IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 2, 2020, accepted March 17, 2020, date of publication March 19, 2020, date of current version March 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982016

MGMDcGAN: Medical Image Fusion Using Multi-Generator Multi-Discriminator Conditional Generative Adversarial Network

JUN HUANG[1,2], ZHULIANG LE[1], YONG MA[1,2], FAN FAN[1,2], HAO ZHANG[1], AND LEI YANG[3], (Member, IEEE)

[1]Electronic Information School, Wuhan University, Wuhan 430072, China
[2]Institute of Aerospace Science and Technology, Wuhan University, Wuhan 430072, China
[3]School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou 450007, China

Corresponding authors: Hao Zhang (zhpersonalbox@gmail.com) and Lei Yang (annieyanglei@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61903279, and in part by the Science and Technology Innovation Team of Colleges and Universities in Henan Province under Grant 18IRTSTHN013.

ABSTRACT In this paper, we propose a novel end-to-end model for fusing medical images characterizing structural information, i.e., I_S, and images characterizing functional information, i.e., I_F, of different resolutions, by using a multi-generator multi-discriminator conditional generative adversarial network (MGMDcGAN). In the first cGAN, the generator aims to generate a real-like fused image based on a specifically designed content loss to fool two discriminators, while the discriminators aim to distinguish the structure differences between the fused image and source images. On this basis, we employ the second cGAN with a mask to enhance the information of dense structure in the final fused image, while preventing the functional information from being weakened. Consequently, the final fused image is forced to concurrently keep the structural information in I_S and the functional information in I_F. In addition, as a unified method, MGMDcGAN can be applied to different kinds of medical image fusion, i.e., MRI-PET, MRI-SPECT, and CT-SPECT, where MRI and CT are two kinds of I_S of high resolution, PET and SPECT are typical kinds of I_F of low resolution. Qualitative and quantitative experiments on publicly available datasets demonstrate the superiority of our MGMDcGAN over the state-of-the-art.

INDEX TERMS Medical image fusion, generative adversarial network, different resolutions, end-to-end, unified method.

I. INTRODUCTION

Medical imaging is a fundamental and powerful tool playing a pivotal role in medical device industry for biomedical research and clinical applications such as medical testing, diagnosis and treatment, which can be divided into structural and functional systems. Magnetic resonance imaging (MRI) and computed tomography (CT) images often provide structural information, where CT shows excellent performance in detecting dense structure, such as bones and implants, and MRI provides texture details and dense structure information. While positron emission tomography (PET) and single-photon emission computed tomography (SPECT) images provide functional information [1]. Each of them conveys different information. However, a single medical imaging modality cannot provide sufficient information for its intended purpose. Owing to the strong complementarity between them, their inherent properties can be almost entirely presented by a fused image by minimizing redundant information while maximizing relevant information [2]. The fused result is more beneficial to human visual perception or automatic detection of the machine [3], [4].

The fusion of medical images is different from other fusion tasks. Specifically, the fusion of medical images and the fusion of infrared and visible images [5] are both the fusion of multi-modal images, but the fusion of medical images mainly extracts more than one kind of information from one source image. For example, in the fusion of medical

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

TABLE 1. The main abbreviations with their original words in this paper.

Original words	medical images characterizing structural information	medical images characterizing functional information	magnetic resonance imaging	computed tomography	positron emission tomography	single-photon emission computed tomography
abbreviations	I_S	I_F	**MRI**	**CT**	**PET**	**SPECT**
Original words	conditional generative adversarial network	intermediate fused image	final fused image	generator in first/second cGAN	discriminator 1/2/3	mask
abbreviations	**cGAN**	I_m	I_f	G_1/G_2	$D_Y/D_M/D_K$	**M**
Original words	entropy	spatial frequency	edge intensity	mean gradient	peak signal-to-noise ratio	structural similarity index measure
abbreviations	**EN**	**SF**	**EI**	**MG**	**PSNR**	**SSIM**

images, in addition to extracting the texture details of MRI image, the intensity information of its bones must also be extracted. On the contrary, multi-exposure and multi-focus images fusion are the fusion of source images taken under different settings in a single modality [6].

For this purpose, many methods have been proposed, which can be divided into six categories according to corresponding schemes including wavelet transformation based methods [7], pyramid methods [8], sparse representation methods [9], neural network based methods [10], salient feature methods [11] and other methods. We will discuss the detailed exposition of these traditional fusion methods and the relationship among them later in Sec. II-A. In these methods, there are in general three key components, i.e., image transform, activity level measurement, and fusion rule design. They typically use the same transform or representation for different source medical images during the fusion process. However, it may not be appropriate for the fusion of multi-modal medical images, as the structural information and the functional information are manifestations of two different phenomena, which results in lower contrast and less texture details in the fused image [12]. Furthermore, the manual designs of complex activity level measurements and fusion rules are required in most existing methods, which is time consuming and becomes more and more complex [13].

The application of deep learning in image analysis has received more and more attention [14], [15]. The detailed exposition of deep learning-based fusion methods will be discussed later in Sec. II-B. These works have provided new ideas for medical image fusion and achieved promising performance. However, there are still some shortages. First, the deep learning framework is generally only applied to a small part, e.g., feature extraction, while the overall fusion process is still in traditional frameworks [16]. Second, the manners to extract features in source images are the same, regardless of the fact that the source images are multi-modal data [12]. Third, existing limited GAN-based methods can only fuse part of the information from source images, causing the loss of other important information [17], [18].

In addition, as a result of the limitations of medical hardware and environments, the medical images characterizing functional information always suffer from lower resolution and more blurred details compared with corresponding medical images characterizing structural information [1], and it is difficult to improve the resolution of medical images characterizing functional information by improving the hardware facilities. To fuse the medical images characterizing functional and structural information of different resolutions, the scheme of up-sampling the images characterizing functional information or down-sampling the images characterizing structural information is bound to spectral distortion or information loss. Therefore, the fusion of multi-modal medical images of different resolutions without loss of important information is significant in practical medical applications.

To address the above challenges, in this paper, we propose a new fusion method via multi-generator multi-discriminator conditional generative adversarial network (MGMDcGAN) to fuse medical images characterizing functional and structural information of different resolutions. For convenience, we abbreviate these images as I_F and I_S, respectively. The main abbreviate words are summarized in Tab. 1. In our method, two adversarial games are established. One is established to acquire the functional information in I_F and texture details in I_S, and the other is for enhancing the dense structure information from I_S to avoid the loss of it, where the inputs are the generated images from the first cGAN and I_S. In addition, for preventing the functional information from being weakened in the final fused image when enhancing the dense structure, the inputs of the discriminator in the second cGAN are both multiplied with a mask obtained from I_S. The generated image of the second cGAN is the final fused image. The principles of two adversarial games are similar: the generators aim to generate a real-like image based on a specifically designed content loss to fool their corresponding discriminators, while the discriminators aim to distinguish the differences between the generated image and source images. Since source images are used as real data, ground-truth images are not required. Moreover, MGMDcGAN is an end-to-end model, with no need of manually designing activity level measurements and fusion rules. In addition, trainable de-convolution layers and content constraints on down-sampled fused images are more suitable for different resolution fusion. Specifically, we perform our methods on fusing the MRI and the PET images (MRI-PET), the MRI and the SPECT images (MRI-SPECT), the CT and the SPECT images (CT-SPECT). In each problem, the former image is of high resolution and the latter one is of low resolution. Both visual effect and quantitative metric results verify the superiority of our method on these three fusion problems.

The main contributions of this paper are summarized in the following three aspects. i) We propose a new end-to-end multi-modal medical image fusion method through the adversarial process between multiple generators and multiple discriminators. ii) The mask is applied to the proposed MGMDcGAN to prevent the functional information from being weakened in the final fused image when enhancing the dense structure information. iii) The proposed MGMDcGAN can be adopted as a unified method for the fusion of MRI-PET, MRI-SPECT and CT-SPECT, which are all of different resolutions.

The remainder of this paper is organized as follows: In Sec. II, we introduce some related work, including an overview of existing traditional and deep learning-based fusion methods, and a theoretical introduction of conditional generative adversarial network. The detailed introduction of our MGMDcGAN is provided in Sec. III. Sec. IV shows the fusion performance of our method on multi-modal medical images of different resolutions including MRI-PET, MRI-SPECT and CT-SPECT, compared with the state-of-the-art in terms of both qualitative visual effect and quantitative metrics. We also conduct the ablation experiments of the second cGAN and mask in this section. Conclusion is given in Sec. V.

II. RELATED WORK

In this section, we give a brief introduction of the existing traditional and deep learning-based fusion methods. Furthermore, since our fusion method is based on conditional generative adversarial network, we also show a basic explanation of conditional generative adversarial network.

A. TRADITIONAL MULTI-MODAL MEDICAL IMAGE FUSION

Due to the great significance of multi-modal medical image fusion and its wide application, many effective medical image fusion methods are constantly proposed. They can be divided into six categories according to corresponding schemes including wavelet transformation based methods, pyramid methods, sparse representation methods, neural network based methods, salient feature methods and other methods. Next, we will briefly present the main ideas of these methods and the relationship among them.

Wavelet transformation based fusion methods for multi-modal medical image fusion [7], [19], [20] can be regarded as multi-scale geometric analysis tools. In general, the fusion methods of medical image based on wavelet transformation comprise three steps: First, the source images are decomposed into low and high frequency components respectively. Then, the different frequency components are fused with different image fusion rules. Finally, the fused image is acquired using inverse transformation. The pyramid methods [8], [21] achieve medical image fusion by the diverse resolutions in the level and the iteration of the images. Wavelet transformation based methods and pyramid methods are the most active field in image fusion. The sparse representation methods [9] are used for the fusion of medical images assuming that the high frequency and low frequency images share the same set of sparse coefficients. The methods can reduce visual artifacts and improve robustness to misregistration by dividing source images into several overlapping patches with a sliding window compared with wavelet transformation based methods and pyramid methods. Neural network-based methods [10] are inspired from the perception behavior of the human brain. An important advantage of the methods is that it can predict, analyze and infer information from given data without going through a rigorous mathematical solution. Therefore, compared with other methods, neural network based methods have the advantages of good adaptability, fault tolerance, and anti-noise capacity. In terms of salient feature methods [11], [22], retained saliency features, shift-invariance and low computational complexity are the most significant advantages of the methods. The edge-preserving filters based on salient feature has a wide range of applications in medical image fusion, such as medical image fusion with guided filter and medical image fusion with multi-scale directional bilateral filter. Firstly, the source images are decomposed into multi-scale representation using edge-preserving filters. Secondly, the base and detailed layer of each source image at different scales is fused with different image fusion rules. Finally, the fused base layer and fused detailed layer are added to reconstruct the fused image. In addition, the biggest advantage of the methods is that they can retain the integrity of salient object regions and improve the visual quality of the fused images. In addition to the above medical image fusion methods, other image fusion methods also provide new ideas and possibilities for medical image fusion, which are based on color space [23], knowledge [24], fuzzy theory [25], total variation [26], *etc*.

B. DEEP LEARNING-BASED MEDICAL IMAGE FUSION

In the past few years, the emergence of deep learning has provided new ideas for multi-modal medical image fusion. The existing multi-modal medical image fusion methods based on deep learning mainly rely on CNN models. Liu *et al.* [27] introduced the CNN for multi-modal medical image fusion, in which the CNN intends to generate a weight map integrating the pixel activity information from two source images, and the fusion process is implemented in a multi-scale manner via image pyramids. Innovatively, Rajalingam and Priya *et al.* [28] adopted a siamese convolutional network to create a weight map which integrates the pixel movement information from two or more multi-modality medical images. Xia *et al.* [29] proposed a novel fusion scheme for multi-modal medical images, which utilizes both the features of the multi-scale transformation and deep convolutional neural network. As for the fusion methods based on GAN, Xu $et al. and Ma et al$. [12], [30] proposed DDcGAN, which employs a generator with two discriminators to acquire the functional information in I_F and texture details in I_S. Besides, Yang *et al.* [31] are committed to applying wasserstein GAN to the fusion of medical images.

Although the above mentioned medical image fusion methods based on deep learning have achieved promising performance, there are still some shortages: (1) The existing methods usually combine deep learning with a traditional framework, and not fully apply the deep learning framework to the entire fusion process. (2) The manners to extract features of the two different types of source images are consistent, regardless of the fact that the multi-modal medical images are manifestations of different phenomena, which is inappropriate for multi-modal medical image fusion. (3) Only part of the information from source images is extracted to participate in fusion, which causes the loss of other important information, *e.g*. the dense structure information from I_S.

C. CONDITIONAL GENERATIVE ADVERSARIAL NETWORK

Generative adversarial network [32] is a generative model proposed by Goodfellow, which contains two adversarial models: the generative model (G) is used to capture the probability distribution and generate new samples, while the discriminator model (D) is used to estimate the probability that a sample is from real data rather than generated sample. The generative model (G) learns the generative distribution P_g on the real data set x by constructing a mapping function $G(z;\theta_g)$ from the prior distribution $P_z(z)$ to the data space. The input of the discriminator model (D) is a real image or a generated image, and a scalar is obtained from $D(x;\theta_d)$ which indicates the probability that the input sample is from the training sample. The optimization of the generative model (G) and the discriminator model (D) can be attributed to a min-max two-player game. The optimization objective function of GAN is expressed as follows:

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

where the generator is continuously trained to fool the discriminator while the discriminator is continuously trained to distinguish the generated data from the real data. The generator and discriminator continue to compete against each other, and finally reach a Nash equilibrium.

One of generative adversarial network's greatest strengths is that it does not require a hypothetical data distribution. It only needs to use a distribution to directly sample to approximate the real data. However, it becomes uncontrollable in the face of more complex applications. The conditional generative adversarial network (cGAN) is an extension of the GAN. Both the generator and the discriminator add an additional condition y to guide the data generation process as part of the input layer, which can be any kind of auxiliary information. The optimization objective function of cGAN is expressed as follows:

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

Similarly, the objective function of the cGAN is a min-max two-player game with conditional probability.

III. PROPOSED METHOD

In this section, we introduce our MGMDcGAN by taking the fusion of high-resolution MRI and low-resolution PET images as an example. With analysis of the characteristics of MRI and PET images, we first explain the color conversion procedure of our MGMDcGAN, and then provide our fusion formulation and the design of loss functions. At the end of this section, the design of network architecture is shown concretely.

A. COLOR CONVERSION

The whole process of color conversion is shown in Fig. 1. The MRI image is a high-resolution grayscale image that provides structural information including texture details and dense structure information. Meanwhile, the PET image is a low-resolution RGB pseudo-color image that represents the uptake of the radiotracer and provides important functional information. Therefore, the de-correlated color model is required and the selection of color model also has a great impact on the fused result. In order to fuse multi-spectral image, we separate the achromatic and chromatic information [33], and the IHS model is widely used to achieve it. After transforming PET image from RGB to IHS space, the fusion process is performed between MRI image and I channel of PET image. However, the color information is seriously distorted [34]. YUV model can effectively solve the above problem, which is adopted in our work. Y is the luminance, which can represent structural details and the brightness variation. We just devote to fusing the Y channel value. U and V are the chrominance or chroma reflecting color and saturation, which should not be changed.

To fuse the MRI and PET images of different resolutions, the resolution of the MRI image is first uniformly set to be 4 × 4 of that of the PET image. Prior to the formal participation in MGMDcGAN fusion, multi-spectral PET image is firstly converted from RGB channels to YUV channels. The conversion process is presented as follows:

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \quad (3)$$

The process of fusion is completed between the high-resolution MRI image and the low-resolution Y channel of the PET image, i.e., Y_{PET}, and they are the inputs of MGMDcGAN. The output is a fused image I_f of high resolution. Because there is no information fused with U and V channels, we use bicubic interpolation as the up-sampling operation for U_{PET} and V_{PET} to retain color and saturation in the PET image. The up-sampled U_{PET} and V_{PET}, i.e., $U_{up-sampled}$ and $V_{up-sampled}$ and the output of MGMDcGAN, i.e., I_f, can be transformed to acquire the fused image in RGB channels according to the inverse conversion. The inverse conversion process is expressed as

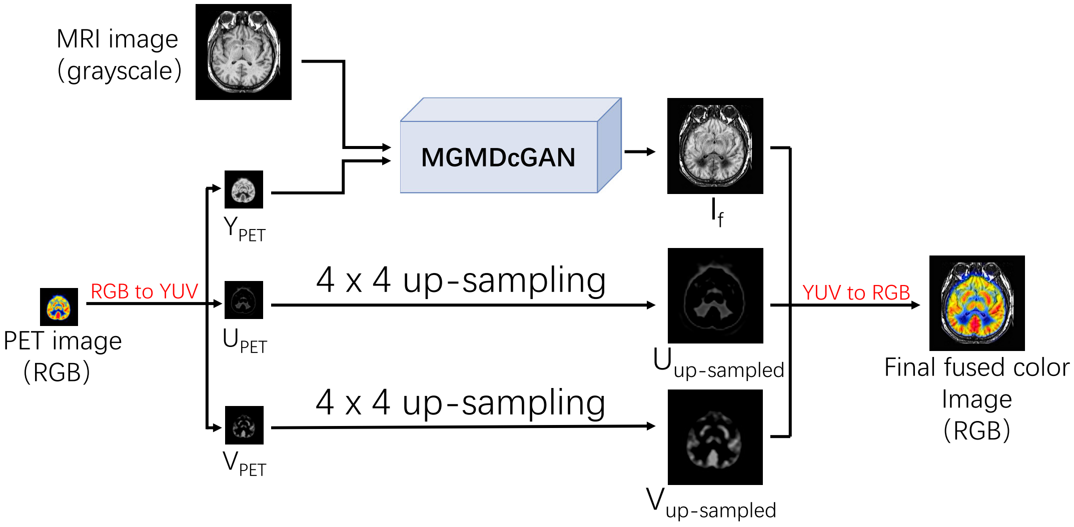


FIGURE 1. The color conversion process of fusion.

follows:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{bmatrix} 0.299 & -0.147 & 0.615 \\ 0.587 & -0.289 & -0.515 \\ 0.114 & 0.436 & -0.100 \end{bmatrix} \begin{pmatrix} I_f \\ U \\ V \end{pmatrix}. \quad (4)$$

B. PROBLEM FORMULATION

We formulate the fusion problem as a conditional GAN model by constructing a conditional generative adversarial network with multi-generator and multi-discriminator. The training procedure of our MGMDcGAN is shown in Fig. 2.

There are two cGANs in our MGMDcGAN, the first of which has two discriminators. The source MRI and Y_{PET} images are used as inputs to the first cGAN, the output of the first cGAN is the intermediate fused image $I_m = G_1(MRI, Y_{PET})$, which contains texture details of MRI image and the functional information of PET image. The intermediate fused image I_m and the source MRI image are used as the inputs of the second cGAN. The output of the second cGAN is the final fused image $I_f = G_2(MRI, I_m)$, which contains texture details and dense structure information of MRI image and functional information of PET image.

In the first cGAN, the generator G_1 is trained with the source MRI and Y_{PET} images as conditions, which is encouraged to be realistic enough to fool the discriminators. Meanwhile, the discriminators D_Y and D_M both generate a scale to estimate the probability of the input from real data rather than G_1. Respectively, D_M aims to distinguish I_m from the source MRI image, while D_Y aims to distinguish the down-sampled I_m from the low-resolution source Y_{PET} image. We employ average-pooling here to down-sample the I_m. Compared to max-pooling, the average-pooling is more appropriate to preserve the background information of the image and the functional information of PET image is mainly presented in this form. Through the adversarial process between the G_1 and two discriminators, the generated sample is continuously approached with the two real data. The optimization objective function of the first cGAN is expressed as follows:

$$\min_{G_1} \max_{D_M, D_Y} \mathbb{E}[\log D_M(MRI)] + \mathbb{E}[\log(1 - D_M(I_m))] + \mathbb{E}[\log D_Y(Y_{PET})] + \mathbb{E}[\log(1 - D_Y(\Upsilon I_m))], \quad (5)$$

where Υ denotes the down-sampling operation and is realized by average-pooling. Through the adversarial process between G_1 and two discriminators, I_m can become closer to two kinds of source images in probability distribution and contains more texture details in MRI and the functional information in Y_{PET}. It is worth noting that only one cGAN will cause the loss of the dense structure information form MRI image. An intuitive method is to enhance the dense structure information of the I_m by adding a discriminator. However, it is difficult to achieve a stable Nash equilibrium between a generator and three discriminators simultaneously, leading to poor fused results. Therefore, we introduce the second cGAN. The influence of the additional cGAN will be analyzed later in Sec. IV-E.1. The intermediate result I_m and MRI are the inputs of the second cGAN. The output of the second generator G_2, i.e., $I_f = G_2(MRI, I_m)$, is the final fused image.

In the second cGAN, the mask M is applied here in order to prevent the functional information from being weakened in I_f when enhancing the dense structure information. Rather than distinguish between whole images, the discriminator D_K only distinguishes the regions extracted by the mask. Thus, the input of D_K can be denoted as $I_f \odot M$ (the fake data) or $MRI \odot M$ (the real data), where $\odot$ denotes the dot product of matrix. The mask M is obtained by setting a threshold to extract the region with high luminance from MRI, which is the representation of the dense structure information. Thus, through the adversarial process between G_2 and D_K, the luminance in this region becomes more similar to each

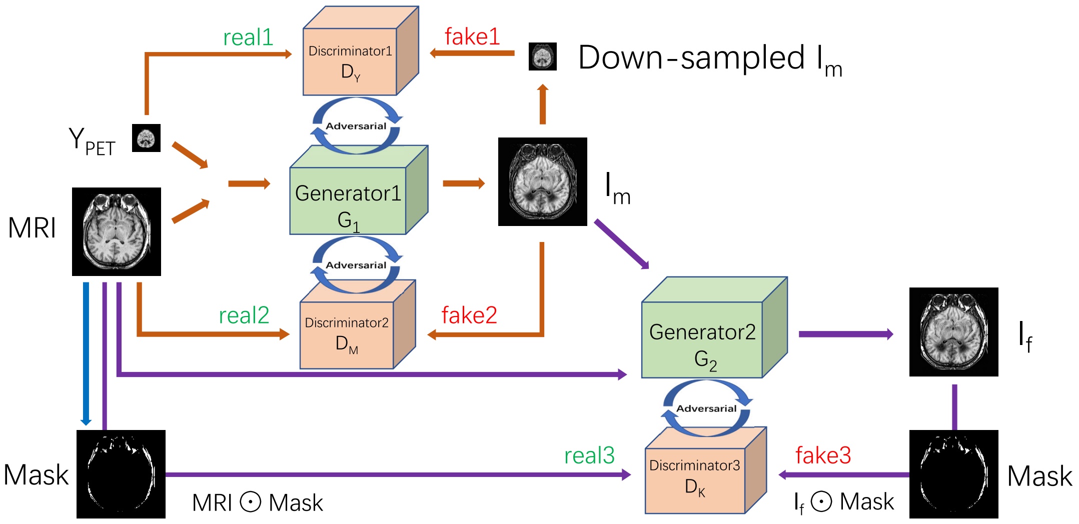


FIGURE 2. The training procedure of MGMDcGAN. The brown lines indicate the operations in the first cGAN, and the purple lines indicate the operations in the second cGAN. The blue line indicates the operation of obtaining the mask.

other and the dense structure information can be enhanced in I_f without weakening the information in other regions. The ablation experiment of the mask M is also conducted later in Sec. IV-E.2. With the mask M, the optimization objective function of the second cGAN is expressed as follows:

$$\min_{G_2}\max_{D_K} \mathbb{E}[\log D_K(MRI \odot M)]+\mathbb{E}[\log(1-D_K(I_f \odot M))]. \tag{6}$$

C. LOSS FUNCTION

GAN is known to be unstable to train and may result in unexpected results [35], especially for multi-generator multi-discriminator conditional generative adversarial network in our work. Therefore, the content loss is introduced to solve the above problem. In our work, the generators are not only trained to fool the discriminators but also satisfy the constraints of the content similarity between the generated image and the source images.

Specifically, in the first cGAN, the loss function of G_1 is composed by the loss from Y_{PET} and the loss from MRI:

$$L_{G_1} = L_{Y_{PET}} + \lambda L_{MRI}, \tag{7}$$

where $L_{Y_{PET}}$ and L_{MRI} are both composed by an adversarial loss and a content loss. More concretely, $L_{Y_{PET}}$ is defined as:

$$L_{Y_{PET}} = L^{adv}_{Y_{PET}} + \alpha L^{con}_{Y_{PET}}. \tag{8}$$

The adversarial loss $L^{adv}_{Y_{PET}}$ denotes the adversarial loss between G_1 and D_Y, which is defined as:

$$L^{adv}_{Y_{PET}} = \mathbb{E}[\log(1 - D_Y(\Upsilon I_m))]. \tag{9}$$

Since the functional information of the Y_{PET} image can be characterized by pixel intensities, we employ the l_1 norm to constrain the down-sampled fused image to have similar pixel intensities with Y_{PET} as the data fidelity term. Thus, the content loss $L^{con}_{Y_{PET}}$ can be expressed as follows:

$$L^{con}_{Y_{PET}} = \mathbb{E}[\|\Upsilon I_m - Y_{PET}\|_1]. \tag{10}$$

The second term L_{MRI} in Eq. (7) reflects the loss from MRI, which is defined as follows:

$$L_{MRI} = L^{adv}_{MRI} + \beta L^{con}_{MRI}, \tag{11}$$

where L^{adv}_{MRI} denotes the adversarial loss between G_1 and D_M, which is defined as follows:

$$L^{adv}_{MRI} = \mathbb{E}[\log(1 - D_M(I_m))]. \tag{12}$$

As for the content loss L^{con}_{MRI}, since the texture details of MRI are mainly characterized by gradient variation, we constrain the fused image to have similar texture details with MRI, and L^{con}_{MRI} can be expressed as:

$$L^{con}_{MRI} = \mathbb{E}[\|\nabla I_m - \nabla MRI\|], \tag{13}$$

where ∇ denotes Laplacian operator.

Discriminators D_M and D_Y in the first cGAN are used to discriminate between source images and I_m, respectively. The distribution of generated samples gets more and more close to that of the real data by minimizing the JS divergence, which is reflected in the loss function of D_M and D_Y:

$$L_{D_M} = \mathbb{E}[-\log(D_M(MRI))]+\mathbb{E}[-\log(1 - D_M(I_m))], \tag{14}$$

$$L_{D_Y} = \mathbb{E}[-\log(D_Y(Y_{PET}))]+\mathbb{E}[-\log(1-D_Y(\Upsilon I_m))]. \tag{15}$$

The role of the second cGAN is to enhance the dense structure information from MRI based on I_m. The loss function of G_2 is composed by the loss from I_m and that from MRI:

$$L_{G_2} = L_{I_m} + \kappa L_{MRI}, \tag{16}$$

where L_{I_m} merely contains a content loss and is defined as:

$$L_{I_m} = \mathbb{E}[\|\nabla I_f - \nabla I_m\| + \delta \|I_f - I_m\|_1]. \tag{17}$$

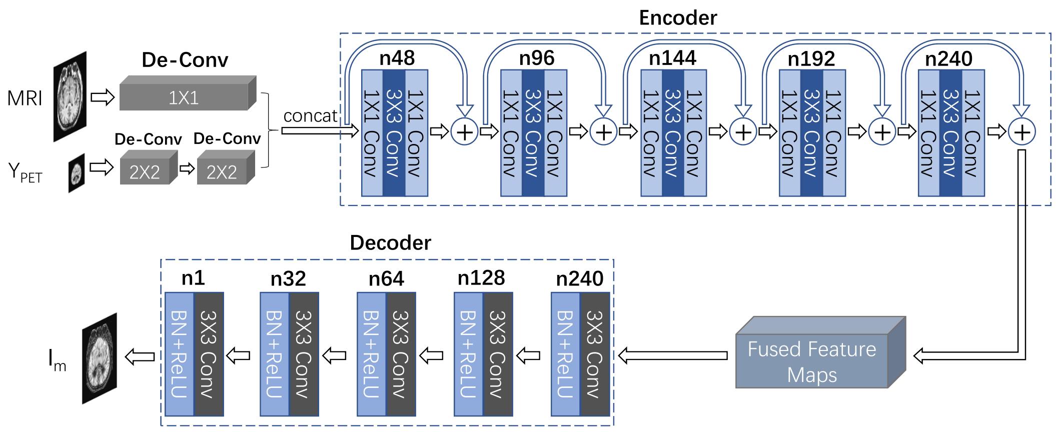


FIGURE 3. The network architecture of G_1.

L_{MRI} in Eq. (16) denotes the loss from MRI:

$$L_{MRI} = L^{adv}_{MRI} + \omega L^{con}_{MRI}, \tag{18}$$

where L^{adv}_{MRI} denotes the adversarial loss between G_2 and D_K, which is defined as follows:

$$L^{adv}_{MRI} = \mathbb{E}[\log(1 - D_K(I_f \odot M))], \tag{19}$$

where M is the mask obtained from MRI. The content loss L^{con}_{MRI} from MRI is no longer characterized by the gradient variation as in the first GAN, but the pixel intensity of the dense structure. L^{con}_{MRI} is represented as follows:

$$L^{con}_{MRI} = \mathbb{E}[\left\| I_f \odot M - MRI \odot M \right\|_1]. \tag{20}$$

Similarly, D_K in the second cGAN is used to discriminate between $MRI \odot M$ and $I_f \odot M$. And the loss function of D_K is defined as follows:

$$L_{D_K} = \mathbb{E}[-\log(D_K(MRI \odot M))] + \mathbb{E}[-\log(1 - D_K(I_f \odot M)]. \tag{21}$$

In the above formulas, λ, α, β, κ, δ and ω are all used to control the trade-off, which are set as 0.6, 1, 1, 1, 1.2 and 10, respectively.

D. NETWORK ARCHITECTURE

1) GENERATOR G_1

The network architecture of G_1 is shown in Fig. 3, including three parts: three de-convolution layers, one encoder, and the corresponding decoder. Since Y_{PET} suffers a lower resolution, a mapping is adopted to increase its resolution. Meanwhile, in order to avoid spectral distortion or information loss, we realize the mapping by de-convolution layers [36] to obtain high-resolution feature maps. The mapping is different from traditional up-sampling and its parameters are automatically obtained by training. Besides, to improve the utilization ratio of information, we employ two de-convolution layers to increase the resolution by 2 times in each layer, rather than directly increase the resolution by 4 times by using one de-convolution layer. Meanwhile, the de-convolution processing for MRI is also performed to obtain a feature map with the same resolution. Feature maps obtained from de-convolution layers are then concatenated as the input to the encoder. The encoder plays the role of feature extraction and fusion, and generates the fused feature maps. Finally, the fused feature maps are reconstructed in the decoder to acquire the high-resolution intermediate result I_m.

The encoder consists of five bottlenecks [37], and the stride of each convolutional layer is set as 1. The decoder is composed of five CNN layers. Batch normalization is used to alleviate gradient exploding/vanishing and accelerate training.

2) GENERATOR G_2

The inputs of generator G_2 are the intermediate fused image I_m and MRI of the same resolution. Also, equivalent de-convolution processing are performed for both the I_m and MRI images to acquire two feature maps with the same resolution. The difference with G_1 is that the de-convolution layers are replaced by convolution layers. It is worth noting that it is undesirable to introduce the additional operation of computing the mask in the testing phase. So we directly feed MRI into G_2 rather than $MRI \odot M$.

3) DISCRIMINATOR D_M, D_Y, D_K

The discriminators are responsible for forming adversarial relationships with corresponding generators in our network. In the first cGAN, either strength or weakness of one discriminator will finally lead to the inefficiency of the other as the training proceeds. Therefore, not only the balance between the discriminators and the generators, but also the balance between D_M and D_Y should be taken into account. We achieve the balance by designing network architectures and training strategy (as discussed in Sec. IV-A.2). The discriminators D_M, D_Y and D_K share the same network architecture, which is shown in Fig. 4. We set the strides of all convolutional layers as 2. In the last layer, the tanh activation function is employed to generate a scalar that estimates the

Algorithm 1 Training Details of MGMDcGAN

```
Parameter descriptions
$N_{G_1}, N_{G_2}, N_M, N_Y, N_K$: The Numbers of Steps to Train $G_1, G_2, D_M, D_Y, D_K$.
$\mathcal{L}_{max}, \mathcal{L}_{min}$ and $\mathcal{L}_{G_1\ max}, \mathcal{L}_{G_2\ max}$ Are Applied to Determining a Range to Uncollapse Training.
$\mathcal{L}_{max}$ and $\mathcal{L}_{min}$ Are for Adversarial Losses of $G_1, G_2, D_M, D_Y$, and $D_K$.
$\mathcal{L}_{G_1\ max}, \mathcal{L}_{G_2\ max}$: The Total Loss of $G_1, G_2$.
We Set $L_{max} = 1.8$, $L_{min} = -1.8$ in the First Batch Empirically in Our Experiments
Initialize $\theta_{D_M}, \theta_{D_Y}$ and $\theta_{D_K}$ for $D_M, D_Y$ and $D_K$; $\theta_{G_1}$ for $G_1$ and $\theta_{G_2}$ for $G_2$;
In each training iteration:
  1) Train Discriminators $D_M, D_Y$ and $D_K$:
      • Sample $n$ $MRI$ patches $\{M^1, \cdots, M^n\}$ and $n$ corresponding $Y_{PET}$ patches $\{Y^1, \cdots, Y^n\}$;
      • Acquire generated data $\{I_m^1, \cdots, I_m^n\}$, $\{I_f^1, \cdots, I_f^n\}$
      • Update Discriminator parameters $\theta_{D_M}$ by RMSPropOptimizer to minimize $\mathcal{L}_{D_M}$ in Eq. (14); (step I)
      • Update Discriminator parameters $\theta_{D_Y}$ by RMSPropOptimizer to minimize $\mathcal{L}_{D_Y}$ in Eq. (15); (step II)
      • Update Discriminator parameters $\theta_{D_K}$ by RMSPropOptimizer to minimize $\mathcal{L}_{D_K}$ in Eq. (21); (step III)
      • While $\mathcal{L}_{D_M} > \mathcal{L}_{max}$ and $N_M < 20$, repeat step I. $N_M \leftarrow N_M + 1$;
      • While $\mathcal{L}_{D_Y} > \mathcal{L}_{max}$ and $N_Y < 30$, repeat step II. $N_Y \leftarrow N_Y + 1$;
      • While $\mathcal{L}_{D_K} > \mathcal{L}_{max}$ and $N_K < 30$, repeat step III. $N_K \leftarrow N_K + 1$;
  2) Train Generators $G_1, G_2$:
      • Sample $n$ $MRI$ patches $\{M^1, \cdots, M^n\}$ and $n$ corresponding $Y_{PET}$ patches $\{Y^1, \cdots, Y^n\}$;
      • Acquire generated data $\{I_m^1, \cdots, I_m^n\}$, $\{I_f^1, \cdots, I_f^n\}$
      • Update parameters $\theta_{G_1}$ by RMSPropOptimizer to minimize $\mathcal{L}_{G_1}$ in Eq. (7); (step IV)
      • Update parameters $\theta_{G_2}$ by RMSPropOptimizer to minimize $\mathcal{L}_{G_2}$ in Eq. (16); (step V)
      • While $(\mathcal{L}_{D_M} < \mathcal{L}_{min}$ or $\mathcal{L}_{D_Y} < \mathcal{L}_{min})$ and $N_{G_1} < 20$, repeat step IV. $N_{G_1} \leftarrow N_{G_1} + 1$;
      • While $\mathcal{L}_{D_K} < \mathcal{L}_{min}$ and $N_{G_2} < 20$, repeat step V. $N_{G_2} \leftarrow N_{G_2} + 1$;
      • While $\mathcal{L}_{G_1} > \mathcal{L}_{G_1\ max}$ and $N_{G_1} < 30$, repeat step IV. $N_{G_1} \leftarrow N_{G_1} + 1$;
      • While $\mathcal{L}_{G_2} > \mathcal{L}_{G_2\ max}$ and $N_{G_2} < 30$, repeat step V. $N_{G_2} \leftarrow N_{G_2} + 1$;
```

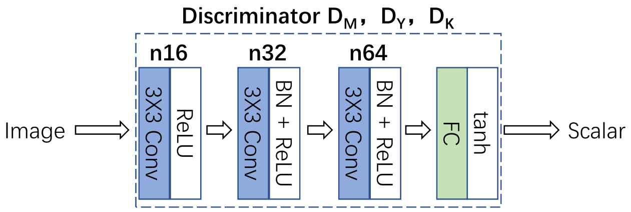


FIGURE 4. The network architecture of discriminator.

probability of the input image from source images rather than generator.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, to verify the effectiveness of our proposed MGMDcGAN, it is firstly compared with 9 state-of-the-art methods on the publicly available datasets by qualitatively for the fusion of MRI-PET, MRI-SPECT and CT-SPECT. Furthermore, 6 metrics are employed to evaluate the fusion results by qualitative comparisons. We also conduct the ablation experiments of the second cGAN and mask.

A. EXPERIMENTAL SETTINGS

1) DATASET

The application of our MGMDcGAN to MRI-PET, MRI-SPECT, and CT-SPECT fusion are all validated on the publicly available Harvard dataset.[1] 83 MRI-PET image pairs, and 19 CT-SPECT image pairs are downloaded to create the training dataset. Y channels of color images are extracted to form 83 MRI-Y_{PET} pairs and 19 CT-Y_{SPECT} pairs. Then, they are cropped into patch pairs of size 84 × 84. In addition, to verify the effectiveness of our MGMDcGAN in the medical image fusion of different resolutions, Y_{PET} and Y_{SPECT} are down-sampled to the size 21 × 21. Finally, 9984 patch pairs of 84 × 84 MRI and 21 × 21 Y_{PET}, and 2176 patch pairs of 84 × 84 CT and 21 × 21 Y_{SPECT} are used as the training set. The image pairs have been aligned in advance, and image registration [38]–[40] is required for unaligned data.

2) TRAINING DETAILS

Alg. 1 summarizes the detailed training process. 9984 patch pairs of 84 × 84 MRI and 21 × 21 Y_{PET} or 2176 patch pairs of 84 × 84 CT and 21 × 21 Y_{SPECT} are employed as the inputs of the training process in the corresponding fusion problem. First of all, The parameters in the generators and discriminators are initialized. In order to form stable adversarial relationships between the generators and the discriminators, and ensure the balance between the D_M and D_Y in order that each one of them is not too weak, the generators and the

[1] http://www.med.harvard.edu/AANLIB/home.html.

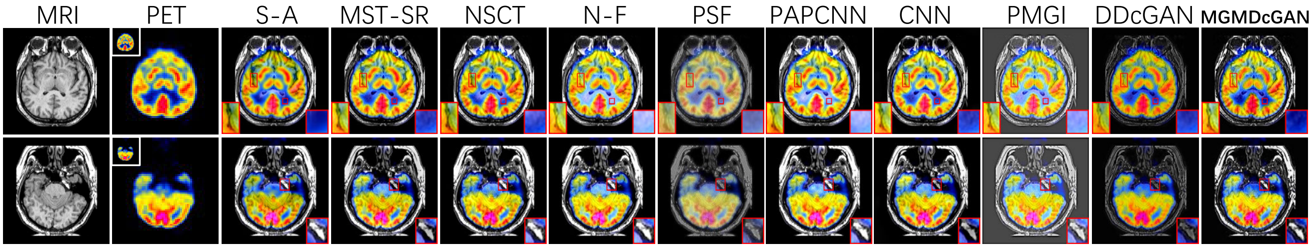


FIGURE 5. Qualitative comparison of our MGMDcGAN with 9 state-of-the-art methods on two typical MRI and PET image pairs of different transaxial sections of the brain-hemispheric. From left to right: high-resolution MRI image, low-resolution PET image, fused images of S-A, MST-SR, NSCT, N-F, PSF, PAPCNN, CNN, PMGI, DDcGAN and our MGMDcGAN.

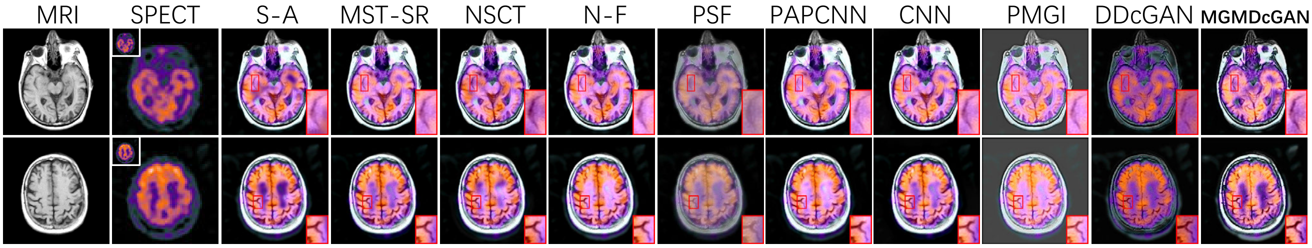


FIGURE 6. Qualitative comparison of our MGMDcGAN with 9 state-of-the-art methods on two typical MRI and SPECT image pairs of different transaxial sections of the brain-hemispheric.

discriminators are not trained once per batch in turn. Actually, in each iteration, the generator is trained more times if it fails to fool the corresponding discriminator and the discriminator is also trained more times if it fails to discriminate the data from the corresponding generator. The parameters in the generators and discriminators are both updated by RMSPropOptimizer. When testing, the final fused image I_f can be generated by the trained generators without discriminators. The fully connected layers are not employed in our generators so that the input source images can be of any size as long as they meet the predefined resolution proportion.

B. COMPARISON ALGORITHMS AND EVALUATION METRICS

1) COMPARISON ALGORITHMS

To give some intuitive results on the fusion performance, we compare our MGMDcGAN with 9 state-of-the-art fusion methods, including S-A [41], MST-SR [42], NSCT [43], N-F [44], PSF [45], PAPCNN [46], CNN [27], PMGI [47] and DDcGAN [12]. Among them, DDcGAN can be directly applied to the multi-modal medical images fusion of different resolutions, while the preprocessing of up-sampling low-resolution source image is necessary for others whose source images share the same resolution. The CNN, PMGI and DDcGAN are the methods based on deep learning, while others are traditional methods. The parameters of these 9 methods are the same as the initial papers.

2) EVALUATION METRICS

In order to have a more accurate evaluation of the experimental results, 6 metrics are used to evaluate the fusion performance of the 10 fusion methods, including entropy (EN) [48], spatial frequency (SF), edge intensity(EI), mean gradient (MG), peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). The EN can measure the amount information contained in the fused image, SF is a metric that reflects the texture details of a image by calculating the gradient distribution, EI reflects the gradient amplitude of edge point, MG is a metric that measures the amount gradient information contained in the fused image, PSNR measures the distortion by the ratio of peak value power and noise power, and the SSIM measures the structure similarity between fused image and source images. The larger the values of the 6 metrics are, the better fusion performance a method achieves. In addition, neither CT nor SPECT mainly characterizes texture details. Therefore, SF and MG do not participate in the quantitative comparison of CT-SPECT fusion.

C. QUALITATIVE COMPARISONS

Figs. 5-7 show typical and intuitive results of 10 methods on the fusion of MRI-PET. MRI-SPECT and CT-SPECT respectively, which are different transaxial sections of the brain-hemispheric.

Compared with the existing 9 fusion methods, there are 4 obvious advantages. First, our results can characterize the functional information of the PET and SPECT images clearly. Second, the abundant texture details can be preserved from the MRI images in our results. Third, our results can clearly reflect the dense structure e.g., bones and implants. Fourth, The functional information in our results is closer to the source PET and SPECT images due to that it does

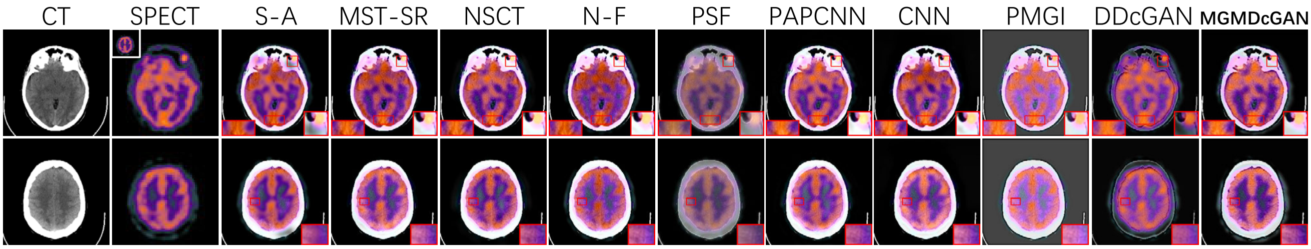


FIGURE 7. Qualitative comparison of our MGMDcGAN with 9 state-of-the-art methods on two typical CT and SPECT image pairs of different transaxial sections of the brain-hemispheric.

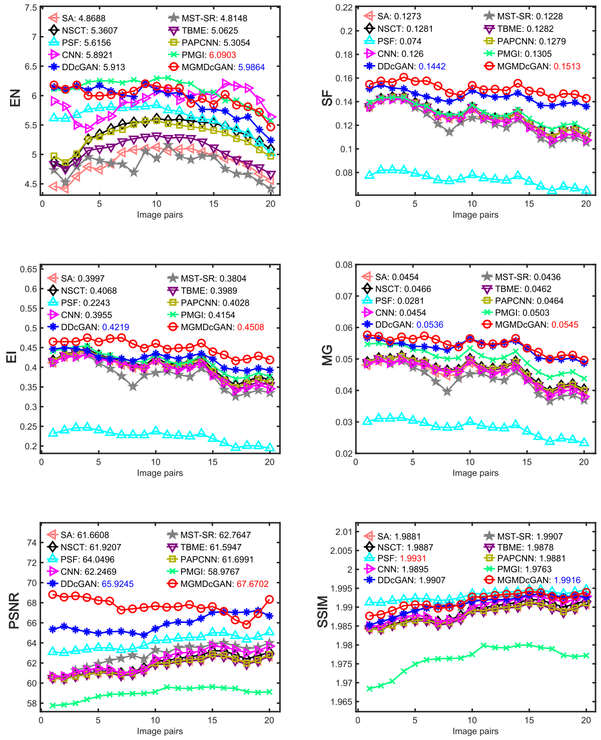


FIGURE 8. Quantitative comparison of our MGMDcGAN for MRI-PET image fusion with 7 state-of-the-art methods. Means of metrics for different methods are shown in the legends. Optimal values are shown in red and suboptimal values in blue.

not suffer from spectral distortion caused by up-sampling the low-resolution PET and SPECT images.

The fusion of MRI-PET can be seen from Fig. 5. S-A, NSCT, N-F, PSF, PAPCNN and PMGI cannot characterize the functional information well, while PSF, CNN, PMGI and DDcGAN cannot obtain abundant texture details. Furthermore, MST-SR, PSF, CNN, PMGI and DDcGAN cannot clearly reflect the dense structure. It is worth noting, compared with DDcGAN, due to the employment of the second cGAN with mask, our MGMDcGAN obtains additional dense structure information, e.g., the skull. In general, the structural and functional information of our MGMDcGAN are the closest to those of source images.

The fusion of MRI-SPECT can be seen from Fig. 6. MRI-SPECT fusion is tested using the trained models of MRI-PET. In contrast, our MGMDcGAN obtains the most functional information without spectral distortion, and presents the clearest texture details. Also the results of MGMDcGAN does not suffer the loss of dense structure information as DDcGAN.

The fusion of CT-SPECT can be seen from Fig. 7. By comparison, MST-SR, NSCT, N-F, PAPCNN, CNN, PMGI and DDcGAN obviously reduce the intensity of color in the SPECT image, leading to the loss of functional information, and the results generated by S-A suffer from spectral distortion. The color of our results is the most similar to that of the source PET images. In terms of the bone information retained from CT images, MST-SR, NSCT, N-F, PAPCNN, CNN, PMGI and MGMDcGAN obtain it well. However, the results of S-A suffer from partial shadow, and DDcGAN almost loses it.

D. QUANTITATIVE COMPARISONS

We randomly selected 20 test pairs of MRI-PET images, 20 test pairs of MRI-SPECT images, and 19 test pairs of CT-SPECT images to further report quantitative comparisons of our MGMDcGAN and the competitors. All test image pairs are of different transaxial sections of the brain-hemispheric. The results of quantitative comparisons on MRI-SPECT, MRI-SPECT and CT-SPECT image fusion are summarized in Figs. 8-10, respectively. In terms of the fusion of MRI-PET, our MGMDcGAN can achieve the optimal values in the SF, EI, MG and PSNR, and the suboptimal values in the EN and SSIM. In the fusion of MRI-SPECT, our MGMDcGAN can achieve the best values in the SF, EI, MG and PSNR, and the metric EN and SSIM also show comparable results, generating the second largest mean values and those mean values merely follow behind those of DDcGAN by 0.0566 and PSF by 0.0019, respectively. As for the fusion of CT-SPECT, the metrics EN and PSNR achieve the best values while EI and SSIM reach the second largest mean value whose mean value merely follows behind those of DDcGAN by 0.0181 and PSF by 0.003. Thus, it can be demonstrated that our MGMDcGAN can preserve the structural and functional information to a great extent at the same time and achieve the best fusion performance.

The mean and standard deviation of runtime in different methods on these 3 kinds of medical image fusion is also

TABLE 2. The mean and standard deviation of running time in different methods. (unit: second).

	S-A	MST-SR	NSCT	N-F	PSF	PAPCNN	CNN	PMGI	DDcGAN	MGMDcGAN
MRI-PET	0.03 ± 0.01	0.03 ± 0.01	3.78 ± 0.03	6.32 ± 0.13	27.65 ± 8.67	3.92 ± 0.19	11.56 ± 0.28	0.03 ± 0.02	0.28 ± 0.22	0.91 ± 0.20
MRI-SPECT	0.03 ± 0.01	0.03 ± 0.01	3.80 ± 0.23	6.33 ± 0.22	28.40 ± 5.66	4.00 ± 0.21	11.06 ± 0.18	0.03 ± 0.03	0.27 ± 0.21	$0.97 + 0.38$
CT-SPECT	0.03 ± 0.01	0.03 ± 0.01	3.81 ± 0.12	6.34 ± 0.16	25.00 ± 5.41	3.88 ± 0.11	11.36 ± 0.22	0.03 ± 0.03	0.28 ± 0.22	0.87 ± 0.24

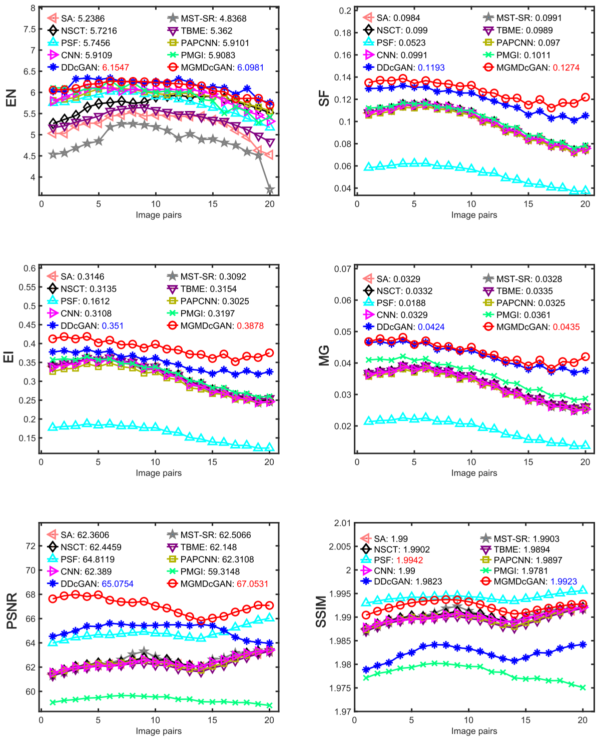


FIGURE 9. Quantitative comparison of our MGMDcGAN for MRI-SPECT image fusion with 7 state-of-the-art methods. Means of metrics for different methods are shown in the legends.

provided in Tab. 2. Comparatively, our MGMDcGAN can still achieve comparable efficiency.

E. ABLATION EXPERIMENTS

1) THE SECOND cGAN

We employed the second cGAN to enhance the information of dense structure from MRI. In order to show the effect of the second cGAN, the following comparative experiments are performed. (a) Only the first cGAN is employed. (b) The second cGAN is employed. The experimental settings of two comparative experiments are the same and the results are shown in the left of Fig. 11. The functional information in PET and texture details in MRI are both preserved in the fused results of (a) and (b). However, the fused results in (a) almost loss the information of dense structure from MRI. By comparison, the fused results of (b) can effectively preserve the information of dense structure from MRI. As a result, this proves that the second cGAN can enhance the information of dense structure from MRI.

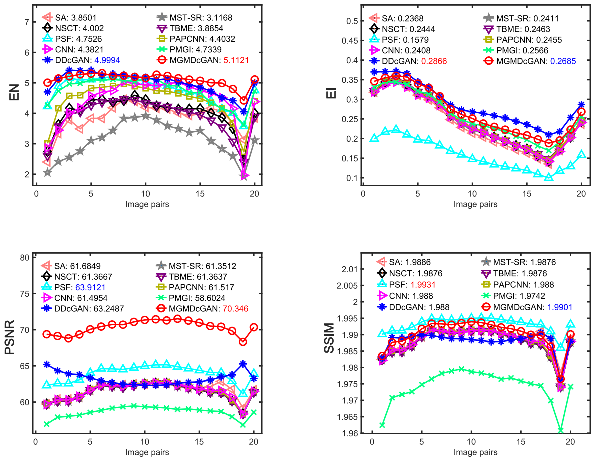


FIGURE 10. Quantitative comparison of our MGMDcGAN for CT-SPECT image fusion with 7 state-of-the-art methods. Means of metrics for different methods are shown in the legends.

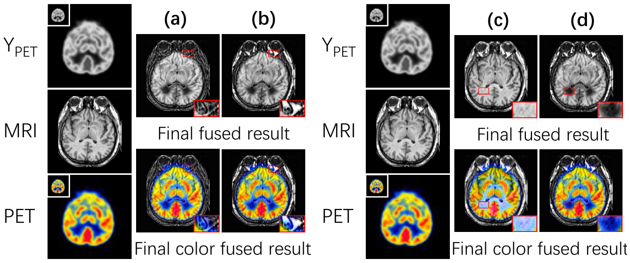


FIGURE 11. Results on whether the second cGAN exists (left) and whether the mask exists (right).

2) MASK

The mask is employed in our MGMDcGAN to prevent the functional information from being weakened in the I_f when enhancing the dense structure information. In order to show the effect of the mask, we perform the following comparative experiments. (c) The mask is not employed, and the input of the D_k is the source MRI image and final fused image I_f. (d) The final fused images I_f are generated by the method proposed in this paper with the mask employed. The experimental settings of two comparative experiments are the same and the comparative fused results are shown in the right of Fig. 11. The fused results of (c) and (d) both preserve the information of dense structure from MRI image. However, in method (c), the final fused image is closer to the MRI image with almost no functional information and the functional information in the final color fused image is seriously weakened. By contrast, method (d) can address above problem, and the fused result can simultaneously preserve the structural information including texture details and dense structure information in MRI image and functional information in PET image. It can be seen that the mask plays an important role in the fusion process.

V. CONCLUSION

In this paper, we proposed a new deep learning-based fusion method for multi-modal medical images of different resolutions, termed as MGMDcGAN. It can simultaneously keep the functional information in I_F and structural information including texture details and dense structure information in I_S without spectral distortion or information loss. Since our method is an end-to-end model, the complex activity level measurements and fusion rules designed in a manual way in traditional fusion strategies are not required. In the first cGAN, the generator generates a real-like fused image to fool two discriminators, while the discriminators aim to distinguish the structure differences between the fused image and source images. The second cGAN with mask is used to enhance the information of dense structure in I_f, while preventing the functional information from being weakened. The adequate experimental results on the fusion of MRI-PET, MRI-SPECT and CT-SPECT indicate that our MGMDcGAN not only presents better visual effects, but also preserves the maximum or approximate maximum amount of information in source images.

REFERENCES

[1] 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.

[2] J. Du, W. Li, and B. Xiao, "Anatomical-functional image fusion by information of interest in local Laplacian filtering domain," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5855–5866, Dec. 2017.

[3] W. Zhao, H. Lu, and D. Wang, "Multisensor image fusion and enhancement in spectral total variation domain," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 866–879, Apr. 2018.

[4] J. Zhang, Z. Zhou, J. Teng, T. Li, and Z. Miao, "Fusion algorithm of functional images and anatomical images based on wavelet transform," in *Proc. 2nd Int. Conf. Biomed. Eng. Informat.*, 2009, pp. 1–5.

[5] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.

[6] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: A unified densely connected network for image fusion," in *Proc. AAAI Conf. Artif. Intell.*, 2020.

[7] F. Zhao, G. Xu, and W. Zhao, "CT and MR image fusion based on adaptive structure decomposition," *IEEE Access*, vol. 7, pp. 44002–44009, 2019.

[8] X. Li, X. Zhang, and M. Ding, "A sum-modified-Laplacian and sparse representation based multimodal medical image fusion in Laplacian pyramid domain," *Med. Biol. Eng. Comput.*, vol. 57, no. 10, pp. 2265–2275, Oct. 2019.

[9] W. Jiang, X. Yang, W. Wu, K. Liu, A. Ahmad, A. K. Sangaiah, and G. Jeon, "Medical images fusion by using weighted least squares filter and sparse representation," *Comput. Electr. Eng.*, vol. 67, pp. 252–266, Apr. 2018.

[10] J. Teng, S. Wang, J. Zhang, and X. Wang, "Neuro-fuzzy logic based fusion algorithm of medical images," in *Proc. 3rd Int. Congr. Image Signal Process.*, Oct. 2010, pp. 1552–1556.

[11] Z. Xu, "Medical image fusion using multi-level local extrema," *Inf. Fusion*, vol. 19, pp. 38–48, Sep. 2014.

[12] H. Xu, P. Liang, W. Yu, J. Jiang, and J. Ma, "Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3954–3960.

[13] 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.

[14] J. Ma, H. Zhang, P. Yi, and Z.-Y. Wang, "SCSCN: A separated channel-spatial convolution net with attention for single-view reconstruction," *IEEE Trans. Ind. Electron.*, early access, Nov. 6, 2019, doi: 10.1109/TIE.2019.2950866.

[15] J. Ma, X. Wang, and J. Jiang, "Image superresolution via dense discriminative network," *IEEE Trans. Ind. Electron.*, vol. 67, no. 7, pp. 5687–5695, Jul. 2020.

[16] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[17] 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.

[18] 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.

[19] G. Bhatnagar, Q. M. J. Wu, and Z. Liu, "Directive contrast based multimodal medical image fusion in NSCT domain," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1014–1024, Aug. 2013.

[20] 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. Biomed. Eng.*, vol. 61, no. 1, pp. 197–206, Jan. 2014.

[21] J. Chen, X. Li, L. Luo, X. Mei, and J. Ma, "Infrared and visible image fusion based on target-enhanced multiscale transform decomposition," *Inf. Sci.*, vol. 508, pp. 64–78, Jan. 2020.

[22] J. Hu and S. Li, "The multiscale directional bilateral filter and its application to multisensor image fusion," *Inf. Fusion*, vol. 13, no. 3, pp. 196–206, Jul. 2012.

[23] C. He, Q. Liu, H. Li, and H. Wang, "Multimodal medical image fusion based on IHS and PCA," *Procedia Eng.*, vol. 7, pp. 280–285, Jan. 2010.

[24] W. Dou, S. Ruan, Q. Liao, D. Bloyet, and J.-M. Constans, "Knowledge based fuzzy information fusion applied to classification of abnormal brain tissues from MRI," in *Proc. 7th Int. Symp. Signal Process. Appl.*, 2003, pp. 681–684.

[25] Y.-P. Wang, J.-W. Dang, Q. Li, and S. Li, "Multimodal medical image fusion using fuzzy radial basis function neural networks," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, 2007, pp. 778–782.

[26] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, Sep. 2016.

[27] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–7.

[28] B. Rajalingam and R. Priya, "Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis," *Int. J. ChemTech Res.*, vol. 11, no. 6, pp. 160–176, 2018.

[29] K.-J. Xia, H.-S. Yin, and J.-Q. Wang, "A novel improved deep convolutional neural network model for medical image fusion," *Cluster Comput.*, vol. 22, no. S1, pp. 1515–1527, Jan. 2019.

[30] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4980–4995, 2020.

[31] Z. Yang, Y. Chen, Z. Le, F. Fan, and E. Pan, "Multi-source medical image fusion based on Wasserstein generative adversarial networks," *IEEE Access*, vol. 7, pp. 175947–175958, 2019.

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

[33] P. Ganasala, V. Kumar, and A. D. Prasad, "Performance evaluation of color models in the fusion of functional and anatomical images," *J. Med. Syst.*, vol. 40, no. 5, p. 122, May 2016.

[34] G. Bhatnagar, Z. Liu, and Q. J. Wu, "Multimodal medical image fusion in NSCT domain," in *Big Data in Multimodal Medical Imaging*. Abingdon, U.K.: Taylor & Francis, 2019, p. 23.

[35] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jun. 3, 2019, doi: 10.1109/TCSVT.2019.2920407.

[36] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2018–2025.

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

[38] J. Ma, J. Jiang, C. Liu, and Y. Li, "Feature guided Gaussian mixture model with semi-supervised EM and local geometric constraint for retinal image registration," *Inf. Sci.*, vol. 417, pp. 128–142, Nov. 2017.

[39] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, May 2019.

[40] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, and Q. Z. Sheng, "Nonrigid point set registration with robust transformation learning under manifold regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3584–3597, Dec. 2019.

[41] W. Li, Y. Xie, H. Zhou, Y. Han, and K. Zhan, "Structure-aware image fusion," *Optik*, vol. 172, pp. 1–11, Nov. 2018.

[42] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, Jul. 2015.

[43] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019.

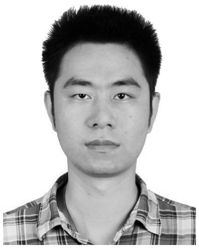
[44] S. Das and M. K. Kundu, "A neuro-fuzzy approach for medical image fusion," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3347–3353, Dec. 2013.

[45] J. Du, W. Li, and B. Xiao, "Fusion of anatomical and functional images using parallel saliency features," *Inf. Sci.*, vols. 430–431, pp. 567–576, Mar. 2018.

[46] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2019.

[47] H. Zhang, H. Xu, Y. Xiao, X. Guo, and J. Ma, "Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity," in *Proc. AAAI Conf. Artif. Intell.*, 2020.

[48] J. Van Aardt, "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.

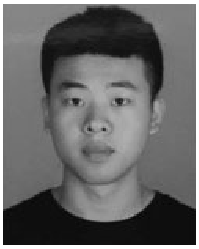


JUN HUANG received the B.S. and Ph.D. degrees from the Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. He is currently an Associate Professor with the Electronic Information School, Wuhan University, China. His main research interests include infrared image and infrared spectrum processing.

ZHULIANG LE received the B.E. degree from the School of Information Engineering, Wuhan University of Technology, Wuhan, China, in 2019. He is currently pursuing the master's degree with the Electronic Information School, Wuhan University. His research interests include computer vision, machine learning, and pattern recognition.

YONG MA received the degree from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, in 1997, and the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003. From 2004 to 2006, he was a Lecturer with the University of the West of England, Bristol, U.K. From 2006 to 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, where he was a Professor of electronics. He is currently a Professor with the Electronic Information School, Wuhan University, Wuhan. His current research interests include signal and systems, remote sensing of the lidar and infrared, infrared image processing, pattern recognition, interface circuits to sensors, and actuators.

FAN FAN received the B.S. degree in communication engineering and the Ph.D. degree in electronic circuit and system, from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2015, respectively. He holds a postdoctoral position with the School of Remote Sensing and Information Engineering, Wuhan University, China. His current research interests include infrared thermal imaging, machine learning, and computer vision.



HAO ZHANG received the B.E. degree from the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan, China, in 2019. He is currently pursuing the master's degree with the Electronic Information School, Wuhan University. His research interests include computer vision, machine learning, and pattern recognition.

LEI YANG (Member, IEEE) received the Ph.D. degree in communication and information system from Tianjin University, Tianjin, China, in 2007. She was a Visiting Researcher with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, from July 2017 to July 2018. She is currently an Associate Professor with the School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou, China. Her research interests include 3D image processing, 3D display, and computer vision.