitations. Moreover, PET data is missing for many subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. To solve this problem, a 3D end-to-end generative adversarial network (named BPGAN) is proposed to synthesize brain PET from MRI scans, which can be used as a potential data completion scheme for multi-modal medical image research.

Methods: We propose BPGAN, which learns an end-to-end mapping function to transform the input MRI scans to their underlying PET scans. First, we design a 3D multiple convolution U-Net (MCU) generator architecture to improve the visual quality of synthetic results while preserving the diverse brain structures of different subjects. By further employing a 3D gradient profile (GP) loss and structural similarity index measure (SSIM) loss, the synthetic PET scans have higher-similarity to the ground truth. In this study, we explore alternative data partitioning ways to study their impact on the performance of the proposed method in different medical scenarios.

Results: We conduct experiments on a publicly available ADNI database. The proposed BPGAN is evaluated by mean absolute error (MAE), peak-signal-to-noise-ratio (PSNR) and SSIM, superior to other compared models in these quantitative evaluation metrics. Qualitative evaluations also validate the effectiveness of our approach. Additionally, combined with MRI and our synthetic PET scans, the accuracies of multi-class AD diagnosis on dataset-A and dataset-B are 85.00% and 56.47%, which have been improved by about 1% and 1%, respectively, compared to the stand-alone MRI.

Conclusions: The experimental results of quantitative measures, qualitative displays, and classification evaluation demonstrate that the synthetic PET images by BPGAN are reasonable and high-quality, which provide complementary information to improve the performance of AD diagnosis. This work provides a valuable reference for multi-modal medical image analysis.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

As the foundation for precision medicine, medical images have developed an imperative component of medical research. With the development of modern medical imaging equipment, medical imaging has arisen in various modalities. Multi-modality medical images assist radiologists and clinicians in the early screening and

diagnosis of diseases more objectively and accurately by providing intuitive insight into the human body's interior. Numerous clinical researches [1,2], such as cancer diagnosis and brain disease diagnosis [3–5], desire high-quality multi-modality medical images to achieve an effective diagnosis because multi-modality imaging reflects various pathologies and achieves the purpose of complementary advantages.

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that seriously affects the normal quality of life for older people and their family caregivers. AD early screening, AD diagnosis, and AD intervention are very significant since AD gets

* Corresponding author.

E-mail address: hxh@scu.edu.cn (X. He).

worse over time [6]. With the rapid development of neuroimaging technology, neuroimaging examination has become an indispensable method for AD diagnosis. Compared with the subjective and preliminary neuropsychological examination using multiple scales, neuroimaging examination can be used not only for the early diagnosis of AD but also to distinguish it from other types of dementia. Recently, various methods based on multi-modal neuroimaging have been proposed to discover biomarkers that contribute to the early diagnosis and prediction of AD [7–9]. In particular, the joint analysis of positron emission tomography (PET) and magnetic resonance imaging (MRI) has been recognized as a useful method for the screening and diagnosis of AD [10–12]. It is evident that the combination of PET and MRI scans provides structural and functional information related to AD, thus improving the effectiveness of diagnosis.

Although PET is a relatively new modality compared to non-invasive MRI, its use is steadily increasing. However, it is not always feasible to acquire both MRI and PET scans for each patient owing to high expense, lack of PET scanner, radiation exposure, and enhanced lifetime cancer risk. Moreover, in the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [13], there are about twice as many subjects with MRI data as subjects with PET data, which is missing significantly. One approach is to discard subjects with stand-alone MRI and use only modality-complete subjects to train diagnostic models. This strategy significantly degrades the number of samples and ignores the useful information provided by data-missing subjects, thus reducing the performance. A more promising strategy is the complement of the missing PET scan, which can make full use of all available subjects. The clinical importance of PET images and the difficulty in providing PET images raise a potential demand for alternative and inexpensive PET images. Recent studies [14–18] attempt to utilize the information from MRI to generate PET scans for the multi-modality data complement. Furthermore, the value of MRI and synthetic PET scans in improving the efficacy of disease diagnosis is further discussed. How to use the existing modal medical images and accurately synthesize the required modal images by computer vision technology is a recent research focus.

Deep learning has made significant strides in many fundamental computer vision (CV) applications, so the existing multi-modality medical images synthesis methods are mainly learning-based. As far as we know, the convolutional neural networks (CNN) model proposed by Cai et al. [19] was the first one to predict the missing PET data from MRI data, which used a network to learn a non-linear relationship between MRI and PET. Gao et al. [20] designed a RIED-Net to learn the mapping between MRI and PET, which improved the generation performance. There are blurry estimations of generated images in these methods since the L_1 or L_2 loss function is usually utilized. Recently, image generation, image restoration, and so on [21–26] have been widely studied and achieved great success, which further promotes the progress of medical imaging synthesis [15,16,27]. Pan et al. [28] proposed a 3D conditional generative adversarial network (cGAN) framework to model bi-directional mappings between MRI and PET scans for AD diagnosis. Yaakub et al. [29] designed a 3D generative adversarial network (GAN) based on residual connection learning the mapping from MRI to PET to support the clinical evaluation of patients with focal epilepsy. To achieve accurate lesion detection, Ben et al. [30] presented an architecture for generating PET images from CT scans by combining a fully convolutional network (FCN) and a cGAN. Pan et al. [31] presented a disease-image-specific deep learning approach to encourage the consistency of the synthetic neuroimages with their respective real neuroimages. Lin et al. [32] designed a reversible generative adversarial network (RevGAN) model to generate the PET data. Sikka et al. [33] explored a global and local aware method with a multi-path archi-

ture that enhanced global structural integrity and local detail fidelity of synthetic PET images.

Although the above frameworks have been proven to be remarkable, they suffer from some problems for the multi-modality medical images synthesis. Firstly, due to the complex spatial structure of medical images, there are obvious gaps between synthetic and real images in terms of semantics, resolution, and edge information of human tissues. Secondly, most of the previous works are based on 2D slice-level [20,34,35], which synthesize images along with one of the axial, coronal, and sagittal planes independently. Since MRI and PET scans are three-dimensional, these methods lead to discontinuous estimation and spatial information loss, unfavorable for medical image synthesis. In addition, the 2D slice-level methods require selecting slices that are utilized as input since not all of them are informative. Some 3D patch-level methods [36–38] obtain the entire 3D estimated image by overlapping the generated image patches averagely. However, these methods are insufficient to learn the global context features among voxels, which particularly influences the generation capability of networks. Furthermore, majority of the previous studies [31–33] only focused on discriminating AD from cognitively normal (CN). However, mild cognitive impairment (MCI) is considered an intermediate state that is crucial for early diagnosis. Further research is needed for the multi-class classification (AD vs. CN vs. MCI) task. To address these issues, a novel 3D end-to-end network BPGAN is proposed for brain MRI-to-PET synthesis, providing technical support for the multi-modal missing medical data completion.

In the proposed work, two data partitioning methods are used to build datasets for model training, validation, and testing, which evaluate insights into the performance of BPGAN in different medical scenarios. On this basis, a novel architecture BPGAN is designed to synthesize PET from MRI by taking full advantage of multiple $1 \times 1 \times 1$ convolutions. The proposed BPGAN can learn the internal relation between MRI and PET, generate high-quality PET scans at pixel and semantic levels, and reduce the fuzziness of the synthetic images. Moreover, a 3D gradient profile (GP) loss is introduced based on the traditional adversarial loss and pixel-wise loss to reconstruct texture information in PET images. At the same time, the structural similarity index measure (SSIM) loss is added to the total loss function to preserve the contrast information of generated images. To verify the superiority of our proposed method, three evaluation metrics, including mean absolute error (MAE), peak-signal-to-noise-ratio (PSNR) and SSIM, are introduced to perform the quantitative evaluation. The quantitative results demonstrate that BPGAN can synthesize high-quality PET scans with three optimal metrics. We make a qualitative evaluation of the synthetic PET scans, which further proves that PET scans generated by our proposed model have a high similarity with ground truth. Furthermore, the AD classification results of stand-alone MRI scan, stand-alone synthetic PET scan, and combination of MRI and synthetic PET images show that MRI-to-PET synthesis is of great significance for multi-modal AD diagnosis.

Overall, the main contributions for this study are as follows.

1. A novel 3D end-to-end generative adversarial network BPGAN is proposed, which can effectively synthesize realistic and diverse brain PET scans from the corresponding brain MRI scans.
2. A hybrid loss combining 3D GP loss, SSIM loss, adversarial loss, KL-divergence constraint, and L_1 loss is introduced to supervise the training process of brain PET scans synthesis on multiple levels.
3. Two alternative data splitting strategies are explored to study the impact on the MRI-to-PET synthesis task and further analyze their applicability in different medical scenarios.
4. The feasibility of using BPGAN as an effective data completion method for multi-modal AD diagnosis is further explored. The

Table 1
Parameter selection specifications for ADNI medical image data.

	MRI	PET
Image	T1-weighted structural MRI, 1.5T, 3D	Fluorodeoxyglucose (FDG) PET
Pre-processed method	Gradwarp, B1 non-uniformity, N3 bias field corrected	Co-registered, Averaged
Project phase	ADNI1, ADNI2, ADNI-GO	ADNI1, ADNI2, ADNI-GO

Table 2
Demographics, MMSE, CDR, and CDR-SOB of dataset-A that split by random sampling.

Split (random sampling)	Diagnosis	Scan-pairs	Age	MMSE	CDR	CDR-SOB
Training	CN	418	75.76(6.18)	29.00(1.87)	0.02(0.16)	0.10(0.37)
	MCI	693	73.64(7.72)	27.48(2.74)	0.48(0.15)	1.57(1.07)
	AD	275	74.66(7.38)	21.91(4.17)	0.89(0.40)	5.27(2.42)
Validation	CN	15	76.83(6.13)	29.06(1.00)	0.03(0.12)	0.10(0.20)
	MCI	22	74.29(6.14)	27.86(1.63)	0.48(0.10)	1.52(0.95)
	AD	9	76.00(5.59)	22.33(3.68)	0.83(0.47)	4.94(2.80)
Testing	CN	98	76.97(6.60)	29.03(1.22)	0.02(0.19)	0.12(0.43)
	MCI	141	73.54(7.89)	27.60(3.11)	0.50(0.10)	1.58(0.99)
	AD	61	75.85(6.99)	21.21(3.93)	0.90(0.42)	5.42(2.40)

Values are reported in terms of mean(standard deviation).

classification results indicate that the high-quality PET images generated by BPGAN contain disease information, which is beneficial to AD diagnosis.

This paper proceeds as follows. Section 2 introduces a detailed description of our approach. Experimental results are depicted in Section 3, which are discussed in the following Section 4. Section 5 gives a conclusion.

2. Methods

2.1. Data acquisition and preprocessing

2.1.1. Data acquisition

AD is a neurological and irreversible brain disease. The largest publicly available neuroimaging dataset for the AD diagnosis is ADNI [13], which comprises T1-weighted structural MRI, Fluorodeoxyglucose (FDG) PET imaging, and other imaging of different modalities.

In total, 1732 image pairs from 873 subjects from the ADNI database are used for our experiments, and their details are shown in Table 1. The data for each subject includes a baseline scan and some follow-up scans at multiple time points after the baseline, with each scan having a subject ID and a session ID. Therefore, each subject has more than one image pair, and each image pair of one subject has the same subject ID and different session ID. The image pairs from subjects are in three categories: cognitively normal (CN), mild cognitive impairment (MCI), and AD. Each subject has images in two modalities (MRI and PET). This paper explores two alternative ways of splitting data into training, validation and testing sets to investigate the performance of the proposed model under various medical scenarios.

Firstly, we consider randomly splitting the data from the scans, ignoring the subject ID, as random sampling. This split method avoids the situation where the model overfits the individual brain structure and ignores critical differences in feature representation among the different disease stages. The demographics, minimal state examination (MMSE), global clinical dementia rating (CDR), and clinical dementia rating scale sum of boxes (CDR-SOB) scores of the dataset, which is named dataset-A, are summarized in Table 2. MMSE, CDR and CDR-SOB are the clinical screening instruments to assess the overall cognitive function, mainly used to ensure that scale distribution between training, validation, and testing sets are not significantly different.

Secondly, we refer to producing the splits by subject ID, named dataset-B. All available scan-pairs from one subject are allocated to only either training or validation or testing set. This strategy is more close to the real time clinical scenario. In practice, new subjects usually have no scan records. Therefore, the data split strategy by subject ID needs to be studied, where each subject in the testing set and validation set has no data in the training set. This strategy is suitable for patients with no recorded PET scans. Summary of the participant demographics, MMSE, CDR, and CDR-SOB scores are demonstrated in Table 3.

2.1.2. Data preprocessing

For data preprocessing, we utilize the Clinica software platform¹ [39] developed by ARAMIS Lab² for data preprocessing, which supports FSL³, Statistical Parametric Mapping⁴ (SPM), and FreeSurfer [40]. First, two modal scans, including MRI and PET, are registered into a Dartel template [41]. Then, the two modalities are spatially aligned to the same standardized Montreal Neurological Institute (MNI) coordinate space to align rigidly with each other, as illustrated in Fig. 1. The inputs of Clinica are the ADNI scans converted to BIDS format. To prevent the interference of redundant information, non-brain tissues are removed from MRI scans and PET scans. Moreover, we resize MRI and PET scans to $128 \times 128 \times 128$ voxels to reduce computational cost. The pixel values range of the images is further normalized to $[-1, 1]$ to avoid the problem of gradients exploding and convergence.

2.2. Overview of proposed method

The goal of PET synthesis from MRI is to generate the perceptually realistic PET scan as close as possible to the real PET scan. Inspired by BicycleGAN [25], we follow the adversarial training strategy and propose a novel 3D end-to-end network BPGAN to improve the quality of synthesized PET images. BPGAN models a mapping function $G(x)$, as represented in Fig. 2. Assuming a dataset where a subject S consists of two modalities images $\{x, y\}$ is given. In the training set, subjects have the paired brain MRI scans x and PET scans y as $\{x_i, y_i\}_{i=1}^N$, while the testing set is composed of subjects with only MRI scans x . The training set contains MRI-PET pairs from subjects in all categories (e.g., AD, CN,

¹ <http://www.clinica.run>

² www.aramislab.fr

³ <https://fsl.fmrib.ox.ac.uk/fsl>

⁴ <https://www.fil.ion.ucl.ac.uk/spm>

Table 3
Demographics, MMSE, CDR, and CDR-SOB of dataset-B that split by subject ID.

Split (subject ID)	Diagnosis	Subjects	Scan-pairs	Age	MMSE	CDR	CDR-SOB
Training	CN	184	411	76.01(6.37)	28.98(1.88)	0.02(0.17)	0.10(0.40)
	MCI	347	673	73.53(7.84)	27.51(2.75)	0.49(0.13)	1.58(1.06)
	AD	146	292	75.24(7.12)	21.72(4.20)	0.90(0.42)	5.30(2.43)
Validation	CN	8	18	76.19(4.03)	29.27(0.80)	0.00(0.00)	0.11(0.21)
	MCI	17	29	72.05(5.36)	27.24(1.92)	0.43(0.28)	1.84(1.06)
	AD	6	9	72.11(4.42)	20.44(4.24)	0.83(0.24)	5.06(2.07)
Testing	CN	53	102	75.98(6.22)	29.05(1.22)	0.02(0.16)	0.09(0.32)
	MCI	82	154	74.48(7.49)	27.52(3.09)	0.48(0.17)	1.49(0.99)
	AD	30	44	73.07(8.49)	22.46(3.35)	0.84(0.35)	5.27(2.45)

Values are reported in terms of mean(standard deviation).

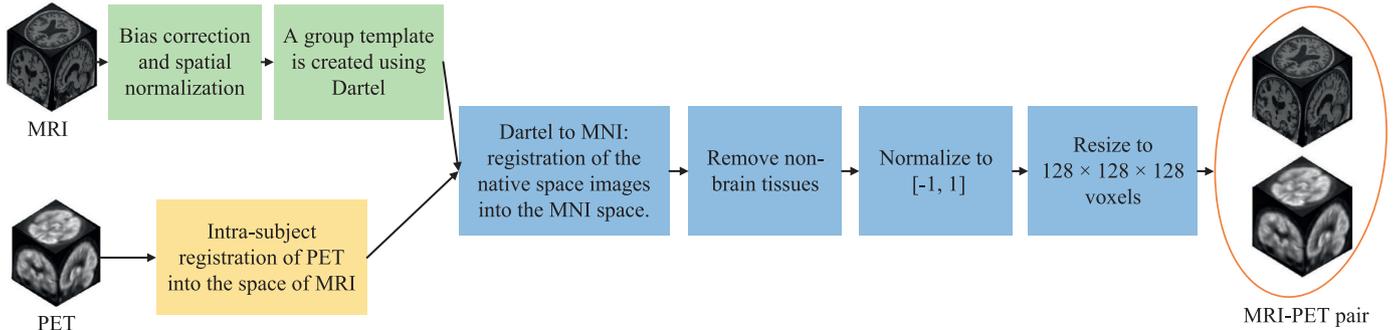


Fig. 1. The illustration of data preprocessing.

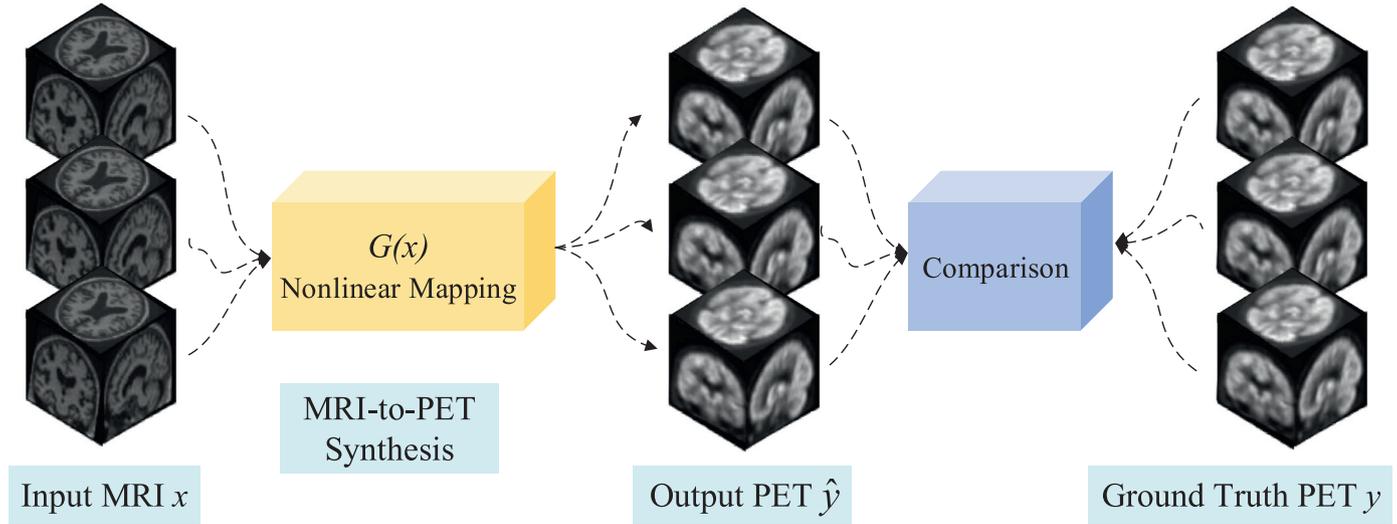


Fig. 2. Illustration of MRI-to-PET synthesis. The left column is input MRI. The synthesized PET scans by G are compared to the ground truth PET scans.

and MCI), so there are no unknown diseases in the inference stage. Our goal is to learn the complex non-linear mapping between x and y in a supervised manner to achieve the multi-modality missing data completion. $G(x)$ can generate the visually appealing PET scans \hat{y} while possessing the diversity. It can be formulated as Eq. 1:

$$\hat{y} = G(x, \theta) \tag{1}$$

The network's parameters θ are optimized by minimizing a particularly proposed objective function.

2.3. Proposed framework

To produce high-quality and diverse results, we introduce bijective consistency [25] between the latent encoding and synthetic PET scans. Latent vectors and the spatial and semantic informa-

tion from PET images are jointly utilized for MRI-to-PET synthesis through conditional variational autoencoder GAN (cVAE-GAN) [42,43] and conditional latent regressor GAN (cLR-GAN) [44,45]. The proposed BPGAN architecture is shown in Fig. 3, whose learning process is as follows.

- The goal of cVAE-GAN is to achieve $y \rightarrow z \rightarrow \hat{y}$. The generator G uses both the distribution $Q(z | y)$ and the input MRI scan x to synthesize the desired PET scan \hat{y} . The ground truth PET scan y is directly encoded to the latent code z by an encoder E to make z useful. This model is briefly regarded as the reconstruction of the real PET scan y . With the latent encode $Q(z | y)$ as a prior condition, the training process is guided in pairs along with the input MRI scan x , desiring the synthetic PET scan \hat{y} close to the real PET scan y . This process resembles an autoencoder [46]. Extending it to a conditional scenario, the

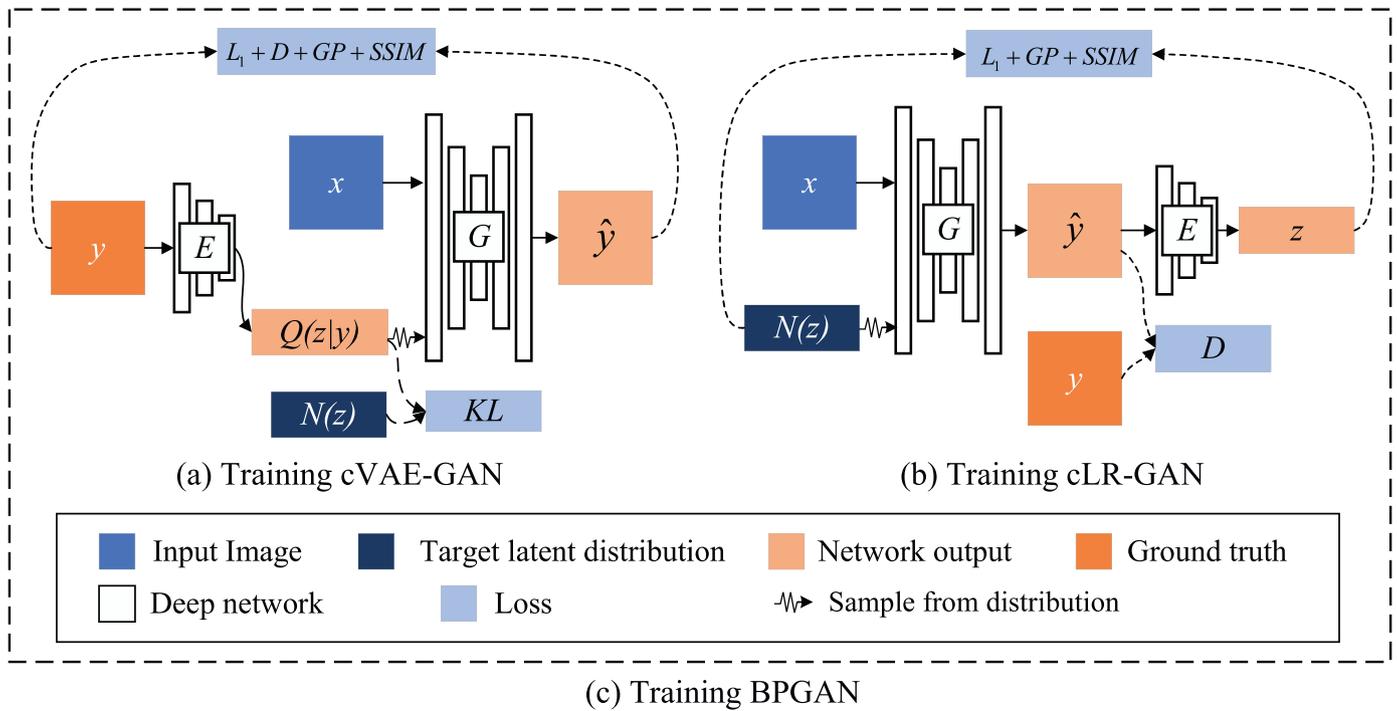


Fig. 3. BPGAN framework.

distribution $Q(z | y)$ of latent code z is a Gaussian assumption [25], $Q(z | y) \triangleq E(y)$. During model learning, sampling $z \sim E(y)$ is permitted to be direct back-propagation [43] with our proposed objective function.

- In cLR-GAN, the generator G uses the latent code z to map input MRI scan x into the synthetic PET scan \hat{y} , where the generated sample \hat{y} is encoded by E to produce a latent vector. This operation can be represented as $z \rightarrow \hat{y} \rightarrow \hat{z}$. The noise vector $N(z)$ starts from a randomly drawn code z and attempts to reconstruct it with $\hat{z} = E(G(x, z))$. Since the noise vector is randomly sampled, the synthetic PET scan \hat{y} is required to be realistic but not necessarily close to the real PET scan y . Note that \hat{z} generated by encoder E is a point estimate rather than a distribution [25]. During the training process, G , D and E are continuously optimized with our objective function.
- BPGAN combines constraints in both directions during the training process [25], aiming to make full use of both cycles.

In terms of network architecture, our proposed BPGAN consists of three components, which are illustrated in Fig. 4: (1) generator, (2) discriminator, and (3) encoder. First, we follow the invertible connection between synthetic output and latent code [25], which facilitates generator G to produce realistic and diverse PET scans. Then we design an advanced multiple convolution U-Net (MCU) generator that associates U-Net with multiple $1 \times 1 \times 1$ convolution sequences to synthesize PET scans with detailed information from the relevant brain MRI scans. It should be emphasized that 3D convolution layers are utilized to optimally capture high-level semantic information, model the spatial structure features of PET scans, and eliminate the slice discontinuity resulting from 2D networks. Moreover, two patch-level discriminators at different scales are modeled with the adversarial learning strategy. Finally, MCU is optimized by incorporating 3D GP loss and SSIM loss into the objective function to improve the visual quality and pixel-level quality of the generated PET scans. The details of the architecture and the extensive loss function are demonstrated in Subsections 2.3 and 2.4, respectively.

2.3.1. Generator

MRI and PET scans are different representations of the same underlying information, and the correspondence between them is learned through a generator. A generator based on the ordinary CNN architecture may lose low-level spatial information with the deeper network, which is not beneficial to the synthesis of PET scans. U-Net [47] addresses the issue effectively by adopting contraction paths to realize the fusion of different levels of features. BicycleGAN [25] and Pix2Pix [23] models based on U-Net [47] perform well in the natural image generation task. Meanwhile, U-Net architecture [47] is feasible for medical pixel-level prediction tasks [28,36,48]. However, the original U-Net [47] has limitations. The spatial resolution of medical images is usually not high. With the increase of the number of down-sampling, the resolution of images gradually decreases, and increasing the network depth at the bottom of U-Net is limited. Because MRI and PET are both three-dimensional, increasing network depth at the top of U-Net results in a large number of parameters and computation time. To balance the depth of the network and the amount of computation, MCU is proposed to capture deeper features without adding a lot of extra complexity.

The main idea of MCU generator diagrammed in Fig. 4 is to combine the advantages of U-Net [47] and multiple convolution sequence. MCU has a multiple convolution module, 7 convolution layers, and 7 transposed convolution layers. All convolution operations except the multiple convolution module are down-sampling or up-sampling by a factor of 2 with the filter kernel size of $4 \times 4 \times 4$. It is noticed that the convolution layer (or deconvolution layer) is followed by instance normalization (IN) rather than Batch Normalization (BN). And the Leaky Rectified Linear Unit (Leaky ReLU) with a slope of 0.2 is selected as the activation function. In the last layer, the Tanh function is utilized without the normalization layer.

The proposed multiple convolution module is inserted to the bottom of U-Net to achieve high-level semantic feature extraction without significantly adding the additional time cost and computational cost, as shown in Fig. 5. Firstly, $1 \times 1 \times 1$ convolution is

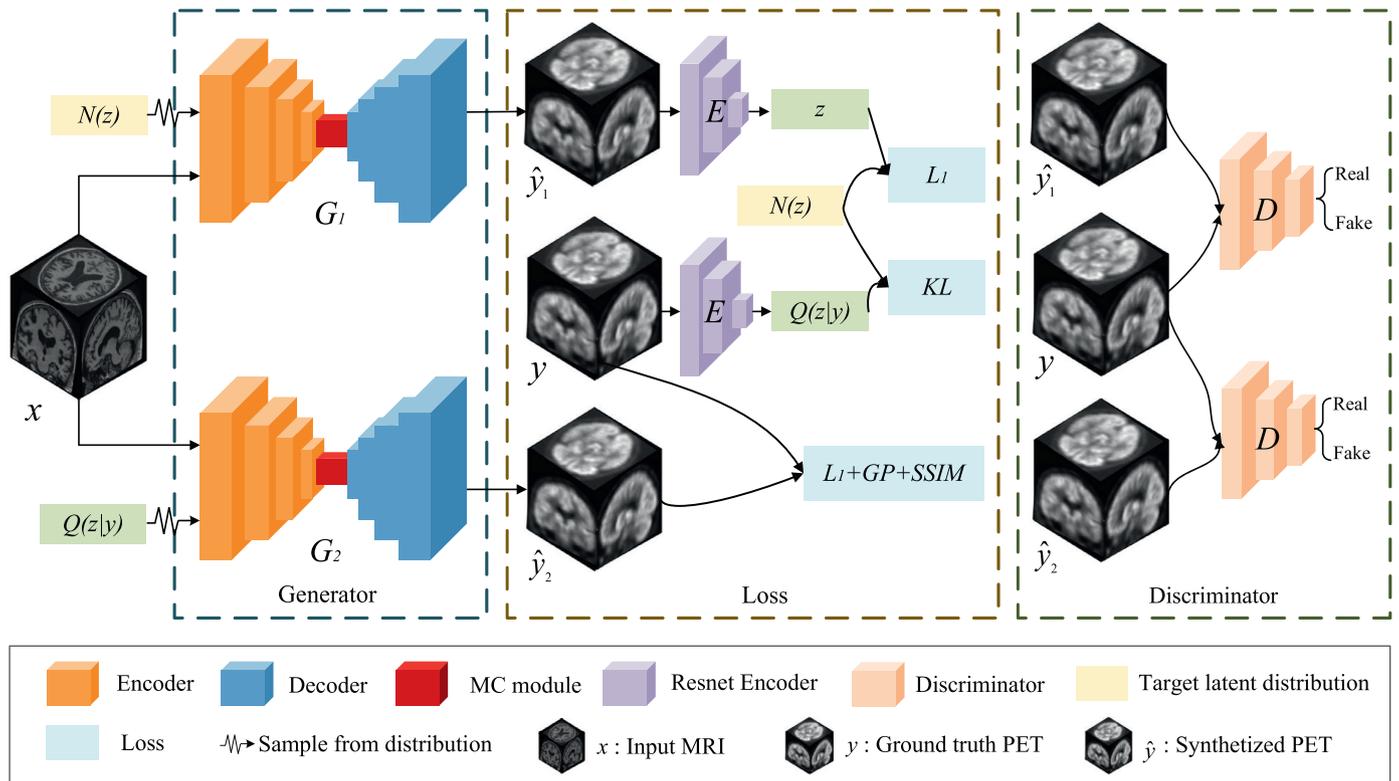


Fig. 4. Overview of the proposed method.

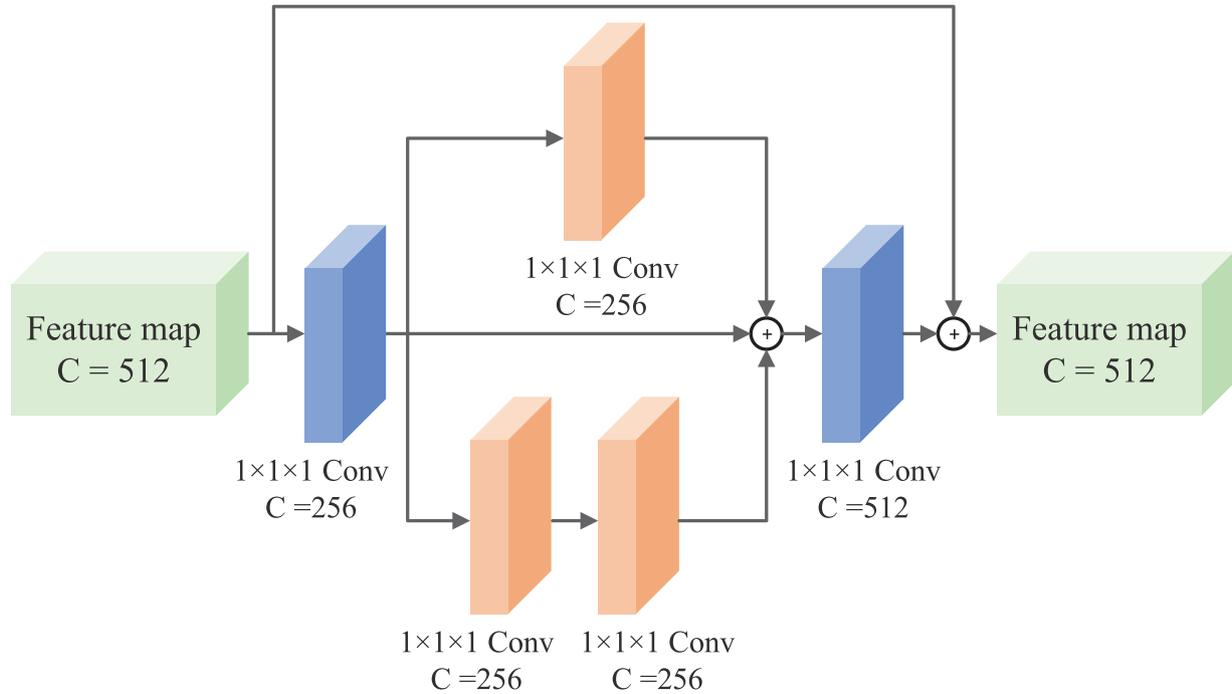


Fig. 5. Multiple convolution module. The \oplus denotes matrix addition and c represents the channel number.

used as a dimension reduction module to remove computational bottlenecks. Multiple $1 \times 1 \times 1$ convolutions are then infused in the cascade and parallel mode. In cascade mode, due to the gradual increase of the number of convolution layers, it can increase the network depth and effectively improve the learning capacity of the network [49]. In parallel mode, the outputs of three cascade branches that accept the same input are added. More diverse information and multi-level feature maps are captured from the image.

Moreover, the projection created by $1 \times 1 \times 1$ convolution directly increases the channel number of feature maps to the same number as input feature maps. And it allows complex and learnable interactions of cross-channel information. Finally, the original feature maps are added directly with other features, like the shortcut mechanism [50], to prevent the gradient from vanishing as follows:

$$X' = X + T(C_1(T(X)) + C_2(T(X)) + C_3(T(X))) \quad (2)$$

where C represents the cascade operation and T is the transformation of cross-channel information. X and X' are the input and output feature maps, respectively.

2.3.2. Discriminator

Discriminator D is a classic CNN architecture that contains 3D convolution operation, 3D normalization operation, Leaky ReLU, and 3D max-pooling operation. For four convolution layers with the filter kernel size of $4 \times 4 \times 4$ in D , the channels are 32, 64, 128 and 256, respectively. We employ two patch-level discriminators at different scales [25] to determine whether each $64 \times 64 \times 64$ and $128 \times 128 \times 128$ patch of input scan is real or not, taking either a real PET scan or a synthesized one as input.

2.3.3. Encoder

Resnet [50] architecture is introduced to encode the images better instead of plain CNN in the encoder E . E consists of multiple 3D convolutional layers followed by 3D normalization and ReLU, 3D average-pooling operation, and several residual connections. It needs to be emphasized that 3D IN is adopted instead of 3D BN because 3D IN ensures the independence of each image instance and is more suitable for the generation task.

2.4. Proposed total loss function

Success in MRI-to-PET synthesis requires semantic reasoning and shape translation. Specifically, although the synthetic PET and MRI scans are quite different in appearance, they should be similar at the semantic level. To generate high-quality and realistic PET scans, an objective function is designed to optimize the proposed model, including five types of components: adversarial loss, KL-divergence constraint (KL loss), L_1 loss, 3D GP loss, and SSIM loss.

The meaning of adversarial loss is to match the distribution of the synthesized PET image \hat{y} with that of the real PET image y . \mathcal{L}_{GAN}^1 and \mathcal{L}_{GAN}^2 , which are the adversarial losses of cVAE-GAN and cLR-GAN [25], are written as follows:

$$\mathcal{L}_{GAN}^1(G, D, E) = \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim p(\mathbf{x}, \mathbf{y})} [\log(D(\mathbf{x}, \mathbf{y}))] + \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim p(\mathbf{x}, \mathbf{y}), \mathbf{z} \sim E(\mathbf{y})} [\log(1 - D(\mathbf{x}, G(\mathbf{x}, \mathbf{z})))] \quad (3)$$

$$\mathcal{L}_{GAN}^2(G, D) = \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim p(\mathbf{x}, \mathbf{y})} [\log(D(\mathbf{x}, \mathbf{y}))] + \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}), \mathbf{z} \sim p(\mathbf{z})} [\log(1 - D(\mathbf{x}, G(\mathbf{x}, \mathbf{z})))] \quad (4)$$

L_1 loss encourages the generated PET scans to match the real PET scans at the pixel level and stabilizes the training. We utilize \mathcal{L}_1^1 and \mathcal{L}_1^2 as follows:

$$\mathcal{L}_1^1(G) = \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim p(\mathbf{x}, \mathbf{y}), \mathbf{z} \sim E(\mathbf{y})} \|\mathbf{y} - G(\mathbf{x}, \mathbf{z})\|_1 \quad (5)$$

$$\mathcal{L}_1^2(G, E) = \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x}), \mathbf{z} \sim p(\mathbf{z})} \|\mathbf{z} - E(G(\mathbf{x}, \mathbf{z}))\|_1 \quad (6)$$

To ensure the encoded vector from the real PET $\mathbf{z} \sim E(\mathbf{y})$ has a similar distribution with the latent vector sampled from a Gaussian distribution, KL loss is enforced following the BicycleGAN [25] to minimize their difference as follows:

$$\mathcal{L}_{KL}(E) = \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y})} [\mathcal{D}_{KL}(E(\mathbf{y}) \|\mathcal{N}(0, I))] \quad (7)$$

where $\mathcal{D}_{KL}(p\|q) = -\int p(z) \log \frac{p(z)}{q(z)} dz$.

It is essential for the PET image synthesis task to preserve high-frequency details while transferring as many low-frequency contents as possible. Unfortunately, although the reconstruction loss can capture the overall structure, it is ineffective in reconstructing high-frequency contents and low-frequency details in the synthetic PET image. To address this issue, we introduce some loss functions.

First, a 3D Gradient Profile (GP) loss is proposed to generate PET images with better structural and perceptual contents while

being more informative, which is an extension of 2D GP loss [51]. The distances in shape between the real and synthetic PET images are approximated by considering the gradient of image patches on the sagittal, coronal, and transverse planes. We regard the pixels in the horizontal and vertical directions of image patches as a vector and calculate the Euclidean distance between the corresponding vectors in real and synthetic PET image patches as the similarity. 3D GP loss considers the spatial profile of a PET image patch as a vector. The similarity between them is measured over each image patch along the three planes in an induced vector space as:

$$\mathcal{L}_D = \frac{1}{D} \sum_D \left(\frac{1}{H} \text{trace}(G(\mathbf{x}; \theta)_D \cdot \mathbf{y}_D^T) + \frac{1}{W} \text{trace}(G(\mathbf{x}; \theta)_D^T \cdot \mathbf{y}_D) \right) \quad (8)$$

$$\mathcal{L}_H = \frac{1}{H} \sum_H \left(\frac{1}{D} \text{trace}(G(\mathbf{x}; \theta)_H \cdot \mathbf{y}_H^T) + \frac{1}{W} \text{trace}(G(\mathbf{x}; \theta)_H^T \cdot \mathbf{y}_H) \right) \quad (9)$$

$$\mathcal{L}_W = \frac{1}{W} \sum_W \left(\frac{1}{H} \text{trace}(G(\mathbf{x}; \theta)_W \cdot \mathbf{y}_W^T) + \frac{1}{D} \text{trace}(G(\mathbf{x}; \theta)_W^T \cdot \mathbf{y}_W) \right) \quad (10)$$

where $(\cdot)^T$ represents transpose and $\text{trace}(\cdot)$ is the sum of the diagonal elements in a square matrix. D , H and W are respectively the depth, height and width of the PET scan. The first and second terms compute similarity among vertical and horizontal profiles of an image patch pair $\{x, y\}$ of size $D \times W$ in Eq. 8, respectively.

The 3D GP loss is computed in the image gradient space with Eq. 8, Eq. 9, and Eq. 10 to drive the content and shape of PET images along with sagittal, coronal, and transverse planes, which is dependent on edge information, like the following:

$$\mathcal{L}(x, y | G(\mathbf{x}; \theta), \mathbf{y}) = \frac{1}{3} (\mathcal{L}_D + \mathcal{L}_H + \mathcal{L}_W) \quad (11)$$

$$\mathcal{L}_{GP}(x, y) = \mathcal{L}(x, y | \nabla G(\mathbf{x}; \theta), \nabla \mathbf{y}) \quad (12)$$

Second, we integrate SSIM loss into our training loss to learn the structural information from three different perspectives: brightness, contrast, and structure [52]. It is worth noting that SSIM needs to take a minus sign as a loss.

$$\mathcal{L}_{SSIM} = -\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (13)$$

where μ_x , μ_y and σ_x , σ_y are the mean and standard deviations of MRI x and PET y respectively, σ_{xy} is their covariance, C_1 and C_2 are used to avoid dividing by zero.

As mentioned above, the aggregate objective function proposed in this paper can be expressed as Eq. 14. Our experimental results show that the proposed loss components effectively minimize the disparity between the distribution of generated images and the distribution of real images in the brain MRI-to-PET synthesis task.

$$G^*, E^* = \arg \min_{G, E} \max_D \mathcal{L}_{GAN}^1(G, D, E) + \mathcal{L}_{GAN}^2(G, D) + \lambda_1 \mathcal{L}_1^1(G) + \lambda_2 \mathcal{L}_1^2(G, E) + \lambda_{KL} \mathcal{L}_{KL}(E) + \lambda_{GP} \mathcal{L}_{GP}(G, E) + \lambda_{SSIM} \mathcal{L}_{SSIM}(G, E) \quad (14)$$

where the hyper-parameters λ_1 , λ_2 , λ_{KL} , λ_{GP} , and λ_{SSIM} is the weight of each part, which is to be excavated in the abundant experiments. They control the relative importance of individual loss terms.

3. Results

3.1. Experimental settings

The hyper-parameters for the total loss function of the proposed model are set as the fixed values before training. Abundant experiments demonstrate that when the hyper-parameters are set as $\lambda_1 = 10.0$, $\lambda_2 = 0.5$, $\lambda_{KL} = 0.01$, $\lambda_{SSIM} = 1.0$, and $\lambda_{GP} = 0.02$, our

method achieves superior performance. Adam [53] optimizer is used with an initial learning rate of 2×10^{-4} . The training process of all models for PET synthesis in this paper lasts for 230 epochs with the learning rate linearly decaying to 0 over the last 30 epochs after maintaining 2×10^{-4} in the first 200 epochs. The batch size is set to 6. Our proposed method is implemented by Python with Pytorch framework on a platform with an NVIDIA GeForce GTX 3090 GPU.

For AD diagnosis, the classification model is trained with Adam [53] optimizer with a batch size of 6. The learning rate is set by using a cosine annealing schedule with the initial learning rate 1×10^{-4} . We train the model for 150 epochs on an NVIDIA GeForce GTX 3090 GPU. Data augmentation via Gaussian blurring with σ uniformly chosen from 0 to 1.5 is performed on MRI or PET scans.

3.2. Quantitative evaluation

To justify the performance of BPGAN, the experimental results are compared to state-of-the-art (SOTA) methods. Quantitative evaluation is done in this work employing three evaluation metrics, including MAE, PSNR, and SSIM. MAE and PSNR measure absolute errors between real and synthetic PET scans, and SSIM represents structural similarities.

MAE is an extensively used metric for the image reconstruction problem. It measures the average absolute difference in pixel values between a synthetic PET scan and a real PET scan. MAE is defined as:

$$\text{MAE}(x, y) = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (15)$$

where x_i and y_i represent the intensity value of index i in the real PET pixels and the synthetic PET pixels, respectively.

PSNR is an expression for the ratio between the maximum possible value and the mean squared error of the synthetic and real image. It is computed as follows:

$$\text{PSNR}(x, y) = 10 \log_{10} \frac{\max^2}{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (16)$$

where x_i and y_i are the intensity values of the real and synthetic PET pixels. n is the number of pixels, and \max is the maximum possible intensity of image pixels.

SSIM evaluates the similarities within pixels of the real and synthetic PET images. It attempts to extract structural information, where the higher value of SSIM indicates the higher visual quality of generated PET images. Let x and y represent the generated PET image patch and the real PET image patch extracted in the same spatial location, respectively. For pixel i , the SSIM score can be calculated as follows:

$$\text{SSIM}(i) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (17)$$

where μ_x , σ_x and σ_{xy} are the average of x , the variance of x , and the co-variance of x and y , respectively. C_1 and C_2 are empirically two constants stabilizing the division with weak denominator.

The quantitative comparisons are conducted on five networks: FCN [57], U-Net [47], Pix2Pix [23], BicycleGAN [25], and BPGAN. As an important variant of the traditional CNN, FCN [57] drives advances in image translation tasks. Many previous works [30,58] have demonstrated the advantages of using FCN structures for image synthesis. Therefore, a typical 3D FCN is utilized as the baseline model. As a representative algorithm for the medical image segmentation, U-Net [47] is chosen for comparison, which has been done specifically for MRI-to-PET synthesis [59]. Finally, Pix2Pix [23] and BicycleGAN [25], which are the SOTA models in the multi-modal generation task, are compared to prove the effectiveness of the proposed model BPGAN. It is worth noting that

Table 4
Quantitative comparison using different methods on dataset-A test set.

Methods	MAE	PSNR	SSIM
FGAN* [54]	0.0803	29.62	0.6817
Adversarial U-Net* [55]	-	25.13	-
HGAN* [18]	0.0757	30.24	0.6945
GAN* [56]	0.0690	-	0.7240
FCN [57]	0.0480	23.46	0.5997
U-Net [47]	0.0467	23.95	0.6720
Pix2Pix [23]	0.0400	25.23	0.6844
BicycleGAN [25]	0.0348	26.19	0.6966
BPGAN	0.0318	26.92	0.7294

The symbol * means just as a reference since the selected experimental data and data preprocessing processes in these methods are different.

Table 5
Quantitative comparison using different methods on dataset-B test set.

Methods	MAE	PSNR	SSIM
FCN [57]	0.0516	22.89	0.5838
U-Net [47]	0.0517	22.92	0.6343
Pix2Pix [23]	0.0514	22.93	0.6061
BicycleGAN [25]	0.0409	24.84	0.6503
BPGAN	0.0396	25.08	0.6646

we have re-implemented these models based on the original paper or original code. We retain the topology of these models while making some necessary changes, such as changing networks from 2D to 3D to accommodate the PET synthesis task.

The results on dataset-A and dataset-B in terms of MAE, PSNR, and SSIM metrics are shown in Table 4 and Table 5. The high PSNR (26.92/25.08) and low MAE (0.0318/0.0396) prove that the PET scans generated by our BPGAN have the highest visual quality, while the high SSIM (0.7294/0.6646) indicates BPGAN can synthesize diverse PET images based on existing MRI scans. Quantitative results suggest that our proposed model can generate plausible and diverse PET images, which is superior to other competitive models.

3.3. Qualitative evaluation

To give an instinctive sense of the visual results of BPGAN compared with other SOTA models, qualitative results are analyzed by our cooperation with clinical experts Yan Liu and Qingyan Cai through visualizing cases. The qualitative observations of three samples from dataset-A and dataset-B with different disease stages are demonstrated in Fig. 6 and Fig. 7.

The results clearly show that, from a visual standpoint, the PET images estimated by the proposed method have higher similarity with the real PET images in axial, coronal, and sagittal planes. It is proved that BPGAN can achieve better-generated performance in the MRI-to-PET synthesis task than others. The results are confirmed by experienced doctors. In the baseline 3D FCN and 3D U-Net, the synthetic scans deviate significantly from the ground truth images, which have undesirable speckle noise marked with yellow lines in Fig. 6 and Fig. 7. The speckle noise could disturb the classification accuracy and lead to poor prediction results in the computer-assisted diagnosis method. At the same time, the contrast of PET images estimated by these methods is low. In comparison, the contrast of our generated PET images is improved, and there is no speckle spot. Although Pix2Pix [23] and BicycleGAN [25] can learn the structural changes, there are still contrast and structural detail errors compared to BPGAN, such as the regions marked with red circles in Fig. 6 and Fig. 7. This region is roughly the temporal lobe, which is marked by us with a discussion with the doctors. The temporal lobe is one of the regions highly associated with AD/MCI in comparison to other regions. The edge structure of the generated image seems to be over-smoothed or fuzzy

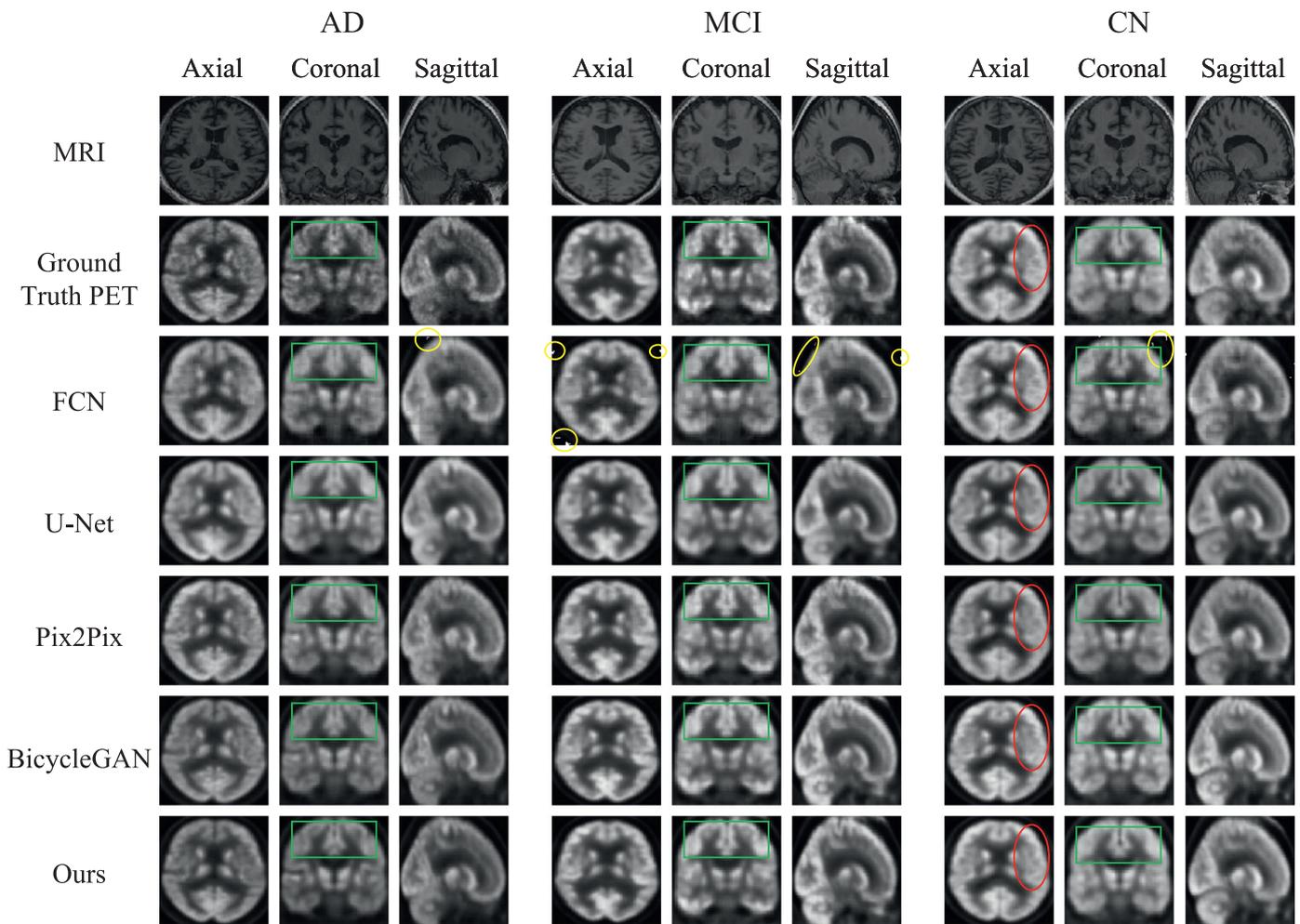


Fig. 6. Qualitative results of PET images synthesized by five methods with their corresponding ground truth PET and MRI images for 3 subjects on dataset-A - one from each group: AD, MCI, and CN in axial (left), coronal (middle), and sagittal (right).

due to the lack of edge information. The edges of brain tissues in PET images estimated by BPGAN are more complete, as shown in green boxes in Fig. 6 and Fig. 7. The green rectangle region is the frontal and parietal lobe, marked after discussion with clinical experts. Studies [60,61] have shown that frontal and parietal lobe hypometabolism, as noted on PET imaging, is the metabolic abnormality associated with AD/MCI. It can be shown that the images synthesized by our BPGAN (7th row) are more consistent with the ground truth (2nd row) than those synthesized by other methods (3rd-6th rows) in Fig. 6 and Fig. 7, particularly in terms of the sulcus width and edge information. Overall, our proposed model can generate diverse results and reflect detailed brain features. Based on qualitative analysis and quantitative assessment, we observe that BPGAN is superior to other methods and has potential application in missing-modal medical image completion.

To study the efficacy of the proposed model, we specially compare the generated PET scans with the real PET scans. The axial, coronal, and sagittal slices of one synthetic 3D PET sample from dataset-A and dataset-B by various approaches are visualized in Fig. 8 and Fig. 9, respectively. Ground truth PET scans and error maps are also illustrated. The pixel value in error maps is calculated from the absolute difference between generated PET scans and ground truth PET scans, which is further highlighted by the pseudo-color processing. We can clearly see that the PET scans synthesized by BPGAN have more low-difference regions than other methods. Moreover, to make a fair comparison, the mean error per pixel in error maps is calculated. In dataset-A and

dataset-B, the mean error per pixel of our PET images is minimum, further demonstrating the effectiveness of the proposed model.

3.4. Experimental results of AD diagnosis

To validate the contribution of generated PET scans as completed medical imaging to multi-modal AD diagnosis, we further discuss the AD classification tasks. The development of Alzheimer's disease includes three stages (CN, MCI, AD), so a multi-class classification task (AD vs. CN vs. MCI) and three types of two-class classification tasks (AD vs. CN, MCI vs. CN, AD vs. MCI) are discussed in this paper. The widely used classification metrics, namely accuracy (ACC), receiver operator characteristic (ROC) curve and area under receiver operating characteristic (AUC), are employed for performance evaluation in AD diagnosis. The classification accuracies of single-modality AD diagnosis and multi-modality AD diagnosis are discussed in subsections 3.4.1 and 3.4.2, respectively. ROC curve and AUC of AD diagnosis are analyzed in subsection 3.4.3.

3.4.1. Classification accuracy of single-modality AD diagnosis

In the single-modality AD diagnosis, 3D-Resnet18, a classical neural network for solving classification problems, is employed as the classifier to evaluate the PET images synthesized by different methods. For a fair comparison, we first train the classifiers based on the real PET images in the training set. Then, we evaluate the classification ACC of real PET images and PET images synthesized by different comparison models in the testing set, respectively.

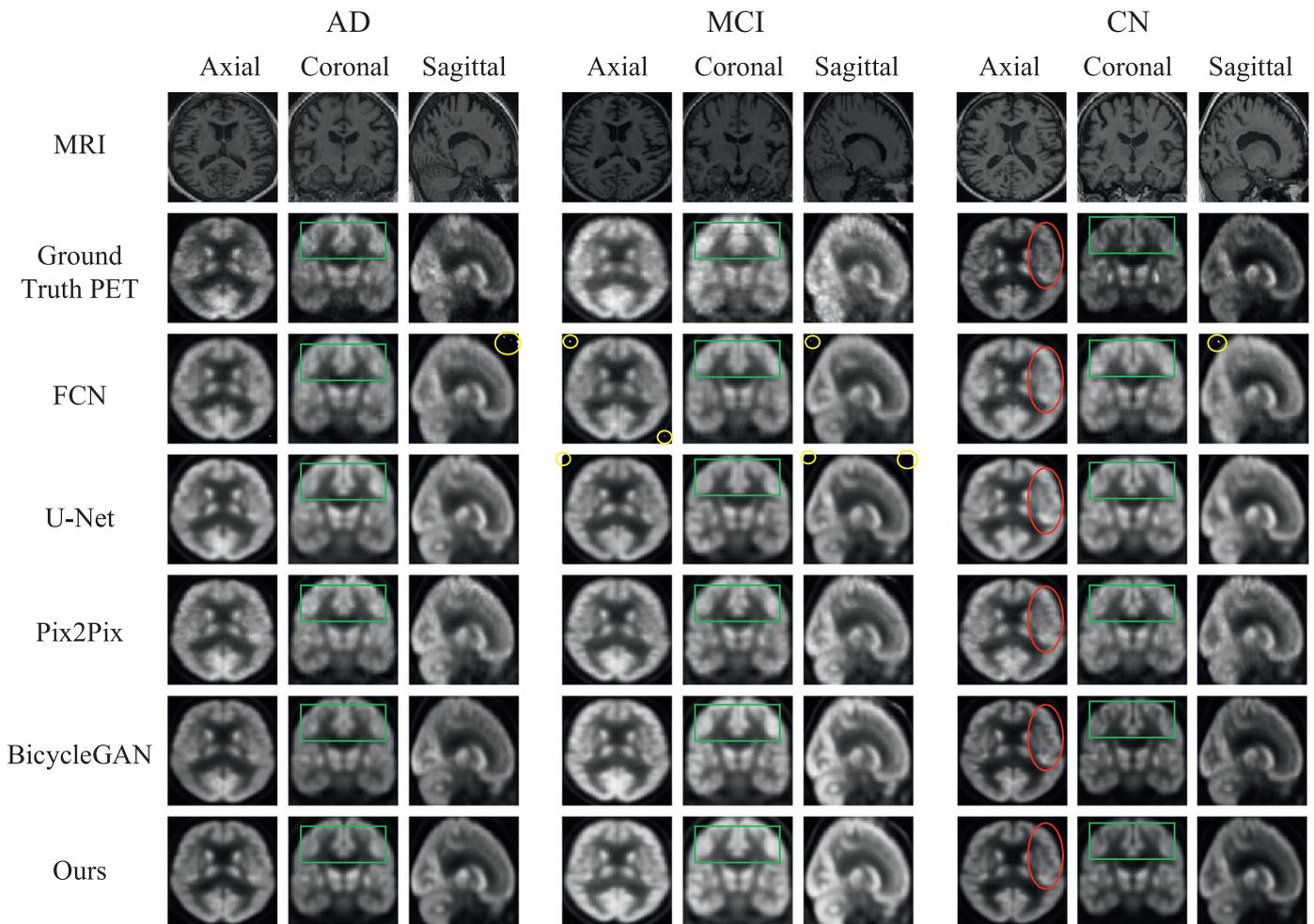


Fig. 7. Examples of the input MRI scans and their corresponding ground truth PET and estimated PET images from patients on dataset-B with AD, MCI, CN, respectively. From left in each subject, estimated PET image in axial plane, coronal plane, and sagittal plane are shown.

The performance of real MRI scans, real PET scans, and generated PET scans in AD classification tasks are presented in Fig. 10 and Fig. 11. If the synthetic PET scans are close to the corresponding real PET scans, their prediction for the disease category should be the same. We can observe from Fig. 10 and Fig. 11 that BPGAN achieves higher performance than other models in multi-class classification, where the accuracies for the dataset-A and dataset-B are 80.34% and 53.49%, respectively. In terms of two-class classification tasks, most of the classification results of the proposed model are close to that of the real PET images, where the MCI vs. CN accuracy of dataset-A is the same as real PET scans. It can be seen from Fig. 10 and Fig. 11 that the classification performance of PET images in dataset-B (including real PET images) is inferior to that in dataset-A. This shows that the complexity of the data affects performance, which is consistent with most AD classification algorithms. The challenge in the disease diagnosis is to distinguish adjacent stages. Since MCI is a transitional state between normal aging and AD, it is difficult to diagnose MCI. Because it does not cause abnormalities in life and social functioning, the behavior of patients with MCI is outside the range of normal aging changes. The results show that the classification accuracy of MCI is significantly lower than that of other stages, especially AD vs. MCI and MCI vs. CN. In classification tasks for adjacent disease stages, our model achieves the best performance in the accuracy of AD vs. MCI on dataset-A, with an accuracy of 88.12%, which is more than 2% higher than that of BicycleGAN [25]. In the challenging MCI vs. CN classification, BPGAN achieves the best performance on dataset-B

with an accuracy of 63.67% compared with other methods. Overall, our generated PET images outperform those generated by other methods in overall accuracy.

3.4.2. Classification accuracy of multi-modality AD diagnosis

In the multi-modality AD diagnosis, since the alignment is pre-processed on MRI and PET scans, pixel-level fusion is performed on them, where simply concatenate together to form a long vector that can then be used for classification. In AD classification tasks using MRI and PET scans, 3D-Resnet18 is also used as the classifier for a fair comparison.

We conduct experiments using the combination of real MRI and synthetic PET scans as well as the combination of real paired MRI and PET scans for multi-modality AD diagnosis. The accuracies of single-modality MRI scans, single-modality PET scans, the combination of real MRI and PET scans, and the combination of MRI and synthetic PET images for AD classification tasks are compared, as shown in Fig. 10 and Fig. 11. As expected, since the two modalities of medical imaging are complementary, we can observe that the multi-modality classification using our synthetic PET imaging achieves substantially improved performance for AD diagnosis compared with the single-modality classification using the stand-alone MRI scan. When adding our generated PET scans, the accuracy of multi-class classification on dataset-A and dataset-B increases by about 1% and 1% separately compared with that of single-modality MRI scans. Especially in the AD vs. CN task on dataset-B, the accuracy of multi-modal medi-

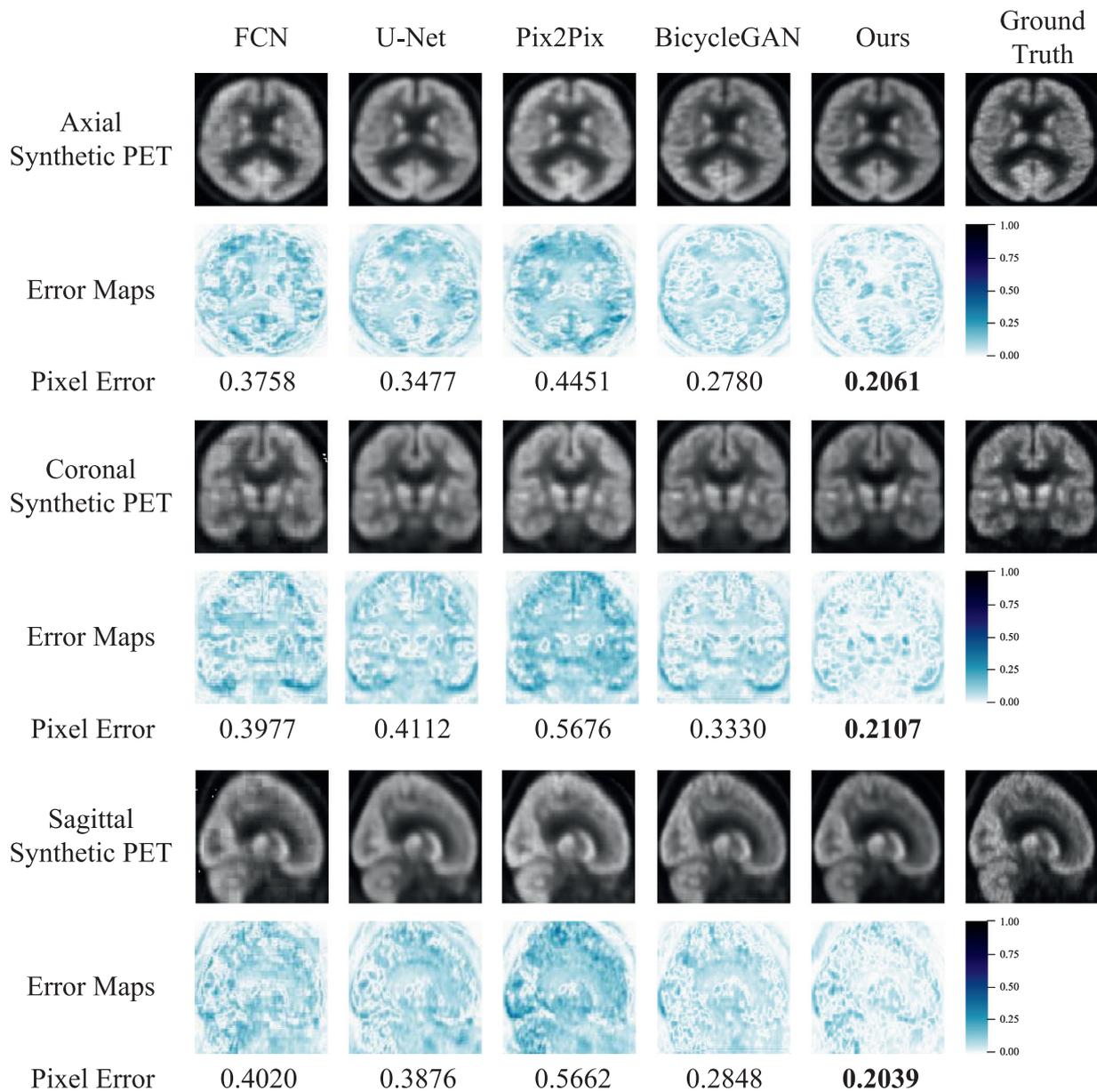


Fig. 8. Illustration of different maps from three planes of one subject in dataset-A, as well as the synthetic PET images and their corresponding real images.

cal imaging with our generated PET scans reaches 85.03%, which increases about 4% than the result of single-modality MRI scans, as shown in Fig. 11. These demonstrates the diagnostic usefulness of generated PET scans for multi-modality diagnosis. With our imputed PET scans, the multi-modality classification achieves close performance for AD diagnosis compared with the multi-modality classification using real MRI and PET scans. In particular, in MCI vs. CN task on dataset-A, the multi-modality classification with our generated PET scans is as accurate as the multi-modality classification using real MRI and PET scans. The results indicate that our proposed method is effective, and the synthesized PET scans can be used as supplementary data for multi-modality AD diagnosis.

3.4.3. ROC curve and AUC of AD diagnosis

In two-class classification tasks, the ROC curve and AUC are further evaluated. ROC curve and AUC represent the model prediction results stably without the fluctuation of evaluation results due to sample imbalance, so they are common evaluation metrics for pre-

dictive accuracy. The experimental results of single-modality MRI scans, single-modality PET scans, the combination of real MRI and synthetic PET scans, and the combination of real paired MRI and PET scans on dataset-A and dataset-B are shown in Fig. 12 and Fig. 13, respectively. By using the PET scans generated by our BPGAN, the AUC of MCI vs. CN classification is 0.64 on dataset-B, higher than other methods. We can observe that the overall performance (in terms of AUCs) of the proposed model is superior to other competitive models.

The macro-averaged and micro-averaged ROC curves and AUC are further employed to evaluate the reliability of synthetic PET images in the multi-class classification, as demonstrated in Fig. 14 and Fig. 15, respectively. Macro-averaged and micro-averaged AUC values of the combination of real MRI and synthetic PET scans equal the ones related to the combination of real paired MRI and PET scans (macro-averaged AUC=0.94, micro-averaged AUC=0.95) for AD vs. CN vs. MCI classification on dataset-A. To sum up, it proves that BPGAN can serve as an effective data completion scheme for multi-modality AD diagnosis.

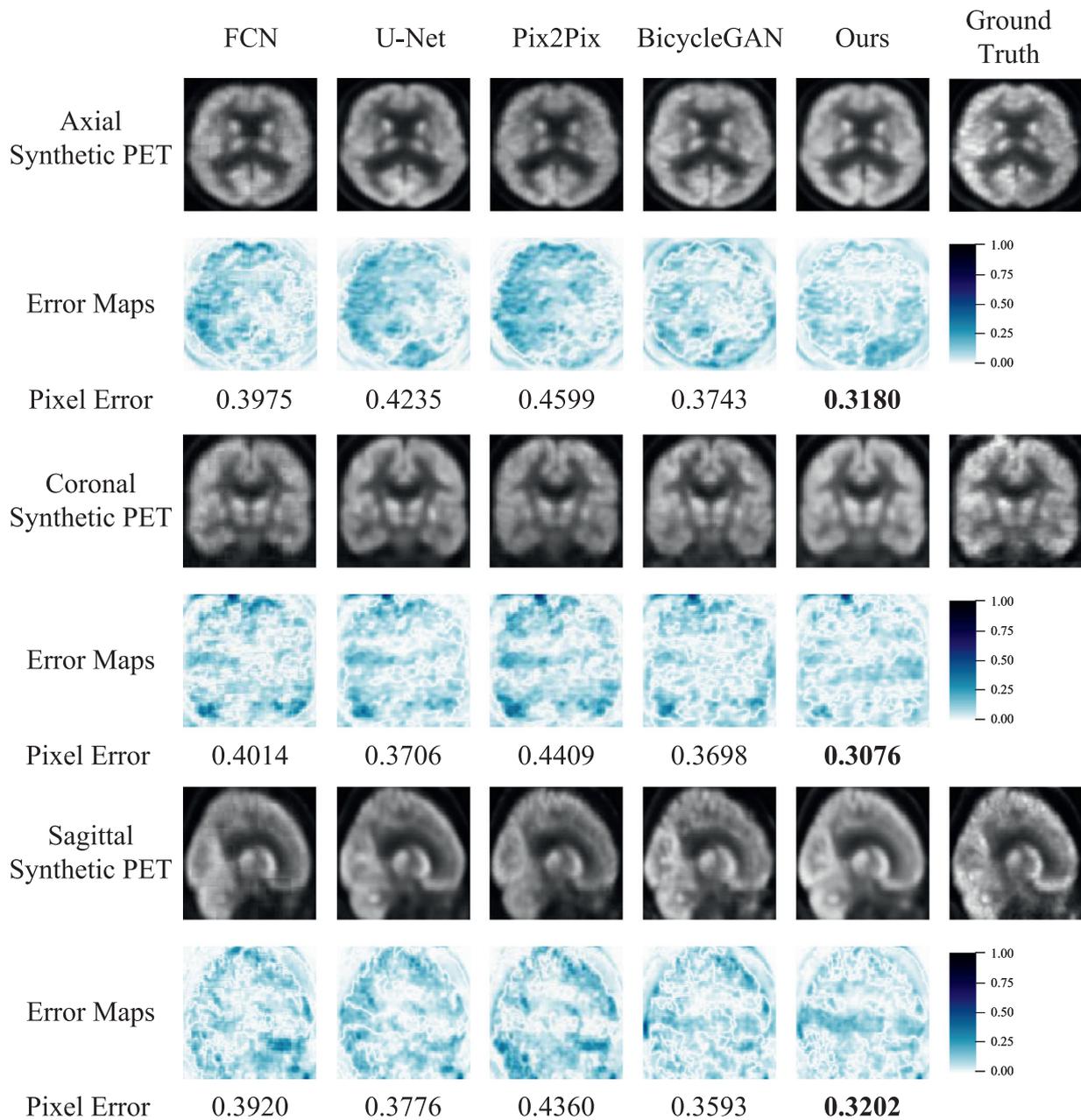


Fig. 9. The difference maps between real and synthetic PET images on dataset-B using five methods are shown.

4. Discussion

4.1. Ablation analysis

We discuss the ablation analysis in this subsection to validate each key component in the proposed method. First, the performance of two generators, i.e., the plain U-Net and MCU, is compared to demonstrate the effectiveness of MCU architecture as a generator. Quantitative comparisons in MAE, PSNR, and SSIM are presented in Table 6 and Table 7. We observe that utilizing the

Table 6
Quantitative comparison between our generator and the simplified generator on dataset-A.

Methods	MAE	PSNR	SSIM
BicycleGAN [25]	0.0348	26.19	0.6966
MCU	0.0337	26.46	0.7088

Table 7
Quantitative comparison between our generator and the simplified generator on dataset-B.

Methods	MAE	PSNR	SSIM
BicycleGAN [25]	0.0409	24.84	0.6503
MCU	0.0404	24.90	0.6536

plain U-Net as the generator results in lower performance, potentially due to the lack of learning capability of the generator. Compared with the original U-Net, generator MCU shows high performance with MAE of 0.0337, PSNR of 26.46, and SSIM of 0.7088 on dataset-A. We conjectured that it is attributed to the proposed multiple convolution module, which captures broader and deeper semantic features. And learnable interactions of cross-channel information encourage minimizing the inconsistency of distributions between MRI and PET scans. The results show that the generator MCU can achieve better-generated performance.

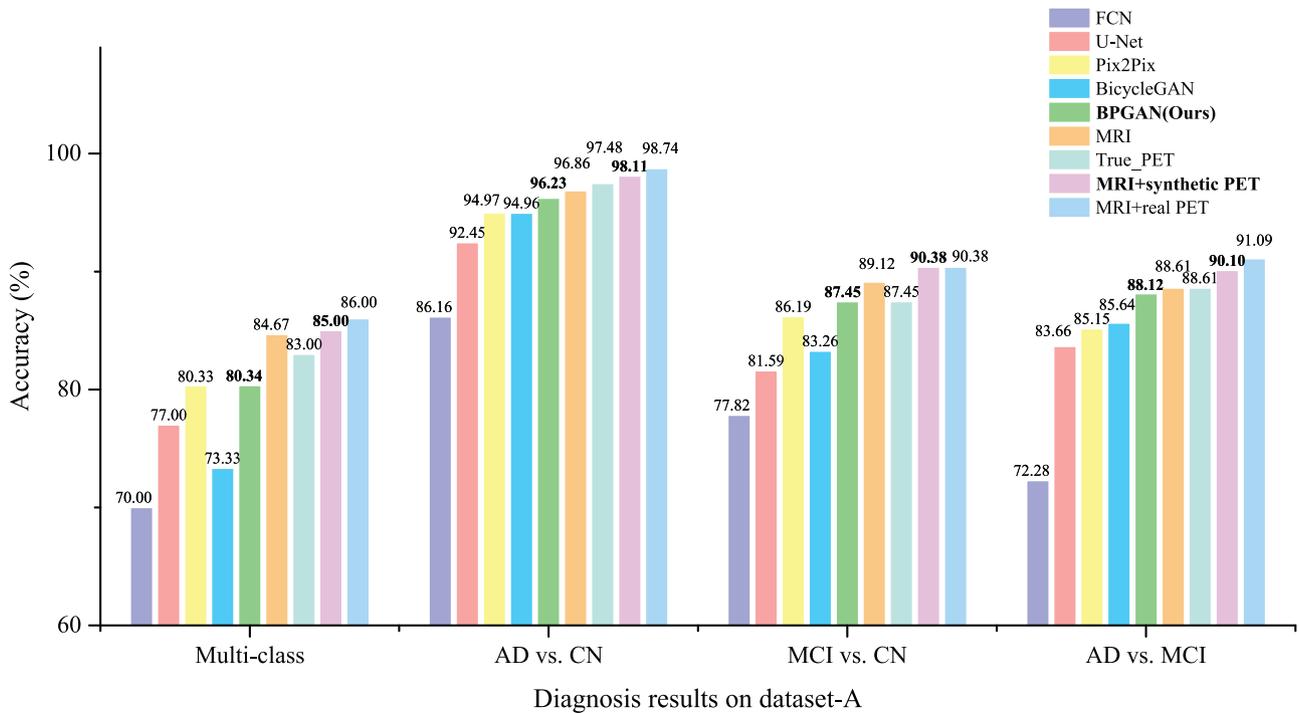


Fig. 10. Comparison of diagnosis results on dataset-A.

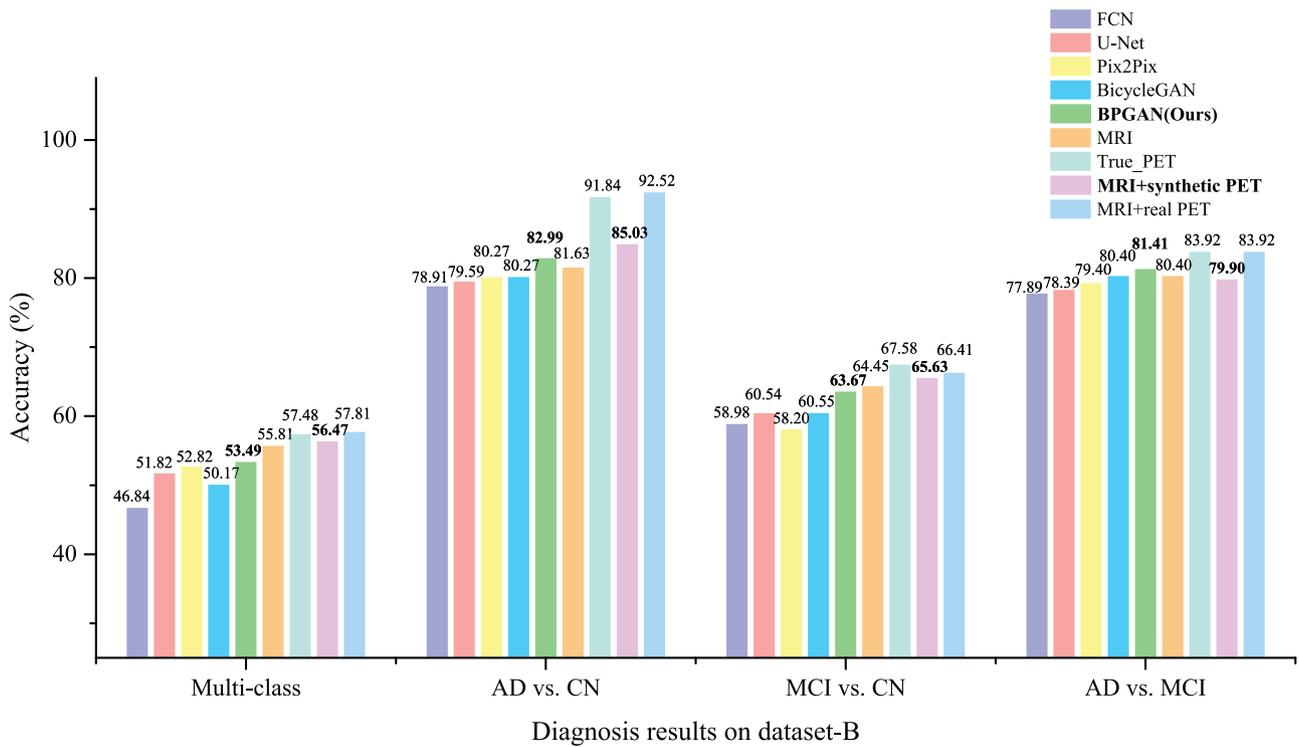


Fig. 11. Comparison of diagnosis results on dataset-B.

An extensive objective function is employed in this paper to optimize BPGAN, including adversarial loss, KL loss, L_1 loss, 3D GP loss, and SSIM loss. To investigate the contribution of each part of the extensive objective function to MRI-to-PET synthesis, we use the following three objective functions for training BPGAN, respectively.

- **w/o SSIM loss and 3D GP loss:** We remove the SSIM loss and 3D GP loss, and the rest remains the same as our final objective function.

- **w/o SSIM loss:** We remove the SSIM loss, and the rest remains the same as our final objective function.

- **our objective function:** Our proposed objective function with five components is utilized.

The ablation experiment results of the objective function on dataset-A and dataset-B are shown in Table 8 and Table 9. The first type of extended loss is 3D GP loss. As can be seen from Table 8 and Table 9, the image gradient loss significantly improves the quality of generated PET scans in all three metrics. By introduc-

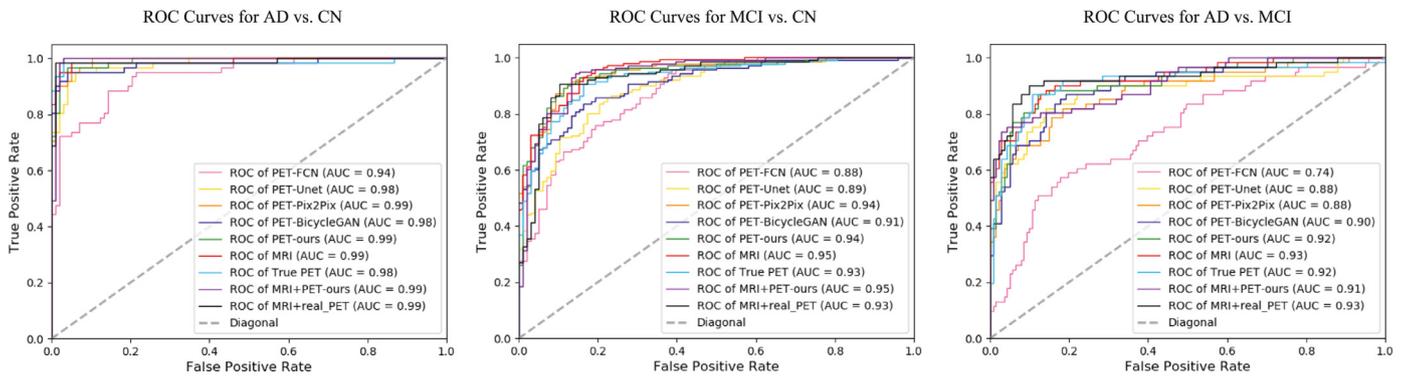


Fig. 12. The ROC curves in the experiments of two-class classification tasks on dataset-A.

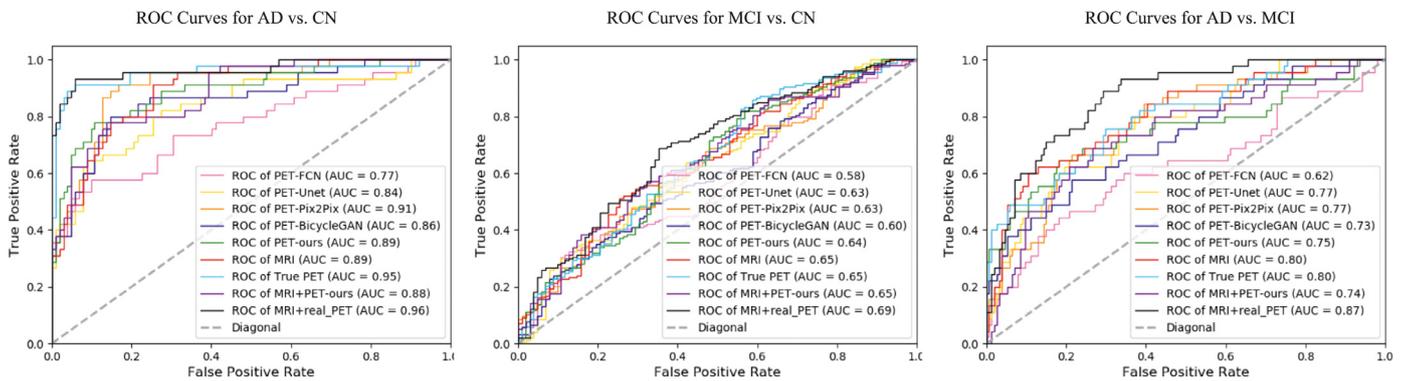


Fig. 13. The ROC curves in the experiments of two-class classification tasks on dataset-B.

Macro-averaged ROC Curves for AD vs. CN vs. MCI

Micro-averaged ROC Curves for AD vs. CN vs. MCI

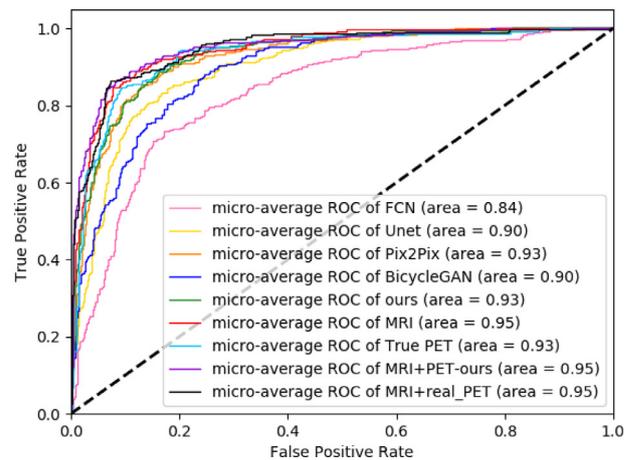
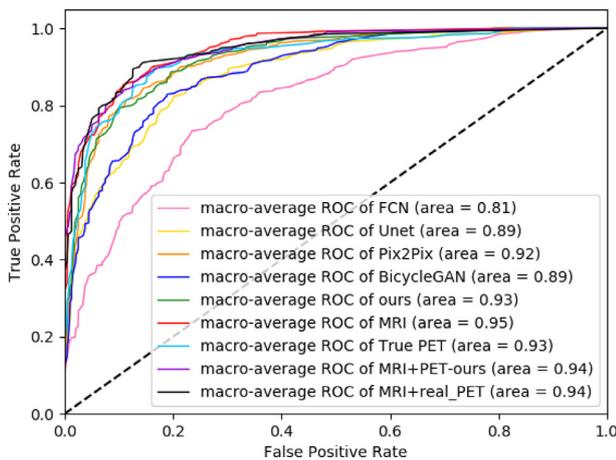


Fig. 14. The ROC curves in the experiments of AD vs. CN vs. MCI classification task on dataset-A.

Table 8
Quantitative comparisons between different objective functions on dataset-A.

Methods	MAE	PSNR	SSIM
w/o SSIM loss and 3D GP loss	0.0337	26.46	0.7088
w/o SSIM loss	0.0324	26.78	0.7247
our objective function	0.0318	26.92	0.7294

Table 9
Quantitative comparisons between different objective functions on dataset-B.

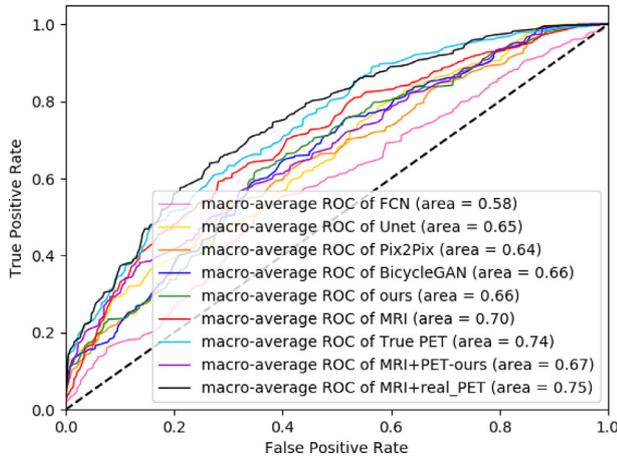
Methods	MAE	PSNR	SSIM
w/o SSIM loss and 3D GP loss	0.0404	24.90	0.6536
w/o SSIM loss	0.0399	25.01	0.6580
our objective function	0.0396	25.08	0.6646

ing 3D GP loss, the edge of local tissues in generated PET scans is clearer, and the local texture of synthetic PET images is more reasonable. The second kind of extended loss is SSIM loss, which results in a modest improvement compared to other objective functions. This is because the SSIM loss can effectively improve the

quality of synthetic images by supervising the brightness, contrast, and structure information between the generated PET scans and real PET scans.

In addition, the qualitative comparison of the three objective functions described above is provided in Fig 16 considering

Macro-averaged ROC Curves for AD vs. CN vs. MCI



Micro-averaged ROC Curves for AD vs. CN vs. MCI

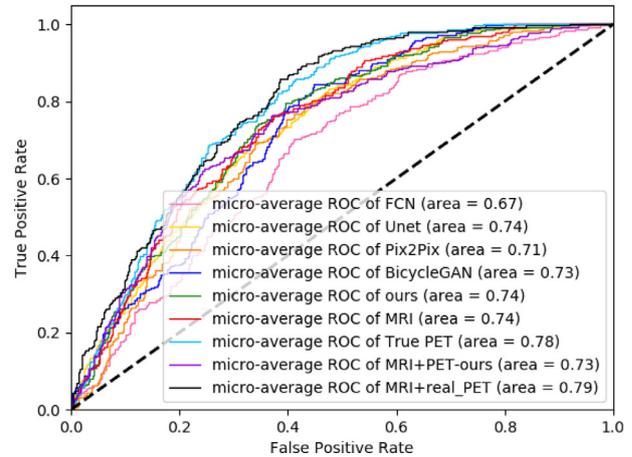


Fig. 15. The ROC curves in the experiments of AD vs. CN vs. MCI classification task on dataset-B.

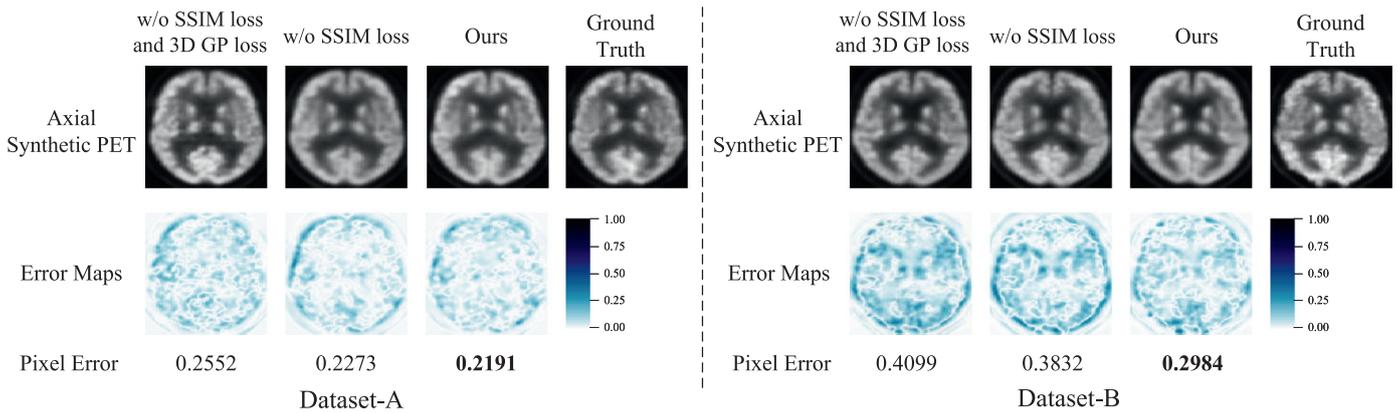


Fig. 16. Qualitative comparison of PET images synthesized with various objective functions and corresponding difference images, as well as the real PET images in the axial plane.

generated PET scans, ground truth PET scans, and error maps. It can be observed that when only considering **w/o SSIM loss and 3D GP loss**, the edge error of brain tissues in the synthesized PET images is obvious. The main reason is that this loss focuses on the macroscopic data distribution and the global information at the pixel level and does not consider the relationship between pixels and the structural information of the synthetic images. And the visualization results show that the synthetic PET images with our loss function are more similar to real PET images than the results of other objective functions. And the mean error per pixel in error maps is reported, which is minimal with our loss function. In summary, with the proposed objective function, BPGAN can perform stable and robust MRI-to-PET synthesis.

4.2. Comparison of different data splitting strategies

To study the impact of different data splitting strategies on the PET image synthesis task, we conduct experiments on the datasets utilizing the splits by the patient ID and random sampling, respectively. Our empirical studies show that the performance of the generator is sensitive to the data splitting methods suitable for different medical scenarios.

First, the data is partitioned randomly on MRI-PET scan-pairs regardless of subjects. The different scan-pairs from the same subject are correlated. As shown in Table 4, when the dataset is randomly divided by scan-pairs, the MAE, PSNR, and SSIM of our

model are 0.0318, 26.92, and 0.7294, respectively. We observe from Fig. 6 that the generated PET scans are close to real PET scans in this case. The underlying reason is that the model relies on patient-level information learned in training to achieve excellent results on the test set. This data partitioning method is suitable for patients with long-term follow-up who have historical PET imaging. Because of the prior information about individual brain structure, the generator focuses more on mapping the structural and functional information of the brain tissue associated with the disease.

In real medical scenarios, a new subject usually has no history of scans. Therefore, the impact of using the dataset split by patient ID needs to be studied. Table 5 and Fig. 7 represent the quantitative measures and qualitative performance based on the subject-separated MRI-PET scan-pairs. The second data splitting strategy is suitable for patients with no recorded PET scans. During the training process, the generator not only learns the structure information such as the skull but also models the mapping between the structural and functional information of brain tissue. Moreover, the differences among the three stages of AD exhibited by a particular scan-pair should be understood. Relative to the generation performance of the first data partition strategy, the second one is somewhat degraded. The potential reason is that the generator is disturbed by the individual brain structure information. It is worth mentioning that the degradation of synthesis performance under the second data splitting strategy is natural, which is anticipated by some AD diagnosis researches [62–64].

5. Conclusion

In this paper, we propose a novel method for brain MRI-to-PET synthesis, which can be regarded as a potential scheme for missing medical imaging complement. We design a 3D end-to-end network BPGAN effectively synthesizes perceptually realistic and diverse PET scans while preserving detailed information of brain issues. Two alternative data splitting strategies are explored to study further and analyze the impact on the MRI-to-PET synthesis in different medical scenarios. We introduce an objective function to supervise the generator during the training process, making the generated PET scans of higher quality. To verify the effectiveness of BPGAN, extensive experiments are conducted on a subset of ADNI. The qualitative and quantitative experimental results demonstrate that the proposed method can effectively generate high-quality and diverse PET scans when given MRI scans, which is superior to other SOTA methods. Moreover, we study the contribution of generated PET scans to AD classification tasks by simulating missing data. The classification results show that reasonable PET scans generated by our method can be utilized significantly for multi-modality AD diagnosis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research in our paper is sponsored by Chengdu Major Technology Application Demonstration Project (2019-YF09-00120-SN), the Sichuan Science and Technology Program (No. 2021YFS0239). We would like to express our thanks and gratitude for the guidance and support of this study provided by chief physician Yan Liu of The Third People's Hospital of Chengdu and chief physician Qingyan Cai of The Fourth People's Hospital of Chengdu.

References

- [1] T. Xu, H. Zhang, X. Huang, S. Zhang, D.N. Metaxas, Multimodal deep learning for cervical dysplasia diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 115–123.
- [2] S. Zhang, D.N. Metaxas, Large-scale medical image analytics: recent methodologies, applications and future directions, *Med Image Anal* 33 (2016) 98–101.
- [3] J. Yuan, X. Ran, K. Liu, C. Yao, Y. Yao, H. Wu, Q. Liu, Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: a review, arXiv preprint arXiv:2102.03336 (2021).
- [4] L. Zhang, M. Wang, M. Liu, D. Zhang, A survey on deep learning for neuroimaging-based brain disorder analysis, *Front Neurosci* 14 (2020).
- [5] M. Weiner, D. Veitch, P. Aisen, L. Beckett, N. Cairns, R.C. Green, D. Harvey, C.R. Jack, W. Jagust, E. Liu, J.C. Morris, R. Petersen, A.J. Saykin, M.E. Schmidt, L. Shaw, J.A. Siuciak, H. Soares, A.W. Toga, J.Q. Trojanowski, The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception, *Alzheimer's & Dementia* 8 (2012) S1–S68.
- [6] S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C. Jack, W. Jagust, J.Q. Trojanowski, A.W. Toga, L. Beckett, The Alzheimer's disease neuroimaging initiative, *Neuroimaging Clin. N. Am.* 15 (4) (2005) 869.
- [7] X. Zhang, Z.L. Yang, G. Lu, G. Yang, L.J. Zhang, Pet/mr imaging: new frontier in Alzheimer's disease and other dementias, *Front Mol Neurosci* 10 (2017).
- [8] M.A. Ebrahimighahnaveh, S. Luo, R. Chiong, Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review, *Comput Methods Programs Biomed* 187 (2020) 105242.
- [9] H. Qiao, L. Chen, Z. Ye, F. Zhu, Early Alzheimer's disease diagnosis with the contrastive loss using paired structural MRIs, *Comput Methods Programs Biomed* 208 (2021) 106282.
- [10] V. Calhoun, 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* 1 3 (2016) 230–244.
- [11] M. Liu, Y. Gao, P. Yap, D. Shen, Multi-hypergraph learning for incomplete multi-modality data, *IEEE J Biomed Health Inform* 22 (2018) 1197–1208.
- [12] Y. Huang, J. Xu, Y. Zhou, T. Tong, X. Zhuang, Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network, *Front Neurosci* 13 (2019) 509.
- [13] C. Jack, M. Bernstein, N. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. Britson, J.L. Whitwell, C. Ward, A. Dale, J. Felmlee, J. Gunter, D. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. DeCarli, G. Krueger, H. Ward, G. Metzger, K.T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. Debbins, A. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, M. Weiner, The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, *J. Magn. Reson. Imaging* 27 (2008).
- [14] H. Kang, J. Park, K. Cho, D.-Y. Kang, Visual and quantitative evaluation of amyloid brain pet image synthesis with generative adversarial network, *Applied Sciences* 10 (2020) 2628.
- [15] S. Hu, B. Lei, Y. Wang, Z. Feng, Y. Shen, S. Wang, Bidirectional mapping generative adversarial networks for brain mr to pet synthesis, arXiv preprint arXiv:2008.03483 (2020).
- [16] A.M. Rossetto, W. Zhou, Gandalf: peptide generation for drug design using sequential and structural generative adversarial networks, *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (2020).
- [17] H.-C. Shin, A. Ihsani, S. Mandava, S.T. Sreenivas, C. Forster, J. Cha, A.D.N. Initiative, Ganbert: generative adversarial networks with bidirectional encoder representations from transformers for mri to pet synthesis, arXiv preprint arXiv:2008.04393 (2020).
- [18] Y. Pan, M. Liu, C. Lian, Y. Xia, D. Shen, Spatially-constrained fisher representation for brain disease identification with incomplete multi-modal neuroimages, *IEEE Trans Med Imaging* 39 (9) (2020) 2965–2975.
- [19] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, S. Ji, Deep learning based imaging data completion for improved brain disease diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2014, pp. 305–312.
- [20] F. Gao, T. Wu, X. Chu, H. Yoon, Y. Xu, B. Patel, Deep residual inception encoder-decoder network for medical imaging synthesis, *IEEE J Biomed Health Inform* 24 (2020) 39–49.
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, *Adv Neural Inf Process Syst* 27 (2014).
- [22] Z. Zha, B. Wen, X. Yuan, J.T. Zhou, J. Zhou, C. Zhu, Triply complementary priors for image restoration, *IEEE Trans. Image Process.* (2021).
- [23] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 5967–5976.
- [24] Z. Zha, X. Yuan, B. Wen, J. Zhou, J. Zhang, C. Zhu, A benchmark for sparse coding: when group sparsity meets rank minimization, *IEEE Trans. Image Process.* 29 (2020) 5094–5109.
- [25] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A.A. Efros, O. Wang, E. Shechtman, Toward multimodal image-to-image translation, in: *Advances in Neural Information Processing Systems*, 2017, pp. 465–476.
- [26] Z. Zha, X. Yuan, B. Wen, J. Zhou, J. Zhang, C. Zhu, From rank estimation to rank approximation: rank residual constraint for image restoration, *IEEE Trans. Image Process.* 29 (2019) 3254–3269.
- [27] E. Moya-Sáez, Ó. Peña-Nogales, R. de Luis-García, C. Alberola-López, A deep learning approach for synthetic MRI based on two routine sequences and training with synthetic data, *Comput Methods Programs Biomed* (2021) 106371.
- [28] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, D. Shen, Synthesizing missing pet from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 455–463.
- [29] S.N. Yaakub, C.J. McGinnity, J.R. Clough, E. Kerfoot, N. Girard, E. Guedj, A. Hamers, Pseudo-normal pet synthesis with generative adversarial networks for localising hypometabolism in epilepsies, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2019, pp. 42–51.
- [30] A. Ben-Cohen, E. Klang, S.P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M.M. Amitai, H. Greenspan, Cross-modality synthesis from ct to pet using fcnet and gan networks for improved automated lesion detection, *Eng Appl Artif Intell* 78 (2019) 186–194.
- [31] Y. Pan, M. Liu, Y. Xia, D. Shen, Disease-image-specific learning for diagnosis-oriented neuroimaging synthesis with incomplete multi-modality data, *IEEE Trans Pattern Anal Mach Intell* (2021).
- [32] W. Lin, W. Lin, G. Chen, H. Zhang, Q. Gao, Y. Huang, T. Tong, M. Du, Bidirectional mapping of brain MRI and pet with 3D reversible GAN for the diagnosis of Alzheimer's disease, *Front Neurosci* 15 (2021).
- [33] A. Sikka, J.S. Virk, D.R. Bathula, et al., Mri to pet cross-modality translation using globally and locally aware gan (GAL-GAN) for multi-modal diagnosis of Alzheimer's disease, arXiv preprint arXiv:2108.02160 (2021).
- [34] J. Islam, Y. Zhang, Gan-based synthetic brain pet image generation, *Brain Informatics* 7 (2020).
- [35] H. Emami, Q. Liu, M. Dong, Frea-unet: frequency-aware u-net for modality transfer, arXiv preprint arXiv:2012.15397 (2020).
- [36] Y. Wang, B. Yu, L. Wang, C. Zu, D. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, L. Zhou, 3D conditional generative adversarial networks for high-quality pet image estimation at low dose, *Neuroimage* 174 (2018) 550–562.
- [37] D. Nie, X. Cao, Y. Gao, L. Wang, D. Shen, Estimating CT image from MRI data using 3D fully convolutional networks, in: *Deep Learning and Data Labeling for Medical Applications*, Springer, 2016, pp. 170–178.
- [38] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with deep convolutional adversarial networks, *IEEE Trans. Biomed. Eng.* 65 (2018) 2720–2730.

- [39] A. Routier, N. Burgos, J. Guillon, J. Samper-González, J. Wen, S. Bottani, A. Marcoux, M. Bacci, S. Fontanella, T. Jacquemont, A. Wild, P. Gori, A. Guyot, P. Lu, M. Díaz, E. Thibeau-Sutre, T. Moreau, M. Teichmann, M. Habert, S. Durrleman, O. Colliot, Clinica: an open source software platform for reproducible clinical neuroscience studies, 2019.
- [40] B. Fischl, *Neuroimage* 62 (2012) 774–781.
- [41] J. Ashburner, A fast diffeomorphic image registration algorithm, *Neuroimage* 38 (2007) 95–113.
- [42] A.B.L. Larsen, S.K. Sønderby, H. Larochelle, O. Winther, Autoencoding beyond pixels using a learned similarity metric, *ArXiv abs/1512.09300* (2016).
- [43] D.P. Kingma, M. Welling, Auto-encoding variational bayes, *arXiv preprint arXiv:1312.6114* (2013).
- [44] J. Donahue, P. Krähenbühl, T. Darrell, Adversarial feature learning, *ArXiv abs/1605.09782* (2017).
- [45] V. Dumoulin, I. Belghazi, B. Poole, A. Lamb, M. Arjovsky, O. Mastropietro, A.C. Courville, Adversarially learned inference, *ArXiv abs/1606.00704* (2017).
- [46] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313 (5786) (2006) 504–507.
- [47] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [48] L. Xiang, Y. Li, W. Lin, Q. Wang, D. Shen, Unpaired deep cross-modality synthesis with fast training, Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S... 11045 (2018) 155–164.
- [49] M. Lin, Q. Chen, S. Yan, Network in network, *arXiv preprint arXiv:1312.4400* (2013).
- [50] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 770–778.
- [51] M. Sarfraz, C. Seibold, H. Khalid, R. Stiefelwagen, Content and colour distillation for learning image translations with the spatial profile loss, *arXiv preprint arXiv:1908.00274* (2019).
- [52] Z. Wang, E. Simoncelli, A. Bovik, Multiscale structural similarity for image quality assessment, in: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, volume 2, IEEE, 2003, pp. 1398–1402.
- [53] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [54] Y. Pan, M. Liu, C. Lian, Y. Xia, D. Shen, Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 137–145.
- [55] S. Hu, J. Yuan, S. Wang, Cross-modality synthesis from MRI to pet using adversarial u-net with different normalization, in: *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, IEEE, 2019, pp. 1–5.
- [56] Y. Wang, G. Yin, J. Lang, P. Wang, J. Li, U-net and gans-based pet synthesis from MRI for soft-tissue sarcomas, *Int. J. Radiat. Oncol. Biol. Phys.* 108 (2020).
- [57] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans Pattern Anal Mach Intell* 39 (2017) 640–651.
- [58] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, Medical image synthesis with context-aware generative adversarial networks, in: *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, 2017, pp. 417–425.
- [59] K.D. Spuhler, J. Gardus, Y. Gao, C. DeLorenzo, R. Parsey, C. Huang, Synthesis of patient-specific transmission data for pet attenuation correction for PET/MRI neuroimaging using a convolutional neural network, *The Journal of Nuclear Medicine* 60 (2019) 555–560.
- [60] C.R. Jack Jr, D.A. Bennett, K. Blennow, M.C. Carrillo, B. Dunn, S.B. Haeberlein, D.M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, et al., NIA-aa research framework: toward a biological definition of alzheimer's disease, *Alzheimer's & Dementia* 14 (4) (2018) 535–562.
- [61] V.K. Shivamurthy, A.K. Tahari, C. Marcus, R.M. Subramaniam, Brain fdg pet and the diagnosis of dementia, *American Journal of Roentgenology* 204 (1) (2015) W76–W85.
- [62] Y. Fung, Z. Guan, R. Kumar, Y. Wu, M. Fiterau, Alzheimer's disease brain MRI classification: challenges and insights, *arXiv preprint arXiv:1906.04231* (2019).
- [63] K. Bäckström, M. Nazari, I.Y.-H. Gu, A.S. Jakola, An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018, pp. 149–153.
- [64] J. Wen, E. Thibeau-Sutre, J. Samper-González, A. Routier, S. Bottani, S. Durrleman, N. Burgos, O. Colliot, Convolutional neural networks for classification of alzheimer's disease: overview and reproducible evaluation, *Med Image Anal* 63 (2020) 101694.



Jin Zhang received her B.S. degree in Electronic Information Engineering from Sichuan University, Chengdu, China, in 2018. She is currently pursuing her Ph.D. degree in Communication and Information Systems, Sichuan University. Her research interests include image processing, deep learning, computer vision and medical image processing.



Xiaohai He received his B.S. and M.S. degrees in electrical engineering from Sichuan University, Chengdu, China, in 1985 and 1991 respectively. In 2002, he received his Ph.D. degree in biomedical engineering from the same university. He is currently a professor with the College of Electronics and Information Engineering, at Sichuan University. His research interests include image processing, pattern recognition, computer vision, image communication, and software engineering. He is a Senior Member of the Chinese Institute of Electronics. He is an editor of the *Journal of Information and Electronic Engineering* and an editor of the *Journal of Data Acquisition & Processing*.



Linbo Qing is an associate professor with the College of Electronics and Information Engineering at Sichuan University. He received his B.S. degree in Electronic Information Science and Technology from Sichuan University, China in 2003. In 2008, he received his Ph.D. degree in Communication and Information System from the same university. His main research interests include image processing, video coding and transmission, and information theory.



Feng Gao received his B.S. degree in Logistics Engineering from Southwest Jiaotong University, China, in 2008 and received his MS. degree in Vehicle Engineering from Xihua University, China, in 2011. He is currently working in External cooperation and liaison office, Southwest Jiaotong University and working as a research assistant in Southwest Jiaotong University National Interdisciplinary Institute on Aging (NIA). His research interests include deep learning, computer vision and medical image processing.



Bin Wang received a bachelor's degree in electronic information science and technology from Chengdu University of Information Technology in 2020. He is currently studying for a master's degree in electronic science and technology at Sichuan University. His research interests include image processing, deep learning and medical image processing.