Group Shuffle and Spectral-Spatial Fusion for Hyperspectral Image Super-Resolution

Xinya Wang, Yingsong Cheng, Xiaoguang Mei, Junjun Jiang, *Senior Member, IEEE*, and Jiayi Ma, *Senior Member, IEEE*

***Abstract*—Recently, super-resolution (SR) tasks for single hyperspectral images have been extensively investigated and significant progress has been made by introducing advanced deep learning-based methods. However, hyperspectral image SR is still a challenging problem because of the numerous narrow and successive spectral bands of hyperspectral images. Existing methods adopt the group reconstruction mode to avoid the unbearable computational complexity brought by the high spectral dimensionality. Nevertheless, the group data lose the spectral responses in other ranges and preserve the information redundancy caused by continuous and similar spectrograms, thus containing too little information. In this paper, we propose a novel single hyperspectral image SR method named GSSR, which pioneers the exploration of tweaking spectral band sequence to improve the reconstruction effect. Specifically, we design the group shuffle that leverages interval sampling to produce new groups for separating adjacent and extremely similar bands. In this way, each group of data has more varied spectral responses and less redundant information. After the group shuffle, the spectral-spatial feature fusion block is employed to exploit the spectral-spatial features. To compensate for the adjustment of spectral order by the group shuffle, the local spectral continuity constraint module is subsequently appended to constrain the features for ensuring the spectral continuity. Experimental results on both natural and remote sensing hyperspectral images demonstrate that the proposed method achieves the best performance compared to the state-of-the-art methods.**

***Index Terms*—Hyperspectral image, super-resolution, group shuffle, spectral-spatial feature fusion block, local spectral continuity constraint module.**

I. Introduction

HYPERSPECTRAL image systems use image technology to obtain dozens or even hundreds of continuous and narrow band spectrograms to compose hyperspectral images. Compared with the multispectral image (MSI) or natural image, hyperspectral images can better reflect the subtle spectral characteristics of the measured materials in detail. Therefore, hyperspectral images significantly enhance the ability to distinguish materials which look similar to humans. It is widely used in computer vision and remote sensing fields, such as mineral exploration [1], medical diagnosis [2] and plant detection [3].

However, due to the limitations of equipment, storage, and so on, there is always a trade-off between spatial and spectral resolution. Hyperspectral images divide more bands within the broad spectral coverage to improve spectral resolution. Thereby, hyperspectral images inevitably have lower spatial resolution compared with the MSI or natural images, which has become the main restriction factor for the application. As the hardware equipment is difficult to upgrade to improve performance, it is necessary to develop software technology to obtain a reliable hyperspectral image with high resolution.

Super-resolution (SR) reconstruction is to transform low-resolution (LR) images into high-resolution (HR) images by signal processing and image processing. Many existing methods leverage high-frequency spatial information from an HR auxiliary image (such as panchromatic image, RGB image, or MSI) to improve the spatial resolution of the observed hyperspectral image, namely fusion-based hyperspectral image super-resolution. Though the considerable performance has been achieved, they assume that the reference image and the input LR hyperspectral image are well co-registered. Unfortunately, obtaining such pairs is extremely arduous in real applications. By contrast, single hyperspectral image super-resolution only requires the LR image to infer the corresponding HR one, which is more practical [4].

To exploit the abundant spectral information among successive bands, several single hyperspectral super-resolution methods aim to design representative features, which are based on the linear combination, sparse representation, low rank, total variation prior, and so on, to recover HR hyperspectral images. However, these sophisticated hand-crafted features only reflect some specific statistical properties of the internal dataset. Recently, the deep convolution neural network (DCNN) gradually prospers in SR because of its extraordinary learning capability of mapping the LR image to the HR image. Nevertheless, it is still very challenging to design a computationally efficient and effective deep network. This is mainly due to the following reasons. On the one hand, hyperspectral images have many continuous and narrow spectral bands. The direct dividing group

Manuscript received 24 August 2022; revised 24 November 2022; accepted 4 January 2023. Date of publication 9 January 2023; date of current version 13 January 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 62276192 and 62071339, and in part by the Key Research and Development Program of Hubei Province under Grants 2020BAB113 and 2021CFB464. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Henry Arguello. *(Xinya Wang and Yingsong Cheng contributed equally to this work.) (Corresponding author: Jiayi Ma.)*

Xinya Wang, Yingsong Cheng, Xiaoguang Mei, and Jiayi Ma are with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: wangxinya@whu.edu.cn; yscheng12138@gmail.com; meixiaoguang@gmail.com; jyma2010@gmail.com).

Junjun Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: jiangjunjun@hit.edu.cn).

Digital Object Identifier 10.1109/TCI.2023.3235153

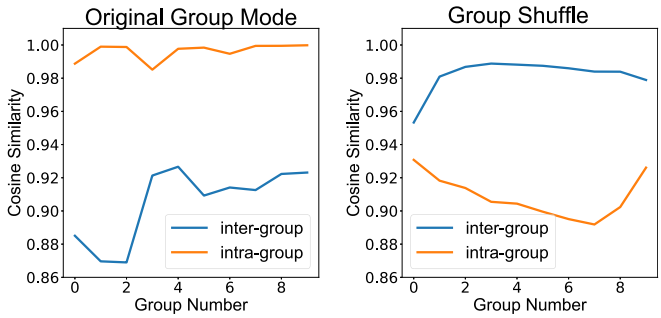


Fig. 1. Average cosine similarity curve of intra-group and inter-group in the CAVE dataset. The intra-group similarity is determined by calculating the average similarity between the bands within the group, while the inter-group similarity is the average similarity between groups. (a) The similarity curve of the original group mode. (b) The similarity curve of our group shuffle.

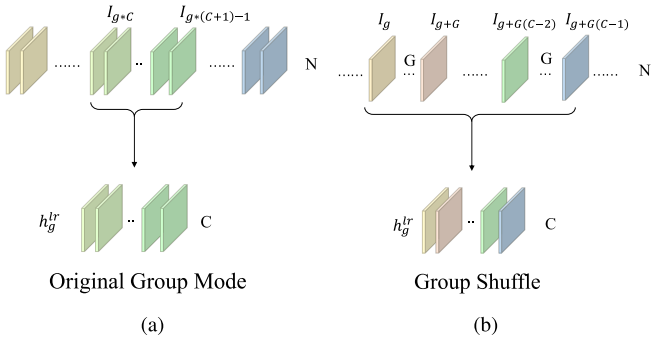


Fig. 3. Comparison of two group strategies, where I_g denotes the g-th band and h_g^{lr} is the g-th group of data. N bands of the hyperspectral data are divided into G groups with C bands in each group. (a) The original group mode, where each group of data is generated by directly cutting a segment of the spectrum. (b) Our group shuffle, where each group of data is generated by sampling at intervals G starting from I_g.

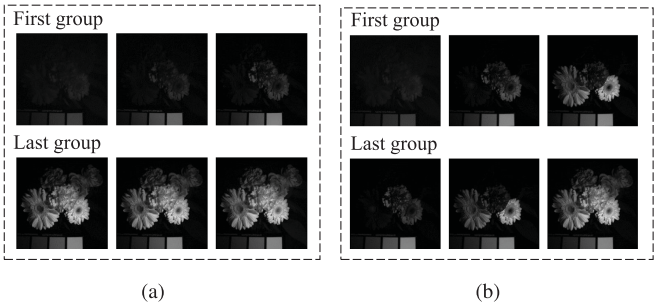


Fig. 2. Comparison of (a) original groups and (b) shuffled groups.

mode in existing group-based methods is insufficient in exploiting rich spectral information, failing in achieving outstanding performance. On the other hand, the existing feature extractors do not fuse the spatial and spectral features adequately.

To address the aforementioned problems, this paper proposes a novel and effective method with the group shuffle for single hyperspectral image SR, termed GSSR, which can leverage the rich spectral information of hyperspectral images. Because of the numerous spectral bands of hyperspectral images, methods that use the whole datacube as input have to limit the dimensionality of the feature maps. With the purpose of fully extracting the spectral information, previous works [5], [6] adopt a group reconstruction mode, which divides hyperspectral images into multiple groups in the spectral dimension for SR separately as shown in Fig. 3(a). In this way, a finite-dimensional feature map can be used to represent groups of data with small spectral dimensions. However, after the bands are divided into groups, we find that the spectrograms within the group are almost identical. As displayed in Fig. 1(a), the intra-group similarities of the original group mode are close to 1. Fig. 2(a) also exhibits the high similarity of the spectral bands in original groups. These quite similar spectrograms carry an enormous amount of redundant information, which is not conducive to image reconstruction. To make matters worse, group data lost the spectral response of other ranges, therefore it presents a locality in the spectrum. Overall, the original excessive intra-group similarity brings about severe information redundancy and localization of spectral responses, which leads to a limited amount of information in the group data. On the other hand, we observe that capturing the same static scene, the different spectral bands would present different levels of texture detail. Therefore, introducing various ranges of spectra in a group can enrich the texture information at different levels while mitigating the spectral locality of the group data. Based on these observations, we introduce the group shuffle, which adjusts spectral sequence to form new groups so that the group data will be more informative for super-resolution. In particular, as shown in Fig. 3(b), we sample the bands at certain interval values to generate the group data. Compared with original groups, shuffled groups comprise more diverse spectral bands. After group shuffle, the decrease of intra-group similarity in Fig. 1(b) could reduce the information redundancy, contributing to learning conducive features for super-resolution. As shown in Fig. 2(b), since the texture of an object may be clearer in some bands than in others, the dissimilar bands to be super-resolved in shuffled group data could complement each other in texture to enhance the super-resolution effect. At the same time, we can observe that due to group shuffle, the inter-group similarity increases and the spectral response between groups are more uniform, which is beneficial for learning repeatable and reliable super-resolution patterns.

After the group shuffle, we further derive an efficient and powerful feature extractor called spectral-spatial feature fusion block (SSFFB). With separable 3D convolution, SSFFB constructs a dual-stream deep structure to extract spectral and spatial features separately. More importantly, SSFFB fuses spectral and spatial features to explore the spectral-spatial coherence among bands, which takes advantage of the complementary properties of the two features. Besides, we also notice that there are no original adjacent spectral bands since the group shuffle changes the order of bands. Therefore, to maintain the spectral continuity of the generated hyperspectral images, we propose the local spectral continuity constraint module (LSCCM) after SSFFB. Because the corresponding adjacency relations between bands of adjacent groups remain present after the group shuffle, LSCCM locally restores the spectral band order and constrains the features close

to those of the previous spectrogram to ensure the continuity of the spectral bands. After acquiring the constrained features, we then upsample them to obtain HR hyperspectral images. Extensive experiments on two public datasets demonstrate that the proposed GSSR method is superior to other state-of-the-art methods. The ablation experiments verify the effectiveness of each component and the group shuffle strategy used in the proposed method.

In summary, the contribution of our work is mainly reflected in the following aspects:

1) We propose a novel and effective SR method with the group shuffle and spectral-spatial feature fusion block, termed GSSR, which achieves the state-of-the-art SR performance.
2) In order to learn complex spectral correlations efficiently and inexpensively, we propose a new group data generation mode called group shuffle. By interval sampling to form new groups, the spectral response within the group is more varied, which makes the group data more informative for super-resolution.
3) We propose an efficient feature extractor block called SSFFB, which fuses spectral and spatial features to exploit the complementary nature of spectral and spatial information, thus obtaining an improvement in model performance.

The rest of this article is organized as follows. Section II introduces some relevant research works. In Section III, we describe our GSSR network structure in detail. Then, adequate experiments and analyses are reported in Section IV. Finally, we make some conclusions in Section V.

II. Related Works

This paper focuses on single image super-resolution (SISR). In this section, we briefly introduce the single gray/RGB image super-resolution and the single hyperspectral image super-resolution.

A. Single Gray/RGB Image Super-Resolution

In recent years, deep learning has been widely used in natural image/video super-resolution and achieved excellent performance [7], [8]. The deep convolutional neural network (DCNN) has a strong ability to learn nonlinear mappings from LR images to HR images in an end-to-end manner. SRCNN [7] employed a three-layer convolutional neural network to solve the SR problem for the first time, and achieved better performance than the traditional SR methods. Where-after, more methods took advantage of the excellent learning ability of the DCNN by introducing some design tricks of network structure [9]. For example, VDSR [10] designed a fairly deep neural network, DRCN [11] introduced a recursive convolution network, and DRRN [12] added residual learning. Batch Normalization also has been removed by EDSR [13], which retains more high-frequency information. At the same time, inspired by generative adversarial networks, some scholars proposed SRGAN [14], which aims to generate more perceptually realistic HR images. But the problem with GAN is that the training process is still not stable. The attention mechanism is also very popular in SR. Some special attention structures, such as channel attention mechanism [15] and cross-scale non-local attention mechanism [16], have achieved effective improvement. Recently, SAN [17] proposed a second-order attention network to extract long-term dependence. To continue to improve the SR performance, researchers noticed that the main error in SR is the drop of high-frequency information, so DFSA [18] focused on the frequency domain to preserve high-frequency information. Meanwhile, there is also Fourier space loss [19], which constrains the distance between the frequency domain of the generated images and HR images. Most recently, HRN [20] developed a cross-modal residual network to enhance spatial nonlinear mapping by extracting frequency domain information.

Although these methods perform remarkably, they are designed for gray/RGB images, which have only one or three spectral channels. Hence the direct application of these SISR methods for the hyperspectral image SR is not a sensible idea. On the one hand, when they super-resolve the spectral image band-by-band, they would neglect the correlation among the spectral bands of the hyperspectral data, leading to unsatisfactory spectral distortion. On the other hand, the vast spectral dimensions of hyperspectral images will result in the multiplication of the model parameters and thus the model will lack enough hyperspectral data for training.

B. Single Hyperspectral Image Super-Resolution

Single hyperspectral image super-resolution does not utilize the auxiliary image, mining only information from the low-resolution image itself to reconstruct the high-resolution image, which has a wider range of application scenarios. Akgun et al. [21] modeled the acquisition process of hyperspectral images of different wavelengths as a linear weighted combination of a few base image planes. Based on sparse representation, Li et al. [22] proposed a hyperspectral super-resolution framework using hybrid analysis and sparsity of spectral-spatial groups. In [23], low-rank and total variation priors are used to regularize the image SR. However, these traditional methods are based on handcraft features, which only reflect the characteristics of a certain aspect of hyperspectral images. And the optimization in the test stage is complicated and time-consuming. Finally, these methods have limited expressive power, so they are not suitable for complex SR problems.

With the popularity of deep learning, some methods based on convolutional networks have also been proposed in hyperspectral SR. Yuan et al. [24] and Xie et al. [25] first applied DCNN and used nonnegative matrix factorization (NMF) as a post-process to keep spectral correlation. Li et al. [26] designed a spectral difference convolution neural network, which still adopted post-process to avoid spectral distortion. In general, these two-stage algorithms separately use DCNN and post-process to extract spatial and spectral information. But post-process, such as NMF, still relies on manual work, which makes the method ineffective and time-consuming.

Recently, several end-to-end deep learning models have been proposed to solve this problem. 3DFCNN [27] used 3D full

convolution to construct the network, but the computational complexity of 3D convolution is very tremendous. Li et al. [28] developed a group deep recursive residual network (GDRRN), which used group convolution to reduce the number of parameters. He et al. [29] proposed a deep Laplacian pyramid network, but this method cannot extract spectral information well, so its performance is poor. Inspired by generative adversarial networks, some researchers proposed 3D-GAN [30], which still used conventional 3D convolution. Jiang et al. [31] proposed the 2D–1D GAN framework, which included two subnetworks, that is, spatial and spectral subnetworks. This decomposition effectively reduces computational complexity, but there is no joint processing of spatial and spectral information, and GAN is still difficult to train. Li et al. [32] used a dual-flow 1D–2D spectral–spatial convolution neural network, which utilized 1D and 2D convolution to learn spectral and spatial features respectively and then fuses these features by changing the size of the feature map. Similarly, Wang et al. [33] further introduced a separable 3D convolution to analyze spatial and spectral information, which reduced the size of the model and retained the processing capability of 3D convolution for the spectral domain. In order to make up for the deficiency of spatial information extraction by separable 3D convolution, MCNet [34], ERCSR [35], and SFCSR [36] used mix 2D/3D convolution networks and shared spatial information to reconstruct spatial details better. Since hyperspectral images have dozens or hundreds of spectral channels, the method of directly rebuilding all spectral bands simultaneously obviously requires a large-scale model, and in fact, it is difficult to capture spectral correlation in numerous bands. Continuing the idea of group convolution from GDRRN, SSPSR [5] took it one step further, and proposed the group reconstruction strategy. Specifically, SSPSR divides the bands of hyperspectral images into groups and reconstructs high-resolution images using branch network with shared parameters, and finally merges the images for processing to learn spectral correlation. This greatly reduces the number of parameters and achieves excellent performance. However, SSPSR uses a progressive two-stage up-sample, which cannot be applied in small upscale factors like 2 or 3. Related to this, RFSR [6] exploited feedback mechanisms to construct interactions between spectral groups and constrain spectral continuity in the HR space by separable 3D convolution. Recently, Interactformer [37] refreshed state-of-the-art metrics by interacting with global and local features extracted by transformer and 3DCNN branches.

However, all of these models neglect the influence of the adjacent and extremely similar spectral images on SR, making it difficult to efficiently utilize the complementary relationship of the spectrum. We design the group shuffle, which separates adjacent bands to increase spectral diversity within the group. In addition, existing feature extractors all have drawbacks, 2D convolution cannot handle both spatial and spectral information, while 3D convolution has high computational complexity. Feature extractors based on separable 3D convolution tend to be very shallow structures. And they do not consider the fusion of spatial and spectral features. Thereby, we introduce SSFFB to better extract and fuse deep spectral-spatial features.

III. The Proposed Method

As shown in Fig. 4, our proposed method consists of the group shuffle and a group spectral band reconstruction module (GSBRM). The former adjusts spectral band sequence to form new groups that have more diverse spectral responses and less redundant information. The latter super-resolves each group of the hyperspectral image individually while maintaining the spectral-spatial structure of the generated HR image in the low spectral space. In this section, we first overview the whole framework of the model and then discuss the group shuffle and GSBRM in detail.

For the input LR hyperspectral image $H^{lr} \in \mathbb{R}^{h\times w\times L}$, we perform the group shuffle first to divide the images into G groups in the spectral dimension. The group strategy reduces the dimensionality of features processed in the network, with shuffle operation further separating adjacent and similar spectral bands. This process can be formulated as

$$[h_1^{lr}, h_2^{lr}, \ldots, h_g^{lr}, \ldots, h_G^{lr}] = GroupShuffle(H^{lr}), \quad (1)$$

where h_g^{lr} is the g-th group input data. Each group has C spectral bands. On the one hand, the group shuffle does not add extra parameters, which keeps the efficiency of the model. On the other hand, it enriches the spectral response within the group, which allows bands with less texture to obtain details from other sharper bands within the group to produce preferable HR images. However, the shuffle operation changes the order of spectral bands. To better maintain spectral continuity, instead of recovering the spectral order in hyperspectral space to constrain the HR images, we attempt to pre-place the spectral correlation constraint module in the GSBRM. Specifically, we adopt a feedback mechanism to learn spectral consecutive relationships through constraining local neighboring features closer to each other. In this way, we effectively reduce the computational complexity of the model while ensuring the effect. Given the previous group of constrained features f_{g-1}^h and LR input h_g^{lr}, the constrained features and HR output of the current group are generated by

$$f_g^h, h_g^{sr} = GSBRM(f_{g-1}^h, h_g^{lr}). \quad (2)$$

Based on the smoothing assumption of the network, spectral continuity between groups h_g^{sr} and h_{g-1}^{sr} should also exist in their features f_{g-1}^h and f_g^h. Thereby, we can constrain features of adjacent bands to have continuity in low spectral space for maintaining spectral correlation of hyperspectral images. Finally, we apply the reshuffle, which is the inverse operation of the group shuffle, to generate the hyperspectral results:

$$H^{sr} = Reshuffle([h_1^{sr}, h_2^{sr}, \ldots, h_g^{sr}, \ldots, h_G^{sr}]), \quad (3)$$

where $H^{sr} \in \mathbb{R}^{H\times W\times L}$ is the corresponding output HR hyperspectral image.

A. Group Shuffle

Hyperspectral data are a large number of continuous and narrow spectral images captured in the same static scene, which makes the adjacent spectral images incredibly similar. As shown

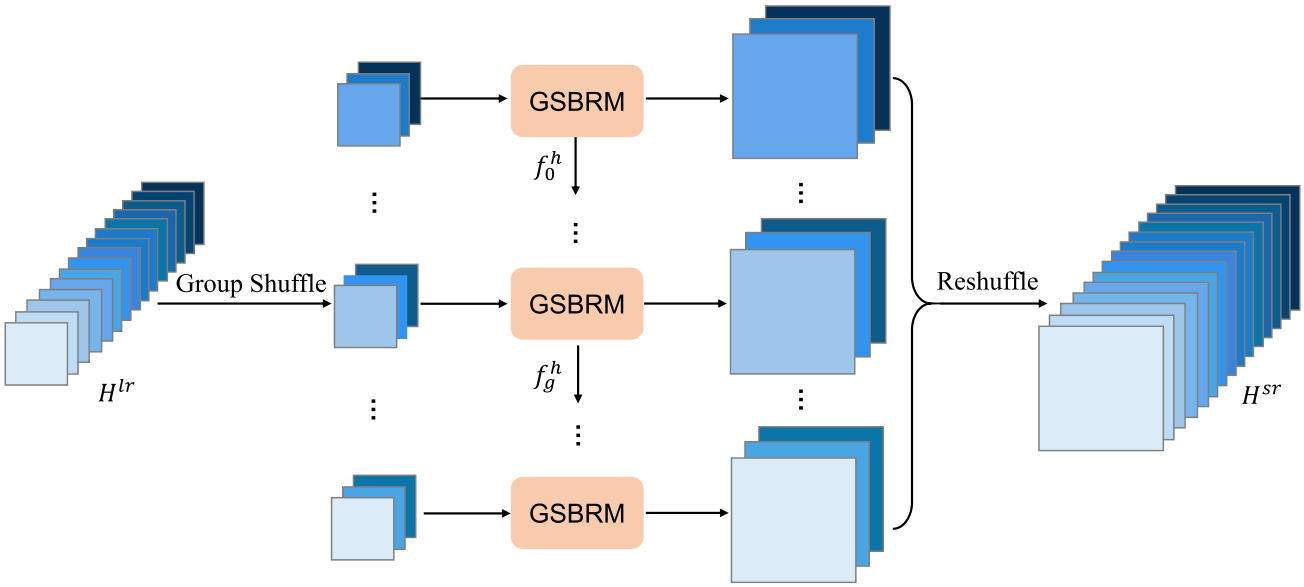


Fig. 4. The whole structure of the proposed GSSR, where the GSBRM stands for group spectral band reconstruction module. To facilitate the display, we set the number of spectral bands N to 15, the number of spectra in each group C to 3 and G to 5.

in Fig. 3(a), the previous methods directly cut a continuous segment of bands as a group to super-resolve, ignoring the information redundancy existing in extremely similar spectral bands.

Besides, group data of the previous method only cover a small spectral range of the reflection characteristics, which results in the spectral locality of the group bands. As shown in Fig. 1(a), the high inter-group similarity reveals the above defects of the original group data. On the other hand, we discover that the texture of an object may be clearer in some bands than in others as shown in Fig. 2(a). The different levels of texture detail in different bands allow certain dissimilar bands to complement each other in texture to enhance the super-resolution effect.

To this end, we propose the group shuffle to exploit the aforementioned band complementarity to tackle the drawbacks of the original group mode. We assume that N spectrograms of hyperspectral images are divided into G groups during group reconstruction, and each group has C bands. For the sake of description, we assume that N is divisible by C, which is $N/C = G$. To ensure the diversity of spectral response within the group, as shown in Fig. 3(b), we generate a group of data by sampling every G band interval. When the interval is G, the C bands within the group basically cover all spectral ranges of the hyperspectral images. Thus, the group shuffle expands the spectral range within the group while reducing information redundancy. Specifically, band i is mapped to the new position i' according to the following formula:

$$i' = f(i, G) = i \; mod \; G \times C + i \div G. \tag{4}$$

The spectral bands in the shuffled group contain broader spectral ranges, which makes the group data more diverse. At the same time, the neighboring spectral bands within the group are spaced by G, which means that they are no longer almost identical. In other words, the group data contain rich information. These can be corroborated by Fig. 1(b) that after group shuffle, the intra-group similarity greatly decreases, and the inter-group similarity increases. More intuitively, Fig. 2(b) also shows that the spectral responses of the spectral images are more diverse, which is advantageous for bands with less detail to obtain textures from other bands.

In summary, the newly proposed group generation mode called group shuffle separates adjacent and similar spectral images. This operation not only reduces the information redundancy of groups but also increases the spectral diversity within the group, greatly improving the effect of super-resolution. After the group shuffle, the spectral and texture information of the group data is richer, with which our model could reconstruct more spatial high-frequency details. Meanwhile, due to group shuffle, the spectral response between groups is more uniform. Therefore, it is easier to learn repeatable and reliable super-resolution patterns.

B. Group Spectral Band Reconstruction

After the group shuffle, the spectral bands of hyperspectral data are reordered. As mentioned above, the spectral bands within the group are spaced G from each other. To exploit the correlation among the bands in groups, we propose the GSBRM, which is dedicated to restoring the spatial structure and spectral correlation of HR hyperspectral images. Therefore, the design of the GSBRM needs to consider two issues: one is to fully extract and process abundant spatial and spectral information, and the other is to maintain spectral continuity in the case that the group shuffle changes the spectral order. To tackle these two problems, we propose SSFFB and LSCCM, respectively. The SSFFB constructs an efficient deep feature extraction and fusion block using separable 3D convolution, while LSCCM

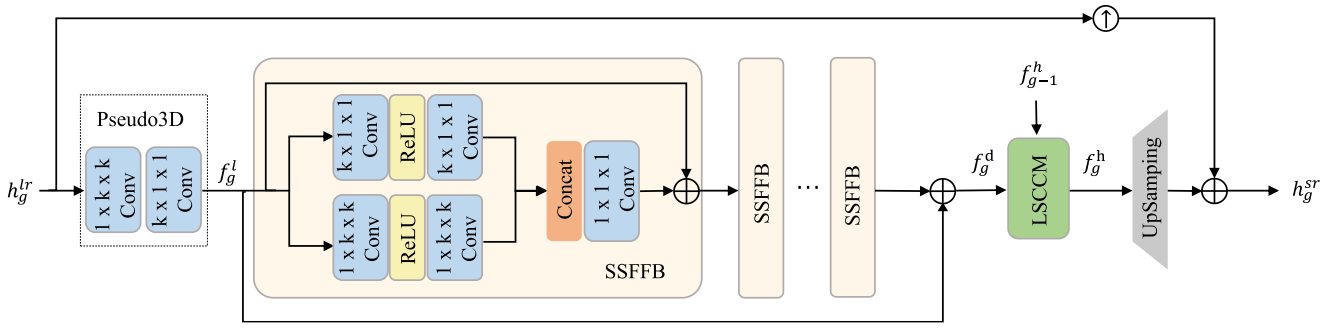


Fig. 5. The overall architecture of the proposed GSBRM.

constrains the local features of adjacent bands closer to maintain the spectral continuity of the hyperspectral images.

The structure of GSBRM is shown in Fig. 5. Let h_g^{lr} represent the target group to be super-resolved. Firstly, a standard separable 3D convolution called Pseudo-3D [38] is applied to obtain the shallow feature f_g^l:

$$f_g^l = Pseudo3D(h_g^{lr}), \tag{5}$$

where $Pseudo3D(\cdot)$ denotes Pseudo-3D convolution operation, which serially places 3D spatial and spectral separable convolution. Then, we stack spectral-spatial feature fusion blocks (SSFFBs) to build the feature extractor. Moreover, we also retain the shallow low-frequency information through the long-term residual connection. Consequently, we further obtain the deep feature f_g^d:

$$f_g^d = SSFFB_M(\cdots SSFFB_m(\cdots(SSFFB_1(f_g^l)))) + f_g^l, \tag{6}$$

where f_g^l denotes the shallow feature, f_g^d is the deep spectral-spatial feature, $SSFFB_m$ refers to m-th SSFFB, and M is the number of SSFFBs. To maintain spectral continuity, LSCCM is employed to interact with the previous group of features to generate the constrained features f_g^h:

$$f_g^h = LSCCM(f_g^d, f_{g-1}^h). \tag{7}$$

The output of LSCCM preserves the spatial structure and spectral correlation of grouped hyperspectral images. Thus, by the upsampling layer, we obtain the HR hyperspectral images which have excellent performance in the retention of spectral-spatial structure. In addition, to retain more low-frequency information, a global residual connection is introduced into GSBRM:

$$h_g^{sr} = U(f_g^h) + h_g^{lr}\uparrow, \tag{8}$$

where $h_g^{lr}\uparrow$ is the result of the bicubic interpolation of the LR input group, and $U(\cdot)$ refers to the upsampling operation which is implemented by PixelShuffle [39]. In the following, we will respectively give details of the SSFFB and the LSCCM.

1) Spectral-Spatial Feature Fusion Block: Hyperspectral images have abundant spectral information, which can provide great help to image SR. Meanwhile, the exploitation of spatial information, especially high-frequency information, has been the key to super-resolution problems. To make full use of the spectral and spatial information, an efficient and powerful feature extractor is essential. However, the existing feature extractors do not sufficiently consider the complementarity of spatial and spectral features, and each has its own shortcomings. 2D convolution cannot process spatial and spectral information at the same time. Compared with 2D convolution, 3D convolution adds an additional dimension to simultaneously process spectral information, but the resulting high computational complexity limits the practical value of the model. Pseudo-3D places 3D spatial (the kernel size is $1 \times k \times k$) and spectral (the kernel size is $k \times 1 \times 1$) separable convolutions in a serial way, which can expand the capacity of the model at the same cost. However, the shallow structure of Pseudo-3D determines that it is not capable of extracting deep features. Furthermore, they ignore the fusion of spectral and spatial features. To address these defects, as displayed in Fig. 5, we design the SSFFB. First, SSFFB contains two branches to extract spatial and spectral features respectively, where each branch consists of two separable 3D convolutions and a ReLU function to build the deep structure to expand the capacity of the model. Then, in order to leverage the complementary characteristics of spatial and spectral features, we concatenate the deep spectral and spatial features and fuse them by pointwise convolution (the kernel size is $1 \times 1 \times 1$). A short-term residual connection is used to keep gradient propagation stable as well. Compared with previous feature extractors, SSFFB has achieved both efficiency and reliability. On the one hand, SSFFB extracts and fuses deep spectral-spatial features based on the characteristics of hyperspectral images, which enhances the capability of the model. On the other hand, SSFFB utilizes separable 3D convolutions to maintain a low computational load. Subsequent experiments demonstrate that SSFFB extracts spectral-spatial features well to improve the performance.

2) Local Spectral Continuity Constraint Module: We extract rich spectral-spatial features by SSFFB, but in order to obtain high-quality hyperspectral HR images, we still need to maintain the spectral continuity of the generated results. After the group shuffle, the original interval of the neighboring spectrograms within the group is G. In other words, there are no original adjacent bands within the group, which is detrimental to preserving the spectral continuity of the generated hyperspectral images. Besides, previous methods constrain the results obtained by merging the outputs of each group in the hyperspectral and high-resolution space to maintain spectral correlation, which

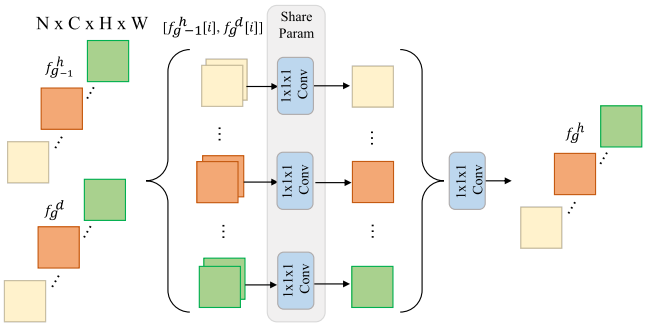


Fig. 6. The structure of the proposed LSCCM, where $f_{g-1}^{h}[i]$ denotes the constrained feature of i-th spectrogram of group $g-1$, $f_{g}^{d}[i]$ denotes the deep spectral-spatial feature of i-th spectrogram of group g.

requires high computational complexity and memory occupation.

Aiming to address these two issues, LSCCM restores the order of features of adjacent groups to constrain spectral correlation in low-resolution and low-spectral space. Specifically, by interval sampling, adjacent relationships occur between features at corresponding positions in adjacent groups. Naturally, it exists a local spectral continuity between the corresponding frequency bands of neighboring groups. In this way, we propose the LSCCM, the structure of which is shown in Fig. 6, where $f_{g-1}^{h}[i]$ denotes the constrained feature of i-th spectrogram of group $g-1$, $f_{g}^{d}[i]$ denotes the deep spectral-spatial feature of i-th spectrogram of group g. We first utilize a feedback mechanism to feed the previous group of constrained features to LSCCM. Then we segment the two groups of features in the spectral dimension and adjust the concatenating way accordingly, with which we recover the adjacency of features locally. Thereafter, we use a pointwise convolution of shared parameters to learn the continuous relationship between features of two adjacent spectral bands. In this context, the pointwise convolution can be regarded as a fusion module, which fuses the features of two consecutive spectral images to generate a more continuous spectral feature. For the deep feature of the current band, it is constrained to be closer to the previous spectral image feature in order to satisfy the continuity relation.

In this way, the global spectral continuity is constrained by restricting the local feature continuity between adjacent groups. To take advantage of the flexibility of neural networks, we do not explicitly constrain the features formally, but use convolution to learn such soft constraints from the data. Thus the learnable LSCCM module can be regarded as a soft constraint for features. Meanwhile, the spectral perception field of the group reconstruction module is expanded, which further improves the effect of the model. Moreover, by performing the neighborhood constraint in the low spectral space, the computational complexity and memory occupation of the model is reduced.

C. Implementation Details

Hyperspectral images have a total of N spectrograms. Each input group has C spectral bands. Suppose $p = N \mod C$, $G = N/C$. If p is equal to 0, that is, C is divisible by N, then after group shuffle, hyperspectral images are divided into G groups with C spectral bands in each group. When p is not equal to 0, that is, there is less than one group of the last p spectral images. Then the last C spectrograms of hyperspectral images are selected as the last group, so there are $G+1$ groups. The overlapping spectral bands will be averaged to obtain the final result after reconstruction. In addition, during the shuffle operation, the last p spectrograms retain the original index. The size k of the convolution kernel in the SSFFB is 3, and zero padding is also used to keep the spatial resolution. The number of SSFFB is M, and the number of feature maps in the convolution process is 64. In the upsampling module, the spectral dimension is squeezed first. ℓ_1 loss is adopted as the loss function in this paper. As for the training process, the Adam optimizer with an initial learning rate of 1e-4 and a weight attenuation rate of 1e-4 is used. The gradient optimization strategy is that the learning rate is halved every 35 training rounds, with a total of 100 training epochs. The model is trained with the Pytorch framework on the Tesla P100 with a batch size of 32.

IV. Experiments and Results

A. Datesets and Experimental Setup

The proposed model is evaluated on two commonly used benchmark datasets: CAVE [40] and Chikusei [41]. The CAVE dataset consists of natural hyperspectral images, while the Chikusei dataset contains one remote sensing hyperspectral image. Nine related super-resolution methods are selected for comparison on the two datasets, including a traditional baseline method such as Bicubic interpolation, a classical single image SR method such as EDSR [13], and seven state-of-the-art hyperspectral super-resolution algorithms, such as 3DFCNN [27], SSPSR [5], MCNet [34], GDRRN [28], RFSR [6], SFCSR [36] and Interactformer [37]. For EDSR, we directly adjust the input and output of the model from RGB to the corresponding spectral dimension. Under a unified training framework and in the same experimental environment, the mentioned comparison methods are fairly implemented on the three scale factors 2, 3, and 4 to compare with our GSSR.

Several widely used super-resolution reference indexes are used to evaluate the effectiveness of the model, including mean PSNR, spectral angle mapper (SAM), and structural similarity index (SSIM). Among them, PSNR and SSIM are usually evaluated by the SR effect of natural images. PSNR and SSIM are taken as the indicators of samples on average in this paper. A higher value of PSNR and SSIM indicates a better result, while SAM is the opposite.

B. Experimental Results on CAVE Dataset

The CAVE dataset includes 512×512 full-spectrum spatial resolution images at 10 nm steps (31 bands) from 400 nm to 700 nm. The CAVE is made up of 32 natural scenes. Among them, 25 scene images are randomly selected as the training set and the rest as the testing set. In view of the small size of the hyperspectral dataset, 24 image patches are randomly

TABLE I
QUANTITATIVE EVALUATION ON THE CAVE DATASET OF STATE-OF-THE-ART HYPERSPECTRAL IMAGE SR ALGORITHMS: AVERAGE PSNR/SSIM/SAM FOR SCALE FACTORS 2, 3, AND 4

Scale	Method	Param	PSNR↑	SSIM↑	SAM↓
×2	Bicubic	-	40.979	0.9635	2.701
	3DFCNN [27]	39k	42.652	0.9692	2.678
	GDRRN [28]	219k	43.391	0.9738	2.332
	EDSR [13]	5532k	43.968	0.9724	2.499
	RFSR [6]	1428k	44.672	0.9738	2.249
	MCNet [34]	1928k	45.369	0.9741	2.210
	Interactformer [37]	2161k	45.432	0.9742	2.209
	SFCSR [36]	1085k	45.321	0.9740	2.209
	Ours	1046k	**45.772**	**0.9747**	**2.202**
×3	Bicubic	-	37.709	0.9337	3.511
	3DFCNN [27]	39k	39.397	0.9433	3.211
	GDRRN [28]	219k	39.892	0.9471	3.171
	EDSR [13]	6270k	40.325	0.9502	3.095
	RFSR [6]	1624k	40.706	0.9521	2.829
	MCNet [34]	2038k	41.249	0.9531	2.785
	Interactformer [37]	2271k	41.298	0.9536	2.781
	SFCSR [36]	1270k	41.262	0.9625	2.804
	Ours	1230k	**41.521**	**0.9554**	**2.769**
×4	Bicubic	-	35.893	0.9085	3.980
	3DFCNN [27]	39k	37.647	0.9212	3.565
	GDRRN [28]	219k	38.027	0.9237	3.624
	EDSR [13]	6122k	38.338	0.9282	3.609
	SSPSR [5]	12875k	38.775	0.9314	3.302
	RFSR [6]	1603k	38.975	0.9314	3.302
	MCNet [34]	2174k	39.235	0.9326	3.198
	Interactformer [37]	2407k	39.313	0.9330	3.218
	SFCSR [36]	1233k	39.236	0.9319	3.216
	Ours	1193k	**39.401**	**0.9339**	**3.183**

The bold represents the best result and underline represents the second best.

selected from the training image and scaled at 1, 0.75, and 0.5 scale factors. Then, these patches are augmented by rotating 90 degrees, horizontally flipped, respectively, to obtain more training samples. Finally, the Bicubic interpolation is used to downsample these patches to a spatial resolution of 32×32 to generate training samples. The original images for testing are treated as ground-truth HR images, and the LR inputs are produced by scaling to the corresponding size. For the 31 spectral bands of the hyperspectral images in CAVE, they are divided into 11 groups with 3 spectra in each group, while the last group is completed by the last three spectral bands. That is, $C = 3$, and $G = 10$. M is set to 8 in this dataset.

Table I shows a comparison of the nine methods on the CAVE testing set, reporting the PSNR, SSIM, and SAM metrics, as well as the number of parameters for each method. For effective feature extraction, 3DFCNN adopts 3D convolution, which may preserve more information on spectral correlation. However, the computational complexity of 3D convolution is huge, especially in a large feature map size caused by pre-upsample mode. GDRRN makes a step forward to improve model efficiency by utilizing group convolution but does not perform as brightly as SSPSR and RFSR. This is because 2D convolution does not have enough ability to extract spectral information while processing spatial information. SISR methods such as EDSR focus on exploiting spatial features and thus ignore the importance of spectral information. MCNet and SFCSR all use a mixture of 2D and separable 3D convolution to extract spectral and spatial features. Interactformer achieves remarkable results consequently by using the transformer and 3D convolution network. However, the large model of Interactformer leads to extremely slow convergence and significantly increases memory requirements. It can be clearly seen that the proposed method performs the best in all the indicators of all the scale factors. Specifically, the average PSNR of our proposed method is 0.2 dB higher than that of the second-best method, and SAM and SSIM values are also significantly better than other methods. Due to group shuffle, our method could sufficiently exploit the spatial and spectral information. And the SSFFB fuses spectral-spatial features to gain a performance boost in an effective way. Thereby, our method could achieve excellent reconstruction performance.

In order to intuitively compare the reconstruction effects of all bands, we calculate the mean absolute error between the generated HR image and the ground-truth image from the spatial and spectral perspectives on the scale factor of 4. In Fig. 7, by averaging the absolute error in the spectral dimension, we obtain the spatial mean error. The visualization results reveal that the proposed method preferably reconstructs the spatial structure of HR hyperspectral images, especially in the processing of details and edge textures. This is mainly because after the group shuffle, the spectral response is richer in the group, thus the model can make full use of the bands with clearer texture to assist the reconstruction of the other bands. To further illustrate the advantages of our method in maintaining the spectral correlation, we produce the mean spectral error map in Fig. 8 on the scale factor of 4 by averaging the errors in spatial dimensions. Fig. 8 shows that the proposed method has a minimum error in all spectral channels. We also show the reconstructed images with an upscale factor of 4 produced by state-of-the-art methods to compare the visual effect of super-resolution in Fig. 9. Since it is impossible to exhibit numerous spectral images of the hyperspectral image simultaneously, we use the 26-th, 16-th, and 6-th spectral bands as R-G-B channels to generate the fake color RGB images. It can be seen that our method reconstructs the spatially detailed texture better.

The proposed GSSR method makes group data richer in texture information through the group shuffle, thus enhancing the ability of the model to extract high-frequency information. At the same time, in the group reconstruction, we design a dual-flow separable 3D convolution to deeply fuse the spectral-spatial features and LSCCM to restrict the continuity of adjacent bands of features to ensure the spectral continuity of hyperspectral images. In this way, our method explores the spectral-spatial correlations of hyperspectral images, leading to superior consequences.

C. Experimental Results on Chikusei Dataset

Chikusei is an aerial hyperspectral dataset taken by the Headwall Hyperspectral-VNIR-C sensor on July 29, 2014, in Tsukikeshi, Japan. It consists of a total of 128 spectral bands with a spectral range of 343-1018 nm. The spatial resolution is 2517×2357, and the spatial resolution of ground objects

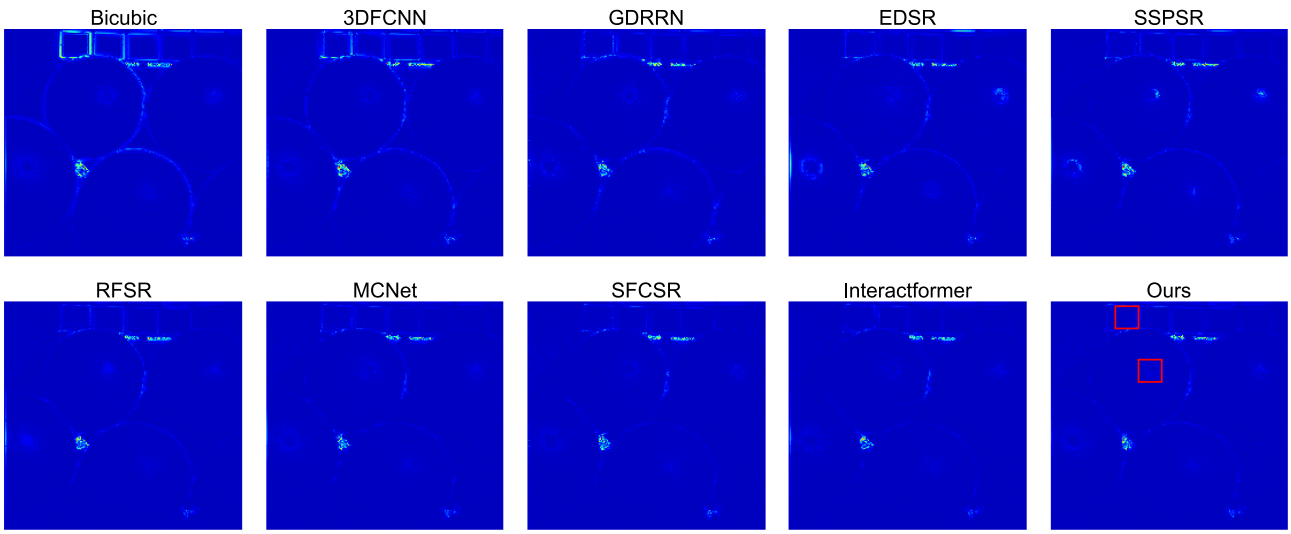


Fig. 7. Mean error maps of a test hyperspectral image in the CAVE dataset at the scale factor 4: balloons.

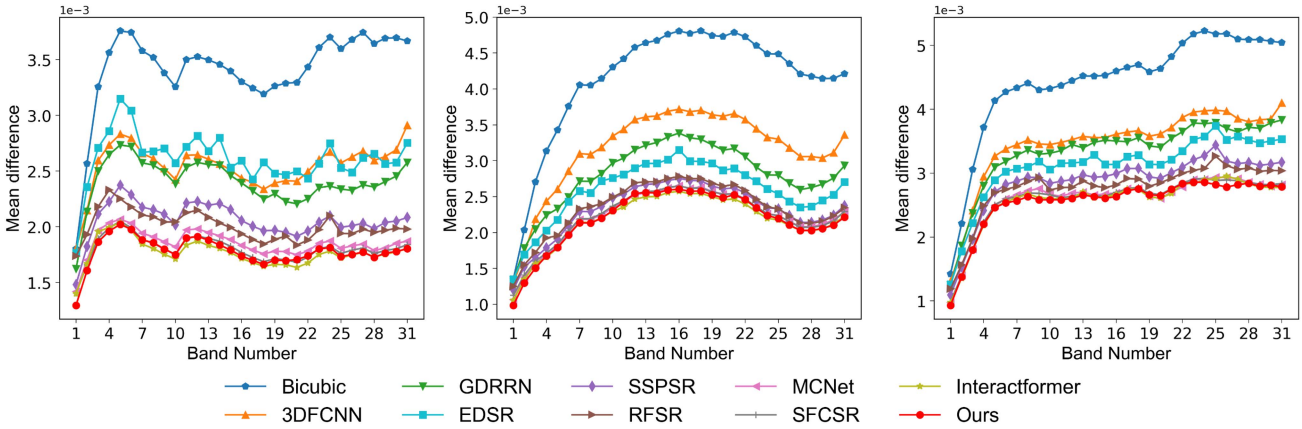


Fig. 8. The mean spectral difference curve of three test hyperspectral images in the CAVE dataset at the scale factor 4.

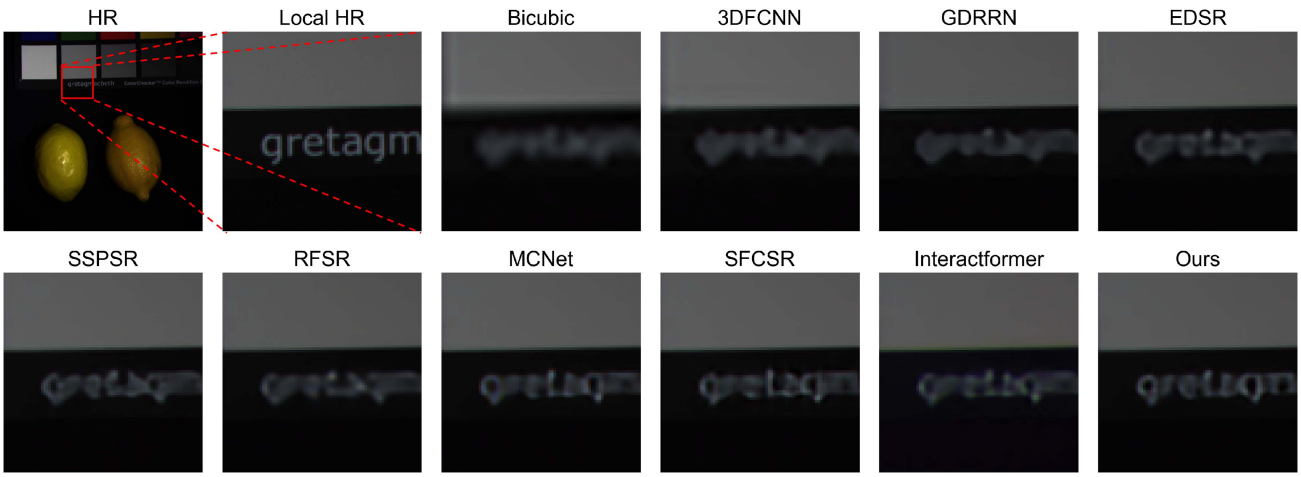
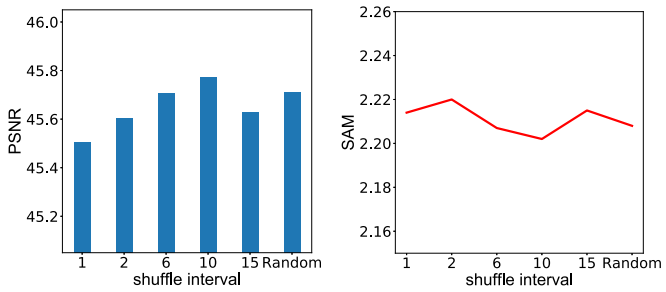


Fig. 9. Reconstructed images of a test hyperspectral image in the CAVE dataset with spectral bands 26-16-6 as R-G-B at the scale factor 4. The first subgraph is the Ground Truth, the second is a part of the Ground Truth in the red box, and the rest is the results of different methods.



(a) (b)

Fig. 10. Experiment results of the model at different shuffle intervals in the CAVE dataset (the band number N is 31) at the scale factor 2. Random on the horizontal axis represents random shuffling of all bands. (a) Histogram for PSNR. (b) Line graph for SAM.

is 2.5 m. Because the edge part of the image is missing, the pixels of $2304 \times 2304 \times 128$ in the center part of the image are first obtained and segmented as the training set and the testing set. In particular, the top region of the image is cut into four non-overlapping $512 \times 512 \times 128$ sub-images as the testing set. For the remaining areas, non-overlapping 32×32 image patches are sequentially segmented and enhanced as training sets, as on the CAVE dataset. For the Chikusei dataset, the number of spectral bands in each group is set as $C = 16$, $G = 8$, and the number of SSFFB is set as 10.

We also conduct a sufficient comparison experiment on the Chikusei dataset. Results compared with nine hyperspectral super-resolution methods are reported in Table II. It can be seen from the table that the proposed method is still the one with the best overall performance, especially in the reconstruction of spatial structure. In addition, as the number of spectral bands increases, methods using the group strategy such as RFSR and SSPSR are more outstanding on remote sensing hyperspectral datasets. On the other hand, methods based on the simultaneous reconstruction of all spectral bands, such as MCNet, 3DFCNN and GDRNN, are all inefficient. Because not only do they require a large memory occupation during training, but also they are incapable to capture the spectral correlation between multiple spectral bands. In this paper, we also adopt a group reconstruction strategy, but we innovatively adjust the spectral sequence to form a new group for reducing the information redundancy of the group data. At the same time, LSCCM is used to constrain local continuity in the neighborhood to maintain spectral continuity, which partially restores the spectral order adjusted by group shuffle.

Similarly, in the space dimension, Fig. 11 shows the mean spatial error map. Despite the overlapping objects and complex texture of remote sensing datasets, it can be seen visually that the method of this paper yields the most outstanding reconstruction. And in the spectral dimension, Fig. 12 shows the mean spectral error map. The error curve of the proposed method is clearly the lowest one, which represents that our method has superiority on all bands. Fig. 13 shows the reconstructed high-resolution hyperspectral image with an upscale factor of 4 by the comparison methods. Specifically, we use the 70-th, 100-th, and 36-th spectral bands as R-G-B channels to get a better visual effect. We can observe that the results of 3DFCNN, GDRNN, and EDSR are fuzzy, while SSPSR and RFSR introduce some artifacts, and the estimations of MCNet and SFCSR are too smooth. Our proposed method preserves the main image structure information with a relatively natural visualization. In general, our GSSR method still performs well by the group shuffle and group reconstruction in remote sensing hyperspectral datasets with more bands, which embodies robustness and generalization of methods.

TABLE II
QUANTITATIVE EVALUATION ON THE CHIKUSEI DATASET OF STATE-OF-THE-ART HYPERSPECTRAL IMAGE SR ALGORITHMS: AVERAGE PSNR/SSIM/SAM FOR SCALE FACTORS 2, 3, AND 4

Scale	Method	Param	PSNR↑	SSIM↑	SAM↓
×2	Bicubic	-	38.500	0.9592	1.758
	3DFCNN [27]	39k	39.974	0.9706	1.575
	GDRRN [28]	442k	41.560	0.9793	1.334
	EDSR [13]	5756k	41.349	0.9782	1.505
	RFSR [6]	725k	41.983	0.9811	1.237
	MCNet [34]	1928k	41.582	0.9797	1.337
	Interactformer [37]	2161k	42.091	0.9822	1.229
	SFCSR [36]	1085k	41.043	0.9777	1.476
	Ours	1320k	**42.537**	**0.9834**	**1.210**
×3	Bicubic	-	34.699	0.9025	2.647
	3DFCNN [27]	39k	35.805	0.9249	2.379
	GDRRN [28]	219k	36.866	0.9406	2.064
	EDSR [13]	6493k	37.088	0.9433	2.148
	RFSR [6]	956k	37.542	0.9487	1.871
	MCNet [34]	2038k	36.979	0.9422	2.078
	Interactformer [37]	2271k	37.592	0.9501	1.909
	SFCSR [36]	1270k	36.588	0.9384	2.279
	Ours	1528k	**37.784**	**0.9517**	**1.868**
×4	Bicubic	-	32.694	0.8494	3.358
	3DFCNN [27]	39k	33.580	0.8760	3.047
	GDRRN [28]	219k	34.278	0.8959	2.708
	EDSR [13]	6346k	34.601	0.9018	2.757
	SSPSR [5]	13546k	35.319	0.9165	**2.375**
	RFSR [6]	983k	34.939	0.9089	2.508
	MCNet [34]	2174k	34.488	0.8994	2.707
	Interactformer [37]	2407k	35.105	0.9124	2.495
	SFCSR [36]	1233k	34.122	0.8946	2.919
	Ours	1467k	**35.396**	**0.9173**	2.470

The bold represents the best result and underline represents the second best.

D. Shuffle Parameters

To explore the influence of the shuffle mode on the hyperspectral super-resolution model, we conduct adequate experiments on the CAVE dataset at the scale factor 2. We first explore the effect of interval in our model by fixing the number of bands in each group C to 3. We mainly change the interval in the shuffle function to verify our idea. In the previous comparison experiments, we set the interval to 10 on the CAVE dataset. Since the number of bands N is 31, we take several special intervals around 10 to observe the results, which are 2, 6 ($\sqrt{N}$ round), and 15 ($N/2$). In particular, we set the interval to 1 to represent the original group strategy. And we also experiment with shuffling all bands randomly to investigate the impact of separating adjacent bands. The experimental results in Fig. 10

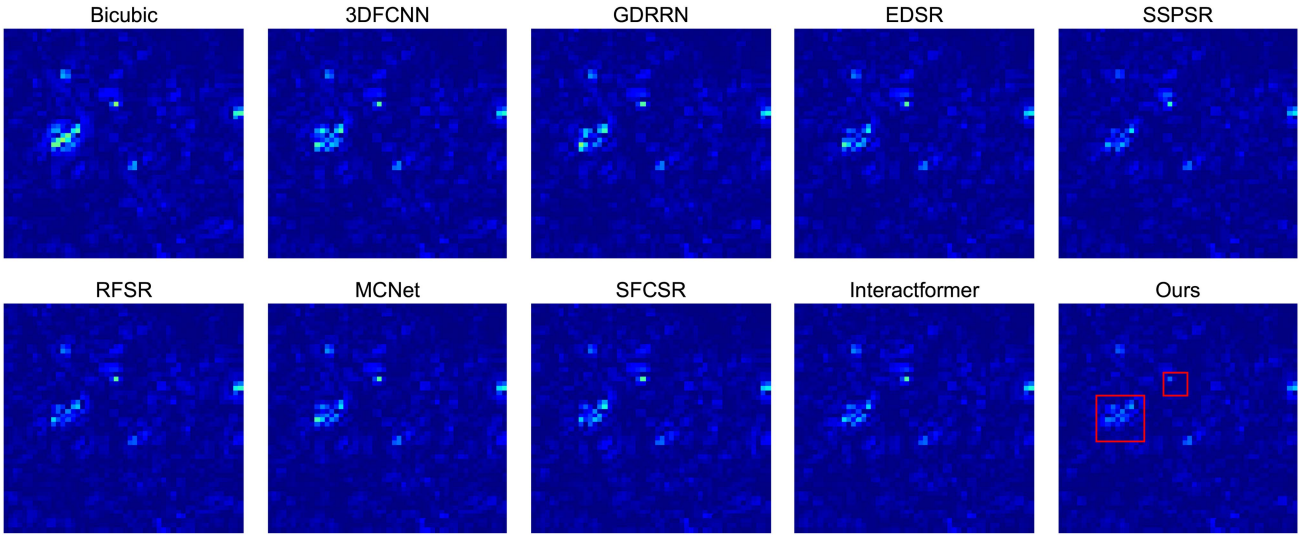


Fig. 11. Mean error map of a test hyperspectral image in the Chikusei dataset at the scale factor 4.

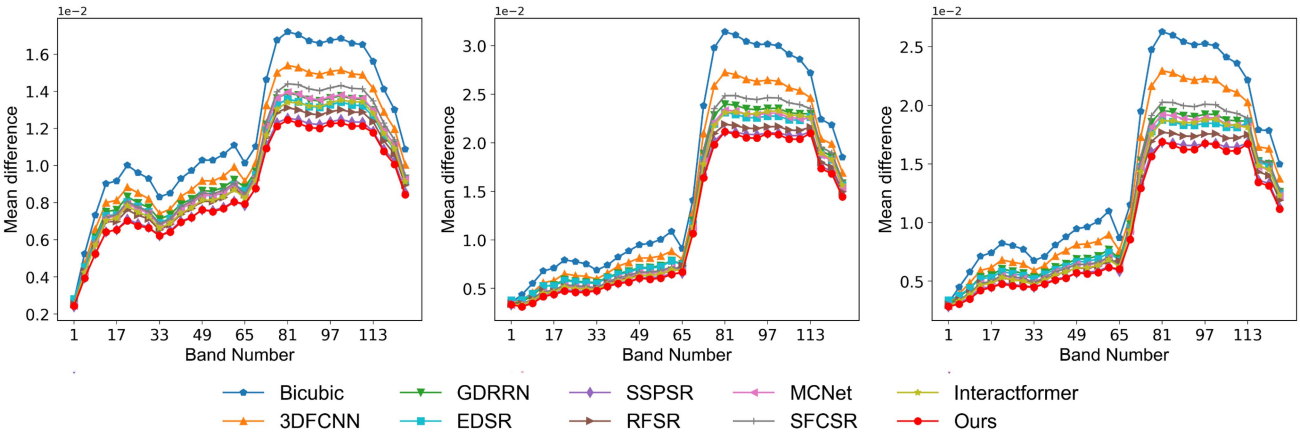


Fig. 12. Mean spectral difference curve of three test hyperspectral images in the Chikusei dataset at the scale factor 4.

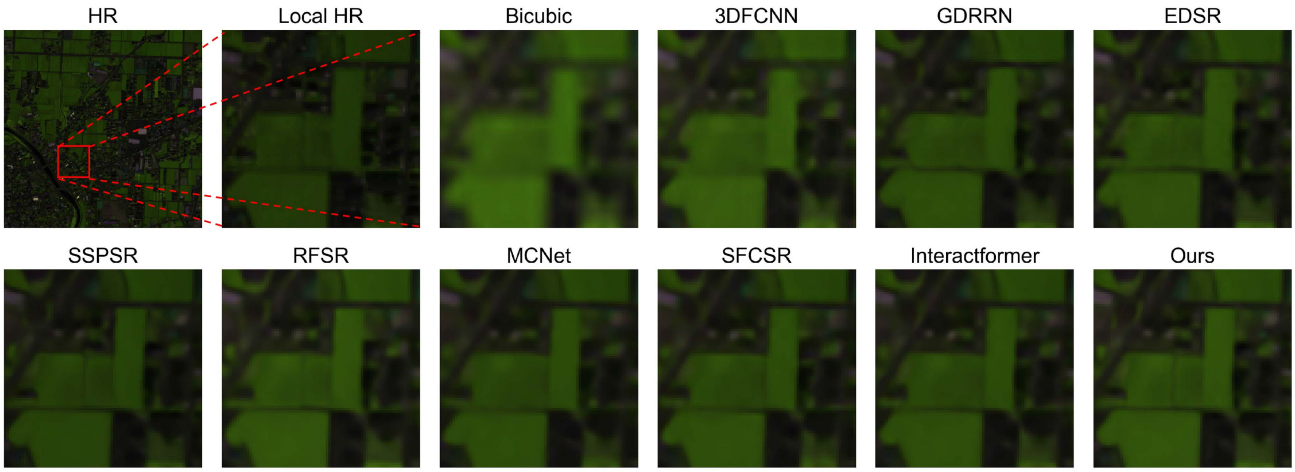


Fig. 13. Reconstructed images of a test hyperspectral image in the Chikusei dataset with spectral bands 70-100-36 as R-G-B at the scale factor 4. The first subgraph is the Ground Truth, the second is a part of the Ground Truth in the red box, and the rest is the results of different methods.

TABLE III
ABLATION STUDY

C	2	3	4	5	31
PSNR	45.609	**45.772**	45.716	45.636	44.827
SSIM	0.9742	**0.9747**	0.9746	0.9745	0.9738
SSAM	2.220	**2.202**	2.207	2.217	2.270

Quantitative comparisons among different C (number of bands in each group) in the CAVE dataset (the band number N is 31) at the scale factor 2.

show that there is a clear peak on the spatial indicator PSNR when the interval is 10. For the spectral index SAM, there is some fluctuation, but it also shows a clear valley at a horizontal coordinate of 10. Too small an interval does not introduce enough variability, while too large an interval means the number of spectra per group is small, which leaves too little information in each group. And from our derivation above, it is clear that when the interval is equal to N dividing C, each group of data is homogeneous and the intra-group spectra are far enough apart. Therefore, the interval value needs to be determined by N and C.

According to Fig. 10, the random shuffle mode achieves appreciable reconstruction results because it reduces the information redundancy and separates the adjacent and similar spectral bands. However, since the process is not reversible, it is difficult for the model to learn the continuity of neighboring spectra, which leads to poor spectral metrics compared to group shuffle. On the one hand, this illustrates that separating highly similar spectral images within groups is conducive to improving performance. On the other hand, we believe the recoverability of shuffle operation in the group shuffle could help the model to learn spectral continuity. In general, the experiment about shuffle mode further verifies our idea, that by adjusting the spectral order to form new groups, the complementary properties of the bands can be better exploited to enhance the effect.

In our method, C is a hyperparameter that varies in different datasets, thus it needs to be determined experimentally. For the CAVE dataset with the scale factor of 2 and band number of 31, we change the value of C and fix the interval in group shuffle to N/C to search for the best C, the result of which are listed in Table III. We find that our model performs best when $C = 3$ in the CAVE dataset. When the number of spectra in the group is too small, the model cannot obtain enough spectral information. And if the number of spectra in a group is too large, the efficiency of the model and the effectiveness of group shuffle are reduced. Especially, when using the whole datacube as input ($C = N$) obtains the worst performance. On the one hand, the number of bands in the whole datacube is too many, leading to a larger computational and spatial complexity. On the other hand, the interval equals 1(interval $= N/C$), which means directly group rather than group shuffle. Therefore, there is a precipitous drop in the super-resolution effect.

E. Ablation Study

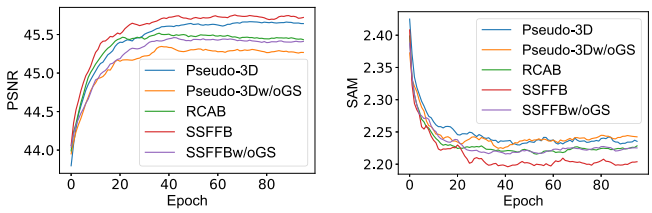
1) Spectral-Spatial Feature Fusion Block: In order to deeply extract and flexibly fuse spatial and spectral information, we develop a spectral-spatial feature fusion block, which introduces a deep dual-flow structure. To verify the effect of SSFFB, we involve multiple sets of experiments on the CAVE dataset at scale factor 2. First, we compare SSFFB with the commonly used 2D feature extractor RCAB [15] and Pseudo-3D [38] to prove the effectiveness of our proposed SSFFB. In addition, we also compare the results of Pseudo-3D and SSFFB without group shuffle for investigating the performance of the 3D feature extractor under different input distributions. Finally, we design three variants of SSFFB to explore the mechanism of SSFFB. To explore the importance of feature fusion, we try to exploit spatial features and spectral features to SR separately. Specifically, spatial-conv stands for reserving $1 \times k \times k$ convolution to extract spatial information, while spectral-conv reserves $k \times 1 \times 1$ convolution. And shallow-SSFFB removes one convolution in each of the two branches to verify the validity of the deep structure. For a fair comparison, we endeavor to keep the parameters of each model roughly in the same order of magnitude by changing the number of modules. Replacing only the feature extractor, we obtain the experimental results in Table IV. To better represent the results, we also plot the convergence curves of the compared variants mentioned above.

As shown in Fig. 14, RCAB converges faster, but the final result is poorer due to the limited capability of 2D convolutional-based feature extractors. Although Pseudo-3D gains its ability to super-resolve spatially by processing both spectral and spatial features using separable 3D convolution, it neglects the fusion of spectral and spatial features, and thus performs relatively poorly

TABLE IV
ABLATION STUDY

Type	Param	PSNR	SSIM	SAM	Num*
RCAB	1084k	45.583	0.9742	2.220	11
Pseudo-3D	1029k	45.693	0.9749	2.219	17
Pseudo-3D w/o GS	1029k	45.379	0.9731	2.230	17
SSFFB w/o GS	1045k	45.510	0.9741	2.212	8
spatial-conv	1004k	45.538	0.9741	2.216	11
spectral-conv	1056k	44.367	0.9730	2.331	35
shallow-SSFFB	1054k	45.719	0.9747	2.210	15
SSFFB	1045k	45.772	0.9747	2.202	8

*The number of modules.
Quantitative comparisons among different types of feature extractors over the testing set of CAVE dataset at the scale factor 2. GS means group shuffle.



(a) (b)

Fig. 14. The convergence curves for part of methods mentioned in the ablation experiments about SSFFB. (a) The curves of PSNR. (b)The curves of SAM.

TABLE V
ABLATION STUDY QUANTITATIVE COMPARISONS AMONG DIFFERENT COMPONENTS OVER THE TESTING SET OF CAVE DATASET AT THE SCALE FACTOR 2

group shuffle	✗	✓	✓
LSCCM	✗	✗	✓
LSCCM w/o RO	✗	✓	✗
PSNR	45.563	45.573	**45.772**
SSIM	0.9740	0.9746	**0.9747**
SAM	2.225	2.226	**2.202**

RO means recover order.

in spectral recovery. Therefore, there is a significant difference between the spectral metrics SAM of Pseudo-3D and SSFFB. In addition, the performance of Pseudo-3D w/o GS is not outstanding, which indicates that Pseudo-3D has excellent PSNR value because the group shuffle operation greatly improves the super-resolution effect. It is also obvious from the convergence curves that the convergence speed and value of Pseudo-3D on the spectral index SAM are not comparable to other methods. On the contrary, SSFFB maintains a better feature extraction ability under different input distributions, especially in keeping the spectral correlation. The convergence of SSFFB also shows a significant advantage in Fig. 14. Besides, according to the results in Table IV, spatial information is the main factor of the reconstruction effect, while spectral information is also contributing to improved performance. The results of the spatial and spectral convolution support the complementary properties of the spatial and spectral features. The comparison with the shallow-SSFFB illustrates that the deeper structure benefits feature extraction, but is not as important as fused features. In conclusion, SSFFB deeply fuses complementary spatial and spectral features to achieve the best performance.

2) Local Spectral Continuity Constraint Module: After the group shuffle, there is no contiguous relationship between the spectral images within the group. As deduced earlier, there is still a local continuity relationship between the groups. Therefore, we propose an LSCCM to learn the spectral continuity locally. LSCCM constrains the local feature between adjacent bands closer to keep the global spectral continuity. On the other hand, the spectral field of our method is broader through feature interaction between groups, which is also accomplished by LSCCM. In this section, we analyze the impact of LSCCM on the performance of the model on the CAVE dataset at scale factor 2. Since the LSCCM has been designed to recover the spectral continuity after group shuffle, we combine them in an experiment to verify their effectiveness. As shown in Table V, group shuffle companying with our proposed LSCCM achieves a significant improvement in terms of PSNR and SAM values. Besides, we have designed a variant of LSCCM that after obtaining the features of the previous group and current group, we do not recover the spectral order to impose the constraint, but directly concatenate them together to process them by convolution. Table V demonstrates that not restoring the order after group shuffle leads to more severe spectral distortion, which indicates that restoring the original spectral order with group shuffle is essential for learning spectral continuity.

V. CONCLUSION

In this paper, we introduce a novel hyperspectral super-resolution network with the group shuffle, named GSSR. In our method, we innovatively separate the adjacent and extremely similar spectral images to form new groups containing richer spectral and the texture information. Furthermore, the LSCCM is proposed to constrain the local features of adjacent groups to ensure spectral continuity. Besides, we design a powerful feature extractor block called SSFFB, which efficiently extracts and fuses spatial and spectral features by separable 3D convolution to improve the reconstruction performance. The qualitative and quantitative experiments on natural and remote sensing hyperspectral datasets demonstrate that our model not only achieves the best performance on objective indicators but also on subjective visualization.

REFERENCES

[1] F. F. Sabins, "Remote sensing for mineral exploration," *Ore Geol. Rev.*, vol. 14, no. 3–4, pp. 157–183, 1999.

[2] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, 2014, Art. no. 010901.

[3] A. Lowe, N. Harrison, and A. P. French, "Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress," *Plant Methods*, vol. 13, no. 1, pp. 1–12, 2017.

[4] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Deep unfolding network for spatiospectral image super-resolution," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 28–40, 2022.

[5] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1082–1096, 2020.

[6] X. Wang, J. Ma, and J. Jiang, "Hyperspectral image super-resolution via recurrent feedback embedding and spatial–spectral consistency regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503113.

[7] 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.

[8] P. Yi, Z. Wang, K. Jiang, J. Jiang, T. Lu, and J. Ma, "A progressive fusion generative adversarial network for realistic and consistent video super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2264–2280, May 2022.

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

[10] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[11] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.

[12] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3147–3155.

[13] 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.*, 2017, pp. 136–144.

[14] 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. 4681–4690.

[15] 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. 286–301.

[16] 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. 5690–5699.

[17] T. Dai, J. Cai, Y. Zhang, S. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11065–11074.

[18] S. A. Magid et al., "Dynamic high-pass filtering and multi-spectral attention for image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4288–4297.
[19] D. Fuoli, L. Van Gool, and R. Timofte, "Fourier space losses for efficient perceptual image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2360–2369.
[20] L. Ji et al., "Cross-domain heterogeneous residual network for single image super-resolution," *Neural Netw.*, vol. 149, pp. 84–94, 2022.
[21] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.
[22] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and spatial–spectral group sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1250–1254, Sep. 2016.
[23] Y. Wang et al., "Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1286.
[24] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.
[25] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6055–6067, Aug. 2019.
[26] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, 2017.
[27] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1139.
[28] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *Proc. IEEE 4th Int. Conf. Multimedia Big Data*, 2018, pp. 1–4.
[29] Z. He and L. Liu, "Hyperspectral image super-resolution inspired by deep laplacian pyramid network," *Remote Sens.*, vol. 10, no. 12, 2018, Art. no. 1939.
[30] J. Li, R. Cui, Y. Li, B. Li, Q. Du, and C. Ge, "Multitemporal hyperspectral image super-resolution through 3D generative adversarial network," in *Proc. 10th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2019, pp. 1–4.
[31] R. Jiang et al., "Learning spectral and spatial features based on generative adversarial network for hyperspectral image super-resolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3161–3164.
[32] J. Li, R. Cui, B. Li, Y. Li, S. Mei, and Q. Du, "Dual 1D-2D spatial-spectral CNN for hyperspectral image super-resolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3113–3116.
[33] Q. Wang, Q. Li, and X. Li, "Spatial-spectral residual network for hyperspectral image super-resolution," 2020, *arXiv:2001.04609*.
[34] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1660.
[35] Q. Li, Q. Wang, and X. Li, "Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8693–8703, Oct. 2021.
[36] Q. Wang, Q. Li, and X. Li, "Hyperspectral image superresolution using spectrum and feature context," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11276–11285, Nov. 2021.
[37] Y. Liu, J. Hu, X. Kang, J. Luo, and S. Fan, "Interactformer: Interactive transformer and CNN for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5531715.
[38] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3D residual networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5533–5541.
[39] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.
[40] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
[41] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over chikusei," Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27, May 2016.

Xinya Wang received the B.S. degree in 2018 from the Electronic Information School, Wuhan University, Wuhan, China, where she is currently working toward the Ph.D. degree with the Multi-Spectral Vision Processing Lab. Her research interests include neural networks, machine learning, and image processing.

Yingsong Cheng received the B.S. degree from the Computer Science and Technology School, Huazhong University of Science and Technology, Wuhan, China, in 2022. He is currently working toward the master's degree with the Electronic Information School, Wuhan University, Wuhan, China. His research interests include computer vision and image processing.

Xiaoguang Mei received the B.S. degree in communication engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2007, the M.S. degree in communications and information systems from Central China Normal University, Wuhan, in 2011, and the Ph.D. degree in circuits and systems from the HUST, in 2016. From 2010 to 2012, he was a Software Engineer with the 722 Research Institute, China Shipbuilding Industry Corporation, Wuhan. He is currently an Associate Professor with Wuhan University. His research interests include hyperspectral image processing, image fusion, and machine learning.

Junjun Jiang (Senior Member, IEEE) received the B.S. degree from the Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His research interests include image processing and computer vision. He was the recipient of the Finalist of the World's FIRST 10 K Best Paper Award at ICME 2017, Best Student Paper Runner-up Award at MMM 2015, 2016 China Computer Federation (CCF) Outstanding Doctoral Dissertation Award, and 2015 ACM Wuhan Doctoral Dissertation Award.

Jiayi Ma (Senior Member, IEEE) received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. He is currently a Professor with the Electronic Information School, Wuhan University. He has authored or coauthored more than 300 refereed journal and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, and ECCV. His research interests include computer vision, machine learning, and pattern recognition. Dr. Ma has been identified in the 2019–2022 Highly Cited Researcher lists from the Web of Science Group. He is the Area Editor of *Information Fusion*, and an Editorial Board Member of *Neurocomputing*.