Hyperspectral Image Super-Resolution via Recurrent Feedback Embedding and Spatial–Spectral Consistency Regularization

Xinya Wang, Jiayi Ma, *Member, IEEE*, and Junjun Jiang, *Member, IEEE*

***Abstract*— Hyperspectral images with tens to hundreds of spectral bands usually suffer from low spatial resolution due to the limitation of the amount of incident energy. Without auxiliary images, the single hyperspectral image super-resolution (SR) method is still a challenging problem because of the high-dimensionality characteristic and special spectral patterns of hyperspectral images. Failing to thoroughly explore the coherence among hyperspectral bands and preserve the spatial–spectral structure of the scene, the performance of existing methods is still limited. In this article, we propose a novel single hyperspectral image SR method termed RFSR, which models the spectrum correlations from a sequence perspective. Specifically, we introduce a recurrent feedback network to fully exploit the complementary and consecutive information among the spectra of the hyperspectral data. With the group strategy, each grouping band is first super-resolved by exploring the consecutive information among groups via feedback embedding. For better preservation of the spatial–spectral structure among hyperspectral data, a regularization network is subsequently appended to enforce spatial–spectral correlations over the intermediate estimation. Experimental results on both natural and remote sensing hyperspectral images demonstrate the advantage of our approach over the state-of-the-art methods.**

***Index Terms*— Feedback embedding, hyperspectral image, recurrent network, super-resolution (SR).**

I. Introduction

HYPERSPECTRAL imaging sensors collect the reflectance information of objects over a certain electromagnetic spectrum [1], and each pixel is dispersed to form tens to hundreds of continuous spectral bands. Compared with the multispectral image (MSI) or natural image, the hyperspectral image possesses a higher data dimension and captures more abundant spectral information, which reflects the subtle spectral properties of different objects. Therefore, the hyperspectral images have a superior diagnostic ability to distinguish visually similar materials. It has been widely used in computer vision and remote sensing fields, such as mineral exploration [2], medical diagnosis [3], and plant detection [4].

However, due to the narrow and dense spectral bands in the hyperspectral imaging systems, the incident energy that can reach the sensor is rather limited. There exists a tradeoff between spatial resolution and spectral resolution. To obtain rich spectral information, the generated hyperspectral images always have a relatively low spatial resolution to reach a compromise, which confines its further application. As the hardware sensors are difficult to improve, it is urgent to develop software technology to improve the spatial resolution of the hyperspectral image.

Super-resolution (SR) is a postprocessing technique to reconstruct the high-resolution (HR) image from one or more low-resolution (LR) inputs instead of modifying the imaging hardware. Most existing hyperspectral image SR methods resort to HR auxiliary images, such as the MSI, the panchromatic (PAN) image, and the RGB image, which are fused with the LR hyperspectral image to improve the spatial resolution. Leveraging Bayesian inference, matrix factorization, sparse representation, or recently advanced deep learning techniques, fusion-based methods have played a dominant role in recent years and achieved considerable performance [5]–[12]. With HR prior information, these fusion methods require that the auxiliary images should be captured at the same scene as the LR hyperspectral image, even be first co-registered which is also challenging [13].

Singe hyperspectral image SR methods have the advantage that the only LR image is needed to infer the corresponding HR one, which is less studied because of the spectral patterns in hyperspectral images. Early, several single image methods rely on the handcrafted prior (self-similarity, sparseness, or low rank) to recover the HR hyperspectral image, which neglects the inherent spatial–spectral properties in hyperspectral images [14]–[17]. Due to the high dimension of hyperspectral data, the reconstruction quality of these methods is quite limited. With the prosperity of the deep neural network (DNN), some two-step methods, combining the collaborative nonnegative matrix factorization (CNMF) with the DNN, or end-to-end DNN methods have been developed to super-resolve the LR

Manuscript received December 22, 2020; revised February 3, 2021; accepted March 4, 2021. Date of publication March 17, 2021; date of current version December 9, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61773295 and Grant 61971165, in part by the Key Research and Development Program of Hubei Province under Grant 2020BAB113, and in part by the Natural Science Foundation of Hubei Province under Grant 2019CFA037. *(Corresponding author: Jiayi Ma.)*

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

Junjun Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: junjun0595@163.com).

Digital Object Identifier 10.1109/TGRS.2021.3064450

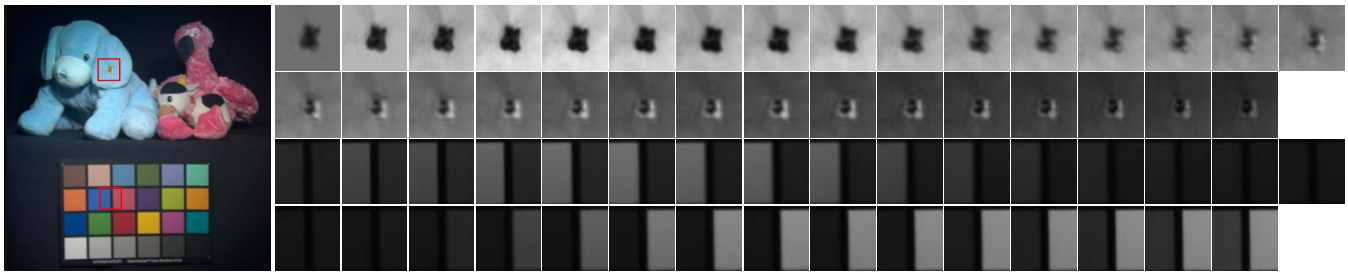


Fig. 1. Example of consecutive changes among spectral bands. (From Left to Right and Top to Bottom) Two small areas from stuffed_toys in the CAVE data set are selected to show the complementarity and continuity among 31 spectral bands.

hyperspectral image. Nevertheless, it is still a challenging task to customize an effective and less complex algorithm for the single hyperspectral image SR. The challenges mainly come from three aspects. First, the available samples in the hyperspectral image data sets are limited, making it difficult to fit a large model. Second, the separate utilization of spatial information and spectral information in the two-step methods will inevitably lead to spatial–spectral inconsistencies, and the performance of CNMF highly depends on the iteration times, the number of the endmembers, and so on. Third, the existing end-to-end DNN methods ignore the continuity among spectral bands, failing to restore high-quality HR hyperspectral images.

To address the above three challenges, this article proposes a novel and effective learning-based method for single hyperspectral image SR, termed RFSR, which can accurately reconstruct hyperspectral images to be spatial–spectral coherent. We observe that capturing from the same static scene, the spectral bands in hyperspectral data can be regarded as a clip of frame sequences with the constant wavelength difference. As shown in Fig. 1, the object exhibits continuous light and dark changes among the hyperspectral bands. There exists a strong correlation among the neighboring spectra due to the narrowbands in the hyperspectral imaging systems. Based on this observation, we introduce a recurrent feedback network (RFNet) that aims to model complementarity and continuity among spectral bands via feedback embedding. Specifically, imitating the sequence modeling, the reconstructed result of the previous spectral bands will be fed back to the current spectral bands for guiding the high-frequency information learning. To improve efficiency, the network is recurrent among spectral band groups, so that the model size will not increase significantly with the number of spectrum bands. After acquiring an intermediate super-resolved hyperspectral image, we further deploy a structural consistency regularization network by efficient Pseudo-3D convolution, which explores the spatial–spectral coherence among the intermediate result and enforces the structural consistency in the HR space. The qualitative and quantitative experiments on various image data demonstrate that the proposed RFSR method can produce spatial–spectral consistent SR results in comparison with other state-of-the-art methods.

In summary, our contributions include the following three aspects: 1) a novel RFNet with the consistency regularization is proposed to address single hyperspectral image SR from a sequence modeling perspective; 2) dividing the hyperspectral spectra into subgroups, our method improves the spatial resolution of each group with the guidance of feedback embedding from the previous group, which could make full use of the complementary and consecutive information; and 3) a subsequent consistency regularization network is appended to explore the spatial–spectral correlations in the HR space, leading to spatial–spectral consistent results.

II. Related Works

SR has been extensively studied in recent years. Here, we briefly review some methods that are most relevant to our work, including single gray/RGB image SR and single hyperspectral image SR.

A. Single Gray/RGB Image SR

Recently, the deep convolutional neural network (DCNN) has shown the extraordinary capability of learning a mapping function between LR and HR image pairs in an end-to-end manner, which has achieved excellent performance in nature images. Dong *et al.* [18] pioneered a three-layer convolutional neural network in RGB image SR task, and the proposed SRCNN method showed better performance than the traditional SR methods. Subsequently, some deep networks that involve skipping connection and residual learning [19] were developed for better performance, including the very deep network for SR (VDSR) [20], deeply recursive convolutional network (DRCN) [21], deeply recursive residual network (DRRN) [22], and enhanced deep SR (EDSR) network [23]. By introducing a generative adversarial network [24], SRGAN was proposed in [25] and [26] for photorealistic SR. To further improve the performance, many strategies were designed to reconstruct more high-frequency information, such as iterative up-and-down samplings in [27], spatial attention mechanism in [28], feedback mechanism in [29], and cross-scale nonlocal attention in [30]. Although the single gray/RGB image SR methods have flourished, achieving considerable performance, most of them cannot be applied to hyperspectral image SR. This is mainly due to the following reasons. On the one hand, when these methods processing one-channel images are used to super-resolve hyperspectral images in a band-by-band manner, they would neglect the spectral correlations among spectra of the hyperspectral data, leading to spectral distortion. On the other hand, a large number of parameters, caused by hundreds to thousands of feature maps for

pursuing good results, lack enough hyperspectral data for training.

B. Single Hyperspectral Image SR

Single hyperspectral image SR does not need any other prior or auxiliary information, which has better feasibility in practice. Akgun *et al.* [14] first modeled the hyperspectral image acquisition process from different wavelengths as weighted linear combinations of a small number of basis image planes. Based on sparse representation, Li *et al.* [31] proposed a novel hyperspectral image SR framework utilizing spectral mixture analysis and spatial–spectral group sparsity. In [15], the low-rank and total variation prior was incorporated to regularize the reconstruction process of hyperspectral image. However, these methods that use the handcrafted priors are complex and time-consuming to optimize, failing to recover more details. In addition, without any external information, these handcrafted priors only leverage the internal example itself, having limited representation power. Recently, several CNN-based methods have been proposed for hyperspectral image SR. To alleviate spectral distortion, a spectral difference convolutional neural network was proposed in [32] and [33] with the combination of a postprocessing strategy. Based on matrix factorization, Yuan *et al.* [34] developed a two-step method that combines the CNMF with the CNN method to enforce collaborations between the LR and HR hyperspectral images. Similarly, deep feature matrix factorization method [35] blended feature matrix extracted by a DNN with CNMF strategy for super-resolving real-scene hyperspectral image. To further exploit spatial–spectral information, Hu *et al.* [36] also integrated the CNMF strategy with an intrafusion operation that conducted on the results from the deep information distillation network. Although these two-step methods have achieved good results, they consider the spatial and spectral information separately and rely heavily on manual processing, such as CNMF and key band selection, which are time-consuming and unstable. Recently, several end-to-end DNN-based methods were proposed. To exploit both the spatial and spectral correlations of neighborhood, Mei *et al.* [37] employed a 3D fully CNN (3DFCNN) method. Although 3D convolution may preserve more information of spectral correlation, the computational complexity is large, especially at a large scaling factor. Due to the correlation among spectral bands, a grouping strategy is proposed. Li *et al.* [38] deployed a grouped deep recursive residual network (GDDRN) [38], in which a grouped recursive module was used with the global residual learning. Sharing the same insight, a spatial–spectral prior SR (SSPSR) method [39] extracted features of each group independently and reconstructed the HR results in a progressive way. However, all of these methods neglect the consecutive information among spectral bands, leading to spatial–spectral inconsistent estimation.

To this end, we imitate the sequence modeling methods to capture the complementary and consecutive information among spectral bands by a recurrent feedback mechanism. In addition, a spatial–spectral regularization network is further employed to enforce the structural consistency in the HR space.

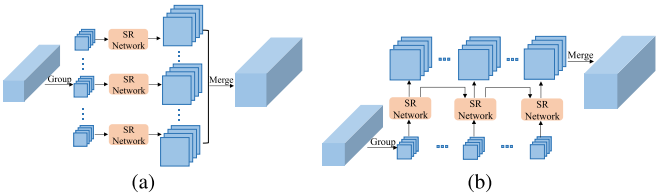


Fig. 2. Comparison of group-based methods. (a) Architecture of the existing group-based methods. (b) Architecture of the proposed method.

III. Motivation

Given an LR hyperspectral image, denoted as $H^{\mathrm{lr}} \in \mathbb{R}^{h\times w\times L}$, hyperspectral image spatial SR aims to reconstruct a super-resolved hyperspectral image, close to the ground-truth HR hyperspectral image $H^{\mathrm{hr}} \in \mathbb{R}^{\alpha h\times \alpha w\times L}$, where $h \times w$ is the spatial resolution, L is the number of the spectral band, and α is the upsampling factor. We believe that the following two aspects that are not fully explored in the existing methods are paramount for high-quality hyperspectral SR: 1) a thorough utilization for the complementarity and continuity among spectral bands and 2) a strict constraint for spatial–spectral correlations. Thus, we will discuss these two aspects, which share the designing insight of our proposed method.

A. Complementarity and Continuity Among Spectral Bands

A hyperspectral image contains rich spectral information from different wavelengths. Due to the narrow spectral bands in the hyperspectral imaging systems, the adjacent bands that look quite similar have quite strong correlations with each other. It has been proved that the redundant information among neighborhood bands would be helpful for high-quality SR [38], [39]. Taken the whole hyperspectral image as input, such as the natural image, the model would require hundreds to thousands of feature maps to achieve the ideal performance, which lacks enough hyperspectral images for training. In existing methods, a popular strategy is to group the spectral bands for efficient reconstruction in an overlapping or nonoverlapping way [35], [39]. Although these group-based methods improve both the reconstruction quality and the computational efficiency, the consecutive information among spectra has not been fully exploited without considering the relationship among band groups. Fig. 2(a) shows the architecture of existing group-based SR methods, in which each group band is reconstructed separately, neglecting the continuous relationship among band groups.

An intuitive way to fully take advantage of the cross-group relationship is by passing on the previous group information to guide high-frequency detail learning for the current group, as shown in Fig. 2(b). In this way, we propose a recurrent feedback mechanism for hyperspectral image SR, which super-resolves each spectrum group by combining the feedback embedding from the previous group. Therefore, the proposed method could not only utilize the intergroup correlation but also consider the cross-group continuity, which is beneficial for the reconstruction process.

B. Spatial–Spectral Consistency in HR Space

As the most important property of a hyperspectral image, the spatial–spectral consistency should be well preserved after

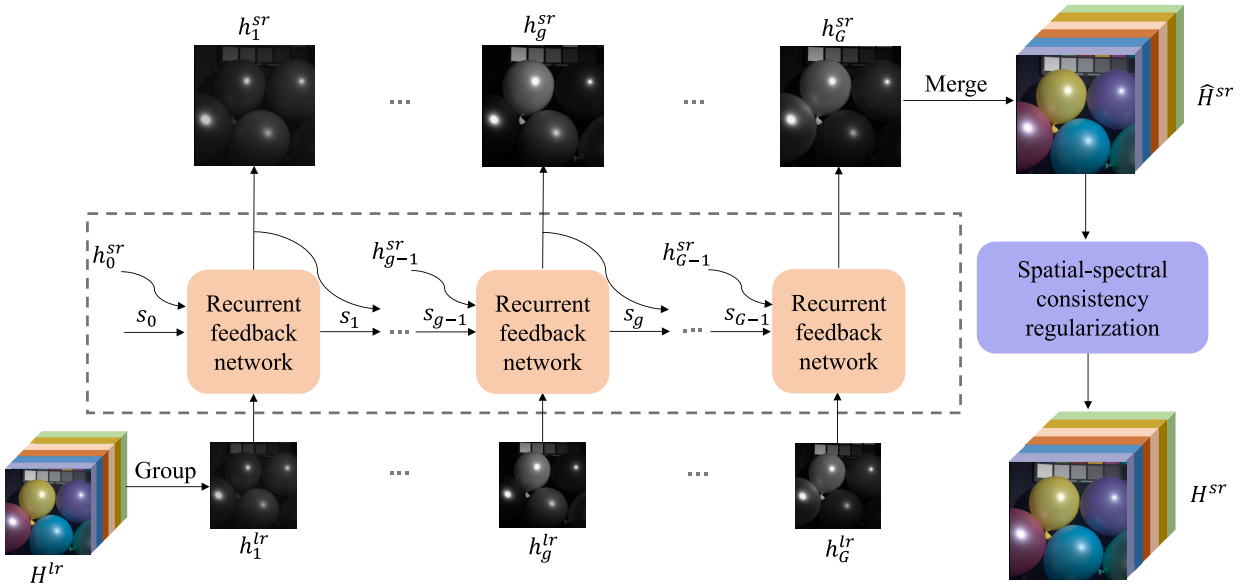


Fig. 3. Whole architecture of the proposed RFSR consisting of the RFNet and the spatial–spectral consistency regularization network.

SR. Generally, existing methods promote the fidelity of such a structure by enforcing the corresponding pixels to share similar intensity values. For most learning-based methods, the spatial–spectral correlations are only exploited in the LR space, while the consistency in the HR space is not well modeled.

We address the challenge of spatial–spectral correlation preservation with a subsequent regularization module on the intermediate HR results. Specifically, an additional network is applied to explore the spatial–spectral geometry coherence in the HR space, which explores the spatial–spectral structure. Moreover, we use a structure-aware loss function, which enforces not only intensity values but also constrains the spatial–spectral consistency in hyperspectral data.

IV. Proposed Method

As shown in Fig. 3, our proposed method consists of an RFNet, super-resolving each group of the hyperspectral image individually by integrating the feedback embedding, and a spatial–spectral consistency regularization network, which enforces the spatial–spectral geometry coherence in the HR space. In the following, we first overview the whole method and then discuss the RFNet and the consistency regularization network in detail.

For the input degraded LR hyperspectral image $H^{\text{lr}} \in \mathbb{R}^{h\times w\times L}$, according to the band correlations, we split spectral bands into nonoverlapping G groups, e.g., $[h_1^{\text{lr}}, \ldots, h_g^{\text{lr}}, \ldots, h_G^{\text{lr}}]$, and each group has c spectral bands. On the one hand, we can not only exploit the correlations among neighboring spectral bands in each group but also reduce the dimension of features processed in the network. On the other hand, the spectral band group in hyperspectral data can be regarded as a clip of frame sequences with the static scene and constant wavelength difference. To fully preserve the continuity of the group sequences, our RFNet focuses on extracting the consecutive information from adjunct groups to assist the SR of the current group by the feedback mechanism. When super-resolving the gth group, the RFNet receives the feedback embedding from the previous group to guide the high-frequency information learning. In other words, the RFNet reconstructs the SR estimation by

$$h_g^{\text{sr}} = \text{RFNet}(s_{g-1}, h_{g-1}^{\text{sr}}, h_g^{\text{lr}}) \tag{1}$$

where h_g^{lr} and h_g^{sr} are the LR group input and the reconstructed result, respectively. The feedback embedding, e.g., h_{g-1}^{sr}, is the super-resolved spectral bands of the previous group. Inspired by the sophisticated structure [40], [41] modeling the sequence, we also attempt to acquire the long-distance dependence for the continuity of the sequence by passing on a hidden state s_{g-1}. By such a feedback mechanism, the RFNet can be concentrated more on the consecutive information among groups, and thus, it generates spatial–spectral consistency results for hyperspectral image SR.

After processing the spectral bands group-by-group, we merge the HR spectral bands into the intermediate result, $\hat{H}^{\text{sr}} = [h_1^{\text{sr}}, \ldots, h_g^{\text{sr}}, \ldots, h_G^{\text{sr}}]$. To preserve spatial–spectral consistency in the HR space, a subsequent regularization network is deployed to model the spatial–spectral correlations among the intermediate hyperspectral images by efficient 3D convolution. Given the rough hyperspectral image, the spatial–spectral consistency regularization network (SCRNet) generates the final SR results by

$$H^{\text{sr}} = \text{SCRNet}(\hat{H}^{\text{sr}}) \tag{2}$$

in which H^{sr} is the final SR result, to be spatial–spectral coherence.

A. Recurrent Feedback Reconstruction

The RFNet is devoted to extracting the complementary information from the feedback embedding to assist in the

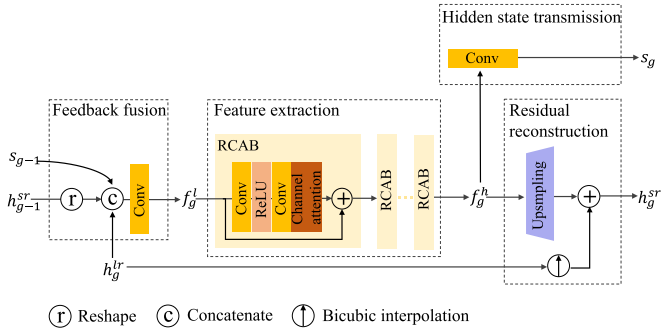


Fig. 4. Diagram of the RFNet involving subphases, i.e., feedback fusion, feature extraction, residual reconstruction, and hidden states transmission.

reconstruction of the reference group. Let h_g^{lr} denote the reference group to be super-resolved. As shown in Fig. 4, this network receives the feedback information (h_{g-1}^{sr} and s_{g-1}), passing the SR result h_g^{sr} and the current hidden state s_g as the input for the next group. There are four subphases involved, i.e., feedback fusion, feature extraction, residual reconstruction, and hidden states transmission.

1) Feedback Fusion: Since the generated SR results, $h_{g-1}^{\text{sr}} \in \mathbb{R}^{H\times W\times c}$, have the target spatial resolution, we first shuffle the spatial pixels into the spectral channels to generate the LR representation, $h_{g-1}^{\prime\text{sr}} \in \mathbb{R}^{h\times w\times c\cdot r^2}$, which could preserve the high-frequency information as much as possible. Then, we fuse the feedback embedding with the reference group to extract the low-level features, denoted as f_g^l, i.e.,

$$f_g^l = \mathcal{C}_{3\times 3}([s_{g-1}, h_{g-1}^{\prime\text{sr}}, h_g^{\text{lr}}]) \tag{3}$$

where $\mathcal{C}_{3\times 3}$ represents the convolutional layer with the kernel size of 3×3. For the first group, the feedback information is initialized by zero values with the same spatial size.

2) Feature Extraction: So far, many SR methods are dedicated to designing the feature extraction block, such as the simplified residual block in [23], channel attention residual block (RCAB) in [28], and nonlocal attention block in [30]. Considering the spectral pattern in hyperspectral data, we extract deep features by the RCAB in [28], which learns residual mappings by exploring cross-channel correlations. Specifically, D RCABs are cascaded to explore the high-level information and can be formulated as

$$f_g^h = \mathcal{R}_{\text{RCAB}_D}(\cdots \mathcal{R}_{\text{RCAB}_d}(\cdots \mathcal{R}_{\text{RCAB}_1}(f_g^l)\cdots)\cdots) \tag{4}$$

where $\mathcal{R}_{\text{RCAB}_d}$ refers to the function of dth RCAB and f_g^h is the obtained high-level features, which is served for delivering the hidden state and reconstructing the SR output of the current group. Specifically, the RCAB consists of two convolutions layers and a channel attention operation that is conducted by a global pooling operation and two convolutions layers (with a reduction ratio to control the number of feature channels) followed by a softmax function.

3) Residual Reconstruction: To upscale the obtained features to the target size, the upscale module is applied here to improve spatial resolution by global residual learning via

$$h_g^{\text{sr}} = \mathcal{U}(f_g^h) + h_g^{\text{lr}} \uparrow \tag{5}$$

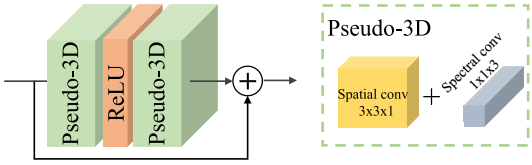


Fig. 5. Structure of the proposed Pseudo-3D residual block, cascading in the spatial–spectral consistency regularization network.

where $h_g^{\text{lr}} \uparrow$ refers to the bicubic upscale version of the input LR hyperspectral image, h_g^{sr} is the generated SR result for the reference group, and $\mathcal{U}$ is the upscale function that is conducted by the PixelShuffle [42] operator followed by a convolutional layer.

4) Hidden State Transmission: According to the existing methods [40], [41] that model sequence successfully, we introduce the hidden state here to capture the long-distance dependence by updating from the high-level features in every iteration, which can be represented by

$$s_g = \mathcal{C}_{3\times 3}(f_g^h) \tag{6}$$

where $\mathcal{C}_{3\times 3}$ represents the convolutional layer with the kernel size of 3×3 and s_g is the current hidden state, passing to the next group.

B. Spatial–Spectral Consistency Regularization

To regularize the spatial–spectral structure in hyperspectral data, an intuitive method is using the 3D convolution. However, the 3D convolution will result in a significant increase in the parameter number and computational complexity. In view of improving the efficiency, but still exploring the spatial–spectral correlations, we adopt the efficient Pseudo-3D convolution [43], which handles the spatial and spectral dimensions with separable convolutions. Specifically, the Pseudo-3D convolution is to replace the 3D convolution with two consecutive convolution layers: one 2D convolution layer to learn spatial features, followed by a 1D convolution layer purely on the spectral axis. This can be implemented by running two 3D convolutions, where the first (spatial) convolution has filter shape $(k, k, 1)$ and the second (temporal) convolution has filter shape $(1, 1, k)$. With a factored version that disentangles the 3D convolutional operation into a spatial part and a spectral part, the Pseudo-3D convolution could explore the spatial–spectral correlations with less computational overhead.

In our regularization network, we use two layers of Pseudo-3D convolution to construct the Pseudo-3D residual block, as shown in Fig. 5. Specifically, for the intermediate results $\hat{H}^{\text{sr}}$, we stack P Pseudo-3D residual blocks to obtain the final HR results

$$H^{\text{sr}} = \mathcal{P}_{P3RB_P}(\cdots \mathcal{P}_{P3RB_p}\cdots(\mathcal{P}_{P3RB_1}(\hat{H}^{\text{sr}})\cdots)\cdots) \tag{7}$$

where $\mathcal{P}_{P3RB_p}$ refers to the function of pth Pseudo-3D residual block. By such a regularization network, our proposed method could preserve the spatial–spectral consistency better.

C. Loss Function

To acquire a high peak signal-to-noise ratio (PSNR), most SR methods are optimized by the pixelwise loss

(e.g., ℓ_1 and ℓ_2) that measures the pixel distances between the SR images and the HR ones. Compared with perceptual and adversarial losses, such pixelwise losses do not impel a deep model to produce fake details, which is undesirable in the remote sensing field. Due to the oversmooth results generated by the ℓ_2 loss function, we adopt the ℓ_1 as context loss to measure the reconstruction accuracy of the network, which can be formulated as

$$\mathcal{L}_{\text{con}}(\Theta) = \frac{1}{N}\sum_{n=1}^{N}\|H_n^{\text{hr}} - H_n^{\text{sr}}\|_1 \tag{8}$$

where H_n^{sr} and H_n^{hr} are the nth reconstructed hyperspectral image and ground-truth hyperspectral image, respectively, N is the number of images in one training batch, and Θ denotes the parameter set of our proposed network. Widely used for the natural SR method, this loss function ignores the special pattern in hyperspectral data, resulting in a spectral distortion. Thereby, we use a spectral measurement to constrain the spectral structure, which can be formulated as

$$\mathcal{L}_{\text{spe}}(\Theta) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\pi}\arccos\Big(\frac{H_n^{\text{hr}}\cdot H_n^{\text{sr}}}{\|H_n^{\text{hr}}\|_2\|H_n^{\text{sr}}\|_2}\Big). \tag{9}$$

However, driven by such loss function, the model fails to recover the spatial–spectral consistency structure since a statistical average of possible HR solutions tends to be given from training data.

In natural image SR works [44], [45], the gradient information is served as a powerful tool to enhance the sharpness of the super-resolved images. Computing the difference between adjacent pixels, we can obtain the gradient map for a hyperspectral image H by

$$\nabla H = (\nabla_h H, \nabla_w H, \nabla_l H) \tag{10}$$

$$\text{M}(H) = \|\nabla H\|_2 \tag{11}$$

where $\text{M}(\cdot)$ stands for the operation to extract gradient map. Considering both spatial and spectral correlations, the gradient map could better reflect the spatial–spectral dependence of a hyperspectral image. Therefore, to alleviate the spatial–spectral distortion, we advocate a gradient constraint to provide additional supervision for better image reconstruction by a gradient loss. We formulate the gradient loss by diminishing the distance between the gradient map extracted from the super-resolved image and the one from the corresponding HR image as follows:

$$\mathcal{L}_{\text{gra}}(\Theta) = \frac{1}{N}\sum_{n=1}^{N}\|\text{M}(H_n^{\text{hr}}) - \text{M}(H_n^{\text{sr}})\|_1. \tag{12}$$

It helps the network focus on neighboring configuration so that the local intensity of sharpness can be inferred more appropriately.

In summary, the final objective loss for the proposed model is a weighted sum of the triple losses

$$\mathcal{L}_{\text{total}}(\Theta) = \mathcal{L}_{\text{con}} + \lambda_1\mathcal{L}_{\text{spe}} + \lambda_2\mathcal{L}_{\text{gra}} \tag{13}$$

where λ denotes the tradeoff parameters of different losses. Specifically, λ_1 controls the weight of the spectral term and a large value may enable the network to attach much importance to the spectral similarity, ignoring the reconstruction of high-frequency details. Thus, we set λ_1 as 0.5. The gradient term considers both spatial and spectral structures, which is constrained by λ_2. To some extent, this parameter balances content loss and spectral loss. In order to avoid the oversharpening result, we set λ_2 to be 0.1.

We investigate the effectiveness of different losses for the reconstruction results and three representative measurements are reported in Table I. Only combining the spectral loss with the context loss, the reconstructed accuracy would slightly decrease to acquire the gain on the spectral similarity. On the contrary, the proposed gradient-based loss would contribute to both spatial and spectral aspects. According to the results in Table I, when using the triple loss, the proposed method could reconstruct high-quality HR estimation. With the supervision in both image and gradient domains, the proposed method not only learns fine high-frequency details but also attaches importance to avoiding spectral geometric distortions, resulting in more structure consistent SR images that the spatial–spectral consistency can be well preserved.

D. Implementation Details

According to the different data sets, the spectral bands of the hyperspectral data are divided into G groups without overlapping. Each group has c spectral bands, and the zero bands are used to pad the last group to make the number of bands equal to c. In the RFNet, the convolution layers are of 3 kernel size and extract 64 feature maps. Thus, the layer in the feedback fusion phase has $(\alpha^2 \cdot c + 64 + c)$ input feature channels. Referring to [28], we set the reduction ratio in channel attention as 16, and the convolutional layer would have four feature maps. To maintain the spatial size of the feature map, zero padding is applied for these 3 convolutional layers. In the regularization network, each 3D convolutional layer with $k = 3$ extracts L feature maps, which equals the number of spectral bands. Zero padding is also used to keep the spatial resolution. For the training procedure, we set $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$ empirically. For the training phase, we empirically choose a minibatch size of 32 and use the Adam optimizer with a weight decay of 1e−4. The initial learning rate is set as 0.001 and is decayed by ten times after 200 epochs, while the total epoch is 300. Our model is implemented by Pytorch on NVIDIA GTX 1080Ti.

V. Experiments and Results

A. Data Sets and Experimental Setup

The proposed method is thoroughly evaluated on three widely used benchmark data sets: CAVE data set [46], Pavia Centre data set,[1] and Chikusei data set [47]. The CAVE data set consists of natural hyperspectral images and the last two data sets are remote sensing hyperspectral image data sets. The results of the proposed method on the data sets are compared with six state-of-the-art methods, including one

[1] http://www.ehu.eus/ccwintco/index.php?title=HyperspectralRemoteSensingScenes

TABLE I

QUANTITATIVE PERFORMANCE (AVERAGE PSNR/SAM/SSIM FOR REPRESENTATION) OF DIFFERENT LOSS FUNCTIONS EVALUATED ON THE TEST DATA SET FROM CAVE DATABASE AT THE SCALE FACTOR 4

$\mathcal{L}_{con}$	$\mathcal{L}_{spe}$	$\mathcal{L}_{gra}$	PSNR↑	SAM↓	SSIM↑
√			39.8724	2.9052	0.9688
√	√		39.8267	2.7850	0.9722
√		√	39.9024	2.8219	0.9694
√	√	√	40.0136	2.6649	0.9735

baseline method, i.e., Bicubic interpolation, two state-of-the-art deep single gray/RGB image SR methods, i.e., VDSR [20] and EDSR [23], and three relevant deep single hyperspectral image SR methods, i.e., 3DFCNN [37], GDRRN [38], and SSPSR [39]. For the single gray/RGB image SR methods, we adjust the channel of the first and last layers adapted to the spectral dimension of the hyperspectral data. We try our best to adjust hyperparameters of the compared methods to achieve their best performance at the scale factors 4 and 8.

Five widely used reference indexes are adopted for performance evaluation, including mean PSNR, spectral angle mapper (SAM), structure similarity index (SSIM) indices, erreur relative globale adimensionnelle de synthese (ERGAS), and root-mean-squared error (RMSE). The best values for these indices are $+\infty, 0, 1, 0$, and 0. As the PSNR, SSIM, and RMSE are commonly used natural image restoration quality indices that are calculated on the single-channel image, we average their values over all spectral bands in hyperspectral images for these indices.

B. Experimental Results on CAVE Data Set

The CAVE data set is gathered by a cooled CCD camera at a 10-nm interval from 400- to 700-nm wavelength. The data set consists of 32 HR hyperspectral scenes, each of which has a dimension of 512×512 with 31 spectral bands. We randomly select 23 scenes from the data set for training, three for validation, and the remaining six hyperspectral images for testing. When the scale factor is 4, we extract 64×64 size patches with the overlapping of 32 pixels for training. At the scale factor 8, training patches of 128×128 pixels are extracted with 64 pixels overlap. The corresponding LR hyperspectral images are generated by bicubic downsampling. The original images for testing are treated as ground-truth HR hyperspectral images, and the LR hyperspectral inputs are produced similar to the training samples. For this data set, we divide the 31 spectral bands into eight groups and pad with zero bands, and each group has four spectral bands, i.e., $G = 8$ and $c = 4$. We use ten residual attention blocks in the RFNet and three Pseudo-3D residual blocks in the regularization network.

Table II reports the objective results of seven methods evaluated on CAVE testing sets, including PSNR, SAM, SSIM, ERGAS, and RMSE measurements. Focus on exploiting the spatial information, natural image SR methods, such as VDSR [20], EDSR [23], can obtain high spatial fidelity, referring to PSNR and SSIM indexes but suffer from spectral distortion. Since the natural image SR methods neglect the spatial–spectral structure, they fail to preserve the spectral information. With the group strategy, GDRRN [38] and SSPSR [39] also achieve comparable performance in spectral similarity. It can be noticed that our proposed method considerably outperforms other algorithms in terms of all evaluation indexes. The average PSNR value of our method is more than 0.6 dB for $4\times$ and 0.4 dB for $8\times$ higher than the second best method. Due to the feedback embedding, our method could excavate the complementary and consecutive information among spectral bands. In the meanwhile, the subsequent regularization network could enforce the spatial–spectral correlation in the HR space. In this way, our method could estimate spatial–spectral consistent hyperspectral images.

TABLE II

QUANTITATIVE EVALUATION ON THE CAVE DATA SET OF STATE-OF-THE-ART HYPERSPECTRAL IMAGE SR ALGORITHMS: AVERAGE PSNR/SAM/SSIM/ERGAS/RMSE FOR SCALE FACTORS 4 AND 8. THE BOLD REPRESENTS THE BEST RESULT AND UNDERLINE REPRESENTS THE SECOND BEST

Scale	Method	PSNR↑	SAM↓	SSIM↑	ERGAS↓	RMSE↓
×4	Bicubic	36.1512	3.4871	0.9498	4.9304	0.0177
	VDSR [20]	38.3049	3.1355	0.9642	4.0469	0.0143
	EDSR [23]	39.0905	3.1651	0.9674	3.6206	0.0131
	GDRRN [38]	38.4440	2.9282	0.9638	3.8670	0.0140
	3DFCNN [37]	38.2389	3.1747	0.9629	4.0026	0.0144
	SSPSR [39]	<u>39.3413</u>	<u>2.8397</u>	<u>0.9686</u>	<u>3.5576</u>	<u>0.0127</u>
	Ours	**40.0136**	**2.6649**	**0.9735**	**3.3509**	**0.0114**
×8	Bicubic	31.6674	4.9947	0.8886	3.9765	0.0289
	VDSR [20]	32.7395	4.8864	0.8986	3.5732	0.0264
	EDSR [23]	34.1514	4.9694	0.9106	3.0413	0.0233
	GDRRN [38]	32.6542	4.6812	0.8948	3.4628	0.0262
	3DFCNN [37]	32.4715	4.9363	0.9038	3.5812	0.0259
	SSPSR [39]	<u>34.4085</u>	<u>4.3509</u>	<u>0.9145</u>	<u>3.1346</u>	<u>0.0231</u>
	Ours	**34.8724**	**4.1465**	**0.9256**	**2.9540**	**0.0214**

In order to compare the reconstructed result over all bands, the mean absolute differences calculated between the reconstructed result and the ground truth are displayed from two aspects, i.e., spatial and spectral. In Fig. 6, we show the mean error maps across all bands of three test images from the CAVE testing data set. At the bottom of these difference maps, we also report their PSNR and SAM values of the reconstructed hyperspectral images to show our considerable advantages. From the visual reconstruction results, we can observe that the proposed RFSR method can achieve the best reconstruction fidelity in recovering the spatial information of the original hyperspectral images. Compared to the SSPSR method, our method behaves well in constructing edges and structures. This is mainly because our method is supervised in both image and gradient domains, leading to more structure consistent SR images. To further show our advantages with respect to spectral information, we display the mean difference of the comparing method along the spectral dimension in Fig. 7 for these images. Rather than displaying the spectral reflectance at several positions, the mean spectral error curve would be more representative. As can be seen in Fig. 7, our method shows similar or a little worse performance to the second-best method for the first few bands and the advantage increases gradually for the subsequent bands. The main reason is that we super-resolve the hyperspectral bands group-by-group with the feedback mechanism, and the improvement in

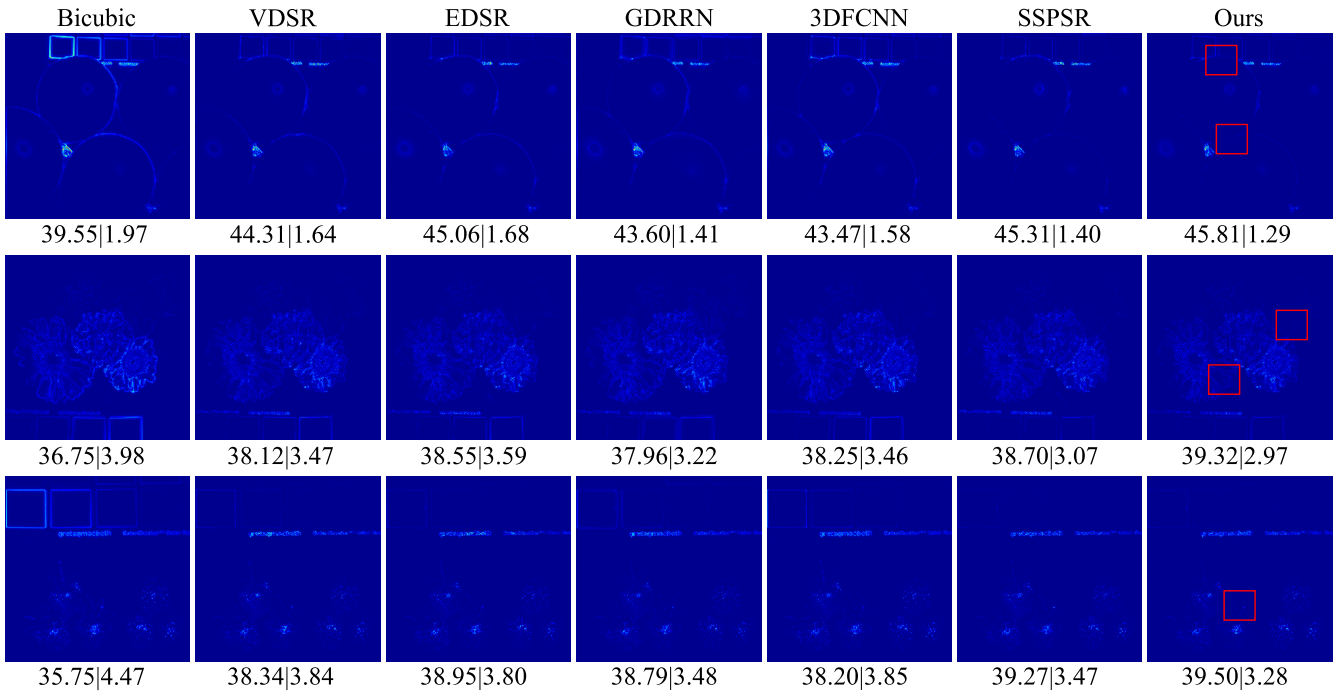


Fig. 6. Mean error maps of three test hyperspectral images in the CAVE data set at the scale factor 4: balloons, flowers, and fake_and_real_strawberries. The corresponding PSNR and SAM values of the comparison method are reported below each map.

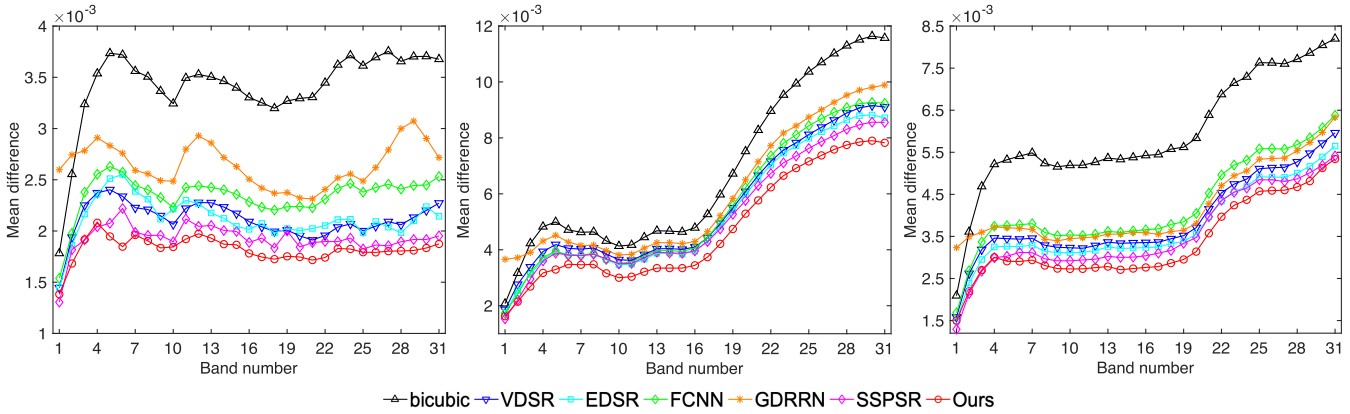


Fig. 7. Mean spectral difference curve of three test hyperspectral images in the CAVE data set at the scale factor 4: balloons, flowers, and fake_and_real_strawberries.

the first several bands would be comparatively lower because of no feedback information. As the number of hyperspectral bands that have been processed increases, the informative band features gradually accumulate, leading to better performance. Our method improves the spatial resolution via the feedback mechanism and the spatial–spectral consistency is regularized on the reconstructed images. Therefore, on the one hand, with the feedback embedding, the continuity in spectral bands can be ensured, resulting in more stable performance from the spectral aspect. On the other hand, due to the consistency regularization network and the structure-aware loss, our method can avoid spectral distortion to some extend.

C. Experimental Results on Pavia Data Set

The Pavia Centre data set is acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, Northern Italy, which is a remote sensing hyperspectral data set. By removing the noisy spectral bands, the number of spectral bands is 102 in the Pavia Centre data set with 1096×1096 pixels in total. Some regions in the center part of the scene, containing no information, are discarded before the analysis, leaving an informative region with 1096×715 spatial size. To evaluate the proposed method, the left part with 1096×224 spatial size is cropped to form the validation and testing data, which has five nonoverlapping hyperspectral images with $216 \times 216 \times 102$ pixels (one for validation and four for testing). Besides, the right part with $1096 \times 491 \times 102$ pixels are used as reference HR hyperspectral images for the training data set, in which the patch size with overlapping is similar to previous settings. For this data set, we set the spectral group $C = 13$ with eight spectral bands in each group. Due to the fewer samples for training, we set $D = 5$ in the RFNet and $P = 1$ in the regularization network.

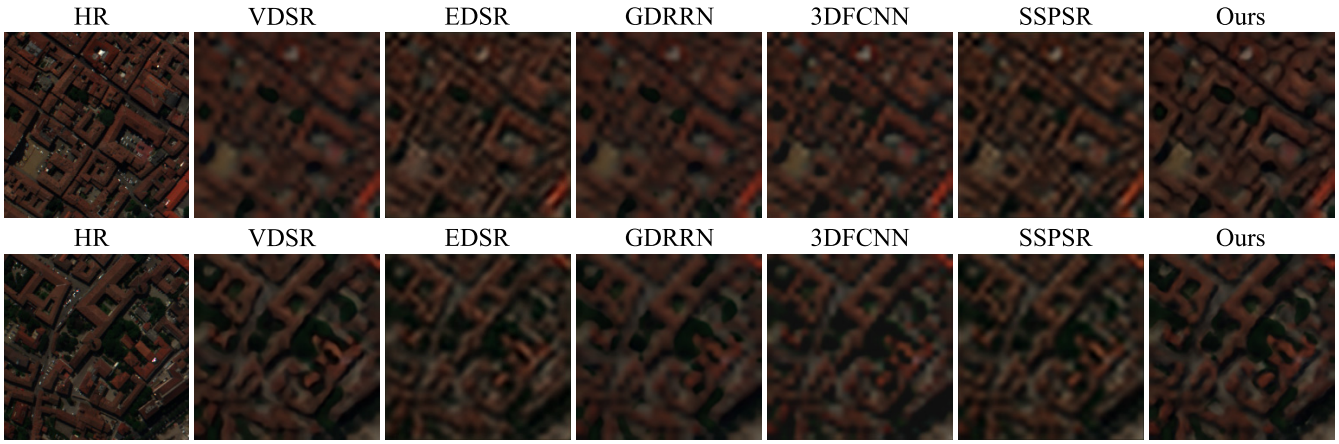


Fig. 8. Reconstructed images of two test hyperspectral images in the Pavia Centre data set with spectral bands 60-31-12 as R-G-B with the scale factor 8. (From Left to Right) Ground truth, results of VDSR [20], EDSR [23], GDRRN [38], 3DFCNN [37], SSPSR [39], and the proposed RFSR method.

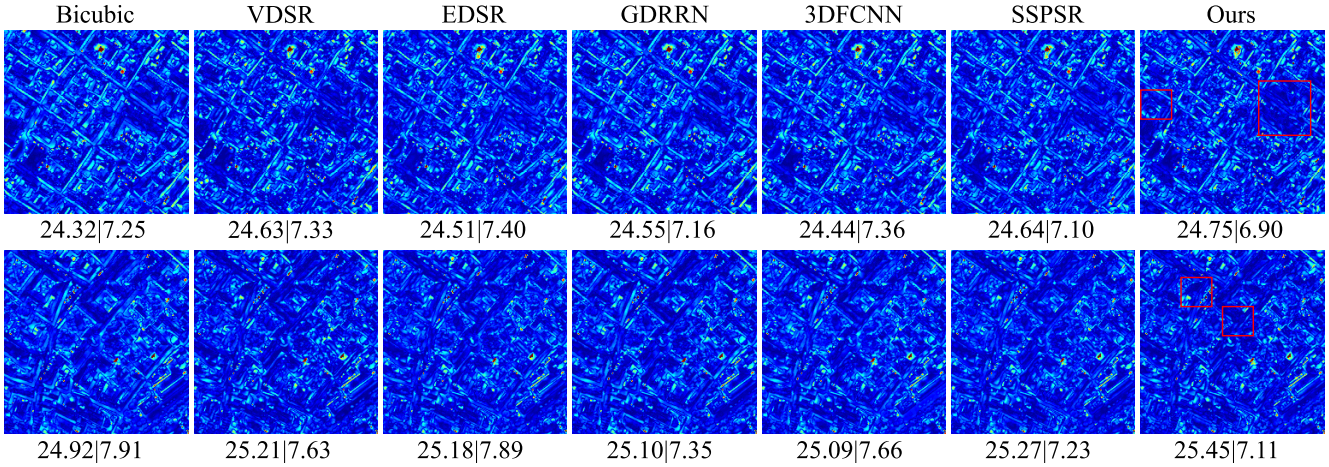


Fig. 9. Mean error maps of two test hyperspectral images in the Pavia data set at the scale factor 8. The corresponding PSNR and SAM values of the comparison method are reported below each map.

TABLE III

QUANTITATIVE EVALUATION ON THE PAVIA DATA SET OF STATE-OF-THE-ART HYPERSPECTRAL IMAGE SR ALGORITHMS: AVERAGE PSNR/SAM/SSIM/ERGAS/RMSE FOR SCALE FACTORS 4 AND 8. THE BOLD REPRESENTS THE BEST RESULT AND UNDERLINE REPRESENTS THE SECOND BEST

Scale	Method	PSNR↑	SAM↓	SSIM↑	ERGAS↓	RMSE↓
×4	Bicubic	27.7268	5.9650	0.7239	6.6828	0.0429
	VDSR [20]	28.8853	5.6188	0.7987	6.0624	0.0372
	EDSR [23]	28.7227	5.6241	0.7896	6.1217	0.0379
	GDRRN [38]	29.0219	5.5490	0.8013	5.9454	0.0366
	3DFCNN [37]	28.5773	5.6311	0.7814	6.2910	0.0387
	SSPSR [39]	<u>29.0150</u>	<u>5.5731</u>	<u>0.8034</u>	<u>5.9803</u>	<u>0.0367</u>
	Ours	**29.1061**	**5.4823**	**0.8076**	**5.9137**	**0.0363**
×8	Bicubic	24.7220	7.6026	0.4744	4.7759	0.0611
	VDSR [20]	24.9350	7.5066	0.5295	4.7164	0.0585
	EDSR [23]	25.0459	7.5273	0.5154	4.6403	0.0592
	GDRRN [38]	24.9776	7.2626	0.4958	4.6892	0.0599
	3DFCNN [37]	25.0054	7.4259	0.5023	4.5357	0.0596
	SSPSR [39]	<u>25.1246</u>	<u>7.290</u>	<u>0.5303</u>	<u>4.6618</u>	<u>0.0581</u>
	Ours	**25.2550**	**7.0675**	**0.5350**	**4.5298**	**0.0579**

Table III tabulates the average quantitative performance in terms of five objective indexes on four testing images of all comparing methods. According to the results, our method is superior to other algorithms at both two scales. Especially, since the small examples in Pavia scenes are used for training, the improvement on the test set is relatively small. On the contrary, this demonstrates that our method can handle challenging data sets better than the state of the art. As the general image SR methods, VDSR and EDSR can acquire pleasurable results. However, their spectral similarity evaluation is not competitive when compared with these single hyperspectral image SR methods, i.e., 3DFCNN [37] and GDRRN [38].

In Fig. 8, we visualize the reconstructed hyperspectral images of two testing images from the Pavia Centre data set of the competitive approaches at the scale factor 8. Specifically, the 60th, 31th, and 12th bands are served as the R-G-B channels for better visualization. We can observe that the estimations of the VDRS [20], GDRRN [38], and 3DFCNN [37] are blurry, while the EDSR [23] and SSPSR [39] methods introduce some artifacts. The proposed method can maintain the main structural information with a pleasurable visual effect. Similarly, we also display the mean difference from two aspects. Fig. 9 shows the error maps of two testing images from the Pavia Centre data set of the competitive approaches at the scale factor 8. As indicated in error maps, some edge contours are not displayed in our results, which indicates that

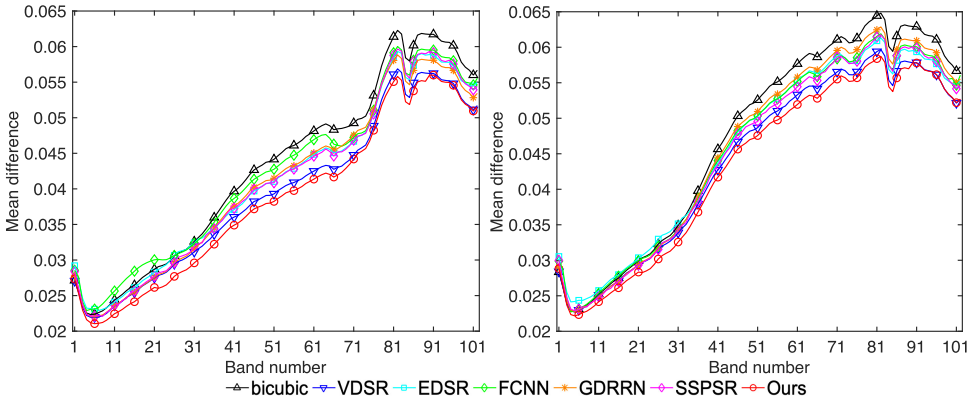


Fig. 10. Mean spectral difference curve of two test hyperspectral images in the Pavia data set at the scale factor 8.

TABLE IV

QUANTITATIVE EVALUATION ON THE CHIKUSEI DATA SET OF STATE-OF-THE-ART HYPERSPECTRAL IMAGE SR ALGORITHMS: AVERAGE PSNR/SAM/SSIM/ERGAS/RMSE FOR SNR (DB): ∞, 40, AND 30 AT SCALE FACTOR 4. THE BOLD REPRESENTS THE BEST RESULT AND UNDERLINE REPRESENTS THE SECOND BEST

SNR	Method	PSNR↑	SAM↓	SSIM↑	ERGAS↓	RMSE↓
∞	Bicubic	50.4952	3.2681	0.9865	6.8669	0.0037
	GDRRN [38]	51.4594	2.7684	0.9889	6.1283	0.0033
	3DFCNN [37]	51.4711	3.0079	0.9891	6.1835	0.0033
	SSPSR [39]	51.9690	2.5620	0.9900	5.7895	0.0032
	Ours	**52.5339**	**2.5113**	**0.9911**	**5.4504**	**0.0030**
40	Bicubic	50.4885	3.2819	0.9864	6.8743	0.0037
	GDRRN [38]	51.4480	2.7703	0.9888	6.1346	0.0033
	3DFCNN [37]	51.4629	3.1268	0.9890	6.1960	0.0034
	SSPSR [39]	51.9582	2.5741	0.9896	5.7915	0.0032
	Ours	**52.5287**	**2.5268**	**0.9906**	**5.4735**	**0.0030**
30	Bicubic	49.9516	4.1763	0.9852	7.5214	0.0038
	GDRRN [38]	51.2683	2.7694	0.9886	6.3950	0.0034
	3DFCNN [37]	51.2277	3.3165	0.9887	6.4630	0.0034
	SSPSR [39]	51.8153	2.7667	0.9896	5.9654	0.0032
	Ours	**52.4902**	**2.6503**	**0.9913**	**5.6109**	**0.0030**

the proposed method performs well in recovering structure information, even at a large scale factor. From the spectral aspect, we plot the mean difference of comparing methods along with the spectral bands in Fig. 10 for two testing images. We can notice that the mean spectral error curve of our method is the lowest one, which indicates that our RFSR method has the advantage over all bands. Since a regularization network is appended on the intermediate results, our method would be more stable with respect to the spatial–spectral consistency.

D. Experimental Results on the Chikusei Data Set

This airborne data set is taken by the Headwall Hyperspec-VNIR-C imaging sensor over agricultural and urban areas in Chikusei, Ibaraki, Japan. The hyperspectral image has 128 bands in the spectral range from 363 to 1018 nm with a 2517 $\times$ 2335 spatial size. Since some edge regions are missing, we extract the informative central region of the original scene with $2000 \times 2000 \times 128$ pixels, which is further divided into training, validation, and test data. Specifically, the top region is cropped into five nonoverlapping hyperspectral subimages with $400 \times 400 \times 128$ pixels and we randomly select three subimages for testing and two subimages for validation. Besides, from the remaining region, we extract overlapping patches as reference HR hyperspectral images for training. At the upscale factor 4, we crop the 64×64 size of patches by the overlapping of 32 pixels. These patches are used as reference HR hyperspectral images and the LR hyperspectral images are generated by Bicubic downsampling. For this data set, we set the spectral group $C = 16$ with eight spectral bands in each group. Since the training samples are fewer, we set $D = 6$ in the RFNet and $P = 2$ in the regularization network. Generally, due to the complicated imaging environment, the hyperspectral images often suffer from noise interference. This remote sensing data set of high quality is selected to verify the robustness of the proposed algorithm to noise. Specifically, the additive Gaussian white noise with three different levels, computed via SNR, is added to the LR hyperspectral images at the scale factor 4.

We compare the proposed RFSR method with some competitors that are designed for hyperspectral image SR, including GDRRN [38], 3DFCNN [37], and SSPSR [39]. The average performance of the PSNR, SAM, SSIM, ERGAS, and RMSE values of competing methods for different noise levels at the scale factor 4 on the Chikusei data set is reported in Table IV. We can observe that our method surpasses other algorithms under different levels of Gaussian white noise in terms of five evaluation indexes. The average PSNR value of our method is more than 0.50 dB higher than that of the SSPSR [39], which ignores the intergroup correlation. Thus, the proposed RFSR algorithm still has outstanding performance irrelevant to the level of SNR.

Fig. 11 shows the reconstructed HR hyperspectral images of the comparing methods with downsampling factor 4 for two SNR levels: ∞ and 30 dB. From the visual reconstruction results, we notice that all comparison algorithms are interfered by noise to some extent, resulting in some distortion. However, our method could reconstruct finer texture details than other comparison methods.

E. Ablation Study

1) Recurrent Feedback Network: In our method, as the spectral bands in hyperspectral data can be regarded as a clip of frame sequences, we capture the spatial–spectral information from a sequence modeling perspective. To improve the

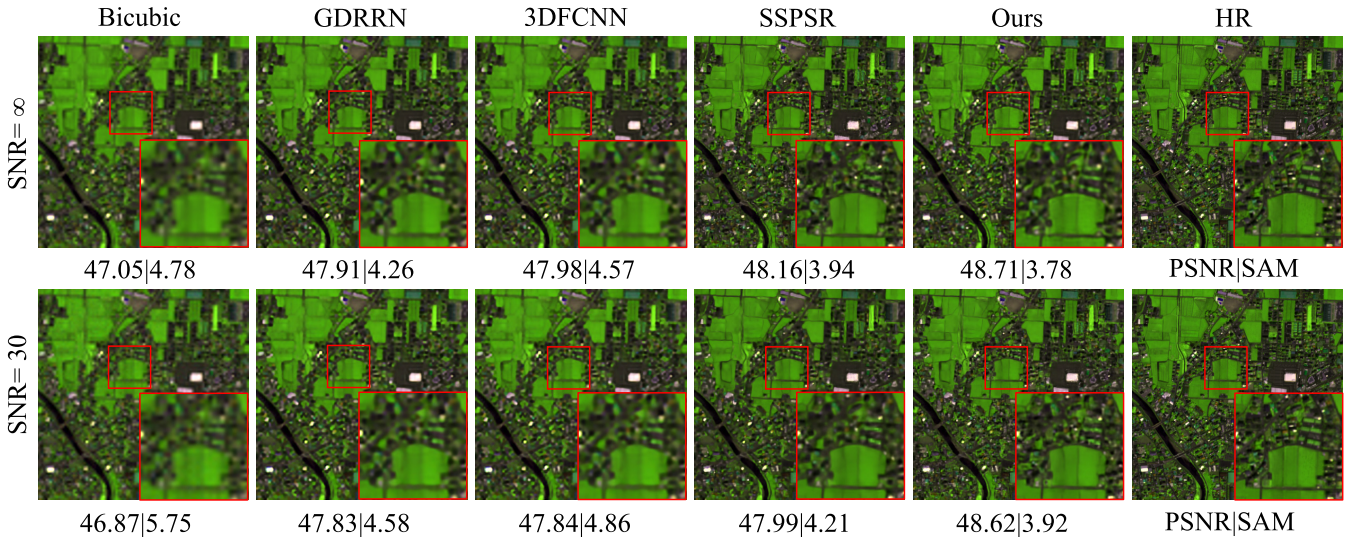


Fig. 11. Reconstructed images of one test hyperspectral image in the Chikusei data set with spectral bands 70-100-36 as R-G-B for two noise levels with the scale factor 4. (From Left to Right) Results of bicubic, GDRRN [38], 3DFCNN [37], SSPSR [39], the proposed RFSR method, and the ground truth.

TABLE V

ABLATION STUDY. QUANTITATIVE COMPARISONS AMONG SOME VARIANTS OF THE PROPOSED METHOD OVER THE TESTING SET OF CAVE DATA SET AT THE SCALE FACTOR 4

Model	Params.($\times 10^6$)	FLOPs($\times 10^9$)	PSNR↑	SAM↓	SSIM↑	ERGAS↓	RMSE↓
Ours	1.1746	371.8541	40.0136	2.6649	0.9735	3.3509	0.0114
Ours w/o GS	1.1435	956.2412	39.8556	2.6468	0.9720	3.4466	0.0122
Ours w/o FE	1.0639	357.3502	39.8219	2.7489	0.9692	3.5092	0.0127
Ours w/o CR	1.1551	213.3475	39.8951	2.7331	0.9710	3.4201	0.0125
Ours with 3D	1.1975	558.3077	39.9481	2.6580	0.9720	3.3679	0.0118

efficiency, the spectral bands are first divided into subgroups and super-resolved group-by-group. Since the group strategy is thoroughly studied in [35] and [39], we do not investigate this strategy here. We just compare the proposed model with and without the group strategy. Discarding the group strategy, Ours w/o GS model super-resolves spectral bands one-by-one. In this way, the complexity of the model is determined by the number of spectral bands, which inevitably results in higher computational overhead. As shown in Table V, referring to the floating point operations (FLOPs) that are calculated on the input size of $128 \times 128 \times 31$ at scale 4, Ours w/o GS model, in which group strategy is discarded, has a higher computational overhead but obtains a worse result except for the SAM index. Although modeling the spectrum correlations band-by-band would preserve the spectral information well, more calculations are required, increasing the training difficulty.

In the RFNet, for capturing the consecutive information, we pass on the feedback embedding from the previous group to guide the high-frequency information learning for the current group. To verify the effectiveness of the feedback mechanism, we remove the feedback embedding transmission and super-resolve the groups by the same feature extraction and residual reconstruction subphases. The variant of the proposed method is represented as Ours w/o FE in Table V, which keeps the same level of parameters as Ours. It can be seen that our method with feedback mechanism achieves better performance on the spatial reconstruction fidelity and the spectral consistency. Due to the sequence modeling method via feedback embedding, our method could make full use of the complementary and consecutive information.

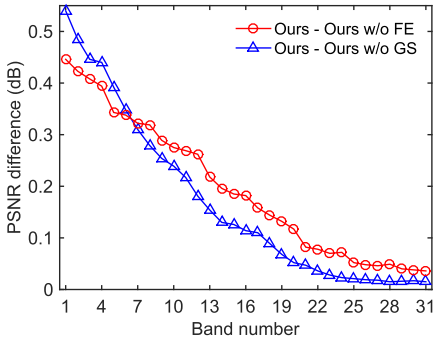


Fig. 12. Mean PSNR difference curve of our model over two variants on test hyperspectral images in the CAVE data set at the scale factor 4.

To further show the advantages of these strategies, we display the average PSNR difference between the proposed model and two variants across all spectral bands on the test hyperspectral images in the CAVE data set in Fig. 12. We can observe that without group strategy, the variant, Ours w/o GS, has a large average PSNR difference, especially for the first several bands, and the mean difference steadily decreases as more bands are processed. Compared to Ours w/o FE in which we remove the feedback embedding, the proposed method could make full use of high-frequency details

from the previous group and performs better for all spectral bands.

2) Spatial–Spectral Consistency Regularization: In order to preserve the spatial–spectral consistency after SR, we deploy a subsequent regularization network to explore the spatial–spectral geometry coherence in the HR space, which is ignored in existing methods. To demonstrate the effect, the consistency regularization network is removed to obtain the variant, Ours w/o CR. According to the results in Table V, with the consistency regularization network, our method has achieved a comparable performance gain compared to Ours w/o CR. By the subsequent regularization network, the spatial–spectral consistency is further explored in the HR space, while the increase in computational complexity is acceptable.

In the consistency regularization network, in view of efficiency, the Pseudo-3D convolution with less computational overhead is chosen as an alternative for modeling the spatial–spectral correlations in the HR space. We replace the Pseudo-3D convolution with the traditional 3D convolution and report the results in Table V, corresponding to Ours with 3D that the model size is similar to Ours. As shown in Table V, our method equipped with Pseudo-3D achieves almost the same performance as Ours with 3D, but the computational complexity is reduced. Consequently, our method would be more efficient.

VI. Conclusion

In this article, we have proposed a novel RFNet with consistency regularization for hyperspectral image SR, called RFSR. In our method, we model the hyperspectral bands in a sequential manner to capture the continuity among spectral bands by dividing them into subgroups. The spatial resolution of each group is improved by fusing the feedback embedding from the previous group, which could fully exploit the complementary and consecutive information. Besides, a subsequent regularization network is deployed to constrain the spatial–spectral structure in the HR space, leading to more spatial–spectral consistent results. Meanwhile, we introduce a structure-aware loss in the gradient domain to avoid spectral geometric distortion. The qualitative and quantitative results on natural and remote sensing hyperspectral scenes have demonstrated the stability and superiority of our method over the state of the art at different scale factors.

References

[1] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.

[2] F. F. Sabins, "Remote sensing for mineral exploration," *Ore Geol. Rev.*, vol. 14, nos. 3–4, pp. 157–183, Sep. 1999.

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

[4] 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, p. 80, Dec. 2017.

[5] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

[6] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[7] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[8] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5344–5353.

[9] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2511–2520.

[10] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by MS/HS fusion net," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1585–1594.

[11] X. Tian, Y. Chen, C. Yang, and J. Ma, "Variational pansharpening by exploiting cartoon-texture similarities," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 15, 2021, doi: 10.1109/TGRS.2020.3048257.

[12] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement," *IEEE Trans. Geosci. Remote Sens.*, early access, Dec. 14, 2020, doi: 10.1109/TGRS.2020.3039534.

[13] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, Jan. 2021.

[14] 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.

[15] Y. Wang, X. Chen, Z. Han, and S. He, "Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization," *Remote Sens.*, vol. 9, no. 12, p. 1286, Dec. 2017.

[16] H. Irmak, G. B. Akar, and S. E. Yuksel, "A MAP-based approach for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2942–2951, Jun. 2018.

[17] H. Huang, J. Yu, and W. Sun, "Super-resolution mapping via multi-dictionary based sparse representation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 3523–3527.

[18] 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.

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

[20] 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. (CVPR)*, Jun. 2016, pp. 1646–1654.

[21] 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. (CVPR)*, Jun. 2016, pp. 1637–1645.

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

[23] 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 (CVPRW)*, Jul. 2017, pp. 136–144.

[24] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[25] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[26] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshop*, 2018, pp. 1–16.

[27] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.

[28] 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.

[29] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3867–3876.

[30] 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. (CVPR)*, Jun. 2020, pp. 5690–5699.

[31] 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.
[32] J. Hu, Y. Li, and W. Xie, "Hyperspectral image super-resolution by spectral difference learning and spatial error correction," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1825–1829, Oct. 2017.
[33] 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, Nov. 2017.
[34] 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.
[35] 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.
[36] J. Hu, M. Zhao, and Y. Li, "Hyperspectral image super-resolution by deep spatial-spectral exploitation," *Remote Sens.*, vol. 11, no. 10, p. 1229, May 2019.
[37] 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, p. 1139, Nov. 2017.
[38] 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 (BigMM)*, Sep. 2018, pp. 1–4.
[39] 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.
[40] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.
[41] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[42] 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. (CVPR)*, Jun. 2016, pp. 1874–1883.
[43] S. Xie, C. Sun, J. Huang, Z. Tu, and K. Murphy, "Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 305–321.
[44] Q. Yan, Y. Xu, X. Yang, and T. Q. Nguyen, "Single image superresolution based on gradient profile sharpness," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3187–3202, Oct. 2015.
[45] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Gradient profile prior and its applications in image super-resolution and enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1529–1542, Jun. 2011.
[46] 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.
[47] 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.

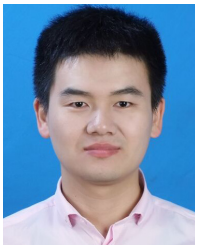
Xinya Wang received the B.S. degree from the School of Electronic Information, Wuhan University, Wuhan, China, in 2018. She is pursuing the Ph.D. degree with the Multi-Spectral Vision Processing Laboratory, Wuhan University.

Her research interests include neural networks, machine learning, and image processing.

Jiayi Ma (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 a Professor with the Electronic Information School, Wuhan University. He has authored or coauthored more than 150 refereed journal articles and conference papers, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, *International Journal of Computer Vision*, CVPR, ICCV, and ECCV. His research interests include computer vision, machine learning, and remote sensing.

Dr. Ma has been identified in the 2020 and 2019 Highly Cited Researcher lists from the Web of Science Group. He is also an Area Editor of *Information Fusion*, an Associate Editor of *Neurocomputing* and *Entropy*, and a Guest Editor of *Remote Sensing*.



Junjun Jiang (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 Science, Wuhan University, Wuhan, China, in 2014.

From 2015 to 2018, he was an Associate Professor with the China University of Geosciences, Wuhan. Since 2016, he has been a Project Researcher with the National Institute of Informatics, Tokyo, Japan. He is 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.

Dr. Jiang received the 2015 ACM Wuhan Doctoral Dissertation Award, the Best Student Paper Runner-up Award at MMM 2015, the 2016 China Computer Federation (CCF) Outstanding Doctoral Dissertation Award, and the Finalist of the World's FIRST 10K Best Paper Award at ICME 2017.