IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 3, 2019, accepted December 29, 2019, date of publication January 3, 2020, date of current version January 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2963741

Fusion of Brain PET and MRI Images Using Tissue-Aware Conditional Generative Adversarial Network With Joint Loss

JIAYIN KANG[1], WU LU[2], AND WENJUAN ZHANG[3]
[1]School of Electronics Engineering, Jiangsu Ocean University, Lianyungang 222005, China
[2]Department of Nuclear Medicine, The First People's Hospital of Lianyungang, Lianyungang 222061, China
[3]School of Computer Engineering, Jiangsu Ocean University, Lianyungang 222005, China

Corresponding author: Jiayin Kang (jiayinkang@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61601194, in part by the Natural Science Foundation of the Jiangsu Higher Education Institution of China under Grant 17KJB520003, in part by the Research and Development Program of Science and Technology Bureau, Lianyungang, China, under Grant SH1508, in part by the Natural Science Foundation of Huaihai Institute of Technology under Grant Z2015009, in part by the Jiangsu University High-Tech Ship Collaborative Innovation Center/Marine Equipment and Technology Institute, Jiangsu University of Science and Technology under Grant HZ20190005, in part by the Lianyungang 521 Project Fund under Grant LYG52105-2018033, and in part by the Lianyungang Haiyan Plan Fund under Grant 2018-QD-011.

ABSTRACT Positron emission tomography (PET) has rich pseudo color information that reflects the functional characteristics of tissue, but lacks structural information and its spatial resolution is low. Magnetic resonance imaging (MRI) has high spatial resolution as well as strong structural information of soft tissue, but lacks color information that shows the functional characteristics of tissue. For the purpose of integrating the color information of PET with the anatomical structures of MRI to help doctors diagnose diseases better, a method for fusing brain PET and MRI images using tissue-aware conditional generative adversarial network (TA-cGAN) is proposed. Specifically, the process of fusing brain PET and MRI images is treated as an adversarial machine between retaining the color information of PET and preserving the anatomical information of MRI. More specifically, the fusion of PET and MRI images can be regarded as a min-max optimization problem with respect to the generator and the discriminator, where the generator attempts to minimize the objective function via generating a fused image mainly contains the color information of PET, whereas the discriminator tries to maximize the objective function through urging the fused image to include more structural information of MRI. Both the generator and the discriminator in TA-cGAN are conditioned on the tissue label map generated from MRI image, and are trained alternatively with joint loss. Extensive experiments demonstrate that the proposed method enhances the anatomical details of the fused image while effectively preserving the color information from the PET. In addition, compared with other state-of-the-art methods, the proposed method achieves better fusion effects both in subjectively visual perception and in objectively quantitative assessment.

INDEX TERMS Positron emission tomography, magnetic resonance imaging, image fusion, generative adversarial network, loss function.

I. INTRODUCTION

Positron emission tomography (PET), a nuclear medicine imaging technology, provides a color image with functional information that reflects the metabolism of different tissues. However, PET image has a low spatial resolution and lacks structural information of tissues [1]. On the other hand, magnetic resonance imaging (MRI), another non-invasive imaging tool, presents strong soft tissue structure information with higher spatial resolution. However, MRI image lacks color information that reflects the metabolic function of specific tissues [2], [3]. Therefore, effectively integrating PET with MRI via image fusion can provide more meaningfully complementary information. In other words,

The associate editor coordinating the review of this manuscript and approving it for publication was Bohui Wang.

the fused image not only retains the spatial structure information of MRI, but also preserves the color information of PET. As a result, this kind of complementary information can assist clinical diagnosis and treatment of diseases better [4], [5].

Over the past few decades, different types of methods for fusing PET and MRI images have been developed. These methods can be roughly categorized into four classes via their implementation mechanisms. The first one is IHS (Intensity–Hue-Saturation) based method which is realized based on the transformation and the replacement strategies [5]–[7]. This type of method first transforms the PET image from RGB color space into IHS color space; then replaces the I component of the transferred PET with the matched MRI (the MRI and the PET need to be registered in advance); finally, inversely transfers the PET with the substituted I component from IHS color space to RGB color space, and then obtains the fused image. This kind of approach usually generates a fused image which contains rich structural information with high resolution, but it generally distorts the color information of PET image due to the substitutive MRI image is rater different from the replaced I component of the PET image. The second type of method for merging PET and MRI images is implemented by the multi-resolution analysis (MRA) strategy [8]–[10]. This kind of method first decomposes PET and MRI images into multi-scale coefficients and transforming bases; then merges the decomposed coefficients according to a certain fusion rule; finally, inversely transforms the fused coefficients and transforming bases so as to get the final fused image. This kind of method can effectively preserve the color information of PET, but it has limitations in enhancing the spatial structure information of fused PET. Moreover, one of key issues confronting by the MRA approach is designing the specific fusion rule, which is very crucial to the fusion effect. The third type of method for fusing PET and MRI images is sparse representation (SR)-based method [11], [12]. This type of approach first solves the sparse representation coefficients both for PET and for MRI images, respectively; then merges the calculated coefficients via specific fusion rule; lastly reconstructs the target image using the fused sparse coefficients and a predefined/learnt over-complete dictionary. Sparse representation has achieved remarkable effects on image fusion. However, in most of the proposed SR-based methods for fusing PET and MRI images, the dictionary is learnt or constructed using the entire image, i.e., extracting the image patches from the entire image to learn or construct a global dictionary. Since the structural similarities among the image patches are not considered while learning or constructing the global dictionary, hence, the sparse coefficients solved by this kind of global dictionary are not very suitable for accurately reconstructing the target image [13]. Furthermore, similar to the MRA-based fusion method, designing the specific fusion rules is also an inevitable issue encountering by the SR-based fusion method. The last but not the least, inspired by other new ideas, the methods for fusing PET and MRI images include such as nonparametric density model-based method [14], ant colony optimization-based method [15] and so on.

Although these recent advanced methods achieve remarkable performance, one of major problems involved in these methods is designing fusion rule. Unfortunately, the fusion rules in the most of existing approaches are manually designed, and become more and more complicated. As a result, the fusion schemes with these complex hand-crafted fusion rules inevitably have the limitations such as implementation difficulty and time-consuming computation.

In recent few years, deep learning (DL) has become one of the most attractive topics in the field of computer vision due to its strong ability to extract image features. Correspondingly, in the field of image fusion, DL has also been successfully applied to various applications, such as remote sensing image fusion [16]–[18], multi-focus image fusion [19]–[21], medical image fusion [22] etc. Liu *et al*. [23] comprehensively summarized DL-based methods for image fusion in details. Actually, most of the proposed DL-based methods for image fusion are realized based on convolutional neural network (CNN), in which a critical prerequisite must be satisfied, i.e., the ground truth should be available in advance. However, in the task of fusing PET and MRI images, it is nearly impossible to establish the ground truth due to defining a standard for final fused images is unrealistic. Moreover, in order to complete the image fusion task, most of the proposed CNN models require additional post-processing procedures because they are not designed in the end-to-end manner [24].

More recently, generative adversarial network (GAN) has drawn a tremendous amount of attention, and has been successfully applied to various applications in the field of computer vision and machine learning, especially to the image synthesis [25]–[27]. In the particular case of image fusion, Ma *et al*. [28] firstly applied the GAN to the image fusion, i.e., fusion of infrared and visual images. Ma *et al*. [29] further improved the image fusion algorithm for infrared and visual images by adding an edge-enhancement constraint. Guo *et al*. [30] proposed an algorithm for multi-focus image fusion using conditional GAN. To the best of our knowledge, there are no reports on the application of GAN and its variants to the fusion of medical images.

According to the above analysis, and inspired by the [28], we propose a method for brain PET and MRI image fusion through the generative adversarial mechanism. Specifically, conditioned on the multiple input images together with the tissue label map generated from the input MRI image, a novel end-to-end tissue-aware framework based on conditional generative adversarial network (TA-cGAN) is proposed. Similar to the original GAN, the training procedure of our proposed TA-cGAN like a two-layer min-max game in which the generator and the discriminator are trained simultaneously with the goal of one beating another, i.e., the generator attempts to output a fused image mainly contains the color information of PET, whereas the discriminator tries to urge the fused image to include more anatomical information of MRI. Furthermore, our proposed TA-cGAN is an

end-to-end model, in which the fused image can be generated automatically from the combining of source images and the tissue label map without manually designing the complicated fusion rules.

The reminder of this paper is organized as follows. In Section II, related works regarding the generative adversarial network and the conditional generative adversarial network are briefly reviewed. Section III details our proposed method. Experiments and analysis are presented in Section IV. The concluding remarks are given in Section V.

II. RELATED WORKS

A. GENERATIVE ADVERSARIAL NETWORK

Generative adversarial network was firstly proposed by Goodfellow *et al*. [31] in 2014, and has drawn appealing attention in the field of machine learning and computer vision. The GAN is a generative model which consists of two adversarial networks namely generator G and discriminator D. The generator attempts to generate fake but plausible samples, whereas the discriminator tries to distinguish between the generated samples and the real samples. Specifically, the generator learns to capture the real data distribution and then generate new plausible samples so as to fool the discriminator, while the discriminator learns to distinguish the model generated distribution from the real data distribution. The two networks are trained against each other until the discriminator be unable to tell whether the generated samples come from the generator or not.

Mathematically, in the original GAN, D and G are trained in a competitive fashion by solving the following min-max optimization problem:

$$\min_G \max_D \mathrm{V}(D, G) = \mathbb{E}_{x\sim P_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim P_z(z)}[\log(1 - D(G(z)))], \quad (1)$$

where x is the real sample from true dataset, and z is the noise; $G(\cdot)$ and $D(\cdot)$ denote the output of the generator G and the discriminator D, respectively; P_{data} denotes the real data distribution, and P_z denotes the prior distribution of noise. G tries to minimize the above objective function as shown in (1) whereas D attempts to maximize it.

B. CONDITIONAL GENERATIVE ADVERSARIAL NETWORK

In the standard GAN, there is no control on modes of the synthesized data. Actually, it is possible to guide the sample synthesis by conditioning the GAN on auxiliary information, such as class label, text information, data from other modalities, et al. Hence, Mirza and Osindero [32] extended the basic GAN framework to the conditional generative adversarial network (cGAN) by feeding the auxiliary information into both the generator and the discriminator as extra input layer.

Mathematically, in the cGAN, D and G are trained by solving the following two-player optimization problem:

$$\min_G \max_D \mathrm{V}(D, G) = \mathbb{E}_{x\sim P_{data}(x)}[\log D(x|y)] + \mathbb{E}_{z\sim P_z(z)}[\log(1 - D(G(z|y)))], \quad (2)$$

where y is the auxiliary input which could be any kind of extra information.

III. PROPOSED METHOD

A. PIPELINE OF PROPOSED METHOD

The main goal of this study is to design a method for fusing a pair of pseudo color PET image and a gray MRI image so as to obtain the fused image with meaningfully complementary information as much as possible. In particular, the conditional generative adversarial network is employed to fulfill the fusion of PET and MRI images. More specifically, we regard the PET and MRI image fusion task as a two-player adversarial game between the generator and the discriminator, where both the generator and the discriminator are conditioned on the tissue label map which is generated from MRI image.

Fig. 1 illustrates the general pipeline of our proposed method that consists of training and testing stages, where I_P stands for the PET image, I_M stands for the MRI image, I_L denotes the tissue label map, and I_F denotes the fusion result.

Assume we have a set of ''pairs'' of PET and MRI images, and all paired PET and MRI images are registered. We further suppose that all MRI images are segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) so as to obtain the tissue label maps.

In the training stage, firstly stack the PET image I_P, the MRI image I_M together with the tissue label map I_L. Then feed the concatenated ''image'' into the generator and obtain the fused image I_F. Next, input the fused image I_F, the MRI image I_M, and the tissue label map I_L into the discriminator whose goal is attempting to distinguish I_F from I_M. It worth noting that both the generator and the discriminator are conditioned on the tissue label map I_L, and trained simultaneously in a competitive fashion, i.e., the generator tries to contain more and more color information from the PET image I_P, while the discriminator urges the fused image I_F include more and more structural details from the MRI image I_M. In this way, the fused image I_F will gradually include more and more anatomical details from the MRI image I_M. Once the generated image (i.e., I_F) produced by the generator cannot be distinguished by the discriminator, the final expected fused image I_F is obtained. In the testing stage, firstly concatenate the PET image I_P, the MRI image I_M together with the corresponding tissue label map I_L, then input the concatenated ''image'' into the trained generator, and finally get the fused image I_F.

B. JOINT LOSS FUNCTION

In the framework of traditional GAN, the generator G aims to generate samples by using random noise that follows a prior probability distribution $z \sim P_z(z)$. In our proposed TA-cGAN, instead of using random noise as the input, we condition the model on multiple images from different modality, i.e., the PET image I_P, the MRI image I_M and its corresponding label map I_L. Furthermore, different from the conventional GAN in which the log likelihood cost is used for the

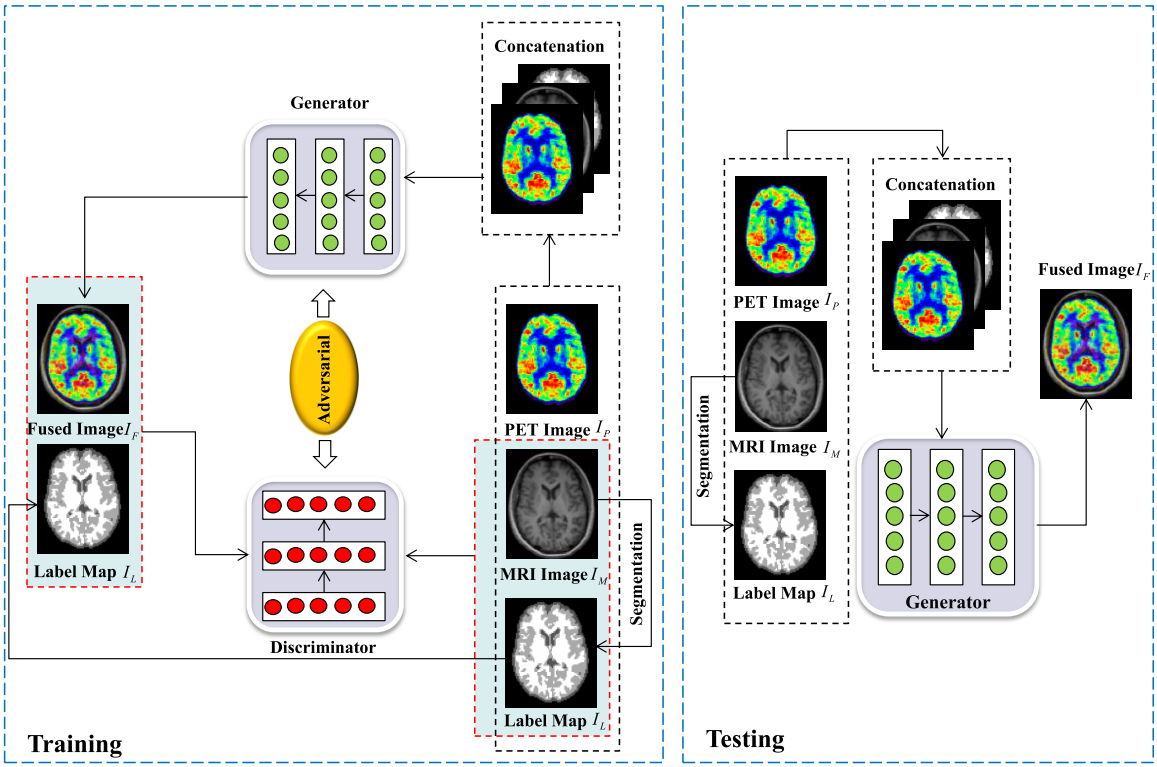


FIGURE 1. Pipeline of the proposed method for fusing brain PET and MRI images.

adversarial loss, we adopt the least square loss which has been proved can boost training stability as well as generate high quality image [33]. During the training process, to satisfy with the PET and MRI image fusion task, except for only using adversarial loss to train the generator G, our proposed TA-cGAN utilizes joint losses including spectral loss L_{Spec}, structural loss L_{Str}, and adversarial loss L_{Adv}. Mathematically, the joint loss used in our work is expressed as follows:

$$L_{Joint} = \lambda_1 L_{Spec} + \lambda_2 L_{Str} + \lambda_3 L_{Adv}, \tag{3}$$

where the spectral loss L_{Spec} urges the fused image to contain similar color information (Characterized by the pixel intensities of PET image) as those of the PET image; The structural loss L_{Str} attempts to make the fused image has similar structure information (Characterized by the gradients of MRI image) as those of the MRI image; The adversarial loss L_{Adv} aims to add more detailed information to the fused image; λ_1, λ_2,and λ_3 are the corresponding weights for spectral loss, structural loss and adversarial loss, respectively.

1) SPECTRAL LOSS

Formally, the spectral loss L_{Spec} is defined based on mean square error (MSE) as follows:

$$L_{Spec} = \frac{1}{MN} ||I_P - I_F||_2^2, \tag{4}$$

where I_P is the original PET image; I_F is fused image generated by the generator G; M and N denote the width and height of the image. The spectral loss mainly tries to make the fused image similar with the PET image in terms of pixel intensities, i.e., to make the fused image I_F preserve the color information contained in the PET image I_P.

2) STRUCTURAL LOSS

The structural loss L_{Str} is defined based on image gradient difference as follows:

$$L_{Str} = \frac{1}{MN}(||\nabla_x(I_M) - \nabla_x(I_F)||_2^2 + ||\nabla_y(I_M) - \nabla_y(I_F)||_2^2), \tag{5}$$

where ∇_x and ∇_y denote the gradient operation of image with respect to the horizontal and vertical direction, respectively. The structural loss attempts to minimize the magnitude difference of the gradients between the MRI image and the fused image, i.e., to make the fused image I_F retain the gradient information contained in the MRI image I_M.

3) ADVERSARIAL LOSS

The adversarial loss L_{Adv} is defined based on the probabilities of the discriminator D over the concatenated training data $I_{Concat} = \{I_P, I_M, I_L\}$ (the PET image I_P, the MRI image I_M and its corresponding label map I_L) as follows:

$$L_{Adv} = [D(G(I_{Concat})) - c]^2, \tag{6}$$

where c denotes the value that the generator G wants the discriminator D to believe for fake data.

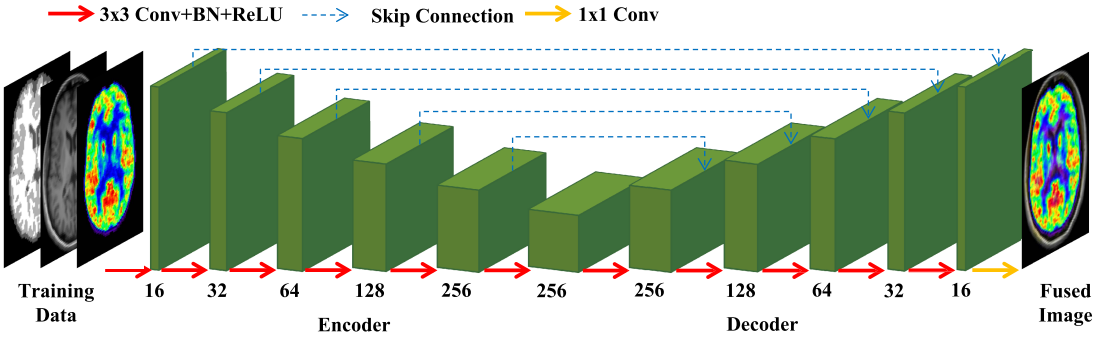


FIGURE 2. U-Net-like network architecture of the generator. Green boxes stand for feature maps with the number of channels indicated under each box. The arrows denote different kinds of operations, where Conv = convolution, BN = batch normalization, and ReLU = rectified linear unit.

C. NETWORK ARCHITECTURE

Similar to the original GAN, the proposed TA-cGAN also consists of two sub-networks, i.e., the generator and the discriminator. However, different from the original GAN which is mainly used for image-to-image translation, our proposed TA-cGAN is designed for images-to-image translation, i.e., input multiple images (PET image I_P, MRI image I_M and its corresponding label map I_L) and output one fused image I_F. The network architectures of the TA-cGAN are detailed as follows.

1) NETWORK ARCHITECTURE OF GENERATOR

In our proposed method, the generator is constructed based on the U-Net [34]. The U-Net utilizes skip connection technique to integrate the low-level feature coming from the shallow encoder layers and the high-level feature coming from the deep decoder layers. Moreover, the skip connection technique can be used to partially solve the problem of gradient vanishing. Due to adopting the idea of skip connection, the U-Net has been successfully applied to many image applications, such as image synthesis [27]. In this work, the network architecture of generator G consists of two parts, i.e., the encoder and the decoder, as shown in Fig. 2. The inputs of the network are the PET image I_P, the MRI image I_M and its corresponding label map I_L; and the output of the network is the fused image I_F.

Specifically, as illustrated in Fig. 2, the entire generator network is composed of 12 convolutional layers. The encoder part consists of 6 down-sample layers that perform convolutions using 3×3 filters with stride 2 in each direction, batch normalization (BN), and rectified linear unit (ReLU) activation operations with slope of negative 0.2. Note that, we do not use pooling operation mainly because it will reduce the spatial resolution of feature maps and will make the network unable to capture fine details in the MRI images. In addition, zero padding with 1×1 in each down-sample layer is employed. The decoder part consists of 6 up-sample layers, where the first five layers perform convolution-BN-ReLU operations, and the last layer only perform convolutional operation using 1×1 filter. In the decoder part, the feature maps in the encoder layers are concatenated with those in the decoder layers using skip connection (as indicated by the dotted arrows in Fig. 2).

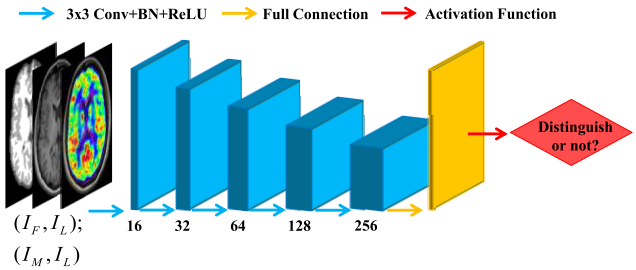


FIGURE 3. Network architecture of the discriminator. The blue boxes denote the convolutional layers with the number of channels indicated below each box; the orange box denotes the fully connected layer. The arrows denote different kinds of operations, where Conv = convolution, BN = batch normalization, and ReLU = rectified linear unit.

2) NETWORK ARCHITECTURE OF DISCRIMINATOR

Different from the generator G, the discriminator D is mainly designed for solving the problem of classification. Specifically, in this study, the major goal of the discriminator D is attempting to distinguish the fused image pair (Fused image I_F and label map I_L) from the MRI image pair (MRI image I_M and label map I_L). Fig.3 illustrates the network architecture of the discriminator D used in our study. As shown in Fig. 3, the inputs of the discriminator D is either the fused image pair or the MRI image pair; and the output of the network is the class label, i.e., distinguished (labeled by 1) or not (labeled by 0).

Briefly, as illustrated in Fig. 3, our network architecture of the discriminator D is a simple convolutional neural network consisting of 5 convolutional layers and 1 fully connected layer followed by a sigmoid activation function. The five convolutional layers, similar to the encoder structure of the generator G, perform the convolution-BN-ReLU operations.

D. TRAINING PARADIGM

Similar to the original GAN, the generator network G and the discriminator D are trained alternatively. Specifically, first fix G to train D for one step according to the joint loss function

as (3), and then fix D to train G for one step too. More intuitively, the training process of the generator G and the discriminator D is just like playing a two-player min-max game, where the generator G aims to minimize the loss function, whereas the discriminator D attempts to maximum it. In this way, the training process will continue, and both the generator G and the discriminator D will gradually become more and more powerful until the termination condition of iteration is satisfied. In the training stage, both G and D are optimized using the Adam solver [35] with $\beta = 0.5$ and learning rate of 0.0002. Note that, the settings for other parameters during the training process as well as the preparation of training samples will be elaborated in the section IV.A.

In the testing stage, first concatenate the PET image I_P, the MRI image I_M, and its corresponding label map I_L; then input the concatenated image into the trained generator G, and output the final fused image I_F.

IV. EXPERIMENTS AND ANALYSIS

In this section, we firstly introduce the experimental settings including experimental data and preprocessing, parameters' setting, compared methods, and evaluation metrics. Then we demonstrate and analyze the experimental results both visually and quantitatively.

A. EXPERIMENTAL SETTINGS

1) EXPERIMENTAL DATA AND PREPROCESSING

In order to validate the performance of our proposed method, we use a publicly available dataset of Whole Brain database (http://www.med.harvard.edu/aanlib/) which is created by the School of Medicine, Harvard University. In this study, 36 pairs of PET and MRI images are collected for the usage of experiments. The collected images include 30 cases of normal control (NC) and 6 cases of mild Alzheimer's disease (AD). For the case of normal control, both the PET and MRI images have the same size of 256 × 256. However, for the case of mild AD, the sizes of the PET images are different from those of the MRI images, i.e., the PET images have the size of 128 × 128, whereas the MRI images have the size of 256 × 256. Therefore, it is necessary to firstly reduce the size of the MRI images into 128 × 128 for the case of mild AD. Furthermore, all the collected MRI images (including the resized MRI images) are segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) by the HMRF-EM algorithm [36].

Normally, large amount of training samples is preferable to train deep neural networks. However, the number of training samples is limited in our study. Hence, in the process of training data preparation, we adopt the data augmentation technique to expand the training samples. Moreover, instead of using the entire images as input, we take the large image patches with the size of 64 × 64 as input. The detailed information of preparing the training data is elaborated as follows:

(1) First, each paired images (PET image, MRI image and its corresponding label map) were flipped from left to right, and then from top to down. Thus, we expand the samples from 36 pairs to 144 pairs which include 120 pairs for the case of normal control and 24 pairs for the case of mild AD.

(2) Next, for all image pairs obtained in the step (1), we crop the entire image without overlap into large image patches with size of 64 × 64, and thus increase the training samples from 144 pairs to 2016 pairs, where 1920 pairs were cropped from the images with normal control, and 96 pairs were cropped from the images with mild AD.

2) PARAMETERS' SETTINGS

In the training stage, the network was trained using the Adam optimizer with an initial learning rate of 0.0002 and a mini-batch size of 10 over 100 epochs. The number of training iterations is set to 100. $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = 0.5$. $c = 0.9$, where c is a label value as shown in (6).

3) COMPARED METHODS

For the purpose of validating the performance of our proposed method, the following five state-of-the-art methods are used to compare with our method: the IHS combined with retina-inspired models (IHS-Retina) method [5], the non-subsampled shearlet transform (NSST) method [10], the low-rank sparse dictionaries learning (LSDL) method [11], the nonparametric density model (NDM) method [14], and the convolutional neural networks (CNNs) method [22].

4) EVALUATION METRICS

It is usually difficult to assess the fusion performance only via visually subjective evaluation. Therefore, it is necessary to choose some quantitative metrics to objectively evaluate the performances of different fusion methods. In this paper, we adopt the following four commonly used metrics to evaluate the performances of different methods: the entropy (EN) [37], the average gradient (AG) [38], the spectral discrepancy (SD) [5], and the $Q^{AB/F}$ [39]. The definitions of these four metrics are sequentially presented as follows:

EN is mainly used to measure the amount of information contained in the fused image. Mathematically, EN is formulated as follows:

$$EN = -\sum_{i=0}^{L-1} P(i)\log_2 P(i), \tag{7}$$

where L denotes the number of gray scale, and it is 256 in our experiments. $P(i)(i = 0, 1, \ldots, L-1)$ is the occurring probability of the pixels with the gray scale $i(i = 0, 1, \ldots, L-1)$ in the fused image. Normally, the larger EN is, the richer information is contained in the fused image, and the better performance is achieved by the fused method.

AG usually reflects the clarity of the fused image, and is mainly used to measure the spatial resolution of the fused

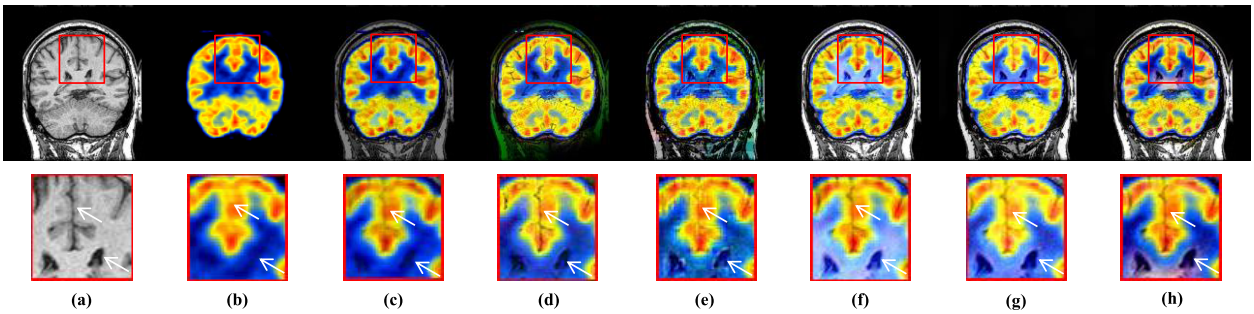


FIGURE 4. Fusion results for a case of normal control. (a) MRI image; (b) PET image; (c) Result using IHS-Retina; (d) Result using NSST; (e) Result using LSDL; (f) Result using NDM; (g) Result using CNNs; (h) Result using proposed method.

image. Formally, AG is defined as follows:

$$AG_k = \frac{1}{(M-1)(N-1)} \times \sum_{x=1}^{M-1}\sum_{y=1}^{N-1}\sqrt{\frac{(\frac{\partial F_k(x,y)}{\partial x})^2 + (\frac{\partial F_k(x,y)}{\partial y})^2}{2}}, \quad k = R, G, B, \tag{8}$$

where $F_k(x, y)$is the pixel value of the fused image at position (x, y). R, G, B are the three components of the fused image with the size of $M \times N$. In this paper, $M = N = 256$ for the case of normal control, and $M = N = 128$ for the case of mild AD. Simply, the larger AG is, the higher spatial resolution fused image has.

SD is mainly used to measure the spectral (color) quality of fused image. Mathematically, SD is expressed as follows:

$$SD_k = \frac{1}{M \cdot N}\sum_{x=1}^{M}\sum_{y=1}^{N}|F_k(x, y) - O_k(x, y)| \quad k = R, G, B, \tag{9}$$

where $F_k(x, y)$ and $O_k(x, y)$ are the pixel values of the fused image and the original PET image at position (x, y), respectively. The meanings of R, G, B and M, N are same as those of (8). A small SD indicates a good fusion result. In other words, smaller AD indicates that the color of the fused image is closer to that of the original PET image.

$Q^{AB/F}$ is mainly used to measure the edge preservation from the source images during the process of fusion. $Q^{AB/F}$ is mathematically defined as follows:

$$Q^{AB/F} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(Q^{AF}(x, y)\omega^A(x, y) + Q^{BF}(x, y)\omega^B(x, y))}{\sum_{x=1}^{M}\sum_{y=1}^{N}(\omega^A(x, y) + \omega^B(x, y))}, \tag{10}$$

where A and B denote the two source images, and F represents the fused image. $Q^{AF}(x, y)$ and $Q^{BF}(x, y)$ are the edge preservation values. $\omega^A(x, y)$ and $\omega^B(x, y)$ are the weights. $Q^{AF}(x, y)$ and $Q^{BF}(x, y)$ are weighted by $\omega^A(x, y)$ and $\omega^B(x, y)$, respectively. Usually, a larger $Q^{AB/F}$ means a good fusion result.

B. EXPERIMENTAL RESULTS WITH VISUAL AND STATISTICAL ANALYSIS

To demonstrate the advantage of our proposed method in terms of fusion effect, in this section, the proposed method is compared with other five competitive methods on two aspects: subjectively visual evaluation and objectively quantitative assessment.

1) SUBJECTIVELY VISUAL EVALUATION

To qualitatively compare the fusion performances of the proposed method with those of the other five state-of-the-art fusion methods mentioned in the section IV. A., we visually demonstrate the fusion results for two cases of PET and MRI images, i.e., the case of normal control as well as the case of mild AD. Subsequently, we analysis the fusion results from two aspects, i.e., the structural details extraction from the original MRI images and the color fidelity preservation from the original PET images.

Fig. 4 shows the fusion results using different fusion methods for a case of normal control. Similarly, the fusion results achieved by different fusion methods for a case of mild AD are displayed in Fig. 5. Note that, for easily observing the differences among the fused images resulted by the different methods, the regions marked by the red rectangles are enlarged and displayed under their corresponding fused image, as shown in Fig. 4 and Fig. 5.

From Fig. 4, it can be seen that the LSDL method fails to preserve the anatomical details [Pointed by the top white arrow as shown in the close-up region of Fig. 4 (e)] from the original MRI image; the anatomical structure [Pointed by the top white arrow as shown in the close-up region of Fig. 4 (c)] is blurred by the IHS-Retina method; Both the NDM method and the CNNs method retain the anatomical structures more or less [Pointed by the top white arrows as shown in the close-up regions of Figs. 4 (f) and (g), respectively] from the original MRI image. In contrast, both the NSST method and the proposed method successfully integrate the structure details from the original MRI image to the fused images [Pointed by the top white arrows as shown in the close-up regions of Figs. 4 (d) and (h), respectively]. Nevertheless, the anatomical structure resulted by the proposed method is clearer than that of the NSST method [Pointed by the right-down

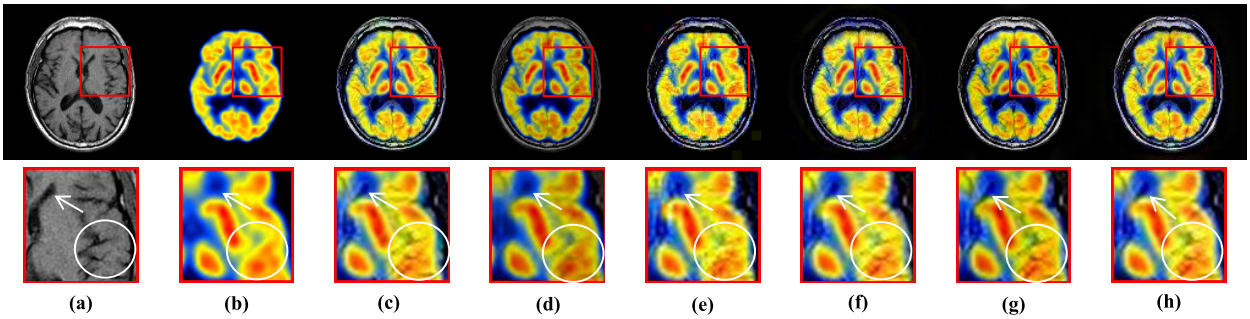


FIGURE 5. Fusion results for a case of mild AD. (a) MRI image; (b) PET image; (c) Result using IHS-Retina; (d) Result using NSST; (e) Result using LSDL; (f) Result using NDM; (g) Result using CNNs; (h) Result using proposed method.

white arrows as shown in the close-up regions of Figs. 4(h) and (d), respectively].

From Fig. 5, it can be observed that both the NSST method and the LSDL method blurred the structural details [Marked by the white circles as shown in the close-up regions of Figs. 5(d) and (e), respectively] coming from the original MRI image; The IHS-Retina method and the NDM method preserved the structural details [Marked by the white circles as shown in the close-up regions of Figs. 5(c) and (f), respectively] more or less from the original MRI image. By contrast, both the CNNs method and the proposed method well retained the anatomical structures [Marked by the white circles as shown in the close-up regions of Figs. 5(g) and (h), respectively] from the original MRI image. Nevertheless, the anatomical structure resulted by the proposed method is closer to the MRI image than that of the CNNs method [Pointed by the top white arrows as shown in the close-up regions of Figs. 5(h) and (g), respectively].

Moreover, in terms of the ability to preserve color fidelity from the original PET image, from Fig. 4, it can be seen that the LSDL method resulted a serious color distortion [See the close-up region of Fig. 4(e)]; The color in the fused images produced by the NDM method and the CNNs method is much lighter [See the blue parts of the fused images as shown in the close-up regions of Figs. 4 (f) and (g), respectively] than that of the original PET image. Compared with the NDM method and the CNNs method, the IHS-Retina method preserve the color information better [See the blue part of the fused image as shown in the close-up region of Fig. 4 (c)] from the original PET image. In contrast, the NSST method and the proposed method successfully preserve the spectral color information from the original PET image [See the blue parts of the fused images as shown in the close-up regions of Figs. 4(d) and (h), respectively]. Nevertheless, compared the red color in the fused image [See the red part of the fused image as shown in the close-up region of Fig. 4(d)] generated by the NSST method, the red color in fused image [See the red part of the fused image as shown in the close-up region of Fig. 4(h)] produced by the proposed method is closer to that of the original PET image.

Similarly, as shown in Fig. 5, we can also see that the LSDL method seriously distorted the spectral color information [See the blue part as shown in the close-up region of Fig. 5(e)]; both the IHS-Retina method and the NDM method preserve the color information more or less [See the blue parts as shown in the close-up regions of Figs. 5(c) and (f)] from the original PET image. By contrast, the NSST method, the CNNs method and the proposed method preserve the color information well [See the blue parts as shown in the close-up regions of Figs. 5(d), (g) and (h)] from the original PET image. However, for one thing, the NSST method seriously distorted the structural details [Marked by the white circle as shown in the close-up region of Fig. 5(d)] in the MRI image; for another, the clarity of the anatomical structure resulted by the CNNs method is worse than that of the proposed method does [Pointed by the top white arrows as shown in the close-up regions of Figs. 5(g) and (h), respectively].

In summary, comprehensively considering both the structural details extraction from the original MRI images and the color fidelity preservation in the original PET image, we can conclude that the proposed method achieves the best fusion results in terms of visual quality than other five fusion methods.

2) OBJECTIVELY QUANTITATIVE ASSESSMENT

To quantitatively compare the fusion performance of the proposed method with those of the competitive fusion methods, we investigate the statistical results of different fusion methods for fusing two types of PET and MRI images, i.e., the images of normal control (NC) and the images of mild AD. In our study, four popular-used objective metrics i.e., EN, AG, SD and $Q^{AB/F}$, are exploited to validate the fusion performance of different methods. The statistical results in terms of EN, AG, SD and $Q^{AB/F}$ are tabulated in Table 1, Table 2, Table 3, and Table 4, respectively. Note that, the best performance is highlighted in bold.

From Table 1, we can see that the proposed method ranks the first place in terms of EN, i.e., it achieves the largest average value of EN over all fused images including the case of normal control and the case of mild AD. This fact implies that the fused images resulted by our proposed method contain more information including spectral colors and anatomical structures. The reason for this is mainly due to incorporating

TABLE 1. Quantitative comparison of different image fusion methods in terms of EN.

Methods	EN		
	NC	Mild AD	Average
IHS-Retina [5]	4.5367	4.0214	4.2791
NSST [10]	5.4900	4.9119	5.2010
LSDL [11]	5.4395	4.5933	5.0164
NDM [14]	4.9434	4.6152	4.7793
CNNs [22]	5.5237	**5.2100**	5.3669
Proposed method	**5.8355**	5.1445	**5.4900**

TABLE 2. Quantitative comparison of different image fusion methods in terms of AG.

Methods	AG		
	NC	Mild AD	Average
IHS-Retina [5]	13.1184	7.3451	10.2318
NSST [10]	13.0233	10.1796	11.6015
LSDL [11]	12.4099	10.5382	11.4741
NDM [14]	11.9157	10.4233	11.1695
CNNs [22]	13.0976	11.2267	12.1622
Proposed method	**13.2410**	**11.7824**	**12.5117**

TABLE 3. Quantitative comparison of different image fusion methods in terms of SD.

Methods	SD		
	NC	Mild AD	Average
IHS-Retina [5]	7.6704	7.0038	7.3371
NSST [10]	5.8379	5.9454	5.8917
LSDL [11]	**5.1625**	9.0478	7.1052
NDM [14]	5.2624	9.5238	7.3931
CNNs [22]	5.2149	5.7318	5.4734
Proposed method	5.1989	**5.4125**	**5.3057**

TABLE 4. Quantitative comparison of different image fusion methods in terms of $Q^{AB/F}$.

Methods	$Q^{AB/F}$		
	NC	Mild AD	Average
IHS-Retina [5]	0.4928	0.4162	0.4545
NSST [10]	0.5903	**0.6278**	0.6091
LSDL [11]	0.5425	0.4935	0.5180
NDM [14]	0.5301	0.4835	0.5068
CNNs [22]	0.6415	0.6098	0.6257
Proposed method	**0.6602**	0.6271	**0.6437**

the spectral loss as well as the structural loss into the loss function as shown in (3).

Similarly, From Table 2, it can be observed that our proposed method achieves the largest value of AG across both the normal control images and the mild AD images. This indicates that the fused images produced by our proposed method have higher spatial resolution than those images generated by other competing methods.

Again, from Table 3, we can observe that our proposed method achieves the smallest average value in terms of SD. This means that compared with other fusion methods, the proposed method preserves more spectral color information from the original PET images. In other words, the spectral colors of the fused images generated by the proposed method are

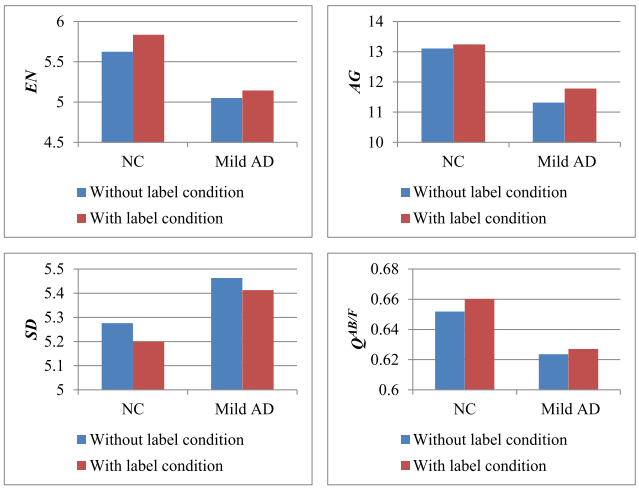


FIGURE 6. Quantitative comparison between no tissue label conditioned model and the tissue label conditioned model, in terms of EN, AG, SD, and $Q^{AB/F}$, respectively. Note that the NC denotes the case of normal control, and the Mild AD denotes the case of mild Alzheimer's disease (AD).

closer to that of the original PET images. The main reason for this is due to incorporating the spectral loss into the loss function as shown in (3).

Also, from Table 4, it is clearly that the proposed method ranks the first place in terms of $Q^{AB/F}$. In other words, the proposed method obtains the highest average value of $Q^{AB/F}$. This reflects that compared with other fusion methods, our proposed method has stronger ability to preserve edge details from the source images, i.e., PET and MRI images.

Overall, from aforementioned four tables, we can conclude that the proposed method achieves the best fusion results in terms of quantitative assessment than other five fusion methods. Specifically, the images fused by the proposed method are more informative, clearer. Moreover, our proposed method can produce the fused images with less color distortion and more structural details.

C. EFFECTIVENESS OF LABEL CONDITION

In our proposed method, TA-cGAN, both the generator G and the discriminator D are conditioned on the tissue label map generated via segmenting the MRI image into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). To verify the contribution of the tissue label condition to performance improvement, we perform comparison experiments on two cases of data using the label conditioned model and the model without label condition, respectively. Fig. 6 visually illustrates the experimental results using two models in terms of previously mentioned four evaluation metrics, i.e., EN, AG, SD and $Q^{AB/F}$.

As shown in Fig. 6, we can find that the tissue label conditioned model performs better on two cases of image fusions than the model trained without label condition.

This fact proves that the tissue label extracted from the MRI image is really helpful for improving the fusion performance in this study.

V. CONCLUSION

In this paper, we propose a novel tissue-aware conditional generative adversarial network called TA-cGAN for fusing the brain PET and MRI images. In our proposed method, both the generator G and the discriminator D are conditioned on the tissue label map generated from the MRI images. In addition, adversarial loss, spectral loss, and the structural loss are incorporated to capture both the spectral colors from the original PET image and the anatomical structures from the original MRI image. Extensive experiments demonstrate that our proposed TA-cGAN outperforms the state-of-the-art fusion methods both in visual perception and in quantitative assessment. Specifically, the fused images generated by our proposed method contain more spectral colors and include more structural details than those images fused by other competing methods. In the future, we will mainly focus on improving the performance of the TA-cGAN via including more cases of PET and MRI images, and extending TA-cGAN to address the general problems of multi-modality medical image fusions.

REFERENCES

[1] A. Mehranian, M. A. Belzunce, C. Prieto, A. Hammers, and A. J. Reader, "Synergistic PET and SENSE MR image reconstruction using joint sparsity regularization," *IEEE Trans. Med. Imag.*, vol. 37, no. 1, pp. 20–34, Jan. 2018.

[2] J. Zhang, D. Chen, J. Liang, H. Xue, J. Lei, Q. Wang, D. Chen, M. Meng, Z. Jin, and J. Tian, "Incorporating MRI structural information into bioluminescence tomography: System, heterogeneous reconstruction and *in vivo* quantification," *Biomed. Opt. Express*, vol. 5, no. 6, pp. 1861–1876, Jun. 2014.

[3] T. Mäkelä, Q. C. Pham, P. Clarysse, J. Nenonen, J. Lötjönen, O. Sipilä, H. Hänninen, K. Lauerma, J. Knuuti, T. Katila, and I. E. Magnin, "A 3-D model-based registration approach for the PET, MR and MCG cardiac data fusion," *Med. Image Anal.*, vol. 7, no. 3, pp. 377–389, Sep. 2003.

[4] M.-L. Jan, K.-S. Chuang, G.-W. Chen, Y.-C. Ni, S. Chen, C.-H. Chang, J. Wu, T.-W. Lee, and Y.-K. Fu, "A three-dimensional registration method for automated fusion of micro PET-CT-SPECT whole-body images," *IEEE Trans. Med. Imag.*, vol. 24, no. 7, pp. 886–893, Jul. 2005.

[5] S. Daneshvar and H. Ghassemian, "MRI and PET image fusion by combining IHS and retina-inspired models," *Inf. Fusion*, vol. 11, no. 2, pp. 114–123, Apr. 2010.

[6] C.-I. Chen, "Fusion of PET and MR brain images based on IHS and log–Gabor transforms," *IEEE Sensors J.*, vol. 17, no. 21, pp. 6995–7010, Nov. 2017.

[7] M. Haddadpour, S. Daneshvar, and H. Seyedarabi, "PET and MRI image fusion based on combination of 2-D Hilbert transform and IHS method," *Biomed. J.*, vol. 40, no. 4, pp. 219–225, Aug. 2017.

[8] P. W. Huang, C. I. Chen, P. Chen, P. L. Lin, and L. P. Hsu, "PET and MRI brain image fusion using wavelet transform with structural information adjustment and spectral information patching," in *Proc. IEEE ISBB*, Chung Li, Taiwan, Apr. 2014, pp. 1–4.

[9] L. Wang, B. Li, and L. Tian, "Multimodal medical, volumetric data fusion using 3-D discrete shearlet transform and global-to-local rule," *IEEE Trans. Bio-Med. Eng.*, vol. 61, no. 1, pp. 197–206, Jan. 2014.

[10] H. Ouerghi, O. Mourali, and E. Zagrouba, "Non-subsampled shearlet transform based MRI and PET brain image fusion using simplified pulse coupled neural network and weight local features in YIQ colour space," *IET Image Process.*, vol. 12, no. 10, pp. 1873–1880, Oct. 2018.

[11] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning," *Pattern Recognit.*, vol. 79, pp. 130–146, Jul. 2018.

[12] Z. Zhu, H. Yin, Y. Chai, Y. Li, and G. Qi, "A novel multi-modality image fusion method based on image decomposition and sparse representation," *Inf. Sci.*, vol. 432, pp. 516–529, Mar. 2018.

[13] S. Liu, G. Zhang, and W. Liu, "Group sparse representation based dictionary learning for SAR image despeckling," *IEEE Access*, vol. 7, pp. 30809–30817, 2019.

[14] Z. Liu, Y. Song, V. S. Sheng, C. Xu, C. Maere, K. Xue, and K. Yang, "MRI and PET image fusion using the nonparametric density model and the theory of variable-weight," *Comput. Methods Programs Biomed.*, vol. 175, pp. 73–82, Jul. 2019.

[15] H. R. Shahdoosti and Z. Tabatabaei, "MRI and PET/SPECT image fusion at feature level using ant colony based segmentation," *Biomed. Signal Process. Control*, vol. 47, pp. 63–74, Jan. 2019.

[16] P. Ghamisi, B. Hofle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[17] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017.

[18] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.

[19] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[20] R. Lai, Y. Li, J. Guan, and A. Xiong, "Multi–scale visual attention deep convolutional neural network for multi–focus image fusion," *IEEE Access*, vol. 7, pp. 114385–114399, 2019.

[21] H. T. Mustafa, J. Yang, and M. Zareapoor, "Multi-scale convolutional neural network for multi-focus image fusion," *Image Vis. Comput.*, vol. 85, pp. 26–35, May 2019.

[22] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *Proc. IEEE ICIF*, Xi'an, China, Jul. 2017.

[23] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.

[24] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020.

[25] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "StackGAN++: Realistic image synthesis with stacked generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1947–1962, Aug. 2019.

[26] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.

[27] Y. Wang, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, and L. Zhou, "3D conditional generative adversarial networks for high-quality PET image estimation at low dose," *NeuroImage*, vol. 174, pp. 550–562, Jul. 2018.

[28] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019.

[29] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, J. Wu, and J. Jiang, "Infrared and visible image fusion via detail preserving adversarial learning," *Inf. Fusion*, vol. 54, pp. 85–98, Feb. 2020.

[30] X. Guo, R. Nie, J. Cao, D. Zhou, L. Mei, and K. He, "FuseGAN: Learning to fuse multi–focus image via conditional generative adversarial network," *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 1982–1996, Aug. 2019.

[31] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, Montreal, QC, Canada, 2014, pp. 2672–2680.

[32] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: https://arxiv.org/abs/1411.1784

[33] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. ICCV*, Venice, Italy, 2017, pp. 2813–2821.

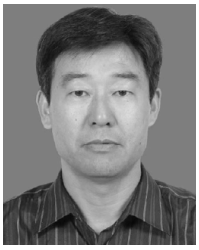
[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Munich, Germany, 2015, pp. 234–241.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[36] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
[37] J. W. Roberts, J. V. Aardt, and F. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, May 2008, Art. no. 023522.
[38] Z. Li, Z. Jing, X. Yang, and S. Sun, "Color transfer based remote sensing image fusion using non-separable wavelet frame transform," *Pattern Recognit. Lett.*, vol. 26, no. 13, pp. 2006–2014, Oct. 2005.
[39] C. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

JIAYIN KANG received the B.S. degree in measurement engineering and the M.S. degree in information system from the Liaoning Institute of Technology, Fuxin, China, in 1998 and 2004, respectively, and the Ph.D. degree in control engineering from the University of Science and Technology Beijing, Beijing, China, in 2008.

From 2008 to 2012, he was an Assistant Professor with the School of Electronics Engineering, Huaihai Institute of Technology, Lianyungang, China. From August 2013 to August 2014, he was a Visiting Scholar with the Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. He is currently an Associate Professor with the School of Electronics Engineering, Jiangsu Ocean University. His research interests include image processing, computer vision, and machine learning.



WU LU received the B.S.Med. degree in clinical medicine from Xuzhou Medical University, Xuzhou, China, in 1989.

From 1989 to 2006, he was the Associate Chief Physician with the Department of Radiology, The First People's Hospital of Lianyungang, Lianyungang, China, where he is currently the Chief Physician with the Department of Nuclear Medicine. His research interests include medical imaging and image analysis, and nuclear medicine.

WENJUAN ZHANG received the B.S. degree in computer science and technology and the M.S. degree in computer engineering from the Liaoning Institute of Technology, Fuxin, China, in 2001 and 2004, respectively.

From 2004 to 2006, she was an Assistant Lecture with the School of Computer Engineering, Huaihai Institute of Technology, Lianyungang, China. She is currently a Lecture with the School of Computer Engineering, Jiangsu Ocean University. Her research interests include image processing and pattern recognition.