Semisupervised Remote Sensing Image Fusion Using Multiscale Conditional Generative Adversarial Network With Siamese Structure

Xin Jin, *Member, IEEE*, Shanshan Huang, Qian Jiang, Shin-Jye Lee, Liwen Wu, and Shaowen Yao, *Member, IEEE*

***Abstract*—Remote sensing image fusion (RSIF) can generate an integrated image with high spatial and spectral resolution. The fused remote sensing image is conducive to applications including disaster monitoring, ecological environment investigation, and dynamic monitoring. However, most existing deep learning based RSIF methods require ground truths (or reference images) to train a model, and the acquisition of ground truths is a difficult problem. To address this, we propose a semisupervised RSIF method based on the multiscale conditional generative adversarial networks by combining the multiskip connection and pseudo-Siamese structure. This new method can simultaneously extract the features of panchromatic and multispectral images to fuse them without a ground truth; the adopted multiskip connection contributes to presenting image details. In addition, we propose a composite loss function, which combines the least squares loss, L1 loss, and peak signal-to-noise ratio loss to train the model; the composite loss function can help to retain the spatial details and spectral information of the source images. Moreover, we verify the proposed method by extensive experiments, and the results show that the new method can achieve outstanding performance without relying on the ground truth.**

***Index Terms*—Conditional generative adversarial network (cGAN), deep learning (DL), image fusion, loss function, remote sensing image fusion (RSIF).**

I. Introduction

IMAGE fusion aims to fuse the complementary information in two or more source images obtained by different sensors so that a new comprehensive image can be generated [1]−[3]. The fused image can present more useful information than any one of the source images. Therefore, image fusion techniques can provide a high-quality fused image to help interpret a scene or target, and related research areas include medical images [5], multifocus images [3], [6], infrared and visible images [7], [8], and remote sensing images [9], [10].

Remote sensing images with high spatial and spectral resolution play an important role in geological exploration, environmental protection, urban planning, marine monitoring, meteorological forecast, disaster relief, and other fields [1], [12]−[15]. However, because of the limitations of imaging mechanisms, panchromatic (PAN) and multispectral (MS) images are obtained by different satellite sensors. A PAN image can be regarded as a PAN-band image, and this kind of image has high spatial resolution of ground objects but lower spectral resolution. The pixels of an MS image are usually represented by red, green, and blue (RGB) values obtained from an MS sensor, and the obtained image has low spatial resolution but high spectral resolution. To understand scenes or ground targets comprehensively, researchers proposed image fusion methods to combine the complementary information of different remote sensing images, i.e., remote sensing image fusion (RSIF). RSIF is usually employed to fuse a gray-scale PAN image and color MS image to obtain an integrated image with high spatial and spectral resolution [13]−[16].

According to the fusion mechanism, the emerged RSIF methods can be regarded as belonging to two classes: conventional image fusion methods and deep learning (DL) based image fusion methods. The conventional image fusion methods can be also divided into three categories: component substitution, multiresolution analysis (MRA), and sparse representation (SR). The component substitution-based methods assume that the geometric details of an MS image exist in its structural components, and these components are obtained by converting the source image into a new space. Then, the spatial information of the PAN image is injected into the MS image by replacing or partially replacing the structural component. Finally, the fused remote sensing image is obtained by using inverse transformation based on the new structural component. However, this kind of method has many problems. For example, intensity–color–saturation transformation always distorts the spectral features of the fused image [12], and the high pass filtering and principal component

Manuscript received December 29, 2020; revised April 15, 2021; accepted June 10, 2021. Date of publication June 21, 2021; date of current version July 26, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 61863036 and 62002313, in part by China Postdoctoral Science Foundation under Grants 2020T130564 and 2019M653507, in part by Key Areas Research Program of Yunnan Province in China under Grant 202001BB050076, in part by Key Laboratory in Software Engineering of Yunnan Province in China under Grant 2020SE408, in part by Postdoctoral Science Foundation of Yunnan Province, and in part by Yunnan University's Research Innovation Fund for Graduate Students under Grants 2020230 and 2020231. *(Corresponding authors: Shanshan Huang and Qian Jiang.)*

Xin Jin, Shanshan Huang, Qian Jiang, Liwen Wu, and Shaowen Yao are with the School of Software, Yunnan University, Kunming 650091, China, and also with the Engineering Research Center of Cyberspace, Yunnan University, Kunming, 650000, China (e-mail: 18487219630@163.com; huangshanshan9633@163.com; jiangqian_1221@163.com; wulw@mail.ynu.edu.cn; yaosw@ynu.edu.cn).

Shin-Jye Lee is with the Institute of Technology Management, National Chiao Tung University, Hsinchu 30010, China (e-mail: camhero@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2021.3090958

analysis (PCA) based RSIF methods may lose the physical features of the source images [17]. Meanwhile, image fusion methods based on MRA can recover the lost spatial information of MS images by the corresponding high-frequency features of PAN images. Although MRA-based methods can preserve more spectral features than component substitution-based methods, they usually suffer from problems of spatial distortion and ringing effects, such as Laplacian pyramid transform (LPT) and wavelet transform (WT) [15], [18]. LPT can reduce the redundant information in the Gaussian pyramid but neglects the correlation of the coefficients between the decomposed layers. WT has good spatial and frequency resolution and can present the frequency information of the images, but it will partially lose some edge information of the source images because it lacks shift invariance [17], [19]. In addition, the SR-based fusion methods can obtain high-resolution MS images by using an overcomplete dictionary that is trained by an image dataset; the spectral distortion with SR-based fusion methods is also smaller. However, most SR-based methods have a complex structure and high computational complexity and usually generate smoothed results, which is not good when we want to preserve the edges of source images [1], [21].

In recent years, and with its continuous development, DL has been gradually applied to image fusion and has shown favorable results. DL methods overcome the difficulties in conventional image fusion research to a certain extent. Existing DL-based image fusion methods can be roughly divided into three categories: methods based on convolutional neural networks (CNNs), convolutional sparse representation (CSR), and generative adversarial networks (GANs). Among these methods, CNN-based methods [22], [23] show great potential in image fusion by virtue of the CNN's strong feature learning ability. The image transformation, activity level measurement, and fusion rules (or a part of them) can be jointly implemented implicitly through learning a CNN. Meanwhile, CSR-based image fusion methods [24], [25] can obtain the SR of the whole image, rather than independently computing the representations for a set of overlapping image patches as in conventional standard SR. Lastly, GANs [22], [30], [31], which have boomed in popularity since 2014, show outstanding performance in image fusion because of their powerful image generation ability [17]. Some image fusion methods based on GANs have been proposed [26]–[28] and show good performance in image fusion. However, these methods still suffer from serious problems in their training because the ground truth or reference images are lacking in RSIF. To address this problem, we propose an RSIF algorithm based on conditional generative adversarial networks (cGANs) with a pseudo-Siamese structure and multiskip connection. The proposed method is an end-to-end image fusion model, and it can simultaneously extract the features of MS and PAN images to achieve good fusion performance.

The contributions of this work are summarized as follows.

1) We propose a novel end-to-end semisupervised image fusion method that does not need the ground truth image (with high spatial and spectral resolution) and can achieve good image fusion performance.
2) We propose a new pseudo-Siamese structure with multiskip connection to extract the unique features of the PAN image and the MS image, and, thus, the fused image contains both the spatial texture information of the PAN image and the spectral information of the MS image.
3 We design a new composite loss function, including the least squares loss, peak signal-to-noise ratio (PSNR) loss, and L1 norm loss, to train the proposed model. In addition, subjective and objective image fusion evaluation indices are used to comprehensively evaluate the fused image's quality.

The rest of this article is organized as follows. Section II introduces the related work. Section III shows the details of the proposed method. Section IV presents the experimental settings, results, and analysis. Finally, Section V concludes this work.

II. Related Work

This section introduces the related knowledge on this subject, including the development of RSIF methods based on DL, the structure of cGANs, and the least squares generative adversarial network's (LSGAN) loss function.

A. Remote Sensing Image Fusion Based on Deep Learning

In remote sensing techniques, it is very important to combine the multisource remote sensing data to obtain complete and reliable information [16]. Therefore, RSIF research began [10], [16]. Popular remote sensing images include PAN images with high spatial resolution/low spectral resolution, and MS images with low spatial resolution/high spectral resolution. These images are often fused to generate a new integrated remote sensing image with high spatial and spectral features. RSIF can maximize the key information of source images and reduce the fuzziness and redundant information in the output images; thus, RSIF can suppress unimportant information to highlight the useful information [32]. As a result, the fused images obtained by RSIF can improve the reliability of remote sensing data.

Due to its powerful learning ability, DL has been extensively researched in image processing, including RSIF [19], [21], [35]. In 2018, Shao *et al.* [19] proposed an RSIF method that introduced a double-branch network structure based on the deep CNN. This double-branch network can capture the significant spectral and spatial features of MS and PAN images. In contrast, Dai *et al.* [20] proposed a spatio-temporal fusion method for low spectral high spatial resolution (LSHT) and high spectral low spatial resolution images based on DL. This method adopted a two-layer fusion strategy based on the CNN. Each layer of the network consists of two steps: first, the LSHT image is processed with the goal of super resolution; second, the processed image is fused by a linear model. In 2019, Liu *et al.* [21] proposed a two-stream fusion network to extract the features of PAN and MS images for fusion and producing a final image. In the same year, Ye *et al.* [35] proposed an image fusion algorithm based on the CNN and using a fusion model with end-to-end attributes, in which the input was a pair of source images and the output was a fused image. Compared with conventional RSIF methods, DL-based methods can extract and fuse features without following any artificial fusion rules. However, most existing DL-based RSIF methods require the ground truth to learn how to fuse the source images, which is a serious limitation because it is difficult to get a mass of reference or labeled images. To solve this problem of limited access to reference images, some researchers began to study unsupervised image fusion methods. In 2019, Tatsumi *et al.* [33] proposed a guided deep decoder network to achieve image fusion. This method allows the network parameters to be optimized in an unsupervised way and without training data.

Recently, GANs have received extensive attention due to their powerful image generation ability, and many GANs with different structures have been proposed [22], [30], [31]. GANs are also promising RSIF models. In 2018, PSGAN proposed by Liu *et al.* [36]. The authors first designed a two-stream fusion structure to generate a high-resolution MS image and then used a full convolution network as a discriminator to distinguish the real and generated (fused) images. Later, Shao *et al.* [32] proposed a residual encoder–decoder conditional generative adversarial network (RED-cGAN) for pan-sharpening to produce more details with sharpened images. Unfortunately, this method needs reference remote sensing images for training, which are difficult to acquire. More recently, an unsupervised framework for pan-sharpening based on GANs was proposed by Ma *et al.* [26]. In this method, the generator separately establishes adversarial games with the spectral discriminator and the spatial discriminator so as to preserve the rich spectral information of MS images and the spatial information of PAN images. To further improve the quality of fused images, we propose a semisupervised RSIF method based on cGANs with a pseudo-Siamese structure.

B. Conditional Generative Adversarial Networks

cGANs [30] were proposed to solve the problem of GANs being too free and uncontrollable. cGANs introduce the conditional variable Y in the modeling of the generator and discriminator, and the additional information of Y will better guide the generation of data. The condition variables Y could be based on a variety of information, such as category labels, multimodal data, and random noise. When the condition is added to the generator, a potential constraint is added to the random distribution of GANs so that more realistic data is generated [22]. As a result, cGANs have been widely used. The objective function of cGANs is

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

where G represents the generator, D represents the discriminator, x and y, respectively, represent the ground truth or reference images and the generated image, and z represents the input of G.

C. Least Squares Generative Adversarial Network

Compared with original GANs, the LSGANs [31] have a different loss function, i.e., the least squares loss function replaces the loss function of original GANs. This change can alleviate the problem of unstable training processes and improve the diversity of the generated images. With the least square loss, the image distribution can be infinitely close to the decision boundary. The loss function is defined as follows:

$$\min_D V_{\mathrm{LSGAN}}(D) = \frac{1}{2} E_{x\sim p_{\mathrm{data}}(x)}\left[(D(x) - b)^2\right] + \frac{1}{2} E_{z\sim p_z(z)}\left[(D(G(z)) - a)^2\right], \quad (2)$$

$$\min_G V_{\mathrm{LSGAN}}(G) = \frac{1}{2} E_{z\sim p_z(z)}\left[(D(G(z)) - c)^2\right] \quad (3)$$

where (2) represents the loss function of generator, and (3) represents the loss function of discriminator. x is the real image, and the random variable z follows a standard normal distribution. The constants a and b, respectively, represent the labels of the real image and the generated image, and c represents the situation where the generated data is determined as true data by the discriminator. Specifically, $a = c = 1$ and $b = 0$ were proved to achieve good performance in [31]. In view of its advantages, the loss function of the LSGAN is used to replace the loss function of cGANs in our proposed model.

III. Proposed Method

To address the limitation of conventional DL-based RSIF methods, we introduce a new semisupervised image fusion model that does not require the ground truth to train. In this method, modified GANs with a pseudo-Siamese network and multiskip connection are proposed based on cGANs. Besides, a composite loss function is designed based on the combination of least squares loss, L1 loss, and PSNR loss to measure the errors between the source images and the generated image. In this section, this new neural network and loss function are reported in detail.

A. Network Structures and Processes

The proposed cGANs model consists of a generator and two discriminators. In the generator, two encoders with the same structure compose a pseudo-Siamese network, but there is only one decoder in this network. To ensure that the fused image could retain more spatial details and richer spectral information, we considered preserving the spectral features of the MS image and preserving the spatial features of the PAN image as two separate tasks; these tasks are completed by the proposed dual-discriminator structure. The two discriminators have the same network structure comprising five convolution modules. The input of the generator is the PAN image and the V channel of the MS image, and the output is the fused V channel. The input of the discriminator $Discriminator\_MS$ is the fused V channel and the V channel of the MS image. This discriminator's purpose is to force the spectral information in the fused image to be as consistent as possible with the spectral information in the MS image. Similarly, the input of the discriminator $Discriminator\_PAN$ is the fused V channel and PAN image. This discriminator can force the detailed texture information in the fused image to be consistent with the spatial information in the PAN image. During model training, once neither of the two discriminators can distinguish their inputs, the desired fused image with both high spatial resolution and high spectral resolution can be obtained. The block diagram of the proposed

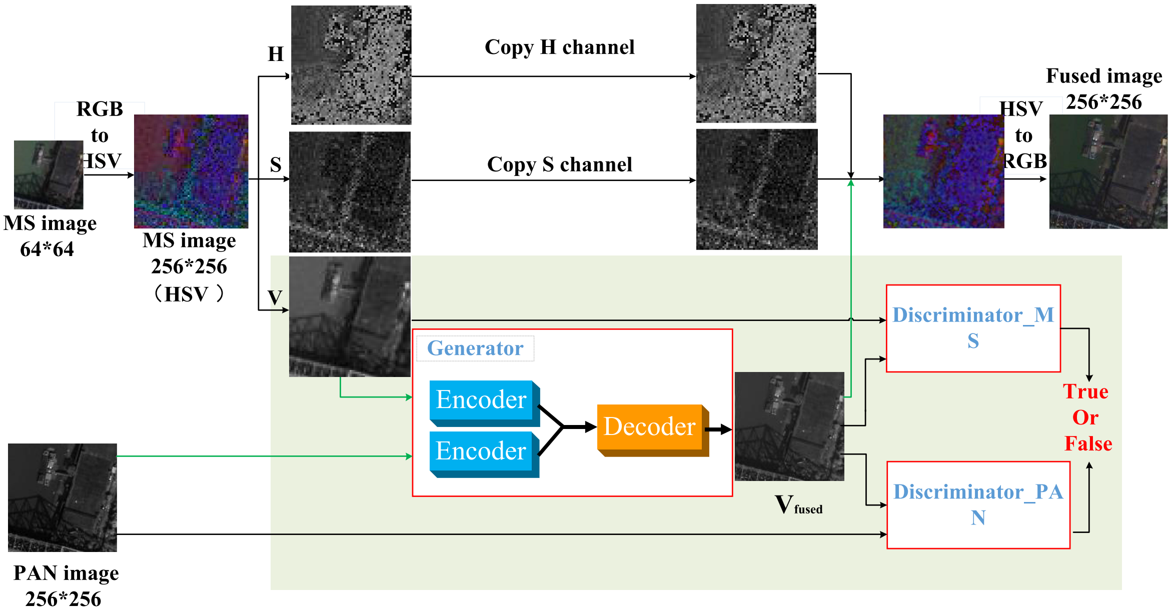

Fig. 1. Framework of the proposed method.

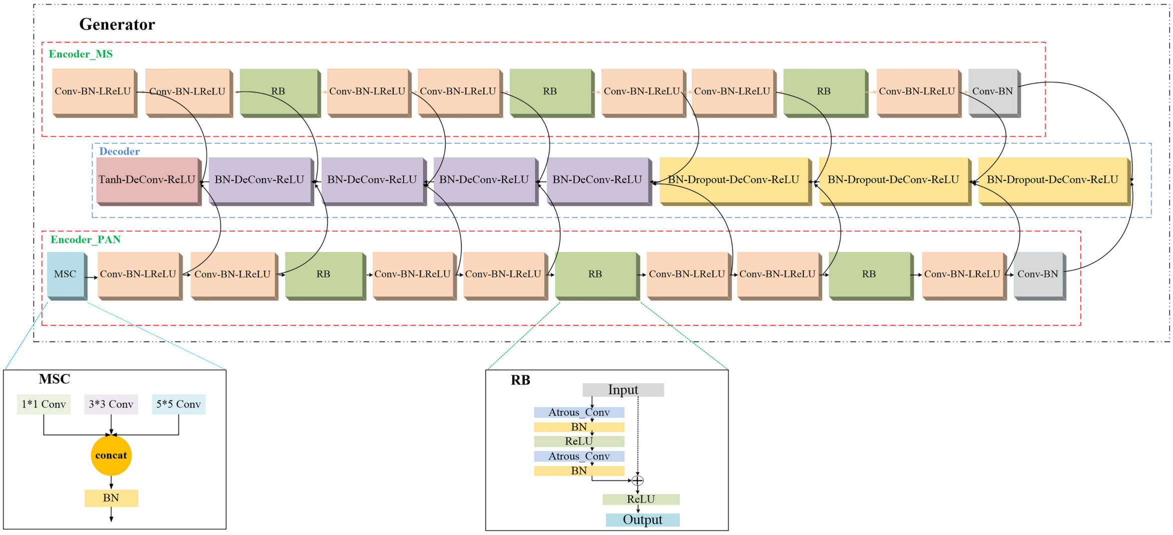

Fig. 2. Detailed structure of the generator.

method is shown in Fig. 1. For a 64∗64 MS image and 256∗256 PAN image pair, the following seven steps are executed. First, the MS image is enlarged to 256∗256 so that it is the same size as the PAN image, and then the enlarged MS image is converted from RGB color space to hue saturation value (HSV) color space. In this work, HSV is used to separate the brightness intelligence from the chrominance information, and V can be regarded as the representation of brightness intelligence that is important to the human visual system. Second, the PAN image and V channel of the MS image are input into the pseudo-Siamese networks to obtain the latent features, and, thus, the detailed features of the MS and PAN images are extracted. Third, the latent features of the two encoders are connected as the input of the decoder that is used to generate the new fused V channel. Fourth, the fused V channel is paired with the PAN image and the V channel of the MS image to input them into two discriminators, and then the discriminated results are fed back to the generator. Sixth, the two discriminators continuously play the zero-sum game with the generator until the discriminators fail to identify the generated V channel; at this point, the optimal fusion result is obtained. Finally, the fused V channel is concatenated with the H and S channels of the MS image, and then we convert them into RGB color space. Thus, we obtain the final fused image.

1) Generator Structure: The structure of our generator is shown in Fig. 2. It consists of two encoders and a decoder. The two encoders have the same network structure, i.e., a pseudo-Siamese network, in which the encoders are both connected with the decoder as inspired by the U-shape neural network. Thus, we can simultaneously extract the latent features of PAN and MS

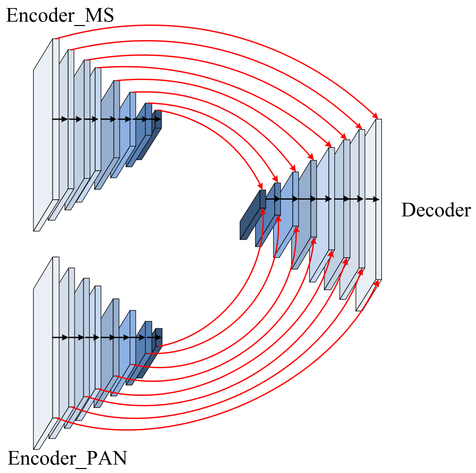

Fig. 3. Adopted skip connection.

images, and then the decoder can generate a new fused image that retains the detailed information of both the PAN image and MS image. The effect of this structure is verified in the experimental section.

a) Pseudo-Siamese Network: In the generator, a new pseudo-Siamese network structure is adopted to simultaneously integrate the features of PAN and MS images. One encoder extracts the spatial features of PAN images, while the other is employed to extract the spectral features of MS images; these encoders are referred to as Encoder_PAN and Encoder_MS, respectively. In the pseudo-Siamese network, we use the residual block (RB), which can deepen the network, to better present the image features and prevent vanishing gradients. After concatenating the extracted features in Encoder_PAN and Encoder_MS layer by layer, we input them into the corresponding layers of the decoder, which has a double-side skip connection as shown in Fig. 3. Through the fusion of low-level features and high-level features between the encoders and decoder, the network can retain abundant features from the PAN and MS images. Thus, the spatial and spectral information can be integrated into the fused image. The structure of our pseudo-Siamese structure with skip connection is shown in Fig. 3.

b) Encoder: The proposed encoder consists of eight convolution modules, as shown in Fig. 2. Except for the last convolution module, which contains a convolution layer and a batch normalization (BN) layer, all modules are composed of a convolution layer and a BN layer, followed by a leaky-rectified linear unit (LReLU) activation function layer. In addition, every two convolution modules have an RB, and, thus, there are three RBs in total. The structure of the RB is also illustrated in Fig. 2. It consists of two convolution modules: The first module consists of an atrous convolution layer, a BN layer, and a rectified linear unit (ReLU) activation function layer; the second module has an atrous convolution layer and a BN layer. The input is added to the output of the second convolution module. Here, compared to the proposed method without RB modules, the RB can better present the image features, and, thus, the performance of the RSIF method is also improved.

Note that Encoder_MS and Encoder_PAN have some structural differences, i.e., whether the multiscale convolution (MSC) module is included; the structure of the MSC module is also shown in Fig. 2.

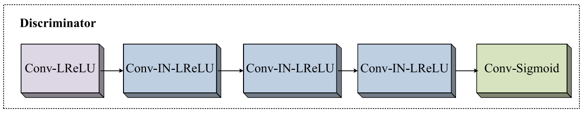

Fig. 4. Structure of the discriminator.

c) Decoder: The decoder is similar to the encoder and consists of eight deconvolution modules. The first three modules are composed of a ReLU activation function layer, a deconvolution layer, a dropout layer, and a BN layer. The middle three modules have the same structure as the encoder but do not contain a dropout layer. The activation function of the last module uses Tanh instead of ReLU. This structure can also be found in Fig. 2.

2) Discriminator Structure: The proposed model has two discriminators that have the same network structure. The discriminator contains five convolution modules. The first modules consist of a convolution layer and an LReLU activation layer. The middle three convolution modules have the same structure: a convolution layer, an instance normalization (IN) layer, and an LReLU activation layer. The last convolution module has only one convolutional layer, and the activation layer uses a Sigmoid activation function. Unlike a common discriminator, we use an IN layer instead of the BN layer. The reason for this is that the BN layer is sensitive to batch size, and the mean and variance of each iteration are calculated on the same batch. As a result, if the batch size is too small, the calculated mean and variance are not enough to represent the distribution of the entire data. Thus, if we set the batch size as 1, the BN will not work. The structure of our discriminator is shown in Fig. 4.

B. Loss Function

In this work, we combine the least square loss, L1 loss, and PSNR loss as the final loss function. Equations (4) and (5) represent the least square loss

$$\begin{aligned} L_{\mathrm{cLSGAN}}\left(G\right) &= E_{(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}) \sim P_{\mathrm{data}}(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}})} \\ &\quad \left[\left(D_{x_{\mathrm{MS}}}\left(G\left(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}\right) \mid x_{\mathrm{MS\_V}}\right) - 1\right)^2\right] \\ &\quad + E_{(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}) \sim P_{\mathrm{data}}(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}})} \\ &\quad \left[\left(D_{x_{\mathrm{PAN}}}\left(G\left(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}\right) \mid x_{\mathrm{PAN}}\right) - 1\right)^2\right], \end{aligned} \tag{4}$$

$$\begin{aligned} L_{\mathrm{cLSGAN}}\left(D_{x_{\mathrm{MS\_V}}}, D_{x_{\mathrm{PAN}}}\right) &= \frac{1}{2} \\ &\left[\begin{array}{l} E_{(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}) \sim P_{\mathrm{data}}(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}})} \\ \quad \left[\left(D_{x_{\mathrm{MS\_V}}}\left(x_{\mathrm{PAN}} \mid x_{\mathrm{MS\_V}}\right) - 1\right)^2\right] \\ + E_{(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}) \sim P_{\mathrm{data}}(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}})} \\ \quad \left[\left(D_{x_{\mathrm{PAN}}}\left(x_{\mathrm{MS\_V}} \mid x_{\mathrm{PAN}}\right) - 1\right)^2\right] \end{array}\right] \\ &+ \frac{1}{2}\left[\begin{array}{l} E_{(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}) \sim P_{\mathrm{data}}(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}})} \\ \quad \left[\left(D_{x_{\mathrm{MS\_V}}}\left(G\left(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}\right) \mid x_{\mathrm{MS\_V}}\right)\right)^2\right] \\ + E_{(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}) \sim P_{\mathrm{data}}(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}})} \\ \quad \left[\left(D_{x_{\mathrm{PAN}}}\left(G\left(x_{\mathrm{MS\_V}}, x_{\mathrm{PAN}}\right) \mid x_{\mathrm{PAN}}\right)\right)^2\right] \end{array}\right] \end{aligned} \tag{5}$$

where G represents the generator, $D_{x_{\mathrm{PAN}}}$ and $D_{x_{\mathrm{MS}}}$ are two discriminators with the same structure, and the input true PAN image and the V channel of the MS image are represented by x_{PAN} and $x_{\mathrm{MS\_V}}$, respectively. In addition, we use $x_{\mathrm{MS\_V}}$ and x_{PAN} to represent the condition of the generator and discriminator; this can guide the generation of the fused image without a ground truth.

L1 loss represents the difference between the real image and the generated image. The addition of L1 loss can improve the clarity of the generated image and thus the quality of the fused image. In this work, the L1 loss between the real MS image and the generated image is denoted by $L_{\mathrm{L1\_MS}}$, and the L1 loss between the real PAN image and the generated image is denoted by $L_{\mathrm{L1\_PAN}}$. The weights of $L_{\mathrm{L1\_MS}}$ and $L_{\mathrm{L1\_PAN}}$ are, respectively, set as 0.3 and 0.7. The equations of L1 loss are shown as follows:

$$L_{\mathrm{L1\_MS}}(G) = E_{(x_{\mathrm{MS}},x_{\mathrm{PAN}})\sim P_{\mathrm{data}}(x_{\mathrm{MS}},x_{\mathrm{PAN}})} \left[\| G(x_{\mathrm{MS}}, x_{\mathrm{PAN}}) - x_{\mathrm{MS}}\|_1\right], \tag{6}$$

$$L_{\mathrm{L1\_PAN}}(G) = E_{(x_{\mathrm{MS}},x_{\mathrm{PAN}})\sim P_{\mathrm{data}}(x_{\mathrm{MS}},x_{\mathrm{PAN}})} \left[\| G(x_{\mathrm{MS}}, x_{\mathrm{PAN}}) - x_{\mathrm{PAN}}\|_1\right], \tag{7}$$

$$L_{\mathrm{L1}}(G) = 0.3 * L_{\mathrm{L1\_MS}}(G) + 0.7 * L_{\mathrm{L1\_PAN}}(G) \tag{8}$$

where L_{L1} is the final L1 loss.

PSNR is a widely used image evaluation index that is based on the errors between corresponding pixels of two images, i.e., the evaluation is an error-sensitivity model. In this work, the PSNR loss between the real MS image and the generated image is denoted by $L_{\mathrm{PSNR\_MS}}$, and the L1 loss between the real PAN image and the generated image is denoted by $L_{\mathrm{PSNR\_PAN}}$. The weights of $L_{\mathrm{PSNR\_MS}}$ and $L_{\mathrm{PSNR\_PAN}}$ are, respectively, set as 0.3 and 0.7. The equations of PSNR loss are expressed as follows:

$$M_{\mathrm{SE\_MS}} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[V_F(i,j) - x_{\mathrm{MS\_V}}(i,j)\right]^2, \tag{9}$$

$$L_{\mathrm{PSNR\_MS}} = 10 * \log_{10}\left(\frac{\mathrm{MAX}_I^2}{M_{\mathrm{SE\_MS}}}\right), \tag{10}$$

$$M_{\mathrm{SE\_PAN}} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[V_F(i,j) - x_{\mathrm{PAN}}(i,j)\right]^2, \tag{11}$$

$$L_{\mathrm{PSNR\_PAN}} = 10 * \log_{10}\left(\frac{\mathrm{MAX}_I^2}{M_{\mathrm{SE\_PAN}}}\right), \tag{12}$$

$$L_{\mathrm{PSNR}} = 0.3 * L_{\mathrm{PSNR\_MS}} + 0.7 * L_{\mathrm{PSNR\_PAN}} \tag{13}$$

where $M_{\mathrm{SE\_MS}}$ is the mean square error (MSE) between the fused V channel and the MS image, and $M_{\mathrm{SE\_PAN}}$ is the MSE between the fused V channel and the PAN image. In (9) and (11), m and n indicate the size of the source image. MAX_I^2represents the maximum pixel value of the image. In this work, the image is normalized, and, thus, $\mathrm{MAX}_I^2 = 1$.

Finally, the loss function of the generator can be expressed as

$$L(\mathrm{G}) = L_{\mathrm{cLSGAN}}(\mathrm{G}) + \lambda L_{\mathrm{L1}}(G) + 0.0001 * L_{\mathrm{PSNR}} \tag{14}$$

where the weight of L_{L1} is λ. In this work, λ = 100, and the weight of L_{PSNR} is set to 0.0001.

IV. Experiments and Analysis

In this section, extensive experiments are conducted to verify the performance of the proposed method with different settings.

A. Dataset

In this work, we adopted a set of remote sensing images from the University of Maryland's website,[1] which contains MS images and their corresponding PAN images. The sizes of the images vary across the datasets. We selected two PAN and MS image pairs in different scenes for experimental training and testing; the sizes of the selected PAN and MS images used in the training set were 1973 × 4096 and 7892 × 16384, respectively. The sizes of PAN and MS images in the test set were 1920 × 4096 and 7680 × 16384, respectively. Because the original remote sensing image is very large, we divided the MS and PAN images into 64 × 64 and 256 × 256 pixels, respectively, to obtain a set of registered image pairs. We used a total of 3840 PAN and MS image pairs in the training set.

B. Experimental Setup

Here we present experiments on our key parameter settings and network structures. In the encoder of the generator, an MSC module is employed, the convolution kernel sizes are set as 1, 3, and 5, and the convolution step is set as 1. The dilated convolution is applied in the RB, whose convolution kernel is set as 3; the step size is 1, and the dilation is 2. The convolution kernel of other convolution modules in the decoder is set as 4, and the step size is 2. In the decoder, the convolution kernel and step size of the deconvolution modules are set the same as in the convolution module in the encoder. In the discriminator, the convolution kernel size of the convolutional layer in all the convolutional modules is set as 4, and the convolution step size in the first three convolutional modules is 2; meanwhile, the step size in the other convolutional modules is set as 1. In the experiment, the number of training epochs is set as 200, the learning rate is 0.00002, and the dropout rate is 0.5. We also propose a composite loss function, in which the weight of the L1 loss function is set as 100, the weight of the PSNR loss is set as 0.0001, and the weight of least square loss is set as 1.

1) Comparison Experiment With Different Skip Connections: Figs. 5–8 show the proposed model's fusion results and image enlargement with different skip connections. We can observe that the fused image can only retain part of the spectral information in the MS image when the proposed method lacks the skip connection. A single-side skip connection, i.e., Proposed_NoPAN_Skip and Proposed_NoMS_Skip, which means either Encoder_MS or Encoder_PAN is connected to the decoder by the skip connection. The first case can retain most of the spectral information of the MS image but only a few details of the PAN image. The second case can retain most of the details of

[1] [Online]. Available: http://glcfapp.glcf.umd.edu:8080/esdi/index.jsp

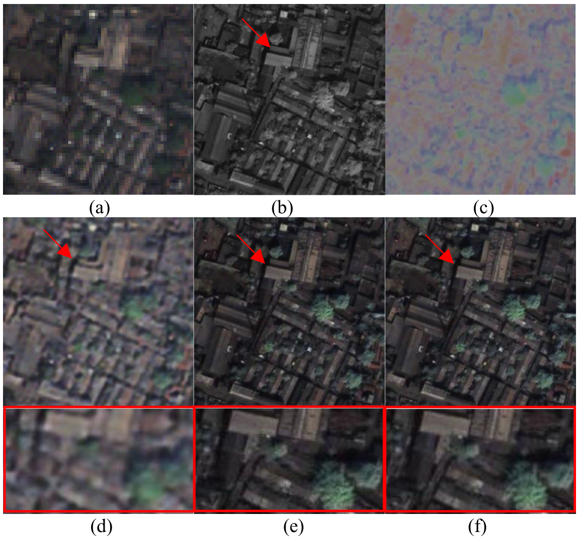

Fig. 5. First group of fused images and enlarged images produced by the proposed method with or without skip connection. (a) MS image. (b) PAN image. (c) Proposed_No_Skip. (d) Proposed_NoPAN_Skip. (e) Proposed_NoMS_Skip. (f) Proposed_Skip.

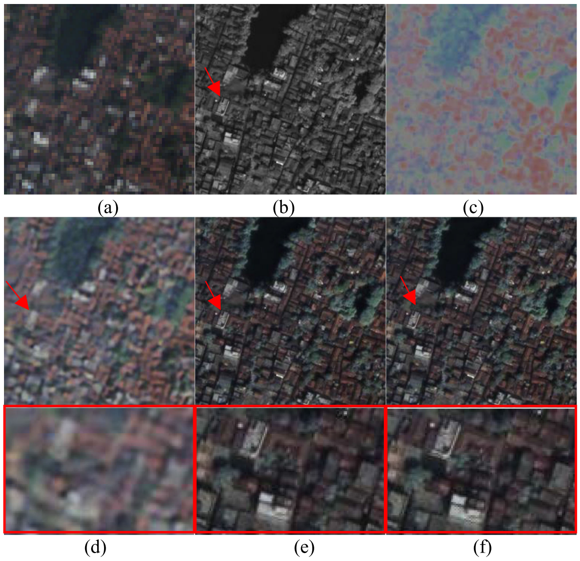

Fig. 6. Second group of fused images and enlarged images produced by the proposed method with or without skip connection. (a) MS image. (b) PAN image. (c) Proposed_No_Skip. (d) Proposed_NoPAN_Skip. (e) Proposed_NoMS_Skip. (f) Proposed_Skip.

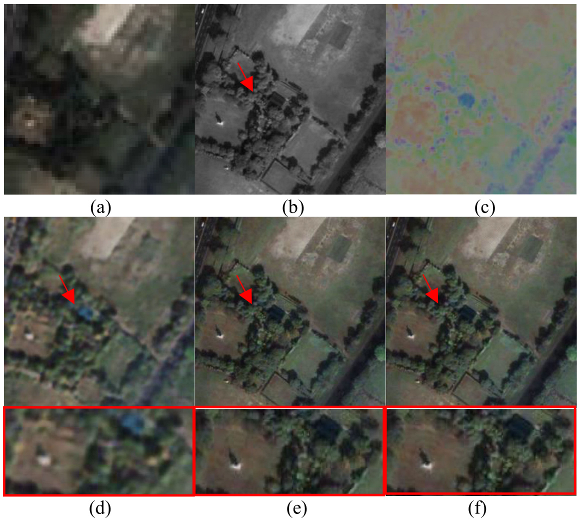

Fig. 7. Third group of fused images and enlarged images produced by the proposed method with or without skip connection. (a) MS image. (b) PAN image. (c) Proposed_No_Skip. (d) Proposed_NoPAN_Skip. (e) Proposed_NoMS_Skip. (f) Proposed_Skip.

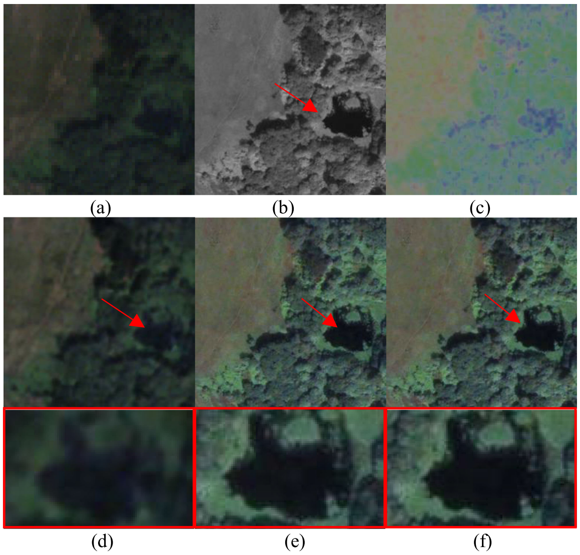

Fig. 8. Fourth group of fused images and enlarged images produced by the proposed method with or without skip connection. (a) MS image. (b) PAN image. (c) Proposed_No_Skip. (d) Proposed_NoPAN_Skip. (e) Proposed_NoMS_Skip. (f) Proposed_Skip.

the PAN image. To concatenate two encoders with a decoder by skip connection, the double-side skip connection is used in the proposed method. In Figs. 5–8, when compared with the first two cases, the fused image obtained by the proposed method with double-side skip connection has a better visual effect and contains more detailed information.

The performance of the proposed model without a skip layer connection (Proposed_No_Skip) and that of the model with a single skip connection (Proposed_NoPAN_Skip) are not satisfactory. Therefore, we only compare the objective indicators of the model with a double-side skip connection (Proposed_Skip) and the model with a single-side skip connection (Proposed_NoMS_Skip). The average values of different metrics are shown in Table I and illustrate that the performance of Proposed_Skip is better than that of Proposed_NoMS_Skip. In general, the proposed method with a double-side skip connection can achieve better performance than that with a nonskip or single-side skip connection.

2) Comparison of the Proposed Method With or Without Multiscale Modules: Figs. 9 –12 show the results of the proposed method with and without multiscale modules. The fused images obtained by our final model have more detailed features in the ocean, and most spectral information of the MS image is retained. In other images, the influence of multiscale

TABLE I
AVERAGED IMAGE EVALUATION INDICES FOR THE PROPOSED METHOD WITH OR WITHOUT SKIP CONNECTION

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
Proposed_NoMS_Skip	6.0888	11.5101	24.2947	0.5434	0.3423	2.9499	0.7603	0.6868	0.8246	1.5088
Proposed_Skip	**6.3260**	**11.9525**	**24.7751**	**0.5529**	**0.3231**	**2.9746**	**0.7689**	**0.6972**	**0.8323**	**1.5723**

The bold values represent the value of the optimal image quality indicator.

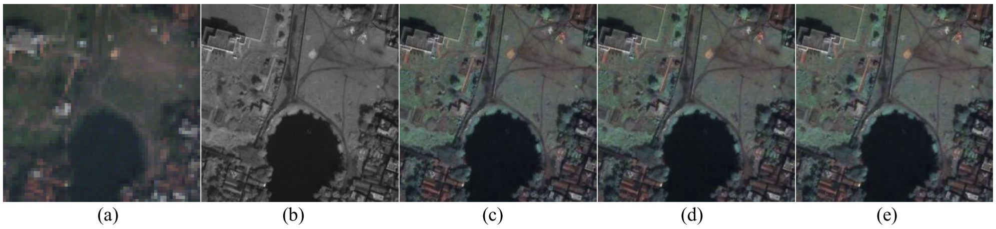

Fig. 9. First group of fused image of the proposed method with or without multiscale modules. (a) MS image. (b) PAN image. (c) Proposed_No_Multi-scale. (d) Proposed_2_ Multi-scale. (e) Proposed_Multiscale.

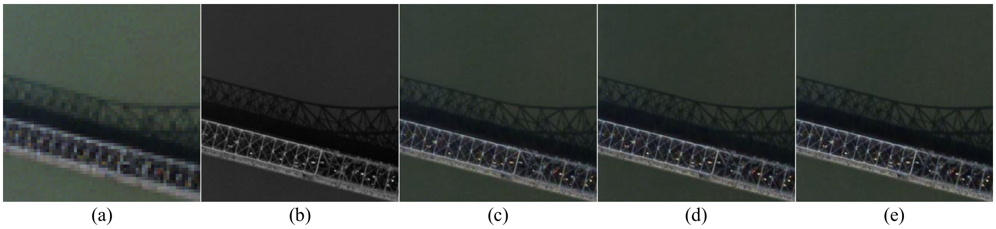

Fig. 10. Second group of fused image of the proposed method with or without multiscale modules. (a) MS image. (b) PAN image. (c) Proposed_No_Multi-scale. (d) Proposed_2_Multi-scale. (e) Proposed_Multiscale.

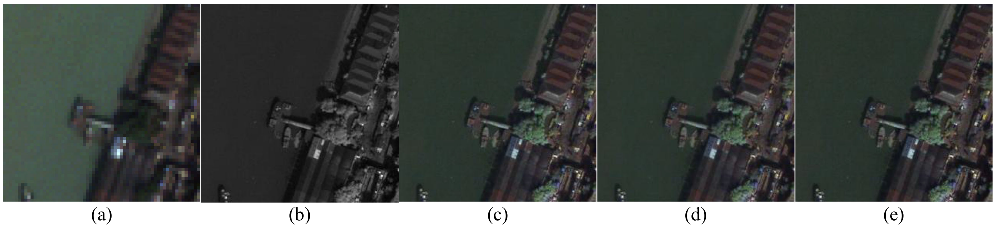

Fig. 11. Third group of fused image of the proposed method with or without multiscale modules. (a) MS image. (b) PAN image. (c) Proposed_No_Multi-scale. (d) Proposed_2_Multi-scale. (e) Proposed_Multiscale.

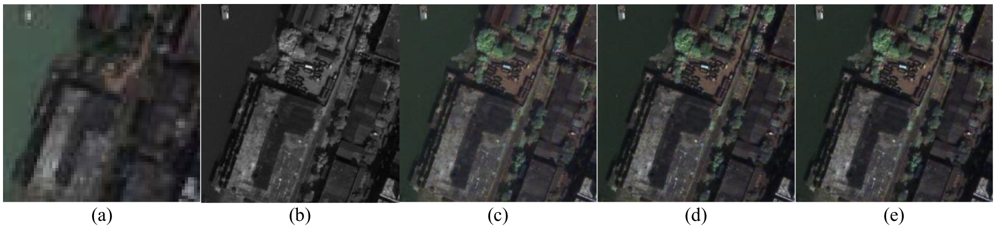

Fig. 12. Fourth group of fused image of the proposed method with or without multiscale modules. (a) MS image. (b) PAN image. (c) Proposed_No_Multi-scale. (d) Proposed_2_Multi-scale. (e) Proposed_Multiscale.

modules on the fused image cannot be easily distinguished by the human eye. Thus, we use ten popular evaluation indicators to analyze the performance of different multiscale modules. The results in Table II reveal that the performance of the MSC module in Encoder_PAN is better than that of the other two ways (Proposed_2_Multi-scale and Proposed_No_Multi-scale).

3) Comparison of Different Loss Functions: The fused results obtained when using different loss functions are shown in Figs. 13 –15. We can observe that the spectral information of the MS image is well retained by our composite loss function

TABLE II
AVERAGED IMAGE EVALUATION INDICES FOR THE PROPOSED METHOD WITH OR WITHOUT MULTISCALE MODULES

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
Proposed_No_Multi-scale	4.5163	10.9087	21.4066	0.5158	0.3855	2.9415	0.5567	0.6373	0.8082	0.8390
Proposed_2_Multi-scale	4.7727	11.7073	22.1461	0.5213	0.3542	2.9396	0.5651	0.6462	0.8146	0.9172
Proposed_Multi-scale	**4.8984**	**11.9989**	**22.7487**	**0.5254**	**0.3413**	**2.9956**	**0.5732**	**0.6508**	**0.8182**	**0.9597**

The bold values represent the value of the optimal image quality indicator.

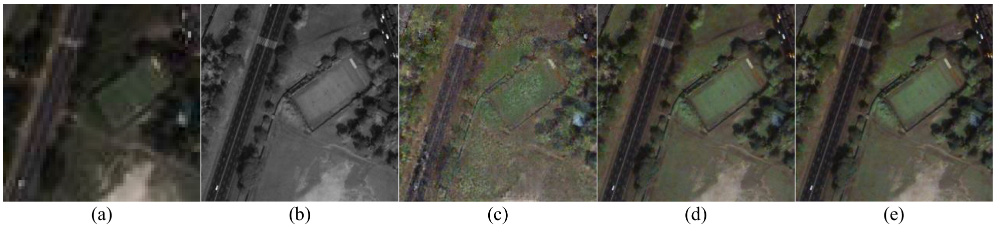

Fig. 13. First group of fused image using different loss functions. (a) MS image. (b) PAN image. (c) LSGAN+PSNR. (d) LSGAN+L1. (e) LSGAN+L1+PSNR.

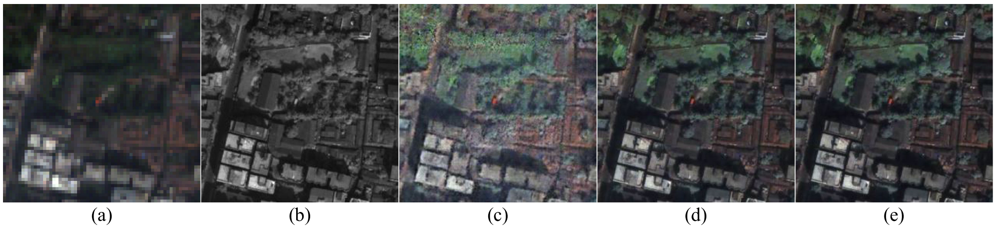

Fig. 14. Second group of fused image using different loss functions. (a) MS image. (b) PAN image. (c) LSGAN+PSNR. (d) LSGAN+L1. (e) LSGAN+L1+PSNR.

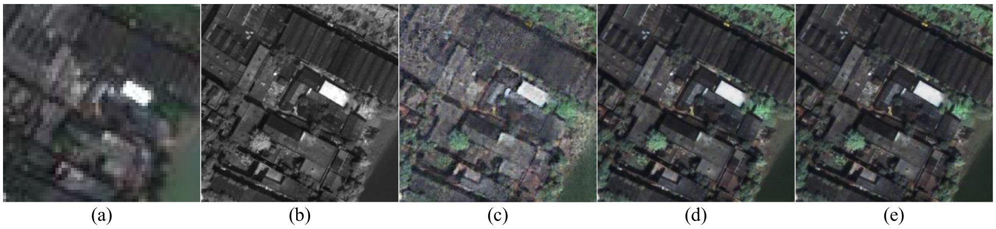

Fig. 15. Third group of fused image using different loss functions. (a) MS image. (b) PAN image. (c) LSGAN+PSNR. (d) LSGAN+L1. (e) LSGAN+L1+PSNR.

TABLE III
AVERAGED IMAGE EVALUATION INDICES FOR THE PROPOSED METHOD WITH DIFFERENT LOSS FUNCTIONS

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
Proposed_LSGAN+L1	6.3871	12.9369	27.8306	0.5399	0.3646	3.0205	0.7636	0.6667	0.8144	0.8252
Proposed_LSGAN+L1+PSNR	**6.6407**	**13.4875**	**28.8063**	**0.5471**	**0.3464**	**3.0661**	**0.7722**	**0.6785**	**0.8236**	**0.8799**

The bold values represent the value of the optimal image quality indicator.

(i.e., the improved LSGAN loss plus PSNR loss), but there is some serious detail loss. When the PSNR loss is modified as L1 loss, the detailed information and visual effect of the fused remote sensing image are greatly improved. When the improved LSGAN loss and L1 loss are combined with PSNR loss, the visual effect of the fused image is similar to that of the combination of LSGAN and L1 loss. Thus, evaluation indices are used to analyze the image quality.

Table III shows the scores of fused images obtained when using different composite loss functions. The values in this table show that the proposed method should adopt the loss function constituted by the improved LSGAN loss function, L1 loss, and PSNR loss. The fused images when using our proposed loss function are better than those obtained when using the combination of LSGAN and L1 loss; this holds for each evaluation indices. The results show that the proposed method with the proposed loss can fuse most spectral information of the MS image and most of the spatial details of the PAN image. Better objective indices are also obtained when using our proposed loss.

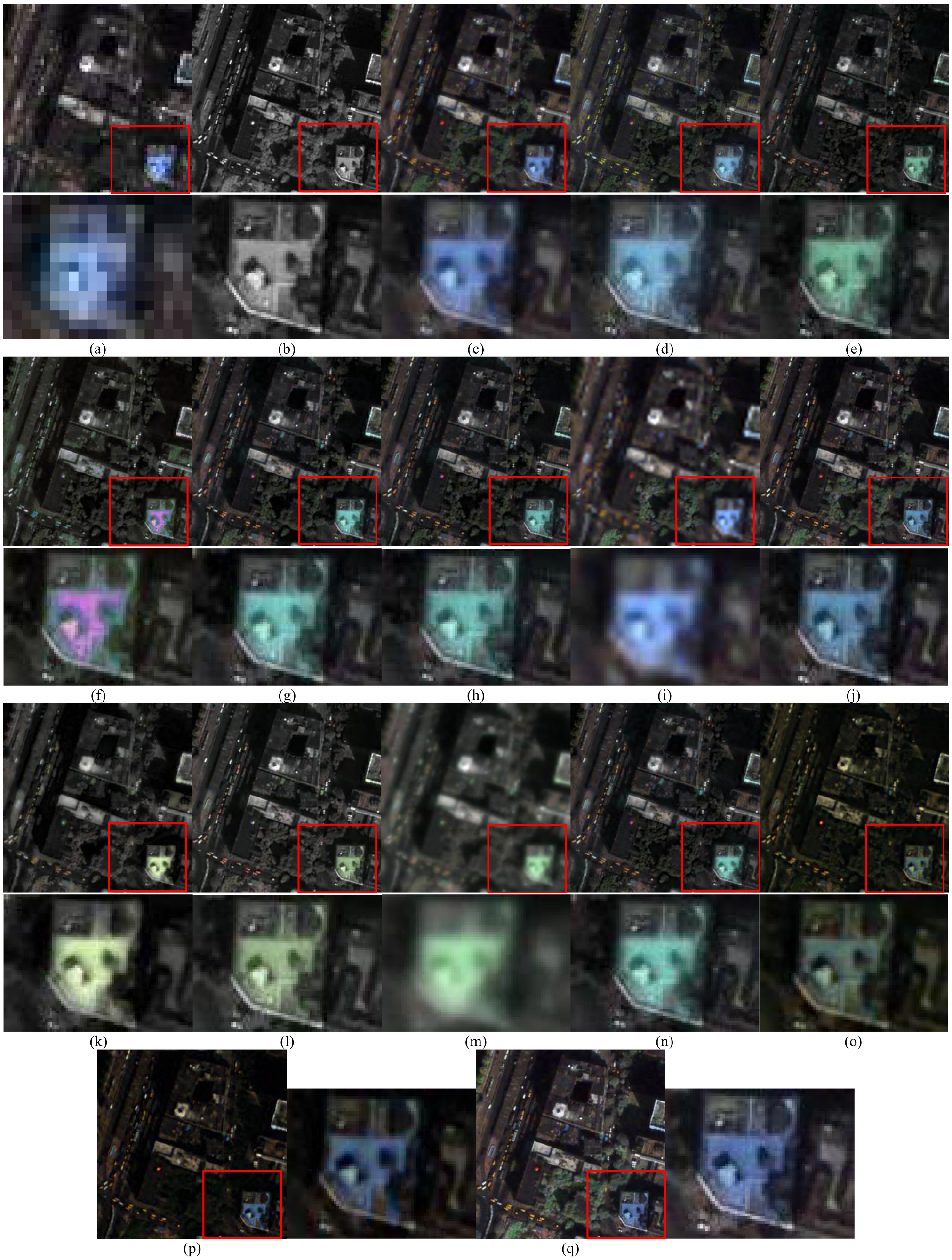

Fig. 16. First group of experimental results of different fusion methods. (a) MS image. (b) PAN image. (c) PCA. (d) GRA. (e) DWT. (f) LAP. (g) DT-DWT. (h) WTSR. (i) FFIF. (j) DCHWT. (k) MGIVF. (l) ASR. (m) CSR. (n) SWT. (o) PNN. (p) PNN+. (q) Proposed_skip.

C. Comparison Experiments

In this subsection, the proposed method is compared with existing image fusion methods through subjective visual analysis and image evaluation indices. Contrast fusion methods include PCA [37], gradient pyramid (GRA) [37], discrete wavelet transform (DWT) [37], Laplacian pyramid (LAP) [37], dual-tree complex wavelet transform (DT-DWT) [38], pan-sharpening method with wavelet transform and sparse representation (WTSR) [18], fast filtering image fusion (FFIF) [39], discrete cosine harmonic wavelet transform (DCHWT) [40], multiscale guided image and video fusion (MGIVF) [41], adaptive sparse representation (ASR) [42], CSR [43], stationary wavelet transform (SWT) [44], pan-sharpening by convolutional neural network (PNN) [45], and target-adaptive CNN-based pan-sharpening (PNN+) [46]. What needs illustration is that all these methods are performed on PAN images and three-band MS images.

Figs. 16–21 show six groups of experimental results of different fusion methods. Although the images in Fig. 16(c), (d), and (i) can retain the spectral information of the MS image,

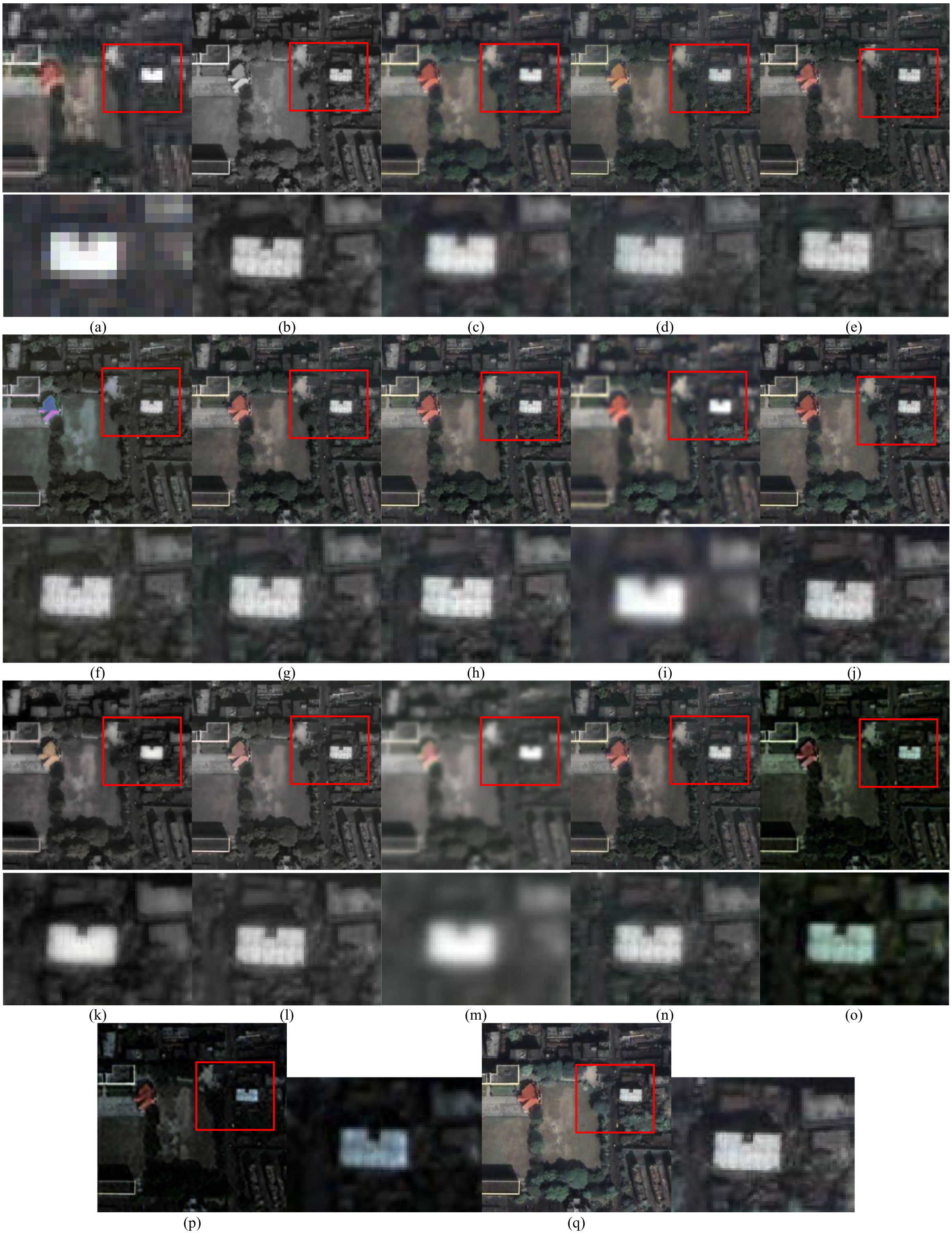

Fig. 17. Second group of experimental results of different fusion methods. (a) MS image. (b) PAN image. (c) PCA. (d) GRA. (e) DWT. (f) LAP. (g) DT-DWT. (h) WTSR. (i) FFIF. (j) DCHWT. (k) MGIVF. (l) ASR. (m) CSR. (n) SWT. (o) PNN. (p) PNN+. (q) Proposed_skip.

the detailed information of the PAN image is seriously lost. In Fig. 16 (e), (f), (g), (h), (l), (n), (o), and (p), most of the details in the PAN image are retained, but the significant spectrum of the MS image is distorted, especially the roof of the building. In Fig. 16(m) and (k), both the spectral information of the MS image and the detailed information of the PAN image are seriously lost, and only part of the features are fused into the final images. The image in Fig. 16(j) has a good fusion effect but suffers from spectral distortion and detail loss. Although images in Fig. 17(c), (e), (g), (h), (j), and (n) present a good visual effect, there are some blurred areas at the roof of the building. Meanwhile, images in Fig. 17(d), (f), (k), (l), (o), and

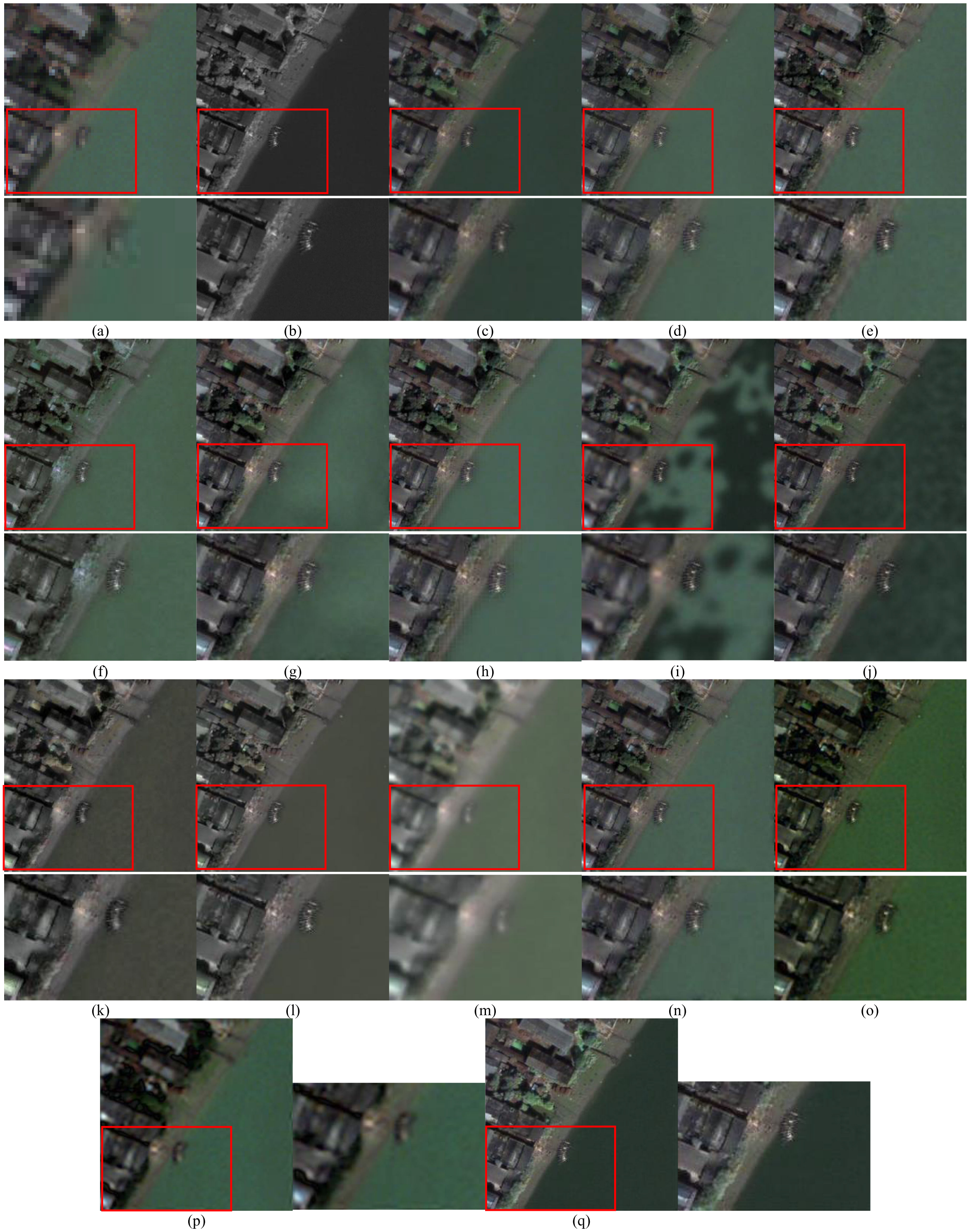

Fig. 18. Third group of experimental results of different fusion methods. (a) MS image. (b) PAN image. (c) PCA. (d) GRA. (e) DWT. (f) LAP. (g) DT-DWT. (h) WTSR. (i) FFIF. (j) DCHWT. (k) MGIVF. (l) ASR. (m) CSR. (n) SWT. (o) PNN. (p) PNN+. (q) Proposed_skip.

(p) have serious spectral distortion, and some objects cannot be distinguished. In Fig. 17(i) and (m), only part of the useful features of the MS and PAN images are fused into the final images. Our proposed method can effectively preserve the details of the source images and has a good visual effect that can distinguish buildings, vegetation, topography, etc. The visual effect of the fused images obtained by the proposed method is closest to the source images, and the results can well present the features of the source images.

Similarly, the images in Fig. 18 show that most image fusion methods can achieve good visual effects. The blurred areas in

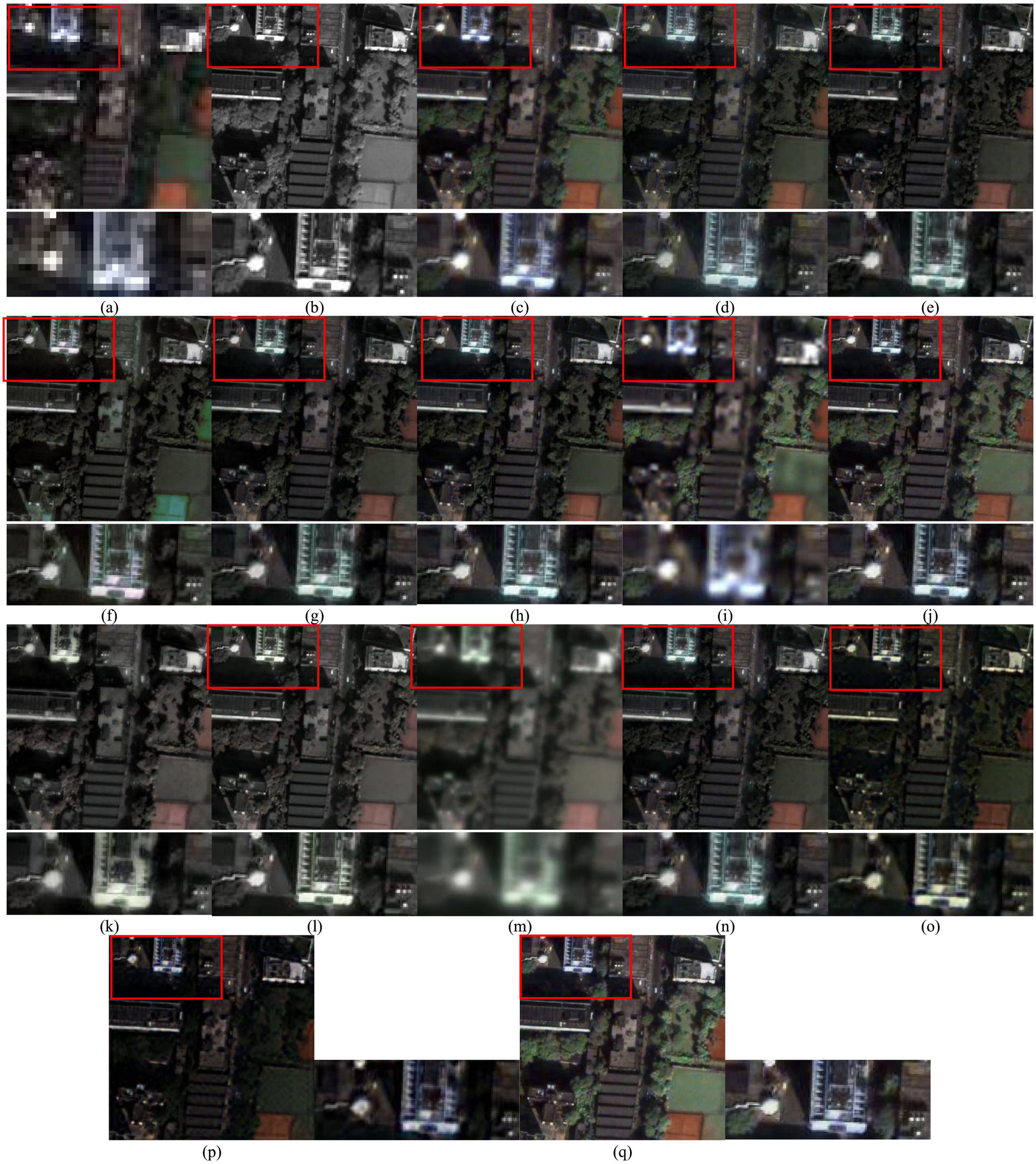

Fig. 19. Fourth group of experimental results of different fusion methods. (a) MS image. (b) PAN image. (c) PCA. (d) GRA. (e) DWT. (f) LAP. (g) DT-DWT. (h) WTSR. (i) FFIF. (j) DCHWT. (k) MGIVF. (l) ASR. (m) CSR. (n) SWT. (o) PNN. (p) PNN+. (q) Proposed_skip.

Fig. 18 (i) and (m) show that these methods [39], [43] fail to fuse the features of MS and PAN images, while only the spectral information is kept. In Fig. 18(k) and (l), most spectral information of the MS image is lost. Besides, the spectral information of the MS image is distorted in Fig. 18(o). Images in Fig. 19(c) and (j) have a good visual effect that is close to the source images but have a certain degree of detail loss. In Fig. 19(d), (e), (g), and (h), the details of source images are well preserved, but the spectrum may have serious distortions, especially the building's roof and vegetation cover. In Fig. 19(o) and (p), some spectrum and detailed features are lost. The fusion performance observed in Fig. 19(f), (k), (l), and (m) is much worse than that of other fusion methods. In Figs. 20 and 21, we can find a similar problem that was present in Figs. 16–19. Our experiments show that the proposed method can effectively fuse the spectral information of the MS image and the details of the PAN image and that the method's performance is competitive compared with other image fusion methods.

Besides, we adopt ten evaluation indices to further analyze the performance of different image fusion methods. These metrics are the average gradient (AG), space frequency (SF), standard deviation (STD), edge based on similarity measure (Q^{abf}), overall information loss (L^{abf}), mutual information (MI), universal image quality index (Q), edge-dependent fusion quality

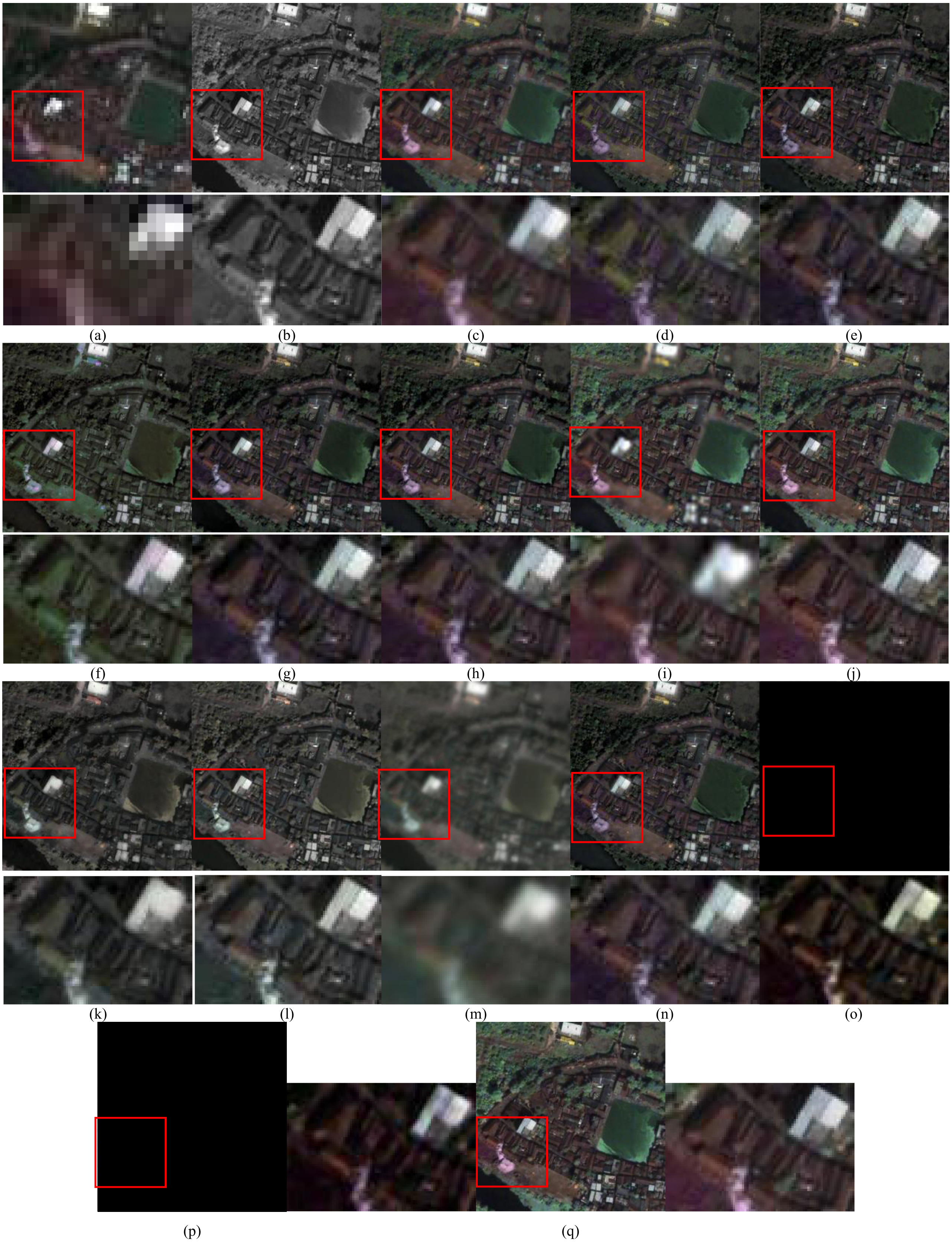

Fig. 20. Fifth group of experimental results of different fusion methods. (a) MS image. (b) PAN image. (c) PCA. (d) GRA. (e) DWT. (f) LAP. (g) DT-DWT. (h) WTSR. (i) FFIF. (j) DCHWT. (k) MGIVF. (l) ASR. (m) CSR. (n) SWT. (o) PNN. (p) PNN+. (q) Proposed_skip.

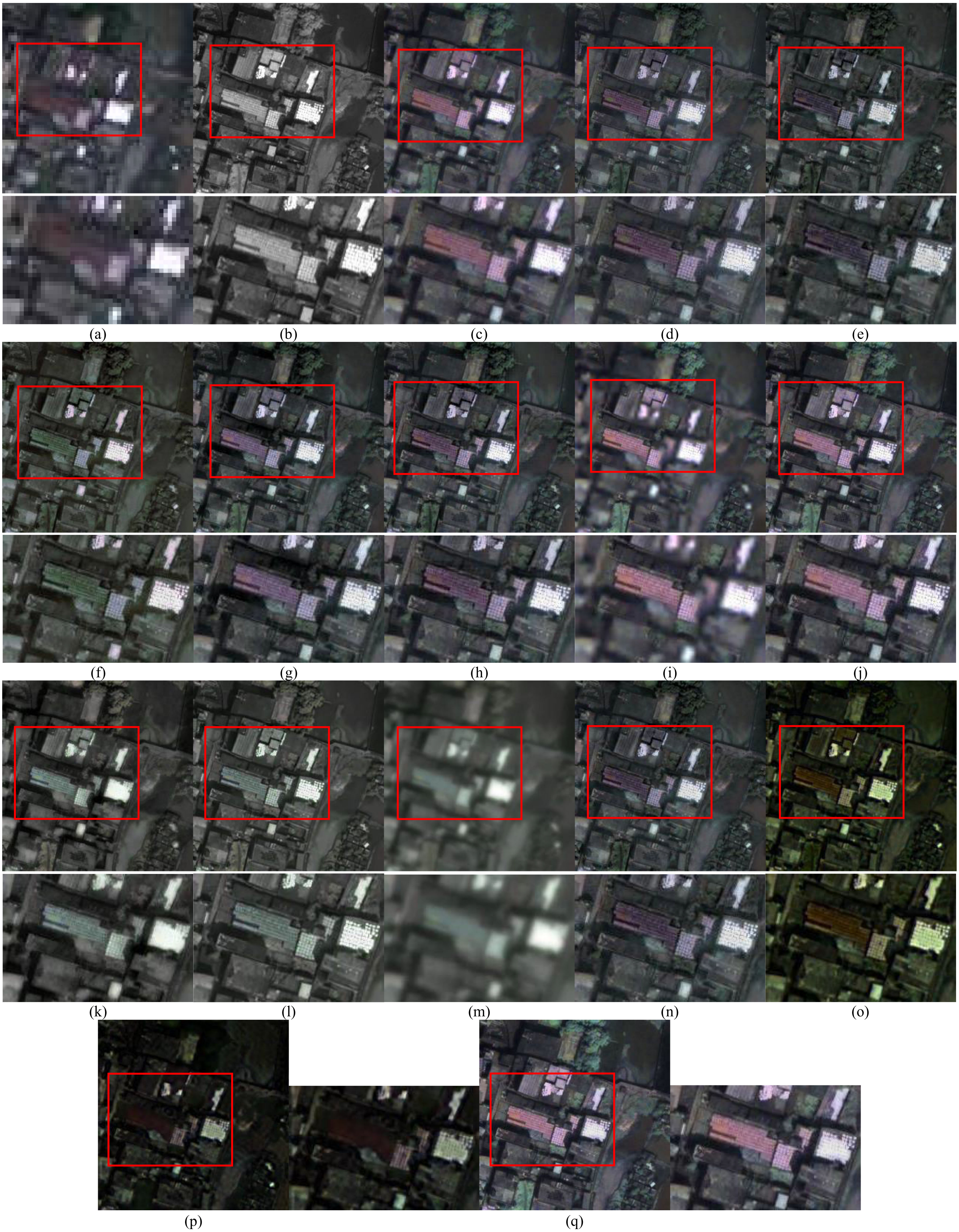

Fig. 21. Sixth group of experimental results of different fusion methods. (a) MS image. (b) PAN image. (c) PCA. (d) GRA. (e) DWT. (f) LAP. (g) DT-DWT. (h) WTSR. (i) FFIF. (j) DCHWT. (k) MGIVF. (l) ASR. (m) CSR. (n) SWT. (o) PNN. (p) PNN+. (q) Proposed_skip.

TABLE IV
EVALUATION INDICES CORRESPONDING TO FIRST GROUP OF FUSION RESULTS

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
PCA	5.6122	12.0158	30.9642	0.4209	0.5527	3.3741	0.7005	0.3452	0.7600	0.7007
GRA	6.7872	15.3459	29.7970	0.4731	0.4972	2.8078	0.6585	0.4297	0.7207	0.5446
DWT	6.5656	14.5042	26.8731	0.4534	0.5088	2.5958	0.5934	0.3786	0.6678	0.5343
LAP	7.4284	16.1966	31.1872	0.5338	0.4190	2.8315	0.7329	0.5114	0.7856	0.7379
DT-DWT	7.3287	16.3992	31.7394	0.5346	0.4240	2.8773	0.7215	0.5375	0.7907	0.7344
WTSR	7.3258	16.4347	29.7135	0.5357	0.4254	2.7217	0.6950	0.5213	0.7556	0.6127
FFIF	4.5985	9.1686	32.7709	0.3732	0.5796	**3.9118**	0.6355	0.1826	0.6819	0.6743
DCHWT	7.9364	17.9801	31.1938	0.5572	0.3934	3.4198	0.7721	0.6160	0.8204	0.7401
MGIVF	6.2304	12.9977	33.1488	0.4272	0.5166	2.9491	0.6717	0.3195	0.7489	0.7789
ASR	8.0027	18.4254	32.1522	**0.5795**	0.3681	3.2690	0.7818	0.6595	0.8330	0.7772
CSR	3.0842	5.9082	32.1412	0.2294	0.7456	3.0914	0.5350	0.0935	0.5734	0.4394
SWT	6.8345	15.1466	29.6364	0.4954	0.4677	2.8328	0.6654	0.4409	0.7380	0.6436
PNN	5.2573	12.1055	21.7103	0.3237	0.6383	2.2920	0.4130	0.2430	0.5004	0.4277
PNN+	5.0799	12.2566	21.8738	0.2851	0.6718	2.3863	0.3507	0.2158	0.4517	0.4151
Proposed	**8.3759**	**18.9787**	**33.3790**	0.5606	**0.3494**	3.3382	**0.7868**	**0.6608**	**0.8330**	**0.8284**

The bold values represent the value of the optimal image quality indicator.

TABLE V
EVALUATION INDICES CORRESPONDING TO SECOND GROUP OF FUSION RESULTS

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
PCA	4.6454	10.2091	32.2395	0.4369	0.5293	3.5820	0.7317	0.4445	0.7936	0.7485
GRA	5.2935	12.5164	30.1523	0.4633	0.5015	3.0631	0.6753	0.5135	0.7410	0.5588
DWT	6.0501	13.6791	30.3038	0.4727	0.4487	2.8746	0.6849	0.5829	0.7690	0.7182
LAP	5.9941	13.4604	31.3165	0.5274	0.4094	3.0149	0.7559	0.6159	0.8189	0.8061
DT-DWT	6.1979	14.1960	32.4720	0.5312	0.4004	3.1243	0.7907	0.6646	0.8446	0.8495
WTSR	6.4656	14.9538	32.6962	0.5245	0.3940	2.9996	0.7798	0.6839	0.8360	0.7989
FFIF	4.0416	7.9391	32.9437	0.3970	0.5437	**3.9300**	0.6867	0.2740	0.7254	0.7195
DCHWT	6.4554	15.1539	32.6639	0.5396	0.3915	3.7070	0.8050	0.7066	0.8480	0.8215
MGIVF	5.3781	11.0228	33.9461	0.4463	0.4664	3.3119	0.7355	0.3987	0.7920	0.9164
ASR	6.0758	14.5420	31.9445	0.5475	0.3976	3.5262	0.7925	0.6993	**0.8503**	0.7952
CSR	2.6562	5.1329	33.6403	0.2594	0.7091	3.2618	0.5586	0.1518	0.6229	0.4763
SWT	5.7129	12.9457	30.4255	0.4952	0.4513	3.1246	0.7034	0.5685	0.7829	0.7127
PNN	5.6006	13.4350	28.3551	0.3682	0.5259	2.7152	0.4565	0.3962	0.5825	0.7635
PNN+	4.6362	11.8754	24.3852	0.2671	0.6467	2.5564	0.3100	0.2530	0.4306	0.5549
Proposed	**7.2058**	**16.3861**	**34.9848**	**0.5497**	**0.3145**	3.7642	**0.8108**	**0.7091**	0.8477	**0.9167**

The bold values represent the value of the optimal image quality indicator.

TABLE VI
EVALUATION INDICES CORRESPONDING TO THIRD GROUP OF FUSION RESULTS

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
PCA	2.7501	6.8151	17.9350	0.4339	0.5286	2.7355	0.4249	0.4603	0.7239	0.6601
GRA	3.1634	7.9065	19.4819	0.4285	0.5205	2.5882	0.3886	0.4901	0.6470	0.5342
DWT	4.1486	9.9700	22.6043	0.4415	0.3952	2.6164	0.4589	0.5546	0.6773	0.8548
LAP	4.0500	9.8943	22.2986	0.4941	0.3740	2.4508	0.4629	0.6268	0.7526	0.9758
DT-DWT	4.1362	9.9278	24.3471	0.4988	0.3590	2.1410	0.4712	0.6147	0.7653	1.0467
WTSR	4.0917	10.1460	22.6016	0.4810	0.3702	2.4202	0.4375	0.6282	0.7555	0.8590
FFIF	3.2695	7.2018	23.9361	0.4212	0.4712	1.9086	0.3875	0.4095	0.7277	0.7663
DCHWT	3.9752	9.6391	19.7532	0.5055	0.3971	2.4856	0.4534	0.5893	0.7907	0.7814
MGIVF	4.1006	9.5621	24.2550	0.4547	0.3650	2.5496	**0.4756**	0.5651	0.7542	1.3207
ASR	3.8540	10.3048	19.9275	0.5105	0.3681	2.7075	0.4289	**0.6625**	**0.8038**	0.9256
CSR	1.7818	4.1354	22.4037	0.2786	0.6429	2.6750	0.3458	0.2791	0.6115	0.6624
SWT	3.9473	9.4274	22.3227	0.4703	0.4017	2.6091	0.4467	0.5510	0.6968	0.8396
PNN	3.9358	9.1407	21.2425	0.3714	0.4387	2.3667	0.3608	0.3722	0.5615	0.8557
PNN+	2.6888	5.9427	24.9055	0.3030	0.5696	2.9820	0.3430	0.2088	0.4642	0.8255
Proposed	**4.4579**	**10.9790**	**25.3109**	**0.5177**	**0.3155**	**2.9162**	0.4734	0.5956	0.7987	**1.0107**

The bold values represent the value of the optimal image quality indicator.

index (Q_e), weighted fusion quality index (Q_w), and structural similarity (SSIM) [47]−[50]. Tables IV–IX show the evaluation index values of images fused by different methods; these tables correspond to Figs. 16–21, respectively. Table X shows the average values of Tables IV–IX.

In Tables IV–IX, most of the scores of our method are better than those of other methods, and only a few values are lower than those of FFIF and ASR. However, the performance of our method is better than that of FFIF and ASR in terms of visual effect; for example, the fused images of ASR in Figs. 18, 20, and 21 are unsatisfactory due to serious spectral distortion. Besides, although some values of MGIVF are higher than those of the proposed method, the visual effect of the proposed method is significantly better than that of MGIVF. In addition, Table VIII combined with Fig. 20 show that the performance of DCHWT is slightly higher than that of the proposed method in terms of AG because of its higher definition. Moreover, although one index of LAP, PNN, and PNN+ is higher than that of the proposed method, the fused images of the proposed method are still competitive.

TABLE VII
EVALUATION INDICES CORRESPONDING TO FOURTH GROUP OF FUSION RESULTS

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
PCA	5.8874	12.9784	35.5269	0.4280	0.5453	3.1793	0.6666	0.4404	0.7802	0.7493
GRA	6.4454	14.6640	31.4610	0.4330	0.5464	2.7408	0.5498	0.3670	0.6555	0.4888
DWT	6.9991	15.8163	31.9116	0.4511	0.5057	2.5647	0.5484	0.4339	0.6804	0.6223
LAP	7.1469	16.0440	33.2439	0.5088	0.4593	2.6804	0.6102	0.4809	0.7325	0.6953
DT-DWT	6.8131	15.5958	32.4373	0.4949	0.4820	2.6909	0.5770	0.4588	0.7132	0.6422
WTSR	7.6237	17.4475	34.0228	0.5200	0.4409	2.6093	0.6398	0.5394	0.7526	0.6863
FFIF	5.9831	12.1097	**37.5514**	0.4250	0.5131	**3.3772**	0.6839	0.3513	0.7819	0.7539
DCHWT	8.0760	18.3712	35.3492	0.5463	0.4070	3.2811	0.7427	0.6401	0.8291	0.7745
MGIVF	6.5163	13.7678	35.9290	0.4381	0.5106	2.9042	0.6680	0.3923	0.7707	0.7462
ASR	7.8712	18.3705	35.2710	**0.5654**	0.3973	3.1168	0.7298	0.6439	0.8188	0.7698
CSR	2.9516	5.8809	33.6917	0.2120	0.7684	3.0167	0.4620	0.1184	0.5480	0.3732
SWT	6.8357	15.3715	32.7533	0.4753	0.4939	2.8011	0.5714	0.4361	0.7077	0.6470
PNN	5.7659	13.8101	27.4741	0.3371	0.6139	2.3398	0.4294	0.3315	0.5843	0.5477
PNN+	5.2760	13.2029	26.1909	0.2688	0.6753	2.4416	0.3318	0.2642	0.4996	0.4802
Proposed	**8.5325**	**19.3396**	37.3715	0.5491	**0.3703**	3.2786	**0.7641**	**0.6708**	**0.8420**	**0.8135**

The bold values represent the value of the optimal image quality indicator.

TABLE VIII
EVALUATION INDICES CORRESPONDING TO FIFTH GROUP OF FUSION RESULTS

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
PCA	5.3836	10.4974	27.8094	0.3928	0.5791	2.4432	0.6022	0.3813	0.7004	0.6919
GRA	6.6097	12.8444	27.1721	0.4428	0.5184	2.0387	0.5622	0.3545	0.6366	0.5268
DWT	7.3283	14.1279	27.8781	0.4496	0.4705	1.9418	0.5704	0.3798	0.6495	0.6779
LAP	7.8481	15.1282	29.8343	0.5370	0.3860	2.0106	0.6657	0.4900	0.7444	**0.8385**
DT-DWT	7.3707	14.3068	30.0255	0.5217	0.4196	2.1230	0.6567	0.4621	0.7424	0.8021
WTSR	7.8882	15.3420	29.6805	0.5340	0.3902	1.9627	0.6637	0.5014	0.7452	0.7460
FFIF	7.2230	13.6093	**31.0164**	0.5099	0.4105	2.5709	0.7446	0.5620	0.8052	0.8168
DCHWT	**8.2215**	16.0046	30.0880	0.5560	0.3635	2.6069	0.7556	0.6456	0.8194	0.8259
MGIVF	6.5068	12.5480	29.3629	0.4606	0.4828	2.3589	0.6763	0.4403	0.7562	0.7725
ASR	8.1726	16.0832	29.4755	**0.5755**	0.3392	2.4342	0.7468	0.6394	0.8154	0.8213
CSR	2.3024	4.5172	24.6620	0.1660	0.8220	2.1795	0.3489	0.0968	0.4466	0.3276
SWT	7.3599	14.0910	28.9516	0.4895	0.4416	2.0895	0.6018	0.4149	0.6850	0.7379
PNN	6.3065	12.9976	27.2696	0.3317	0.5617	1.8933	0.4241	0.2405	0.5292	0.7699
PNN+	5.0220	11.1664	25.6835	0.2365	0.6720	2.0634	0.3164	0.1436	0.4375	0.6676
Proposed	8.1935	**16.1419**	30.4905	0.5625	**0.3471**	**2.6550**	**0.7665**	**0.6677**	**0.8273**	0.8292

The bold values represent the value of the optimal image quality indicator.

TABLE IX
EVALUATION INDICES CORRESPONDING TO THE SIXTH GROUP OF FUSION RESULTS

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
PCA	5.5489	11.4799	33.1795	0.4544	0.5110	3.1026	0.6822	0.4744	0.7667	0.7644
GRA	6.2096	13.0224	28.0879	0.4548	0.5122	2.4905	0.5894	0.3865	0.6370	0.4976
DWT	6.7911	13.9445	28.3166	0.4464	0.4804	2.2833	0.5857	0.3963	0.6338	0.5861
LAP	7.1682	14.8160	31.0551	0.5359	0.4017	2.4249	0.6895	0.5205	0.7502	0.7678
DT-DWT	7.2048	14.9332	33.8476	0.5446	0.3923	2.5361	0.7137	0.5638	0.7939	0.8223
WTSR	7.4862	15.5295	31.5288	0.5378	0.3874	2.3981	0.6984	0.5487	0.7649	0.7287
FFIF	5.4899	11.2606	33.9537	0.4437	0.5019	3.0963	0.6803	0.4208	0.7694	0.7416
DCHWT	7.6570	16.0654	35.2375	0.5621	0.3697	3.3182	0.7589	0.6428	0.8378	0.8348
MGIVF	6.3830	12.9873	36.0012	0.4724	0.4547	2.8170	0.7115	0.4942	0.7973	0.8958
ASR	7.4544	**16.1505**	33.0260	**0.5753**	0.3664	3.0095	0.7355	0.6358	0.8158	0.7894
CSR	2.4370	4.9526	29.0906	0.2098	0.7782	2.6474	0.4245	0.1432	0.5217	0.3662
SWT	6.7953	13.8236	29.5055	0.4903	0.4495	2.4842	0.6233	0.4362	0.6835	0.6453
PNN	6.2212	13.4171	27.7614	0.3480	0.5510	2.1640	0.4282	0.2693	0.5057	**1.0906**
PNN+	4.4522	9.9525	22.6289	0.2170	0.7084	2.2102	0.2781	0.1237	0.3471	0.6779
Proposed	**7.7360**	15.9971	**36.3684**	0.5666	**0.3449**	**3.4192**	**0.7685**	**0.6557**	**0.8451**	0.8396

The bold values represent the value of the optimal image quality indicator.

In Table X, although the SSIM scores of MGIVF are higher than those of the proposed method, most average scores of the proposed method are higher; thus, our proposed method can fuse more information from the source images into the fused image. Besides, the metrics of each experiment are provided by six tables in the supporting document, in which the scores of the proposed method are competitive when compared with other methods.

The experimental results in Figs. 16–21 and metrics in Tables IV–IX show that the proposed method can effectively integrate the spectral information of the MS image and the details of the PAN image into the fused images. These results reveal that the proposed method can achieve good performance in both visual effect and objective indices, and, thus, it is a competitive RSIF method.

V. CONCLUSION

In this work, a new semisupervised RSIF method based on cGANs is proposed to solve the problem existing in most DL-based fusion methods that need a ground truth or reference

TABLE X
AVERAGE VALUES OF TABLES IV–IX

	AG	SF	STD	Q^{abf}	L^{abf}	MI	Q	Q_e	Q_w	SSIM
PCA	4.9713	10.6660	29.6091	0.4278	0.5410	3.0695	0.6347	0.4244	0.7541	0.7192
GRA	5.7515	12.7166	27.6920	0.4493	0.5160	2.6215	0.5706	0.4236	0.6730	0.5251
DWT	6.1877	13.4799	27.9560	0.4509	0.4770	2.5109	0.5720	0.4428	0.6750	0.6390
LAP	6.6060	14.2566	29.8226	0.5228	0.4082	2.5689	0.6529	0.5409	0.7640	0.8036
DT-DWT	6.5642	14.3012	30.9459	0.5233	0.4091	2.5804	0.6607	0.5539	0.7782	0.8251
WTSR	6.8135	14.9756	30.0406	0.5222	0.4014	2.5186	0.6524	0.5705	0.7683	0.7386
FFIF	5.1009	10.2149	32.0287	0.4283	0.5033	3.1325	0.6364	0.3667	0.7486	0.7454
DCHWT	7.0536	15.5357	30.7143	0.5445	0.3870	3.1364	0.7146	0.6401	0.8242	0.7964
MGIVF	5.8525	12.1476	32.1072	0.4499	0.4660	2.8151	0.6564	0.4350	0.7699	**0.9051**
ASR	6.9051	15.6461	30.2995	0.5590	0.3728	3.0105	0.7026	0.6567	0.8229	0.8131
CSR	2.5355	5.0879	29.2716	0.2259	0.7444	2.8120	0.4458	0.1471	0.5540	0.4409
SWT	6.2476	13.4676	28.9325	0.4860	0.4510	2.6569	0.6020	0.4746	0.7157	0.7044
PNN	5.5146	12.4843	25.6355	0.3467	0.5549	2.2952	0.4187	0.3088	0.5440	0.6737
PNN+	4.5258	10.7327	24.2780	0.2629	0.6573	2.4400	0.3217	0.2015	0.4385	0.5630
Proposed	**7.4169**	**16.3037**	**32.9842**	**0.5510**	**0.3403**	**3.2286**	**0.7284**	**0.6600**	**0.8323**	0.8730

The bold values represent the value of the optimal image quality indicator.

image (with high spatial and spectral resolution) for training. This method is an end-to-end image fusion model that only uses the PAN image and MS image as the input to generate the high-quality fused image. In addition, an encoder with a pseudo-Siamese structure is proposed by combining a multiscale module and multiskip connection; thus, we can extract the unique features of the PAN and MS images simultaneously. Therefore, the fused image can retain both the spatial details of the PAN image and the spectral information of the MS image. We also propose a compound loss function that uses the least square loss function as adversarial loss, and we combine it with the L1 loss and PSNR loss to measure the errors between the fused image and the source images. Experiments show that the proposed method can achieve satisfactory performance, and that it has obvious advantages over most existing methods.

REFERENCES

[1] M. Bikash *et al.*, "A survey on region based image fusion methods," *Inf. Fusion*, vol. 48, pp. 119–132, 2019.

[2] S. Li *et al.*, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, 2017.

[3] R. F. Beltran, J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "Remote sensing image fusion using hierarchical multimodal probabilistic latent semantic analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4982–4993, Dec. 2018.

[4] W. Zhao, D. Wang, and H. Lu, "Multi-focus image fusion with a natural enhancement via a joint multi-level deeply supervised convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 4, pp. 1102–1115, Apr. 2019.

[5] X. Jin *et al.*, "Brain medical image fusion using L2-norm-based features and fuzzy-weighted measurements in 2D Littlewood-Paley EWT domain," vol. 69, no. 8, pp. 5900–5913, 2020.

[6] C. Du *et al.*, "Multi-focus image fusion using deep support value convolutional neural network," *Optik*, vol. 176, pp. 567–578, 2019.

[7] C. Son and X. Zhang, "Near-infrared fusion via color regularization for haze and color distortion removals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3111–3126, Nov. 2018.

[8] J. Ma *et al.*, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, 2019.

[9] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Pyramid fully convolutional network for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 5, pp. 1549–1558, May 2019.

[10] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 348–362, Jan. 2020.

[11] Z. Wu *et al.*, "A new variational approach based on proximal deep injection and gradient intensity similarity for spatio spectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6277–6290, Oct. 2020.

[12] Y. Yang, W. Wan, S. Huang, F. Yuan, S. Yang, and Y. Que, "Remote sensing image fusion based on adaptive IHS and multiscale guided filter," *IEEE Access*, vol. 4, pp. 4573–4582, 2016.

[13] C. Han, H. Zhang, C. Gao, C. Jiang, N. Sang, and L. Zhang, "A remote sensing image fusion method based on the analysis sparse model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 439–453, Jan. 2016.

[14] C. Deng *et al.*, "An improved remote sensing image fusion algorithm based on IHS transformation," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 3, pp. 1633–1649, 2017.

[15] S. Xu and Z. Li, "Remote sensing image fusion using intensity-hue-saturation transform and steerable pyramid transform," in *Proc. Chin. Autom. Congr.*, 2017, pp. 4959–4962.

[16] M. Daily *et al.*, "Geologic interpretation from composited radar and Landsat imagery," *Photogramm. Eng. Remote Sens.*, vol. 45, no. 8, pp. 1109–1116, 1979.

[17] Y. Liu *et al.*, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, 2018.

[18] Y. Liu and Z. Wang, "A practical pan-sharpening method with wavelet transform and sparse representation," in *Proc. 10th IEEE Int. Conf. Imag. Syst. Techn.*, 2013, pp. 288–293.

[19] 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. 1–14, May 2018.

[20] P. Dai *et al.*, "A remote sensing spatiotemporal fusion model of Landsat and MODIS data via deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 7030–7033.

[21] X. Liu, Y. Wang, and Q. Liu, "Remote sensing image fusion based on two-stream fusion network," in *Proc. Int. Conf. Multimedia Model.*, vol. 10704, 2019, pp. 428–439.

[22] Z. Wang *et al.*, "Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform," *Expert Syst. Appl.*, vol. 171, 2021, Art. no. 114574.

[23] H. Li *et al.*, "Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural network," *Pattern Recognit. Lett.*, vol. 141, pp. 45–53, 2021.

[24] S. Nirmalraj and G. Nagarajan, "Fusion of visible and infrared image via compressive sensing using convolutional sparse representation," *ICT Express*, 2020, to be published, doi: 10.1016/j.icte.2020.11.006.

[25] C. Xing *et al.*, "Using Taylor expansion and convolutional sparse representation for image fusion," *Neurocomputing*, vol. 402, pp. 437–455, 2020.

[26] J. Ma *et al.*, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, 2020.

[27] H. Zhang *et al.*, "MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion, *Inf. Fusion*, vol. 66, pp. 40−53, 2021.

[28] S. Huang *et al.*, "Semi-supervised remote sensing image fusion method combining Siamese structure with generative adversarial networks," *J. Comput.-Aided Des. Comput. Graph.*, vol. 33, no. 1, pp. 92–105, 2021.

[29] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. 28th Conf. Neural Inf. Process. Syst.*, Montréal, Canada: Palais des Congrès de Montréal, Dec. 2014, pp. 2672–2680.

[30] M. Mehdi and O. Simon, "Conditional generative adversarial nets," 2014. [Online]. Available: arxiv.org/abs/1411.1784

[31] X. Mao *et al.*, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2813–2821.

[32] Z. Shao, Z. Lu, M. Ran, L. Fang, J. Zhou, and Y. Zhang, "Residual encoder-decoder conditional generative adversarial network for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1573–1577, Sep. 2020.

[33] U. Tatsumi, Y. Naoto, and W. He, "Guided deep decoder: unsupervised image pair fusion," *ECCV*, vol. 12351, pp. 87–102, 2020.

[34] S. Lohit *et al.*, "Unrolled projected gradient descent for multi-spectral image fusion," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 7725–7729.

[35] F. Ye, X. Li, and X. Zhang. "FusionCNN: A remote sensing image fusion algorithm based on deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 14683–14703, 2019.

[36] X. Liu, Y. Wang, and Q. Liu, "PSGAN: A generative adversarial network for remote sensing image pan-sharpening," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 873–877.

[37] R. Oliver, "Pixel-level image fusion and the image fusion toolbox," Sep. 30, 1999. [Online]. Available: http://www.metapix/toolbox.htm<./bib>

[38] D. Renza, E. Martinez, and A. Arquero. "Quality assessment by region in spot images fused by means dual-tree complex wavelet transform," *Adv. Space Res.*, vol. 48, no. 8, pp. 1377–1391, 2011.

[39] K. Zhan *et al.*, "Fast filtering image fusion," *J. Electron. Imag.*, vol. 26, no. 6, 2017, Art. no. 063004.

[40] B. Shreyamsha, "Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform," *Signal, Image Video Process.*, vol. 7, no. 6, pp. 1125–1143, 2013.

[41] D. Bavirisetti *et al.*, "Multi-scale guided image and video fusion: A fast and efficient approach," *Circuits, Syst., Signal Process.*, vol. 38, no. 12, pp. 5576–5605, 2019.

[42] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Process.*, vol. 9, no. 5, pp. 347–357, 2014.

[43] Y. Liu *et al.*, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[44] K. Liu, L. Guo, and H. Li, "Image fusion algorithm using stationary wavelet transform," *Comput. Eng. Appl.*, vol. 43, no. 12, pp. 59–61, 2017.

[45] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.

[46] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[47] C. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

[48] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, 2002, Art. no. 313.

[49] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. Int. Conf. Image Process.*, 2003, pp. 173–176.

[50] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

Xin Jin (Member, IEEE) received the B.S. degree in electronics and information engineering from the Henan Normal University, Xinxiang, China, in 2013, and the Ph.D. degree in communication and information systems from Yunnan University, Kunming, China, in 2018.

He is an Associate Professor with the School of Software, Yunnan University. His current research interests include pulse coupled neural networks theory and its applications, image processing, optimization algorithm, and bio-informatics.

Shanshan Huang received the B.S. degree in software engineering from Nantong University, Nantong, China, in 2018, and the M.S. degree in software engineering from the School of Software, Yunnan University, Kunming, China, in 2021. She is currently working toward the Ph.D. degree in software engineering at the School of Big Data & Software Engineering, Chongqing University, Chongqing, China.

Her research interests include deep learning, computer vision, and image processing.

Qian Jiang received the B.S. degree in thermal energy and power engineering and the M.S. degree in power engineering and engineering thermo-physics from Central South University (CSU), Changsha, China, in 2012 and 2015, respectively.

She is a Lecturer with the School of Information, Yunnan University, Kunming, China. Her research interests include machine learning, fuzzy set theory, bio-informatics, and image processing.

Shin-Jye Lee received the M.Sc. degree in engineering from the Department of Computer Science, University of Sheffield, Sheffield, U.K., in 2001, the M.Phil. degree in science and technology policy from the Judge Business School, University of Cambridge, Cambridge, U.K., in 2012, and the Ph.D. degree in computer science from the School of Computer Science, University of Manchester, Manchester, U.K., in 2011.

He is currently an Associate Professor with the National Chiao Tung University, Hsinchu, Taiwan, and he also made his academic career in Poland and Taiwan successively. In addition, he also had practical experiences in Fujitsu and Microsoft, from 2002 to 2005. Further, his research interests primarily comprise machine learning, computational intelligence and decision support system, operational research, and technology policy, especially for the climate change issues and energy prediction.

Liwen Wu received the B.Eng. degree in network engineering from the National Pilot School of Software, Yunnan University, Kunming, China, in 2016. She is currently working toward the Ph.D. degree in communication and information systems at the School of Information Science and Engineering, Yunnan University. In the Ph.D. study period, she has been committed to the prediction of protein function and participated in many scientific research projects.

Her research interests include bioinformatics, neural network theory and applications, and privacy protection of machine learning.

Shaowen Yao (Member, IEEE) received the B.S. and M.S. degrees in telecommunication engineering from the Yunnan University, Kunming, China, in 1988 and 1991, respectively, and the Ph.D. degree in computer application technology from University of Electronic Science and Technology of China, Chengdu, China, in 2002.

He is a Professor with the School of Software, Yunnan University. His current research interests include neural network theory and applications, cloud computing, and big data.