A Restoration Scheme for Spatial and Spectral Resolution of the Panchromatic Image Using the Convolutional Neural Network

Xin Jin, Ling Liu, Xiaoxuan Ren, Qian Jiang, Shin-Jye Lee, Jun Zhang, and Shaowen Yao

***Abstract*—Remote sensing images are the product of information obtained by various sensors, and the higher the resolution of the image, the more information it contains. Therefore, improving the resolution of the remote sensing image is conducive to identify Earth resources from the remote sensing image. In this article, we present a multiple-branch panchromatic image resolution restoration network based on the convolutional neural network to improve the spatial and spectral resolution of the panchromatic image simultaneously, named MBPRR-Net. Specifically, we adopt a multibranch structure to extract abundant features and utilize a feature channel mixing block to enhance the interaction of adjacent channels between features. Feature aggregation in our method is used to learn more effective features from each branch, and then a cubic filter is utilized to enhance the aggregated features. After feature extraction, we use a recovery architecture to generate the final image. Moreover, we utilize image super-resolution to restore spatial resolution and image colorization to restore the spectral resolution so that we can compare it with some image colorization and super-resolution methods to verify the proposed method. Experiments show that the performance of our method is outstanding in terms of visual effects and objective evaluation metrics compared with some existing excellent image super-resolution and colorization methods.**

***Index Terms*—Artificial neural network, deep learning, multispectral (MS) image, panchromatic (PAN) image, remote sensing image processing.**

Manuscript received 5 October 2023; revised 16 December 2023; accepted 2 January 2024. Date of publication 9 January 2024; date of current version 23 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62101481 and Grant 62261060; in part by Basic Research Project of Yunnan Province under Grant 202201AT070112, Grant 202301AW070007, Grant 202201AU070033, and Grant 202301AU070210; in part by Major Scientific and Technological Project of Yunnan Province under Grant 202202AD080002; and in part byYunnan Province Expert Workstations under Grant 202305AF150078. *(Xin Jin and Ling Liu contributed equally to this work.) (Corresponding author: Qian Jiang.)*

Xin Jin, Ling Liu, Xiaoxuan Ren, Qian Jiang, and Shaowen Yao are with the Engineering Research Center of Cyberspace, Yunnan University, Kunming 650000, China, and also with the School of Software, Yunnan University, Kunming 650000, China (e-mail: xinxin jin@163.com; liuling@mail.ynu.edu.cn; 17221140@bjtu.edu.cn; jiangqian_1221@163.com; yaosw@ynu.edu.cn).

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

Jun Zhang is with the China Mobile Communications Group, Yunnan Company Ltd., Kunming 650000, China (e-mail: 18214590842@139.com).

The code is available at https://github.com/jinxinhuo/MBPRR-Net/

Digital Object Identifier 10.1109/JSTARS.2024.3351854

I. Introduction

REMOTE sensing images play an important role in environmental monitoring [1], Earth resource census [2], agricultural production [3], disaster prevention and mitigation [4], and urban planning and construction [5]. Some satellites, such as Quickbird and WorldView-II, III, IV, are equipped with panchromatic (PAN) cameras and multispectral (MS) cameras. PAN cameras can shoot PAN images. The band of PAN image refers to all visible light bands 0.38–0.76 μm, and the PAN image is the mixed image with all these bands. The PAN image has a high spatial resolution, but because it is just a mixed-band image displayed as gray level, the spectral resolution is low. MS images taken by MS cameras can make up for the deficiency of PAN images in spectral information. Compared with the PAN image, the MS image has higher spectral resolution, but because of the limitation of physical devices, the MS image has lower spatial resolution. The improvement of remote sensing images' spatial and spectral resolution is conducive to interpret the remote sensing scenes, so as to better play the role of remote sensing images in Earth resource acquisition [6], [7]. *Considering the advantages of the PAN and MS images, we design a multiple-branch PAN image resolution restoration network (named MBPRR-Net) to enlarge the size of the PAN image and the number of channels, so as to improve the spatial and spectral resolution simultaneously.* An example of our idea is shown in Fig. 1(a), and Fig. 1(b) is used for visual contrast.

The goal of image super-resolution is to generate a high-spatial-resolution (HSR) image from a low-spatial-resolution (LSR) image. Thus, image super-resolution can used to restore the spatial resolution of the PAN image. It play an important role in computer vision to improve the spatial resolution of an image. Super-resolution is helpful for image compression, for example, we can use the small image to transmit or save, and use the large image to view, which can reduce the amount of data transferred and saved. In recent years, deep learning has received widespread attention in the field of image processing, some image super-resolution methods that are based on deep learning [8], [9], [10], [11], [12], [13] have made great progress. Among many methods, SRCNN model [8] first introduces deep learning into image super-resolution, which uses CNNs to achieve a transition between LSR and HSR images. Dong et al. [9] presented FSRCNN to enhance SRCNN, accelerate the inference speed of model, and improve the quality of the output

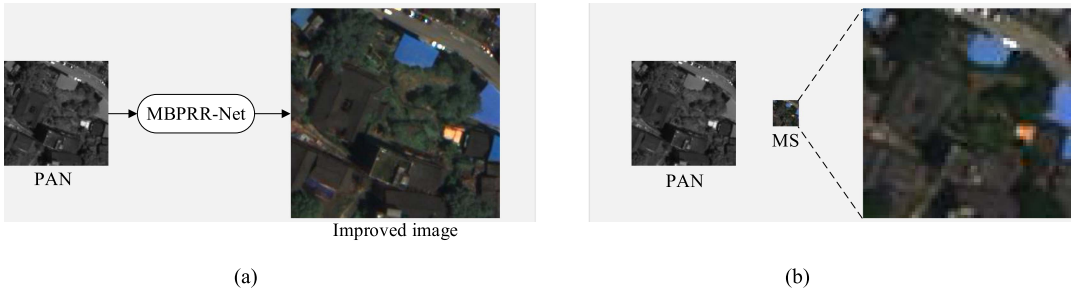

Fig. 1. Example of our idea for remote sensing image restoration. (a) An example of our idea. (b) A pair of MS and PAN images for visual contrast.

images. Recently, there are many other methods optimize the performance of image super-resolution from the perspective of the model. For example, Zhang et al. [10] introduced a channel attention mechanism. Jiang et al. [11] added the dense module to the network to generate multiscale features, and then, fused the features and reconstructed HSR images using pixel-shuffle operations. Similarly, Mei et al. [12] improve the super-resolution effect by designing a module that used cross-scale features. Using mean square error (MSE) or mean absolute error (MAE) as the loss function in image super-resolution often produce fuzzy images. To avoid generating blurred images, Ledig et al. [13] first applied adversarial loss to image super-resolution, which greatly improved the reality of reconstructed images. However, there are still some problems with the aforementioned methods, such as that it is difficult to express all the textures in the GANs, so in some places, complex textures will generate the wrong textures.

The goal of image colorization is to generate a colorized image with three channels from a gray-scale image with one channel. In general, a color image (such as MS image) has more abundant spectral or color information than that of gray image (such as PAN image). Thus, image colorization can be used to restore the spectral resolution of the PAN image. Conventional image colorization methods depend on user's guide, such as scribbles [14], [15], [16], reference images [17], [18], [19], and semantic information [20], [21]. User-guided colorization methods have limitations that require user interaction. It can improve the limitation of the user-guided colorization methods when using deep learning. The gray-scale image colorization algorithm based on convolutional neural networks (CNNs) [22], [23], [24], [25] belongs to supervised learning, and the label of image colorization training phase is color image. Generative adversarial networks (GANs) can be regarded as unsupervised learning networks, which consist of a generator and a discriminant. The parameter's update of the generator is guided by the discriminant. Therefore, GANs-based image colorization [26], [27], [28], [29] can generate colorized images. Existing studies show that deep learning plays an important role in image colorization, but there are some limitations to these methods, such as it fails to solve multimodal problems [22], relies on color distribution [23] or object detection algorithm [25], needs amounts of data to train [26], etc. Therefore, image colorization is still a challenging problem and awaits further exploration.

Some existing methods combine pansharpening to improve the spectral resolution of PAN images. For instance, Ozcelik et al. [30] borrowed the idea of pansharpening to input PAN images and MS images into a guided colorization model to improve the spectral resolution of PAN images. Lu et al. [31] refer to hypersharpening and pansharpening techniques, while reconstructing the HSR image with a two-stage cascaded CNN. However, pansharpening methods usually need to input PAN images and MS images to the model at the same time, and do not involve the improvement of spatial resolution.

In this article, we introduce a novel MBPRR-Net model to restore the spatial and spectral resolution of PAN image based on the characteristics of the PAN image. First, there are large differences between the same class and small differences between different classes in a remote sensing, that is, the spectral information between objects of the same type is very different, and the spectral information between objects of different types are very close, which will lead to large deviation between the reconstructed spectral information and the true spectral information. To solve this problem, we consider to design a network, which is powerful in feature extraction. Second, the PAN image has a complex background and many small objects, which will reduce the quality of the image when the spatial resolution is increased. We consider to use a feature enhancement method to solve this problem. Third, due to the complementarity of PAN and MS images, we consider to use the original PAN image as the brightness label to improve the spatial resolution, and utilize MS image's chroma information as the chroma label to predict the spectral information.

The main contributions of this article are as follows.

1) We propose an MBPRR-Net to reconstruct an image with both high spatial and spectral resolution, and the MBPRR-Net only needs to input PAN image.
2) We design a feature channel mixing block (FCMB) to enhance the channel correlation between features. Experiments show that the FCMB can extract effective features and facilitate the feature recovery.
3) We employ a cubic filter to enhance the extracted features to help the model generate a color image with a clear texture, and also design a self-convolution block (SCB) to recover the extracted features.
4) We train our model to predict PAN image's spectral information according to the MS image so that the spectral

resolution reconstruction result of the PAN image is consistent with the real scene.

The rest of this article is organized as follows. Section II reviews some papers about image super-resolution and image colorization methods. Section III give an account of the proposed method. Section IV discusses the performance of our method. Finally, Section V concludes this article.

II. Related Work

This section gives a review of the existing image super-resolution and colorization methods, respectively.

A. Image Super-Resolution

Image super-resolution restores clear texture from an LSR image, and finally, gets an HSR image. In recent years, thanks to the powerful learning ability of deep-learning technology, image super-resolution has made remarkable progress. LSR images are typically obtained through a series of degradation operations, which not only lose a lot of details, but also introduce a series of noise. In essence, the super-resolution process based on deep learning belongs to a supervised learning, which uses a pair of HSR and LSR images to train the network model, and then, an inverse operation of the aforementioned degradation operation is used to reconstruct the HSR image.

Dong et al. [8] first applied CNNs to the image super-resolution technology, and achieved a great improvement in the quality of generated image compared with the conventional interpolation and optimization algorithms. In particular, this method was further optimized by the FSRCNN model [9], which greatly accelerated the speed of training. Ledig et al. [13] used GANs to solve the problem of super-resolution. They proposed that it could improve the peak signal-to-noise ratio (PSNR) value when using MSE as the loss function in network training, but the details information of recovered images were lost usually. Therefore, Ledig et al. [13] used perceptual loss and adversarial loss to improve the authenticity of the recovered images. Lim et al. [32] removed unnecessary modules from the conventional resnet [33] structure and improved the super-resolution reconstruction performance. To solve the problem that the deep network is difficult to train, Zhang et al. [10] proposed a new structure named RCANs, which used a channel attention mechanism to give channel-wise different attention so as to increase the interdependence among channels. To improve the performance of super-resolution, some effective models, such as in [10], [13], and [32], are designed by increasing the complexity of the model. However, these methods still have problems with a large solution space, which leads to limit the performance of super-resolution and cannot produce satisfactory texture. To reduce solution space of mapping function between LSR and HSR, Guo et al. [34] proposed a network named DRN, which used a dual regression structure to improve the performance of the super-resolution model.

B. Image Colorization

1) User-Guided Colorization: Most of the early image colorization methods were guided, such as scribble-based colorization [14], [15], [16], example-based colorization [17], [18], [19], and text-based colorization [20], [21].

a) Scribble-based colorization: Scribble-based colorization methods need users to provide scribbles to guide the colorization process. For instance, Anagnostopoulos et al. [14] began with the user drawn abstract color indications, and then, splashed multiple colors on the sketch to get the color image, at last, refined the color image to realize colorization. Ci et al. [16] and Sangkloy et al. [15] utilized GANs in scribble-based hand-draw sketch colorization.

b) Example-based colorization: It is crucial to use appropriate reference images to realize example-based colorization methods. Reference images usually come from the user's [18], [19] or the Internet [17]. Lee et al. [19] utilized the image with identical geometric distortion as a reference image, and then, an attention mechanism is used to transfer color from the reference image to the sketch. To select the reference image efficiently, He et al. [18] proposed the deep learning approach to predict colors from an aligned reference image to a gray-scale image directly. To avoid providing reference images from the user directly, Alex et al. [17] designed a colorization system that searched for reference images on the Internet; the user was required to provide semantic text labels or semantic cues for searching for reference images on the Internet.

c) Text-based colorization: To complete image colorization with the help of textual messages, text-based colorization method uses textual messages and images as the input. Bahng et al. [20] and Kim et al. [21] captured the semantics of the input to generate color information. There is a serious problem with these methods, that is, it is difficult to recognize the text semantics.

All of these methods need numerous user's interactions, and whether or not these methods can produce a realistic color image depends on the guidance provided by the user. In brief, user-guided colorization methods have significant limitations.

2) Fully Automatic Colorization: Fully automatic colorization methods are proposed to address the limitations of user-guided colorization methods. Fully automatic colorization based on deep learning has achieved great achievement. The existing deep-learning-based colorization methods can be divided into two types: CNNs-based methods [22], [23], [24], [25] and GANs-based methods [26], [27], [28], [29].

a) CNNs-based colorization: Iizuka et al. [22] proposed a fully automatic colorization method to combine global information and local features. To address the multimodal problem, Zhang et al. [23] argued that image colorization task is a kind of classification task, and they rearranged class distribution during model training phase to increase the diversity of the color. Özbulak et al. [24] proposed a capsule network with generative and segmentation capabilities; this method is easy to cause the problem of color discontinuity between two adjacent small blocks. Su et al. [25] used an existing object detection algorithm to extract object of the input image, then used a color network

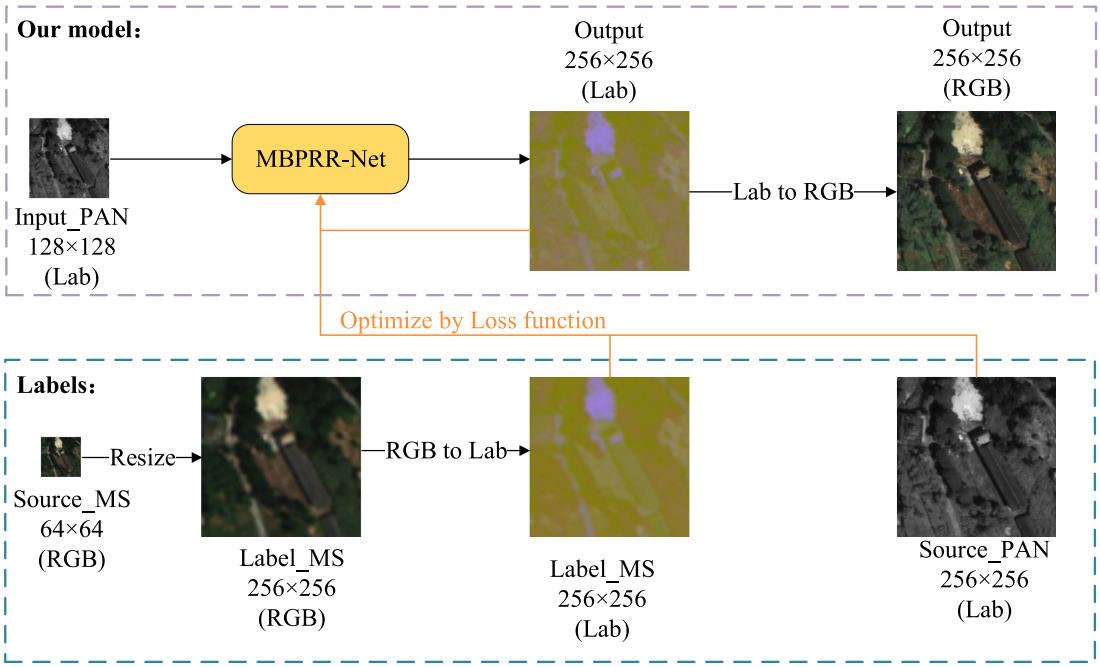

Fig. 2. Proposed framework of the PAN image spatial and spectral resolution restoration.

to color object and full-image except object, respectively, and finally, they fused object and full-image except object to generate the final colorized image.

b) GANs-based colorization: GANs-based image colorization methods mainly solve multimodal problems [28] and model training problems that require large amounts of data [27], [29]. CNNs require a lot of manual labor when designing effective loss functions, whereas GANs can automatically learn a loss function suitable for achieving the image colorization task. Isola et al. [26] designed a network to learn a mapping from the input image to the output image, at same time, they designed a loss function to optimize the mapping network to achieve image translation tasks involving the image colorization. Vitoria et al. [28] proposed an adversarial strategy to capture geometric, perceptual, and semantic information from input image, and they utilized these information to realize image colorization. To solve the problem that needs a lot of data to train, Yoo et al. [27] proposed a colorization model, which is added a memory-augmented unit to produce colorized images with a small amount of data. However, this model only focused on major color transfers and ignored minor color transfers. Since the pix2pix method is restricted by pairs of images, which may be difficult to obtain for SAR images, Ji et al. [29] used a multidomain cycle-consistency generative adversarial network for SAR image colorization, which did not use pairs of images to train the model.

III. Proposed Method

To restore the spatial and spectral resolution of the PAN image simultaneously, we introduce a new PAN image colorization and super-resolution model based on CNNs. In this method, VGG19 [35] is invoked as the backbone to extract image features. Moreover, multiple branches are applied to extract rich features. A feature aggregation module is proposed to further utilize the extracted features from various branches. Finally, we employ a feature recovery module to generate the colored and super-resolution images. Besides, we designed a loss function that is composed of $L1$ and structural similarity (SSIM) to measure the errors. The errors consists of two sets: the first set is the error between the source PAN image and the brightness channel of the generated image, and the second set is the error between the *ab* channels of the resized source MS image and the *ab* channels of the output image. In this section, we describe the details of the proposed network structure, loss function, and parameters setting.

A. Proposed Framework

The proposed PAN image spatial and spectral restoration framework is shown in Fig. 2. Source images are pairs of gray PAN images and MS images with B ($B > 3$) spectral bands. We denote the PAN image as $S_{\text{PAN}} \in \mathbb{R}^{H\times W}$ and the MS images as $S_{\text{MS}} \in \mathbb{R}^{\frac{H}{\text{scale}}\times\frac{W}{\text{scale}}\times B}$. The MBPRR-Net model mainly include three modules: feature extraction module, feature aggregation module, and feature recovery module. The input of MBPRR-Net is the PAN image. To realize image super-resolution, we resize S_{PAN} to $X_{\text{PAN}} \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}}$ as input image. The output of MBPRR-Net is the colorized HSR images in CIELab color space; we denote it as $\hat{Y}_{\text{colored\_Lab}}$. To facilitate visual analysis, the color space of model output images will be converted from CIELab into RGB; we denote the output image in RGB color space as $\hat{Y}_{\text{colored}}$.

There are three considerations during the training phase that differ from the testing phase. First, we considered using MS images as the label of chrominance information to overcome the limitation of PAN images (lack of spectral information).

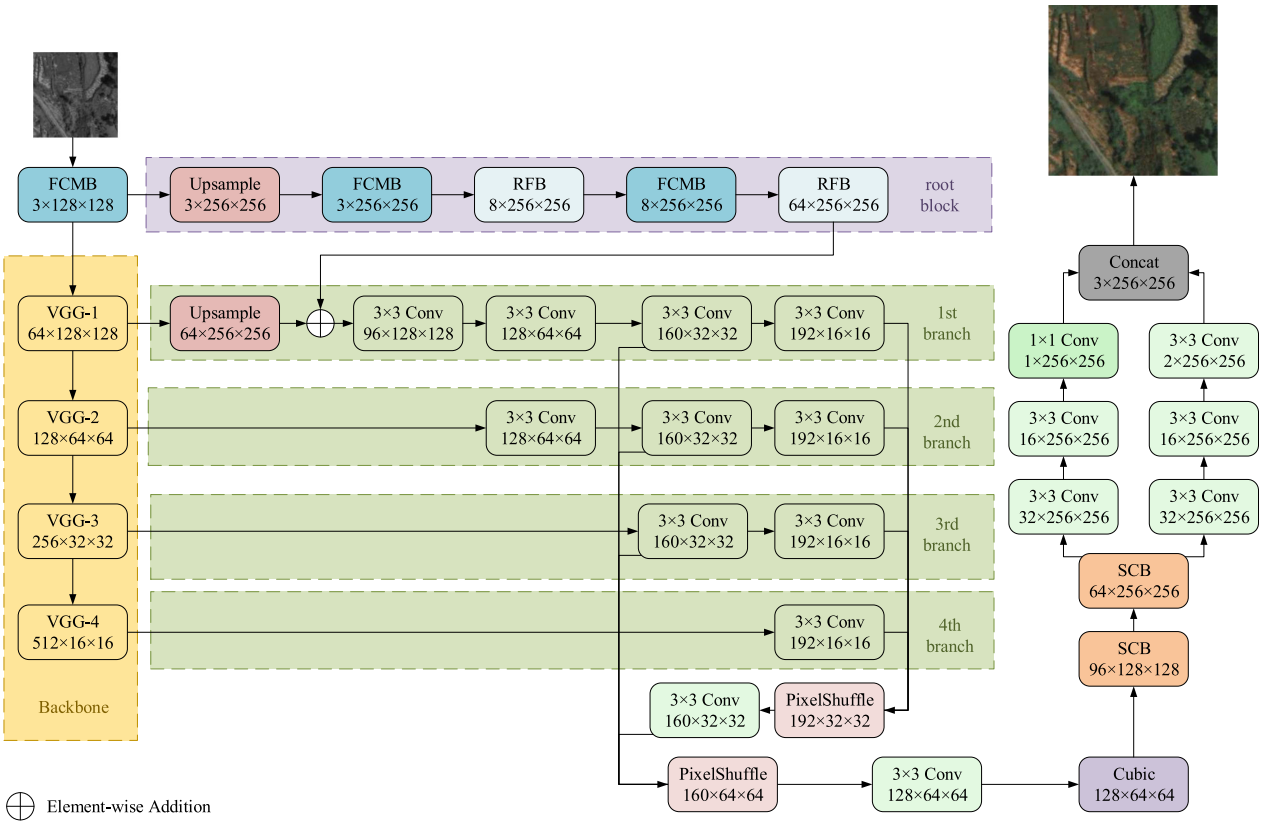

Fig. 3. Proposed MBPRR-Net model structure.

Meanwhile, to achieve the separation of luminance information and chromaticity information, we convert the colorspace of MS image from RGB to CIELab because that L channel represents luminance information and *ab* channels represent chromaticity information in CIELab color space. Second, because PAN images have richer spatial details than MS images, we utilize source PAN images as the label of luminance information. Finally, as source MS images' resolution is only a quarter of the source PAN images, the MS images must be resized to match the source PAN images. In a word, *ab* channels of resized MS images $Y_{\mathrm{MS}} \in \mathbb{R}^{H\times W\times B}$ in CIELab color space $Y_{\mathrm{MS\_Lab}} \in \mathbb{R}^{H\times W\times B}$ is the label of $\hat{Y}_{\mathrm{colored\_Lab}}$'s *ab* channels, meanwhile, S_{PAN} is the label of $\hat{Y}_{\mathrm{colored\_Lab}}$'s L channel. All parameters can be automatically updated by minimizing the loss among the output of MBPRR-Net $\hat{Y}_{\mathrm{colored\_Lab}}$ and a pair of images (S_{PAN} and $Y_{\mathrm{MS\_Lab}}$). In short, our framework is represented as:

$$\hat{Y}_{\mathrm{colored\_Lab}} = \mathcal{M}(X_{\mathrm{PAN}}) \tag{1}$$

where $\mathcal{M}(.)$ is the MBPRR-Net model.

B. Multiple-Branch PAN-Image Resolution Restoration Network (MBPRR-Net)

The architecture of our MBPRR-Net model is represented in Fig. 3. It is made up of feature extraction, feature aggregation, and feature recovery. The overall network structure is inspired by the U-shape neural network. The Feature extraction module and feature aggregation module can be regarded as encoders, and feature recovery module can be seen as a decoder. First, we design a module to extract the features of the PAN image. Then, we utilize a operation named pixel-shuffle to aggregate these extracted features. At last, we design a decoder to generate a new colorized HSR image that improve the spatial resolution of the PAN image and learn the rich spectral information of the MS image.

1) Feature Extraction Module: Inspired by [36], the feature extraction module uses VGG19 as a backbone block, and we only use the first 12 layers before the max-pooling layers and remove the last seven layers of VGG19. In addition, there are four branches to get rich features. We extract features in different scales before each max-pooling layers of VGG19 and input to four branches, respectively. The choice of the number of branches will be discussed later in the ablation study. In order to realize image super-resolution, an FCMB block is used at the beginning of this model, up-sample operations are used at the beginning of root branch and the first branch. Inspired by [37], we add receptive field block (RFB) in the root block to expand the receptive field and improve feature extraction capability of the proposed network. The RFB uses four branches to extract features of different scales. In each branch, 1×1 convolution is used to change the number of channels to reduce the number of parameters. At the same time, a residual structure is used in each branch to prevent vanishing gradient problem. Fig. 4 has shown the RFB structure. The FCMB improves the interdependence between image channels by mixing features of different channels and adding attention between adjacent channels. Our FCMB

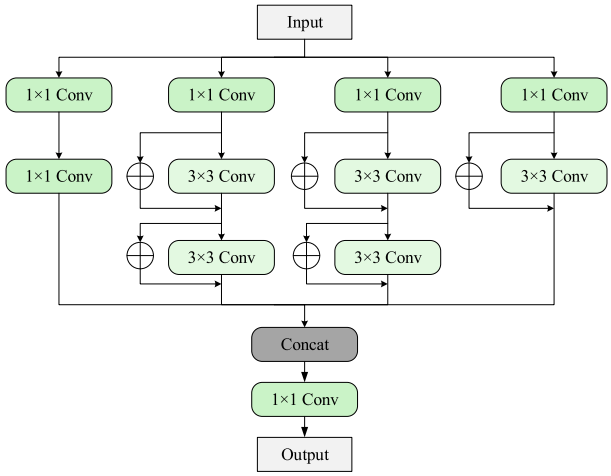

Fig. 4. RFB structure detail.

diagram is displayed in Fig. 5. First, the input image is divided into *C* feature maps (*C* represents input image's channel number), these feature maps are denoted as u_i. Second, we split u_i in half to get u_{i1} and u_{i2}. Third, we use a dual-involution (DInv) block inspired by [38] to enhance the correlation between two adjacent channels, and generate a mixing feature map denoted as v_i. Fourth, we use a concat operation to combine all v_i features. Fifth, the combined features go through a convolution block, a dual-squeeze-and-excitation (DSE) block inspired by [39] to get the output. We formulate the FCMB as

$$v_i = \begin{cases} \mathcal{K}(I_{(i,1)}, I_{(i+1,2)}), & i \neq c \\ \mathcal{K}(I_{(1,2)}, I_{(i,1)}), & i = c \end{cases} \tag{2}$$

$$\text{out}_{(p,q)} = \mathcal{S}\left(\Phi(\mathcal{C}(v_1, v_2, \ldots, v_c), W_{(p,q)}), X_{(p,q)}\right) \tag{3}$$

where $I_{(i,j)}$ represents the *j*th part of the *i*th channel of the input feature map, the value of *i* is from 1 to *c* (*c* is the input feature's channel number), and the value of *j* is 1 or 2. $\text{out}_{(p,q)}$ is the value at point (p, q) on the output feature map of the FCMB, $\mathcal{K}(.)$ represents a dual-involution block, $W_{(p,q)}$ represents the convolution weight of pixel (p, q), $X_{(p,q)}$ denotes the value at point (p, q) on the input feature map of FCMB, $\mathcal{C}(.)$ is concat operation, $\Phi(.)$ is convolution operation, and $\mathcal{S}(.)$ is the DSE block. Each branch finally gets a feature map of the same size to facilitate later feature aggregation.

2) Feature Aggregation Module: The proposed feature aggregation module consists of two pixel-shuffle blocks and a cubic filter, which is inspired by [40]. First, we concatenate the last and penultimate features of each branch of the feature extraction module, respectively. Second, these two concatenated features through a pixel-shuffle operation become two aggregated features, and then, a convolution operation is used to produce a feature map with a specific number of channels. Finally, a cubic filter is utilized to improve the global connectivity of aggregated features. The graphical representation of cubic filter is shown in Fig. 6. We formulate the cubic filter as

$$O(x, y) = C(I(x, y), Q(x, y)) \odot P(X, Y) \oplus I(x, y) \tag{4}$$

where $\odot$ is a matrix dot product, and $\oplus$ is a matrix addition. $O(x, y)$ is the value at pixel (x, y) on the output feature of the cubic filter, and $I(x, y)$ is the value at pixel (x, y) on the input feature of the cubic filter. $Q(x, y)$ denotes the weight of the filter for pixel (x, y). $C(.)$ is the Conv-Pool-FullConnection module, which including three part: the first part consists of three layers of alternating convolution and maxpooling operations, and followed by a convolutional layer to adjust the number of channels; the second part is a global average pooling layer to expand the feature into a vector; and the final part is a full-connection (FC) layer to adjust the vector size. $P(X, Y)$ is a polynomial from [40]. X is an $H \times H$ Vandermonde matrix, and Y is a transpose of $W \times W$ Vandermonde matrix. Since the output of FC is a vector but the output of $P(X, Y)$ is a matrix, we need to reconstruct the dimensions of FC's output to ensure that the output of FC and the output of P can carry out $\odot$ operation.

3) Feature Recovery Module: The feature recovery module consists of two SCBs and six convolution layers, which are used to generate the final colorized HSR image. First, we use an SCB to adjust the size of the feature maps. The SCB consists of two standard convolution layers, an involution [38] layer, and a pixel-shuffle layer. The SCB utilizes the involution layer to reduce kernel redundancy on the channel. The standard convolution kernels have a remarkable property, that is, the convolution kernel is shared in the spatial dimension but independent in the channel dimension. Involution has the characteristic of symmetric inversion compared with standard convolution. In the feature recovery stage, enlarging the size of the feature map is a key step, we utilize a pixel-shuffle layer in the SCB to enlarge the size of feature map. Then, we divide six convolution layers into two groups to generate the *L* channel and *ab* channels of output image, respectively. Finally, we employ a concate operation to merge the *L* channel and *ab* channels and generate the image with high spatial and spectral resolution.

C. Loss Function

In this work, we adopt the $\mathcal{L}_1$ and SSIM as the final loss function, and it is formulated as follows:

$$\mathcal{L} = \alpha\mathcal{L}_1\left(\hat{Y}_{\text{colored\_}L}, S_{\text{PAN}}\right) + \beta\mathcal{L}_1\left(\hat{Y}_{\text{colored\_}ab}, Y_{\text{MS\_}ab}\right) + \eta\text{SSIM}\left(\hat{Y}_{\text{colored\_Lab}}, \mathcal{C}(S_{\text{PAN}}, Y_{\text{MS\_}ab})\right) \tag{5}$$

where $\mathcal{L}_1(.)$ represents the $\mathcal{L}_1$ loss and its formula refers to (6), and SSIM(.) is the SSIM loss and its formula refers to (7). $\mathcal{C}(.)$ is a concat operation, we use $\mathcal{C}(.)$ to concatenate the original PAN image and the *ab* channels of $Y_{\text{MS\_}ab}$ to get a label image denoted as $Y_{\text{label\_Lab}}$ with three channels in the CIELab color space. $\hat{Y}_{\text{colored\_}L}$ and $\hat{Y}_{\text{colored\_}ab}$ are the *L* channel and *ab* channels of our MBPRR-Net model's output $\hat{Y}_{\text{colored\_Lab}}$, respectively, $Y_{\text{MS\_}ab}$ is *ab* channels of the resized MS image. In this work, we calculate three-part loss, the first part is $\mathcal{L}_1$ loss between $\hat{Y}_{\text{colored\_}L}$ and S_{PAN}, the second part is $\mathcal{L}_1$ loss between $\hat{Y}_{\text{colored\_}ab}$ and $Y_{\text{MS\_}ab}$, and the third part is SSIM loss between $\hat{Y}_{\text{colored\_Lab}}$ and $Y_{\text{label\_Lab}}$. α, β, and η are the weights of the aforementioned three-part loss,

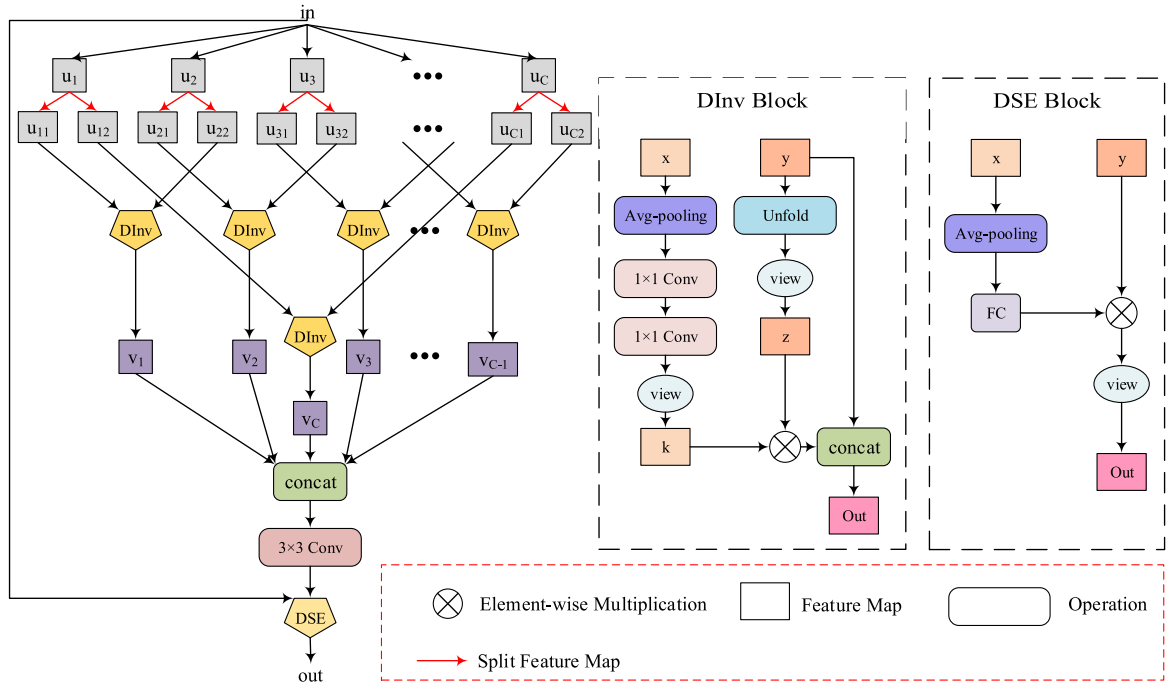

Fig. 5. Structure of the FCMB.

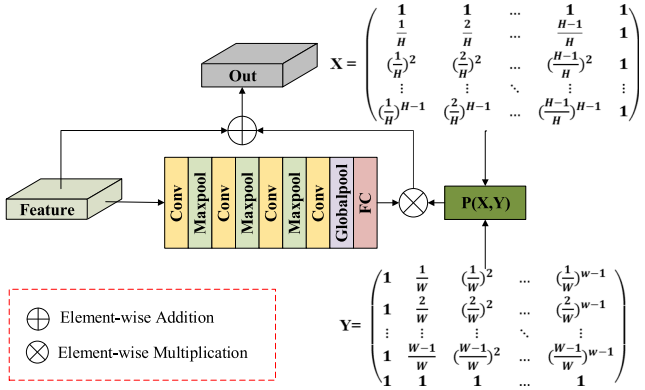

Fig. 6. Diagram of the cubic filter.

we set α, β, and η as 1.

$$\mathcal{L}_1(x_1, x_2) = \sum_{i=1}^{n} |{x_1}^{(i)} - {x_2}^{(i)}| \tag{6}$$

$$\mathrm{SSIM}(p, q) = [l(p, q)]^{\rho}[c(p, q)]^{\tau}[s(p, q)]^{\varphi} \tag{7}$$

where SSIM is composed of three contrast modules: luminance contrast function $l(p, q)$, contrast contrast function $c(p, q)$, and structural contrast function $s(p, q)$.

D. Parameters Setting

We present the key parameters setting of our proposed MBPRR-Net, including hyperparameters setting, model training, and testing details.

TABLE I
MODEL TRAINING PARAMETERS SETTING

Parameters	Name		Value
Input	degraded PAN image		128×128
Batchsize	-		6
Epoch	-		100
Learning rate	-		1×10^{-4}
Learning schedule	StepLR	adjust epoch	10
		reduction rate	0.8

TABLE II
QUANTITATIVE RESULTS OF THE PAN IMAGE SUPER-RESOLUTION TASK

Method	PSNR (↑)	CC (↑)	UIQI (↑)	SAM (↓)
SRCNN [8]	37.02	0.983	0.982	0.049
SRGAN [13]	36.74	0.982	0.981	0.051
EDSR [32]	37.85	0.984	0.983	0.044
LapSR [42]	36.14	0.976	0.975	0.055
DRN [34]	30.30	0.923	0.899	0.111
MHAN [43]	38.27	0.984	0.984	0.044
Ours	38.34	0.985	0.984	0.043

Red text marks the best performance.

All convolution layers use a convolution kernel of size 3×3 except the last convolution of the SCB is 1×1. We set the convolution stride of the feature extraction module as 2, and the rest of the modules as 1. We use a zero-padding strategy in all 3×3 convolution layers to keep the image size fixed. All activation functions in our module are using Mish function [41], the formula for the Mish function refers to (8).

$$\mathrm{Mish}(x) = x \cdot \tanh(ln(1 + e^x)). \tag{8}$$

All model training parameters in Table I remained the same in all experiments. We initialize the layers of backbone with the weights of the pretrained VGG19 on ImageNet and randomly initialize the rest layers.

For model testing, we use three test strategies. The first one is that we input source PAN image into our trained MBPRR-Net directly, and the model output image's spatial resolution is 512 × 512. The second one is that we degrade the source PAN images to 128 × 128 as input, and the model output image's spatial resolution is 256 × 256. The last one is that we clip the source PAN image to 128 × 128 as input, and the model output image's spatial resolution is 256 × 256. These strategies are used to demonstrate that our MBPRR-Net model can improve arbitrary-size PAN images' spatial and spectral resolution, whether PAN images are degraded or not.

Both model training and testing of the proposed MBPRR-Net are implemented by using an NVIDIA GeForce GTX 2080Ti GPU based on the PyTorch framework.

IV. Experiments and Analysis

In this section, we use a lot of experimental analysis to verify the effectiveness of the MBPRR-Net model. First, the selection of datasets in our works is described. Second, the evaluation metrics are described. Third, our method is compared with some representative image colorization and image super-resolution methods. Finally, some ablation studies are used to validate our model design selection. Moreover, we also perform our method in semantic segmentation applications and the results are shown in supporting document.

A. Datasets

In this work, our images come from QuickBird satellite and WorldView-II satellite. The QuickBird dataset contains PAN images with spatial resolution of 0.61–0.72 m and MS images with spatial resolution of 2.44–2.88 m. The WorldView-II dataset contains PAN images with spatial resolution of 0.5 m and MS images with spatial resolution of 1.8 m. For a pair of PAN and MS images taken by one satellite in the same scene, the width and height of the MS image are both a quarter of the PAN image. Our training dataset has 2800 pairs of PAN images and MS images, including 1400 paired images from QuickBird and 1400 paired images from WorldView-II, and the test dataset has 240 pairs of PAN and MS images, including 120 paired images from QuickBird and 120 paired images from WorldView-II.

The selection of dataset includes three steps. First, we select PAN images and the corresponding MS images from QuickBird or WorldView-II. Second, we clip PAN and MS images into small image patches in which the size of the PAN image patch is 256 × 256 and MS image patch is 64 × 64. Finally, we randomly shuffle image patches, and then, select PAN and MS image pairs as training set or test set.

B. Evaluation Metrics

We utilize a variety of image quality evaluation metrics to conduct objective analysis of the MBPRR-Net model's output images, including PSNR, cross correlation (CC), universal image quality index (UIQI), and spectral angle mapper (SAM). The higher metrics represent the better the image quality except SAM. In this section, we denote the image generated by our MBPRR-Net model as $\hat{Y} \in \mathbb{R}^{C\times H\times W}$, label image is expressed as $Y \in \mathbb{R}^{C\times H\times W}$, $\hat{y}_j$ and y_j denote the jth columns of $\hat{Y}$ and Y, respectively, and $\hat{y}^i$ and y^i denote the ith bands of $\hat{Y}$ and Y, respectively.

1) Mean Square Error (MSE): The MSE is the average of the square of the error between true image x_1 and processed image x_2 that can be calculated as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left({x_1}^{(i)} - {x_2}^{(i)}\right)^2. \tag{9}$$

2) Peak Signal-to-Noise Ratio (PSNR): The PSNR is the logarithm of the highest pixel value of the processed image x_2 divided by the MSE of the true image x_1 and the processed image x_2, the PSNR is formulated as follows:

$$\mathrm{PSNR} = 10 \times \lg\left(\frac{\mathrm{MAX}_{x_2}^2}{\mathrm{MSE}(x_1, x_2)}\right) \tag{10}$$

where $\mathrm{MAX}_{x_2}^2$ represents the maximum pixel value of x_2, and MSE(.) is calculated by (9). The higher the PSNR value of the processed image, the better the processing effect of the image (the lower the distortion of the image). Note that the unit of PSNR is decibel.

3) Cross Correlation (CC): CC is defined as the characteristic of geometric distortion, which can be computed as follows:

$$\mathrm{CC} = \frac{1}{c}\sum_{i=1}^{c}\frac{\sum_{j=1}^{n}\left(\hat{y}_j^i - \mu_{\hat{y}^i}\right)\left(y_j^i - \mu_{y^i}\right)}{\sqrt{\sum_{j=1}^{n}\left(\hat{y}_j^i - \mu_{\hat{y}^i}\right)^2\sum_{j=1}^{n}\left(y_j^i - \mu_{y^i}\right)^2}} \tag{11}$$

where μ is the mean value of image. The closer CC to 1, the more correlated the two images are.

4) Universal Image Quality Index (UIQI): The UIQI considers correlation loss, brightness distortion, and contrast distortion. The UIQI is formulated as follows:

$$\mathrm{UIQI} = \frac{\sigma_{\hat{y}y}}{\sigma_{\hat{y}}\sigma_y}\frac{2\mu_{\hat{y}}\mu_y}{\mu_{\hat{y}}^2 + \mu_y^2}\frac{2\sigma_{\hat{y}}\sigma_y}{\sigma_{\hat{y}}^2 + \sigma_y^2}. \tag{12}$$

5) Spectral Angle Mapper (SAM): In an image, the SAM employ a vector to represent the spectrum of each pixel. The SAM is used to represent the similarity between two spectra, and utilizes the angle between two spectral vectors to calculate the similarity. Note that the SAM is smaller, the angle of two spectral vectors is smaller, and the higher the degree of similarity between the two spectra.

$$\mathrm{SAM} = \cos^{-1}\left(\frac{\hat{y}^T y}{\sqrt{\hat{y}^T\hat{y}}\sqrt{y^T y}}\right) \tag{13}$$

C. Experimental Comparison and Discussions

In this section, we analyze the experimental results of our method, and compare them with some existing colorization

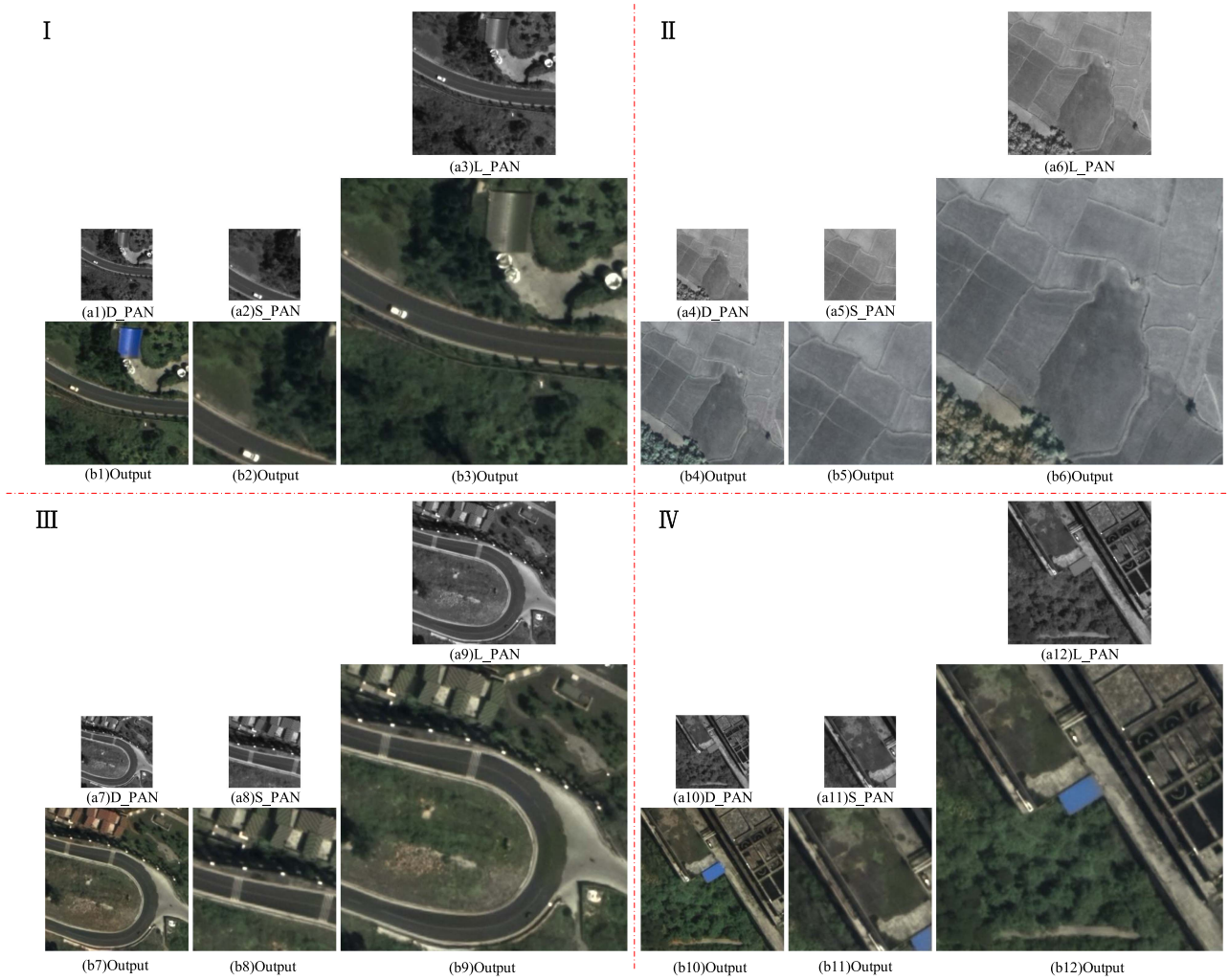

Fig. 7. Some results of our MBPRR-Net when inputting images of different sizes. (a) Inputs, D_PAN $\in \mathbb{R}^{128\times128}$ is the degraded PAN image, S_PAN $\in \mathbb{R}^{128\times128}$ is the small-size PAN images, and L_PAN $\in \mathbb{R}^{256\times256}$ is the large-size PAN images. (b) Outputs of our MBPRR-Net model.

and super-resolution methods in terms of objective data and subjective vision.

First, we evaluate the model performance under different types of input images; we show four groups of model outputs in Fig. 7 when input different kinds of images. The input images include the following three types: the degraded PAN images (D_PAN $\in \mathbb{R}^{128\times128}$), the small-size PAN images (S_PAN $\in \mathbb{R}^{128\times128}$), and the large-size PAN images (L_PAN $\in \mathbb{R}^{256\times256}$). From Fig. 7, we can see that the output images have higher spatial resolution than the input PAN images, and have color information. In general, the model can generate near realistic images for any size input images.

Second, we show the experimental results by category in Figs. 8–11, mainly showing land, building, river, and other categories that are not easy to classify. Experiments show that the results of our model can improve the spatial resolution (the MBPRR-Net model's output image is wider and higher than the MBPRR-Net model's input image), and also improve the spectral resolution (the color information of the MBPRR-Net model's output image is same as the MS image). Figs. 8–11 show that our output image has no artifacts or blurred edges, which illustrates that our model is successful in restoring the spatial resolution. Figs. 8–10 show that our model successfully solves the problem of large intraclass differences in remote

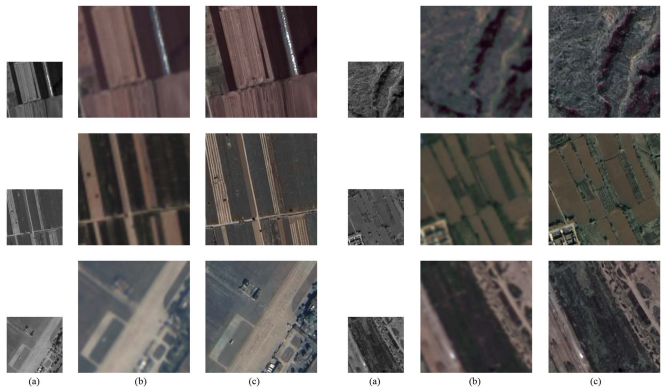

Fig. 8. Experimental results on land scenes. (a) Input. (b) Resized MS image. (c) Output.

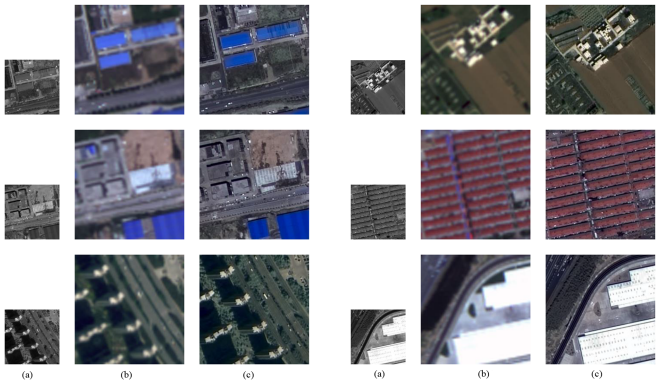

Fig. 9. Experimental results on building scenes. (a) Input. (b) Resized MS image. (c) Output.

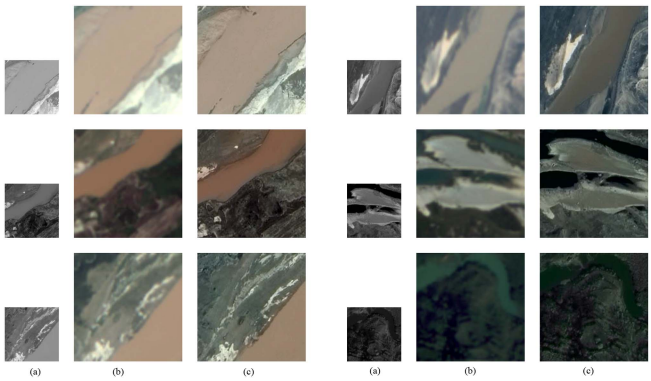

Fig. 10. Experimental results on river scenes. (a) Input. (b) Resized MS image. (c) Output.

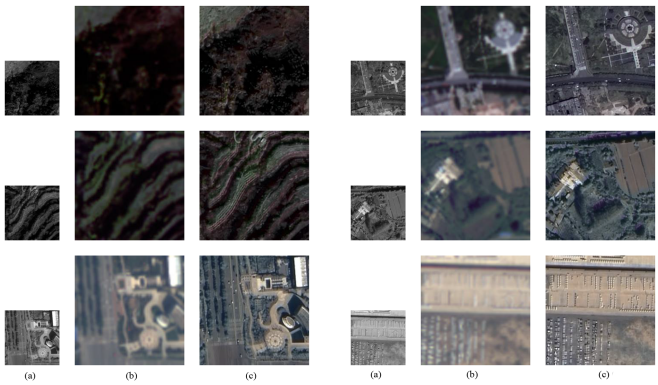

Fig. 11. Experimental results on some scenes that are not easy to classify. (a) Input image. (b) Resized MS image. (c) Output image.

sensing images. For the same kind of objects, our model can recover different spectral or color information according to MS images. For example, the land in Fig. 8 is brownish-yellow, green, or gray, and the building in Fig. 9 is blue, silver, white, or dark gray, and the river in Fig. 10 is blackish green, green, or brownish-yellow. Our model also solves the problem of small interclass differences in remote sensing images. For example, the land in Fig. 8 appears brownish-yellow, and the river in Fig. 10 also appears brownish-yellow, which realizes the reconstruction

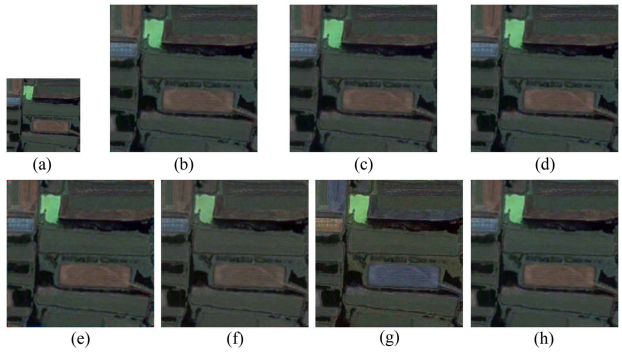

Fig. 12. Image super-resolution results of our MBPRR-Net and comparison algorithm. (a) Input. (b) SRCNN [8]. (c) SRGAN [13]. (d) EDSR [32]. (e) LapSR [42]. (f) DRN [34]. (g) MHAN [43]. (h) Ours.

TABLE III
QUANTITATIVE RESULTS OF THE PAN IMAGE COLORIZATION TASK

Method	PSNR (↑)	CC (↑)	UIQI (↑)	SAM (↓)
Iizuka [22]	35.38	0.977	0.975	0.061
Zhang [23]	34.06	0.969	0.966	0.073
Isola [26]	33.90	0.970	0.966	0.075
Yoo [27]	32.71	0.968	0.963	0.085
Vitoria [28]	35.02	0.975	0.973	0.064
Ours	36.06	0.979	0.977	0.056

Red text marks the best performance.

of the same spectral or color information for different types of objects.

Third, we compare the effect of the MBPRR-Net model with some excellent super-resolution models including SRCNN [8], SRGAN [13], EDSR [32], LapSR [42], DRN [34], and MHAN [43]. The results of different methods are shown in Fig. 12. The output images produced by our method are comparable to the image generated by other super-resolution methods, visually. In particular, our method does not produce noticeable forged artifacts. The experiment result of Table II shows that our method has higher value in objective indicators than other methods, which further illustrates the effectiveness of our MBPRR-Net model.

Fourth, we compare our method with other excellent colorization models. All the colorization models use the same input features, these models are trained with the same iterations. The colorization visual results of PAN image of our algorithm (in Section III) and the comparison algorithms (Iizuka [22], Zhang [23], Isola [26], Yoo [27], and Vitoria [28]) in Dataset-I are shown in Fig. 13. Fig. 13(a) is the MS image $\in \mathbb{R}^{C\times 64\times 64}$; for visual contrast, the size of the MS image is enlarged to 256 $\times$ 256. Red rectangular boxes are used to mark some small areas where the color detail is significantly incorrect compared to our method. We can observe the following.

1) Our proposed method can generate an image with color information that is closer to the MS image than other methods. For example, the first row of Fig. 13 shows that our method can colorize the roof, land, and vegetation correctly, while other methods have some wrong colors (i.e., the roof in the upper left corner is gray, whereas in

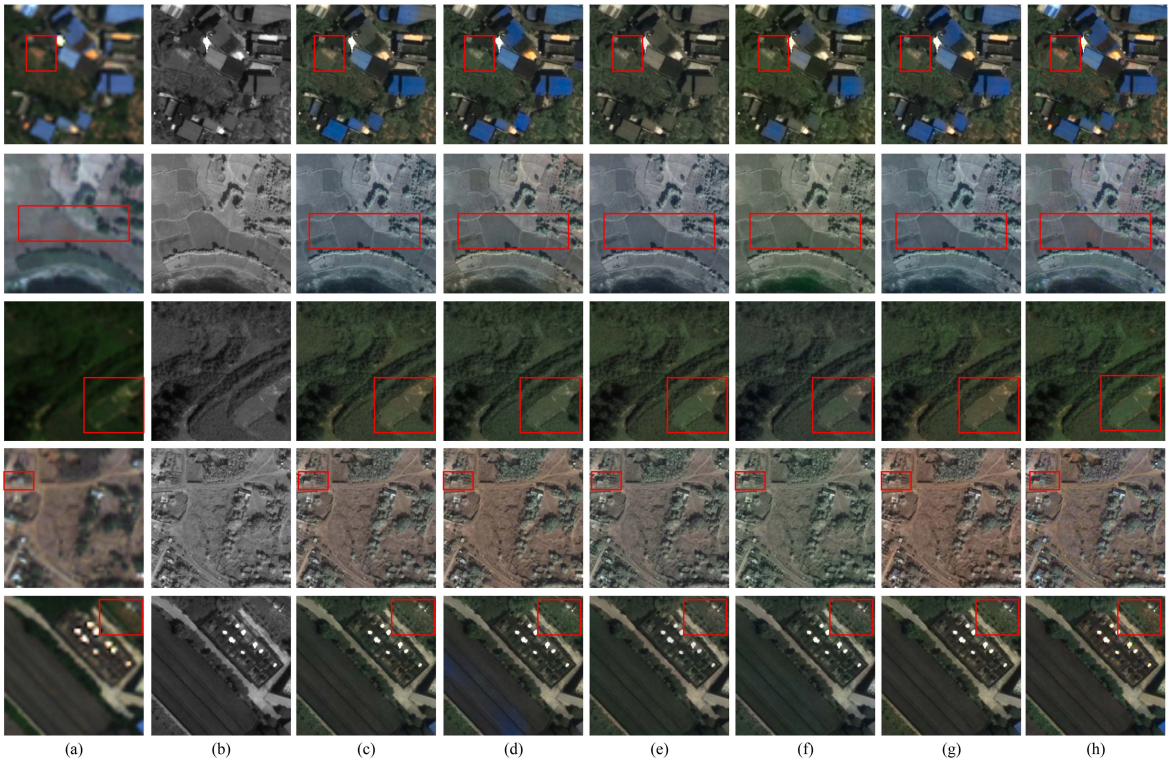

Fig. 13. Colorization results of our MBPRR-Net and comparison algorithm. (a) Resized MS image. (b) PAN image. (c) Iizuka [22]. (d) Zhang [23]. (e) Isola [26]. (f) Yoo [27]. (g) Vitoria [28]. (h) Ours.

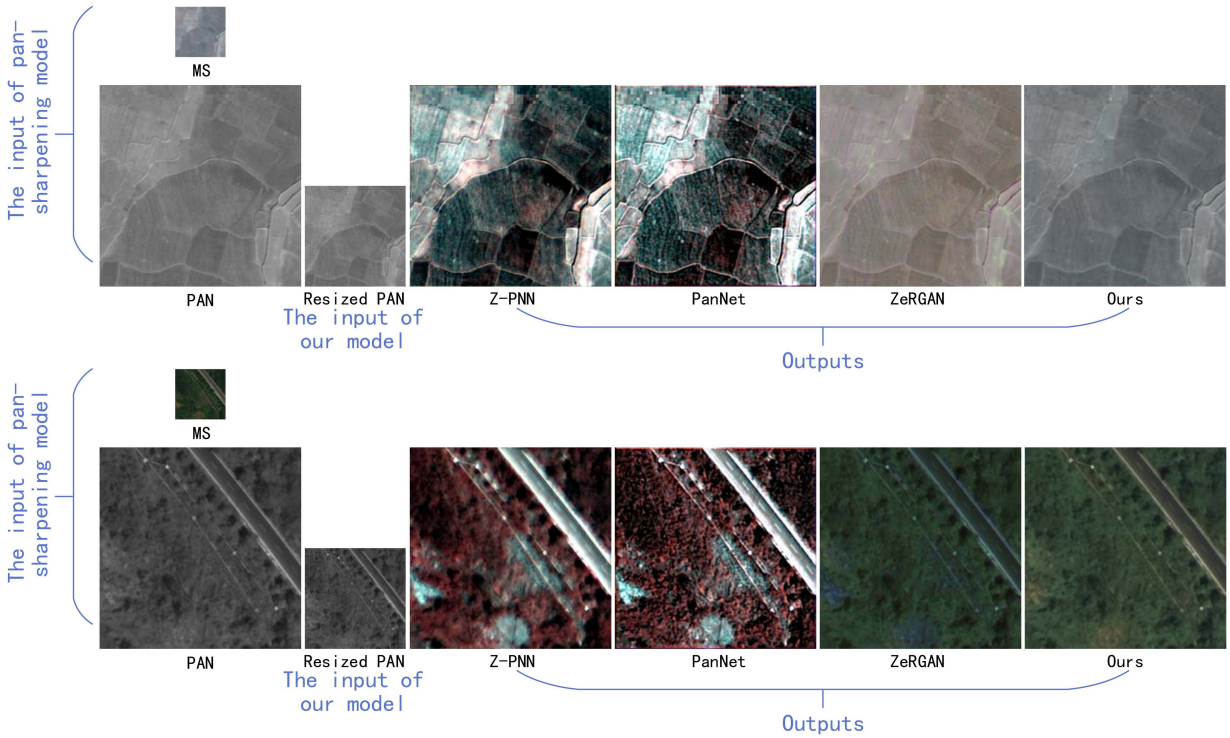

Fig. 14. Results of our MBPRR-Net and pan-sharpening algorithm.

TABLE IV
QUANTITATIVE RESULTS OF OUR MODEL, COLORIZATION, AND SR COMBINED MODELS (FIRST USE THE SR MODEL, AND THEN, USE THE COLORIZATION MODEL)

SR	Colorization	PSNR (↑)	CC (↑)	UIOI (↑)	SAM (↓)
SRGAN [13]	Iizuka [22]	23.63	0.754	0.646	0.208
	Yoo [27]	32.68	0.869	0.859	0.082
	Zhang [23]	31.97	0.784	0.775	0.089
	Vitoria [28]	30.56	0.760	0.761	0.127
	Isola [26]	30.75	0.933	0.925	0.110
LAPSR [42]	Iizuka [22]	28.78	0.956	0.937	0.086
	Yoo [27]	33.26	0.959	0.955	0.089
	Zhang [23]	31.30	0.960	0.947	0.106
	Vitoria [28]	32.74	0.941	0.935	0.102
	Isola [26]	32.27	0.954	0.949	0.090
EDSR [32]	Iizuka [22]	29.37	0.961	0.943	0.081
	Yoo [27]	33.01	0.958	0.953	0.089
	Zhang [23]	31.97	0.955	0.948	0.101
	Vitoria [28]	32.88	0.957	0.952	0.091
	Isola [26]	32.63	0.957	0.953	0.087
MHAN [43]	Iizuka [22]	29.26	0.960	0.942	0.083
	Yoo [27]	33.00	0.959	0.954	0.090
	Zhang [23]	33.73	0.960	0.956	0.083
	Vitoria [28]	29.06	0.939	0.927	0.099
	Isola [26]	32.44	0.957	0.952	0.089
SRCNN [8]	Iizuka [22]	29.11	0.959	0.942	0.083
	Yoo [27]	33.41	0.959	0.955	0.087
	Zhang [23]	28.60	0.954	0.937	0.092
	Vitoria [28]	28.88	0.941	0.928	0.098
	Isola [26]	32.41	0.956	0.951	0.089
DRN [34]	Iizuka [22]	32.25	0.923	0.897	0.133
	Yoo [27]	31.35	0.910	0.893	0.152
	Zhang [23]	31.50	0.921	0.905	0.140
	Vitoria [28]	31.25	0.912	0.894	0.155
	Isola [26]	20.17	0.715	0.608	0.221
Our		33.87	0.963	0.961	0.072

Red text marks the best performance.

TABLE V
QUANTITATIVE RESULTS OF OUR MODEL, COLORIZATION, AND SR COMBINED MODELS (FIRST USE COLORIZATION MODEL, AND THEN, USE THE SR MODEL)

Colorization	SR	PSNR (↑)	CC (↑)	UIOI (↑)	SAM (↓)
Iizuka [22]	SRCNN [8]	32.15	0.951	0.948	0.088
	SRGAN [13]	30.54	0.900	0.848	0.076
	LAPSR [42]	31.58	0.944	0.941	0.093
	EDSR [32]	32.30	0.951	0.949	0.087
	DRN [34]	28.40	0.942	0.927	0.103
	MHAN [43]	29.10	0.917	0.821	0.080
Yoo [27]	SRCNN [8]	31.56	0.951	0.947	0.094
	SRGAN [13]	26.50	0.937	0.770	0.160
	LAPSR [42]	31.41	0.949	0.945	0.097
	EDSR [32]	31.39	0.948	0.944	0.096
	DRN [34]	26.18	0.843	0.832	0.162
	MHAN [43]	24.92	0.805	0.556	0.217
Zhang [23]	SRCNN [8]	30.10	0.953	0.941	0.110
	SRGAN [13]	27.49	0.937	0.811	0.150
	LAPSR [42]	29.76	0.947	0.935	0.113
	EDSR [32]	30.32	0.955	0.943	0.108
	DRN [34]	28.59	0.940	0.927	0.107
	MHAN [43]	24.94	0.880	0.588	0.214
Vitoria [28]	SRCNN [8]	32.65	0.958	0.955	0.083
	SRGAN [13]	27.34	0.915	0.783	0.153
	LAPSR [42]	32.24	0.951	0.948	0.086
	EDSR [32]	33.26	0.959	0.957	0.077
	DRN [34]	29.05	0.945	0.933	0.098
	MHAN [43]	25.10	0.805	0.562	0.212
Isola [26]	SRCNN [8]	32.49	0.955	0.952	0.087
	SRGAN [13]	27.16	0.917	0.769	0.155
	LAPSR [42]	32.03	0.950	0.946	0.092
	EDSR [32]	32.74	0.957	0.953	0.085
	DRN [34]	28.49	0.939	0.925	0.108
	MHAN [43]	24.86	0.786	0.539	0.216
Our		33.87	0.963	0.961	0.072

Red text marks the best performance.

the MS image, it is blue; the land in the red rectangular box is green, whereas in the MS image it is brown).

2) Our model is especially good at colorizing rare instances. For example, in the fourth row, the MS image has a green plant in the red box, and only our method colorizes the plant green, whereas all other methods colorize the plant brown.
3) Our method does not have the problem of color overflow. For example, in the last row, the road in Fig. 13(d) appears blue and the road in Fig. 13(f) and (g) appear green, but blue and green are both colors that should not be on the road. However, our method gives the road normal color, which is same as the MS image.

In addition to visual analysis, we also evaluate colorization quality quantitatively. We present the average evaluation metrics on the testing dataset in Table III. According to the aforementioned analysis, we can infer that the proposed method is better than the compared algorithms in both visual effect and objective evaluation metrics.

Fifth, we combine SR models and colorization models to compare with our model. Tables IV and V show the quantitative results of two groups of comparative experiments, respectively. By comparing Tables IV and V, it can be seen that the effect of our model is better than the combination of the SR model and colorization model. At the same time, it is not difficult to see that the result of using the SR model first and then the colorization model is better than that of using the colorization model first and then the SR model on the whole, which indicates that the sequence of models will affect the final experimental results. However, our model does not have the problem of sequencing because we use one model to achieve the effect of SR and colorization models at the same time. Therefore, our model not only has the best experimental effect, but also does not need to consider the order of the SR model and colorization model, which can be implemented in one step.

Finally, we compared our method with the pan-sharpening model, the experimental results of the Z-PNN model proposed by Ciotola et al. [44], the PanNet model proposed by Yang et al. [45], and the ZeRGAN model proposed by Diao et al. [46] are shown in Fig. 14. The pan-sharpening model requires two inputs, MS and PAN, while our model requires only one input. It can be seen from Fig. 14 that the spectral information of the images output by PNN and PanNet models is rich, but it is obviously inconsistent with the spectral information of MS, that is, these two models can restore the spectrum, but the restored spectrum does not conform to the real situation. Compared with the ZeRGAN model, the output images of our model have clear texture, and the spectral information is closer to MS.

TABLE VI
ABLATION STUDY ON THE NUMBER OF EFFECTIVE BRANCHES

Branch number	PSNR (↑)	CC (↑)	UIQI (↑)	SAM (↓)
1	27.33	0.836	0.828	0.154
2	33.41	0.961	0.959	0.076
3	33.75	0.962	0.960	0.072
5	33.74	0.962	0.960	0.071
4 (Our)	33.87	0.963	0.961	0.072

Red text marks the best performance.

TABLE VII
AN ABLATION STUDY ON DIFFERENT COMBINATIONS OF FCMB, CUBIC, AND SCB

	1	2	3	4	5	6	7	8
FCMB	✗	✓	✗	✗	✓	✓	✗	✓
Cubic	✗	✗	✓	✗	✓	✗	✓	✓
SCB	✗	✗	✗	✓	✗	✓	✓	✓
PSNR	33.67	33.75	33.85	27.35	33.78	33.78	33.65	33.87

Red text marks the best performance.

Thus, we can infer that the proposed method is better than the compared algorithms in both visual effect and objective evaluation metrics.

D. Ablation Study

To determine the appropriate number of branches of MBPRR-Net and prove the effectiveness of modules in MBPRR-Net, we performed two ablation experiments, respectively.

1) Selection of Branch Number: To determine the number of branches of our MBPRR-Net model, we experimented with single branch, two branches, three branches, four branches, and five branches. It can be seen from Table VI that the performance of MBPRR-Net is the best when four branches are used. There are three possible reasons why using four branches performs well. First, the PAN image has the problem of large differences between same class and small differences between different classes. This problem can be solved by designing a module to extract more useful features. Using multiple branches is helpful to extract more useful features from the input image. Therefore, it is possible to adopt a model with four branches, which will be better than a model with single branch, two branches, or three branches. Second, the fewer branches are used, the fewer feature maps can be used for aggregation. The feature aggregation module can use the extracted feature map to make the aggregated feature map wider and higher so that the extracted features can be fully utilized to reconstruct an image. Therefore, having more branches will help the model improve the performance of the aggregation module. Third, the effect of the MBPRR-Net does not always increase with the increase of the number of branches, that is, when the number of branches increases to a certain extent, and then, continues to increase, the effect of the MBPRR-Net decreases. This is because the number of branches increases, the depth of the model will increase, and may cause vanishing gradient problem. At the same time, the width and height of the extracted feature map will become smaller, while the too small feature map has little useful information. Therefore, the back branches cannot improve the effect of the MBPRR-Net. Through experiments, we find that MBPRR-Net has the best effect when the number of branches is four.

2) Effectiveness of Modules in MBPRR-Net: To validate the effect of several important modules in our MBPRR-Net model, we do an ablation study by combining different modules. The average PSNR of different combinations is illustrated in Table VII. First, we can see that the performance of the baseline without FCMB, cubic, and SCB is not very well (PSNR = 33.67 dB) from first group in Table VII. Second, we add one of FCMB, Cubic, and SCB to the baseline, respectively (from second to fourth groups in Table VII). We can validate that adding each module to the baseline can improve the effect of the baseline except adding SCB to the baseline. This is because the SCB module is used for feature recovery, which is based on the condition that abundant features are extracted. Thus, the SCB module would not be able to perform its function without the FCMB module to extract features and cubic filter to enhance features. Then, we add two modules to the baseline (from fifth to seventh groups in Table VII). By comparing the fourth and sixth groups, we can see that using both SCB module and FCMB module is better than using only SCB module. Since Cubic performs well (PSNR = 33.85 dB), the model with FCMB and SCB also performs well (PSNR = 33.78 dB). Finally, we combine these three modules together to get the eight group in Table VII, and three components perform the best (PSNR = 33.87 dB). These quantitative analyses show that our proposed FCMB, Cubic, and SCB modules are useful.

V. CONCLUSION

In this article, we proposed a novel model named MBPRR-Net to restore the spatial and spectral resolution of the PAN image simultaneously. To exploit the correlation between the feature channels, we design a module FCMB to fuse neighboring features in channel dimension. We use a multibranch structure to obtain the features in different scales, and then, we aggregate the features of each branch by pixel shuffling to make full use of the features extracted by each branch. We use the cubic filter to weight the extracted features according to their importance. At the same time, we add an attention mechanism to the SCB module to assign weights to the features to be recovered so that the important information will not be ignored. To recover the spatial resolution and spectral resolution, we use two branches to generate the L channel and ab channel of the image, respectively. We verified the spectral resolution improvement by comparing it with the colorized model, and the spatial resolution improvement by comparing it with the super-resolution model. Extensive experiments and visualization demonstrate that our MBPRR-Net outperforms most existing colorization and super-resolution methods.

Since our model needs to improve the spatial and spectral resolution of remote sensing images at the same time, it is more complex than the super-resolution model that only improves the spatial resolution and the colorization model that only improves the spectral resolution. In the future, we will optimize the model and speed up the training and inference speed without reducing the effect of the model.

REFERENCES

[1] H. Li et al., "Temperature-based and radiance-based validation of the collection 6 MYD11 and MYD21 land surface temperature products over barren surfaces in northwestern China," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1794–1807, Feb. 2021.

[2] L. Jiao, L. Huo, C. Hu, and P. Tang, "Refined UNet v3: Efficient end-to-end patch-wise network for cloud and shadow segmentation with multi-channel spectral features," *Neural Netw.*, vol. 143, pp. 767–782, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608021003130

[3] A. Sharma, X. Liu, and X. Yang, "Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks," *Neural Netw.*, vol. 105, pp. 346–355, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608018301813

[4] M. G. Gale, G. J. Cary, A. I. Van Dijk, and M. Yebra, "Forest fire fuel through the lens of remote sensing: Review of approaches, challenges and future directions in the remote sensing of biotic determinants of fire behaviour," *Remote Sens. Environ.*, vol. 255, 2021, Art. no. 112282. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425720306556

[5] K. Indrajit and R. Moumita, "Deep neural network-based heterogeneous domain adaptation using ensemble decision making in land cover classification," *IEEE Trans. Artif. Intell.*, vol. 1, no. 2, pp. 167–180, Oct. 2020.

[6] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Netw.*, vol. 95, pp. 19–28, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608017301806

[7] H. Hosseinpour, F. Samadzadegan, and F. D. Javan, "CMGFNet: A deep cross-modal gated fusion network for building extraction from very high-resolution remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 96–115, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271621003294

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[9] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 391–407.

[10] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 294–310.

[11] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *Pattern Recognit.*, vol. 107, 2020, Art. no. 107475. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320320302788

[12] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. S. Huang, and H. Shi, "Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5689–5698.

[13] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 105–114.

[14] N. Anagnostopoulos, C. Iakovidou, A. Amanatiadis, Y. Boutalis, and S. A. Chatzichristofis, "Two-staged image colorization based on salient contours," in *Proc. IEEE Int. Conf. Imag. Syst. Techn.*, 2014, pp. 381–385.

[15] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6836–6845.

[16] Y. Ci, X. Ma, Z. Wang, H. Li, and Z. Luo, "User-guided deep anime line art colorization with conditional adversarial networks," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 1536–1544, doi: 10.1145/3240508.3240661.

[17] A. Y.-S. Chia et al., "Semantic colorization with Internet images," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 1–8, Dec. 2011, doi: 10.1145/2070781.2024190.

[18] M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan, "Deep exemplar-based colorization," *ACM Trans. Graph.*, vol. 37, no. 4, Jul. 2018, doi: 10.1145/3197517.3201365.

[19] J. Lee, E. Kim, Y. Lee, D. Kim, J. Chang, and J. Choo, "Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5800–5809.

[20] H. Bahng et al., "Coloring with words: Guiding image colorization through text-based palette generation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 443–459.

[21] H. Kim, H. Y. Jhoo, E. Park, and S. Yoo, "Tag2Pix: Line art colorization using text tag with SECat and changing loss," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9055–9064.

[22] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, no. 4, Jul. 2016, doi: 10.1145/2897824.2925974.

[23] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 649–666.

[24] G. Özbulak, "Image colorization by capsule networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 2150–2158.

[25] J.-W. Su, H.-K. Chu, and J.-B. Huang, "Instance-aware image colorization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7965–7974.

[26] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.

[27] S. Yoo, H. Bahng, S. Chung, J. Lee, J. Chang, and J. Choo, "Coloring with limited data: Few-shot colorization via memory augmented networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11 275–11 284.

[28] P. Vitoria, L. Raad, and C. Ballester, "ChromaGAN: Adversarial picture colorization with semantic class distribution," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2020, pp. 2434–2443.

[29] G. Ji, Z. Wang, L. Zhou, Y. Xia, S. Zhong, and S. Gong, "SAR image colorization using multidomain cycle-consistency generative adversarial network," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 296–300, Feb. 2021.

[30] F. Ozcelik, U. Alganci, E. Sertel, and G. Unal, "Rethinking CNN-based pansharpening: Guided colorization of panchromatic images via GANs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3486–3501, 2021.

[31] X. Lu, J. Zhang, D. Yang, L. Xu, and F. Jia, "Cascaded convolutional neural network-based hyperspectral image resolution enhancement via an auxiliary panchromatic image," *IEEE Trans. Image Process.*, vol. 30, pp. 6815–6828, 2021.

[32] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1132–1140.

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

[34] Y. Guo et al., "Closed-loop matters: Dual regression networks for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5406–5415.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[36] J. An, K. G. Kpeyiton, and Q. Shi, "Grayscale images colorization with convolutional neural networks," *Soft Comput.*, vol. 24, no. 3, pp. 4751–4758, 2020, doi: 10.1007/s00500-020-04711-3.

[37] S. Liu, D. Huang, and Y. Wang, "Receptive field block net for accurate and fast object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 404–419.

[38] D. LI et al., "Involution: Inverting the inherence of convolution for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 12321–12330.

[39] H. Jie, S. Li, S. Gang, and S. Albanie, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[40] S. Moran, P. Marza, S. Mcdonagh, S. Parisot, and G. Slabaugh, "DeepLPF: Deep local parametric filters for image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12823–12832.

[41] D. Misra, "Mish: A self regularized non-monotonic neural activation function," in *Proc. Br. Mach. Vis. Conf.*, 2020, pp. 1–14.

[42] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5835–5843.

[43] D. Zhang, J. Shao, X. Li, and H. T. Shen, "Remote sensing image super-resolution via mixed high-order attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5183–5196, Jun. 2021.

[44] M. Ciotola, S. Vitale, A. Mazza, G. Poggi, and G. Scarpa, "Pansharpening by convolutional neural networks in the full resolution framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5408717.
[45] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5449–5457.
[46] W. Diao, F. Zhang, J. Sun, Y. Xing, K. Zhang, and L. Bruzzone, "ZeRGAN: Zero-reference GAN for fusion of multispectral and panchromatic images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8195–8209, Nov. 2023.

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

From 2018 to 2020, he was a Postdoctoral Fellow with the School of Software, Yunnan University, where he is currently an Associate Professor. His research interests include pulse coupled neural networks and its applications, image processing, information fusion, optimization algorithm, and fuzzy set theory.

Ling Liu received the B.E. degree in software engineering, in 2019, from the Yunnan University, Kunming, China, where she is currently working toward the master's degree in software engineering with the School of Software, Yunnan University, Kunming, China, in 2023.

Her research interests include deep neural networks, image colorization, and image super-resolution.

Xiaoxuan Ren received the B.E. degree in computer science and technology from Beijing Jiaotong University, Beijing, China, in 2021. She is currently working toward the master's degree in artificial intelligence and machine learning with the School of Software, Yunnan University, Kunming, China.

Her current research interests include image processing, deep learning, and image super-resolution.

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, Changsha, China, in 2012 and 2015, respectively, and the Ph.D. degree in communication and information systems from Yunnan University, Kunming, China, in 2019.

From 2019 to 2021, she was a Postdoctoral Fellow with the School of Software, Yunnan University, where she is currently an Associate Professor. Her research interests include deep neural networks, fuzzy set theory, bioinformatics, image processing, and information fusion.

Shin-Jye Lee received the M.Sc. (Eng.) degree in advanced software engineering from the Department of Computer Science, The University of Sheffield, Sheffield, U.K., in 2001, the M.Phil. degree in 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, The University of Manchester, Manchester, U.K., in 2011 .

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

Jun Zhang received the B.S. degree in communication engineering from the Hubei University of Technology, Wuhan, China, in 2009, and the master's degree in signal and information systems from Yunnan University, Kunming, China, in 2016 .

He is a Engineer with China Mobile Communications Group Yunnan Co., Ltd., Kunming. His current research interests include artificial intelligence, communication network, and data mining.

Shaowen Yao 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 the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2002.

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