Dual-Stage Approach Toward Hyperspectral Image Super-Resolution

Qiang Li, *Student Member, IEEE,* Yuan Yuan, *Senior Member, IEEE,* Xiuping Jia, *Fellow, IEEE,* and Qi Wang, *Senior Member, IEEE*

***Abstract*—Hyperspectral image produces high spectral resolution at the sacrifice of spatial resolution. Without reducing the spectral resolution, improving the resolution in the spatial domain is a very challenging problem. Motivated by the discovery that hyperspectral image exhibits high similarity between adjacent bands in a large spectral range, in this paper, we explore a new structure for hyperspectral image super-resolution (DualSR), leading to a dual-stage design, i.e., coarse stage and fine stage. In coarse stage, five bands with high similarity in a certain spectral range are divided into three groups, and the current band is guided to study the potential knowledge. Under the action of alternative spectral fusion mechanism, the coarse SR image is super-resolved in band-by-band. In order to build model from a global perspective, an enhanced back-projection method via spectral angle constraint is developed in fine stage to learn the content of spatial-spectral consistency, dramatically improving the performance gain. Extensive experiments demonstrate the effectiveness of the proposed coarse stage and fine stage. Besides, our network produces state-of-the-art results against existing works in terms of spatial reconstruction and spectral fidelity. Our code is publicly available at https://github.com/qianngli/DualSR.**

***Index Terms*—Hyperspectral image, super-resolution (SR), band partition, group fusion, back-projection**

I. Introduction

HYPERSPECTAL image is widely utilized in various fields because of its rich spectral information. Generally, several discriminant bands in hyperspectral image are selected for subsequent analysis according to the spectral characteristics of the objects of interest. [1]. However, hyperspectral image produces high spectral resolution at the sacrifice of spatial resolution [2], [3], which cannot meet the requirements of some scene applications. Considering this dilemma, the researchers propose hyperspectral image super-resolution (SR). Without reducing the number of bands, hyperspectral image SR aims to find a high-resolution (HR) image with better visual quality and refined details from counterpart low-resolution (LR) version in spatial domain. SR is a greatly challenging task in computer vision, because it is an ill-posed inverse problem, i.e. there are multiple solutions for the same LR image.

This work was supported by the National Natural Science Foundation of China under Grant U21B2041, U1864204, and 61825603.

Qiang Li, Yuan Yuan, and Qi Wang are with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, P. R. China. (e-mail: liqmges@gmail.com, y.yuan1.ieee@gmail.com, crabwq@gmail.com).

Xiuping Jia is with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: x.jia@adfa.edu.au) (*Corresponding author: Qi Wang*)

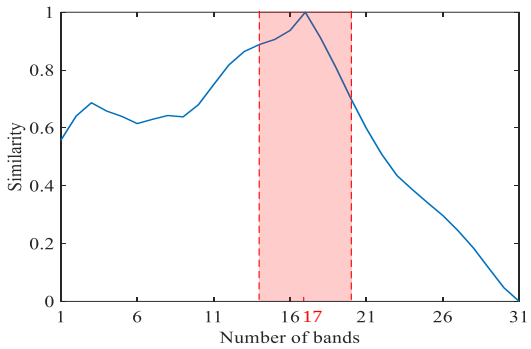


Fig. 1. Illustration of the existence of high similarity in adjacent bands. We provide the similarity values between the 17-*th* band and the other bands. The similarity value here is calculated by the reciprocal of Euclidean distance.

Currently, the existing techniques can be divided into two categories, including single hyperspectral image SR without any auxiliary information [4]–[7] and single hyperspectral image SR using a given HR multispectral image [2], [8]–[11]. In our paper, we employ the former to build the model. As for this type, previous approaches are roughly classified into three groups. They are *2D convolutional neural network (CNN)* [5], [12], *3D CNN* [5], [13]–[15], and *2D/3D CNN* [4], [16]. With respect to the first group, it usually regards each band in hyperspectral image as an image to design the model using 2D convolution. This way is similar to natural image SR. Thus, the approaches for natural image SR [17], [18] can directly be applied to this field. As mentioned above, the spectrum curve of the target is usually used to analyze whether there is a corresponding target in practical applications. This means that the reconstructed spectral curve maintains the original curve characteristics as much as possible, so that the target can be found according to the spectral curve. This is also one of key differences between hyperspectral image SR and natural image SR. Nevertheless, the first group solely investigates spatial knowledge, which results in inferior spectral fidelity.

Hyperspectral image shows rich spectral knowledge, which is helpful to enhance the performance of super-resolved image in spatial aspect. Considering this fact, various hyperspectral image SR algorithms based on 3D convolution are developed [5], [6], [13], [14], [19]. By contrast, the performance of these studies is superior for the first kind of method, which benefits from the exploration of spectral dimension. Among these algorithms, the regular 3D convolution has been abandoned due to the large number of parameters and turns to separable 3D convolution [6], [13], [20]. As for hyperspectral image SR,

using the abundant spectra mainly attempts to enhance spatial performance. In this respect, we should treat the analysis of spectral and spatial domain differently. However, these works ignore this point.

To cope with the above problem, the algorithms using 2D/3D convolution are proposed [4], [16], [21]. In contradiction to existing works, this type of approach adopts mixed 2D/3D convolution to establish the model. On the premise of obtaining spectral information, more 2D convolutions are attached with model to heighten the feature learning ability in space. Meanwhile, it dramatically decreases the parameters. The above approaches fixedly inputs all the bands of the hyperspectral image into the model. Since the hyperspectral image contains dozens or even hundreds of bands, this inevitably requires more GPU memory. Inspired by high similarity between adjacent bands [22], [23], Wang *et al.* [16] create a novel dual-channel structure to establish network. Impressively, unlike existing methods, it integrates the information of single LR band and two adjacent LR bands to achieve super-resolved band. At present, there is extremely little research using this novel input mode.

In fact, within a certain spectral range, relatively distant bands can also explicitly assist the current band to reconstruction, because these bands are also similar, but the similarity is relatively small (see Fig. 1). If more adjacent bands within a relatively large spectral range are utilized, it is beneficial to supplement the missing knowledge during the reconstruction of current band. Therefore, the key problem is *how to effectively use the adjacent bands to boost performance*. Motivated by these discoveries, we propose a dual-stage approach toward hyperspectral image SR (DualSR). In coarse stage, we first select current band and its four bands with high similarity within a certain spectral range. Then these bands are divided into three groups to study the potential knowledge for current band. To effectively utilize these bands, an adjacent spectral fusion mechanism is developed to generate complementary content from intra/inter-groups, improving representation learning. Then each initial SR band in hyperspecral image is obtained by recurrent manner. Due to the lack of exploration of a larger perspective in spectral dimension, an enhanced back-projection under spectral angle constraint is designed to further refine result in fine stage. Experiments demonstrate that the proposed dual-stage SR algorithm outperforms the existing works.

In summary, the contributions of this paper are four-fold:

- A novel dual-stage structure to hyperspectral image SR is proposed. In coarse stage, the coarse SR image is obtained in band-by-band using supervised manner, which achieves excellent performance. In fine stage, the coarse result is refined in a global perspective using unsupervised manner, dramatically improving the performance gain.
- Neighboring band partition (NBP) is presented to group five bands with high similarity in a certain spectral range into three groups, which modulates the current band to study potential knowledge.
- An adjacent spectral fusion mechanism based on intra/inter-groups is proposed, which generates complementary content from adjacent bands. It enables the model to learn more missing details in process of single band reconstruction.
- An enhanced back-projection method via spectral angle constraint is developed. Under the condition of spectral angle constraint, the result is optimized by calculating the reconstruction error. It belongs to a plug-and-play method, which is suitable for any hyperspectral image SR algorithms.

This paper is an extension of our previous work on hyperspectral image SR [24] in the IEEE International Conference on Acoustic, Speech, and Signal Processing. Compared with the conference version, this paper has some extensions as follows:

- **Methodology:** After obtaining super-resolved image in coarse stage, we add a plug-and-play method as post-processing to further prompt the performance by unsupervised manner, outperforming others by a non-negligible margin.
- **Experiments:** More experiments are conducted to verify the effectiveness of the proposed model on two datasets, such as more scale factors, improvement of performance, generalizability to various blur kernels, etc. Note that since we modify the parameters of the model in coarse stage, such as batchsize, rate learning, etc, the obtained results are higher than that of the conference version.

The rest of this paper is organized as follows. Section II reviews related hyperspectral image SR algorithms. Section III mainly describes the details in terms of coarse stage and fine stage. The experimental results are analyzed and discussed in Section IV. Finally, this work is summarized in Section V.

II. Related Works

While hyperspectral image SR has an extensive history, in this section, we mainly describe the methods based on CNN, which involves the input mode, the type of convolution used, and some two-stage methods.

A. Input Mode

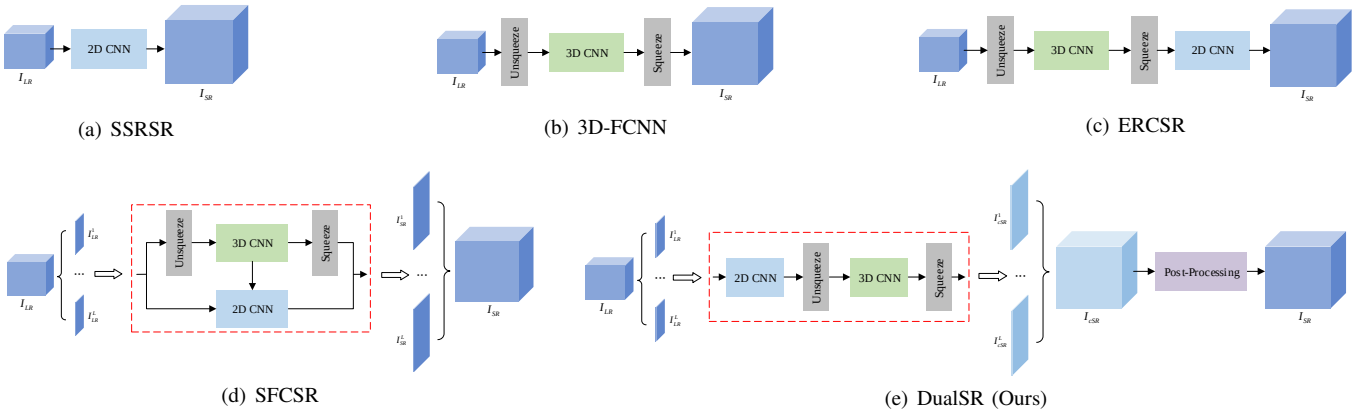
Unlike natural images, hyperspectral images have multiple bands, which can effectively identify objects. As for hyperspectral image SR, whether CNN is built with 2D or 3D convolution, the existing methods [4], [7], [14], [25], [26] are mainly to input all bands into the model for study. The simplified structures are shown in Figs. 2(a)-2(c). It obviously requires more GPU memory than the RGB image SR task. Under the condition of limited hardware resources, GPU memory is still a major constraint to build a deep network.

To alleviate this trouble, a natural improvement is to change input mode. Concretely, all the bands of hyperspectral image are input into the model in band-by-band. Typical work is SFCSR [16]. The structure of its network is displayed in Fig. 2(d). Inspired by this new structure, in our paper, we adopt this manner to design the backbone in coarse stage.

B. CNN Using Different Convolutions

In this section, we introduce several representative works using different types of convolution.

(a) **SSPSR:** For hyperspectral image SR and natural image SR, there is no difference except for the type of input image. Therefore, Jiang *et al.* [25] refer to single RGB image SR



(a) SSRSR (b) 3D-FCNN (c) ERCSR

(d) SFCSR (e) DualSR (Ours)

Fig. 2. Simplified structure of several methods. I_{LR} denotes LR hyperspectral image. I_{SR} is super-resolved hyperspectral image. I^i_{LR} $(1 \leq i \leq L)$ represents i-*th* band of LR hyperspectral image, where L is the total number of bands in hyperspectral image cube. I^i_{cSR} $(1 \leq i \leq L)$ represents i-*th* band of coarse SR hyperspectral image.

approaches, and introduce spatial-spectral prior network using 2D convolution to design model, whose simplified flowchart is shown in Fig. 2(a). Since only 2D convolution cannot discover spectral knowledge, this type of method is hardly employed.

(b) **3D-FCNN:** The regular 3D convolution is first utilized to explore both spatial context and spectral correlation in neighboring bands [14]. The structure is shown in Fig. 2(b). As we mentioned previously, the regular 3D convolution produces large parameters, thus making the network not easy to learn. To relieve this drawback, further improvements include the use of separable 3D convolution [5], [6], [13], [19] and reducing the number of 3D convolution [4], [21].

(c) **ERCSR:** Li *et al.* state that MCNet [21] designs the main modules in parallel, which does not effectively combine 2D/3D convolution, resulting in structural redundancy. For this reason, a novel alternate structure with 2D/3D convolutions is developed [4] (see Fig. 2(c)). It not only solves module redundancy. Besides, it also improves the representation learning in spatial domain. Currently, ERCSR is by far the best performance among the existing techniques.

(d) **SFCSR:** The approach [16] first applies novel input mode to analyze the information from both single band and its two adjacent bands (see Fig. 2(d)). Although it availably saves memory footprint, it does not take into account more neighboring bands. Importantly, such single band SR inevitably leads to serious spectral distortion during recurrent strategy. Furthermore, the recurrent mode ignores the overall learning and optimization of hyperspectral image. Considering this issue, after hyperspectral image is reconstructed in band-by-band, we add an enhanced back-projection strategy as post-processing to further optimize result. The simplified overview is shown in Fig. 2(e). The strategy refines the result globally after hyperspectral image is reconstructed in band-by-band. It can maintain the spectral curve well. Meanwhile, the performance of other metrics has also been improved to a certain extent.

C. Two-Stage Methods

At present, there are few hyperspectral image SR models based on two-stage strategy. Now, we review existing coarse-to-fine architectures. For instance, Yuan *et al.* [27] propose a framework by transfer learning from natural image domain to hyperspectral image domain. Its aim is to learn the common properties of the two modes. Then collaborative nonnegative matrix factorization is introduce to enforce collaborations between the observed LR hyperspectral image and the transferred HR hyperspectral image. Although it can refine transferred results, it requires many iterations and is very time consuming to calculate. Wang *et al.* [28] first develop a recurrent feedback network to generate each band in hyperspectral image. Then the Pseudo-3D convolution is utilized to further optimize the initial result by supervised manner. Zhang *et al.* [29] propose a coarse-to-fine scheme, including fusion module and adaptation module. They are utilized to respectively general image prior and specific hyperspectral image. The adaptive module here is made up of multiple convolutional layers and requires many iterations to refine the results. Different from these works, to obtain better super-resolved result, in our paper, an enhanced back-projection method is designed to optimize by unsupervised manner. Importantly, the proposed approach does require iteration.

III. Proposed Method

The overall scheme of our proposed DualSR is shown in Fig. 3. As depicted, our method consists of two steps: coarse stage and fine stage. In coarse stage, the coarse hyperspectral image is restored in band-by-band using supervised way. Note that the method is our conference version [24]. For ease of description, the method is named CoarSR. During this process, the current band is assisted to enhance the exploration ability through four adjacent bands. After obtaining coarse result, we adopt unsupervised manner in fine stage to globally learn the information, which further optimizes result.

A. Coarse Stage with Supervised Manner

Previous works [5], [12], [21] almost adopt this way to build model, i.e., all the bands are processed synchronously. Actually, it is not necessary because the reconstruction of the current band is mainly influenced by adjacent bands. Inspired

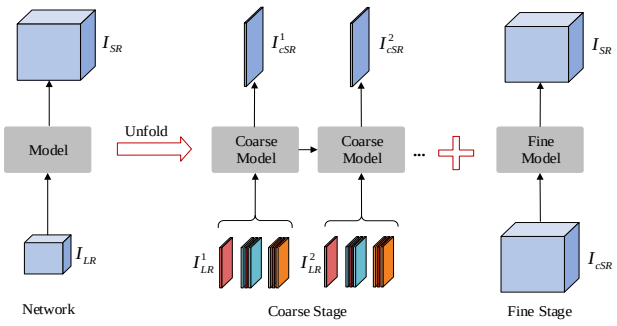


Fig. 3. Overview of the proposed DualSR for hyperspectral image SR.

by the pipeline in [16], we propose a novel structure for hyperspectral image SR via adjacent spectral fusion, whose flowchart is shown in Fig. 4. The overview of coarse stage mainly covers three modules, involving neighboring band partition (NBP), adjacent spectral fusion mechanism (ASFM), and feature context fusion (FCF).

1) Neighboring Band Partition: SFCSR employs two adjacent bands to discuss, where non-adjacent bands relative to the current band position are not considered. Fig. 1 shows that these bands are also greatly similar within a certain spectral range. In fact, relatively distant bands can also explicitly serve the current band. Therefore, our method adopts single band I^i_{LR} and its four adjacent bands for analysis, i.e.,

$$P(i) = \begin{cases} \left[I^1_{LR}, I^2_{LR}, I^3_{LR}, I^4_{LR}, I^5_{LR}\right], & 1 \le i < 3 \\ \left[I^{i-2}_{LR}, I^{i-1}_{LR}, I^i_{LR}, I^{i+1}_{LR}, I^{i+2}_{LR}\right], & 3 \le i \le L-3. \\ \left[I^{L-4}_{LR}, I^{L-3}_{LR}, I^{L-2}_{LR}, I^{L-1}_{LR}, I^L_{LR}\right], & L-3 < i \le L \end{cases} \tag{1}$$

To assist current band through neighboring bands, in our paper, according to the similarity with the current band, the five bands are partitioned into three groups by neighboring band partition (NBP). As for the case $3 \le i \le L-3$, it can be denoted as

$$f_{BP}(P(i)) = \begin{cases} I^i_{LR}, \\ \left[I^{i-1}_{LR}, I^i_{LR}, I^{i+1}_{LR}\right], \\ \left[I^{i-2}_{LR}, I^i_{LR}, I^{i+2}_{LR}\right], \end{cases} 3 \le i \le L-3. \tag{2}$$

With respect to those bands with larger or smaller index, this is different from the above case. They are partitioned by

$$f_{BP}(P(i)) = \begin{cases} I^i_{LR}, \\ \left[I^1_{LR}, I^2_{LR}, I^3_{LR}\right], \\ \left[I^4_{LR}, I^i_{LR}, I^5_{LR}\right], \end{cases} 1 \le i < 3, \tag{3}$$

$$f_{BP}(P(i)) = \begin{cases} I^i_{LR}, \\ \left[I^{L-2}_{LR}, I^{L-1}_{LR}, I^L_{LR}\right], \\ \left[I^{L-3}_{LR}, I^i_{LR}, I^{L-4}_{LR}\right], \end{cases} L-3 < i \le L. \tag{4}$$

Note that the current band I^i_{LR} appears in each group.

2) Adjacent Spectral Fusion Mechanism: Adjacent bands not only exhibit high similarity, but also provide the difference in diverse texture in some regions. It implies the missing information in the current band can be obtained from other bands. Existing works [4], [30], [31] apply small convolution kernel directly to hyperspectral image. Such the features of non-adjacent bands cannot be extracted, which hinders the borrowing of complementary content from relatively distant adjacent bands. Considering this issue, an alternative spectral fusion mechanism (ASFM) is developed to modulate the network to yield more complementary information from intra/inter-groups, which is depicted in Fig. 5.

Intra-group Fusion: After partitioning these bands, the initial features F^1_0, F^2_0, F^3_0 from three groups are obtained by 2D convolution layer. To sufficiently explore the potential information in each group, three parallel branches are performed to separately map the three inputs into deep features. For simplicity, we take the first subnet as an example to describe in details. Since this module only depends on spatial content, it is handled in much the same way as natural image SR. Therefore, to learn the spatial features of these bands, we employ the main residual module commonly used in natural image SR algorithms [32], [33] for reference to construct residual unit. Concretely, the residual unit in the intra-group part is composed of two identical blocks. Each block contains two convolution layers with kernel 3×3, ReLU function, and skip connection. Moreover, to borrow complementary information within different groups, the local features from other two groups are concatenated in current group to fuse. Under the action of intra-group fusion, the features of current band I^i_{LR} in the spatial domain are fully developed.

Inter-group Fusion: Only applying 2D convolution in intra-group fusion part cannot study content except for spatial dimension. Hence, to capture spectral features across different groups, we exploit 3D convolution to integrate those through inter-group fusion scheme. However, the direct use of regular 3D convolution significantly produces a large number of network parameters. For this reason, two convolution layers with kernels, i.e., $3\times1\times1$ and $1\times3\times3$, are attached with network to investigate spectral and spatial features, respectively. Then, the information between two domains is availably fused through an addition operation. With respect to hyperspectral image SR, its aim is to heighten spatial resolution while obtaining high spectral fidelity. To achieve this end, the fusion is followed by the convolution layer with kernel $1 \times 3 \times 3$ to further enhance the learning ability in spatial domain. By doing so, the inter-group fusion part aggregates the information between different groups. It realizes the knowledge complementarity across bands in the spectral and spatial domain.

3) Feature Context Fusion: In the process of SR bands, we only pay attention to the knowledge of multiple adjacent bands, i.e., spectral context. Actually, we can also use intermediate features to construct feature contexts because adjacent bands are similar. At present, feature context has been successfully applied in [16], which has shown to help model representation learning. Since our network is implemented through recurrent strategy in coarse stage, we also adopt this manner to achieve information interaction between bands, namely feature context fusion (FCF).

Let F^i_D and F^{i-1}_D denote the generated intermediate features for current bands i and previous band $i-1$. To fuse both features, the concatenation operation is introduced to dynamically generate fusion data through weights w_1 and w_2. Then, it is

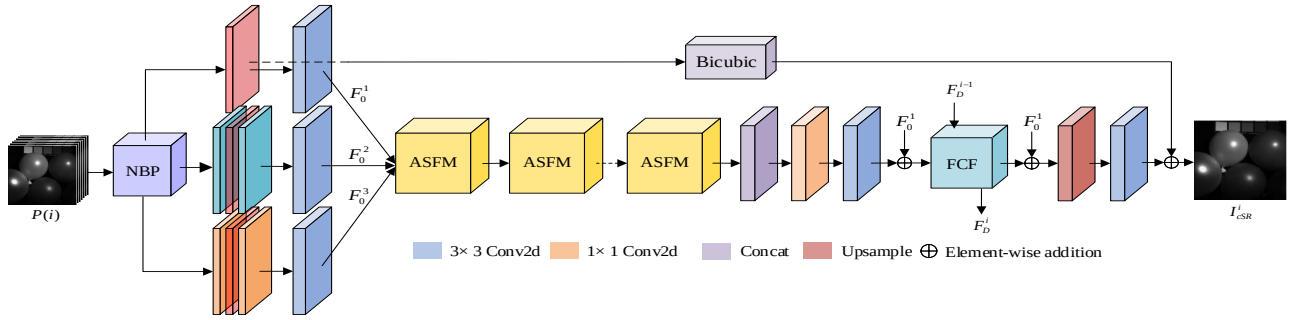


Fig. 4. Overview of the proposed CoarSR for hyperspectral image SR in coarse stage.

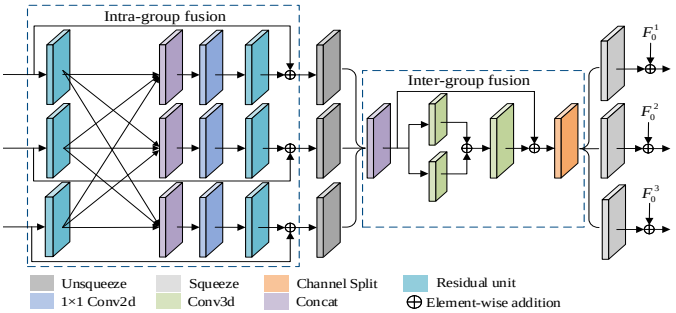


Fig. 5. Adjacent spectral fusion mechanism.

followed by convolution layer with kernel 1×1 to reduce channels, i.e.,

$$F = Conv([w_1 * F_D^i, w_2 * F_D^{i-1}]). \tag{5}$$

Note that FCF is not performed for first band reconstruction. Intuitively, this mechanism is very similar to RNN [34], except that the input band varies with the index increment.

Algorithm 1: Dual-stage hyperspectral image SR algorithm (DualSR)

Input: Hyperspectral image dataset containg LR-HR pair, scale factor s
Output: Super-resolved hyperspectral image I_{SR}

```
Randomly initialize coarse model parameters θ ;
while not converged do
    Sample LR-HR batch;
    while i≤ L do
        Partiton bands into three groups by Eqs. 2-4;
        Update θ by excuting coarse model;
        i ← i + 1;
    end
end
Generate coarse model parameters θ_c;
Obtain coarse SR results U and V in terms of scale factor s;
Compute the reconstruction error under spectral angle constrain by Eq. 11;
Obtain fine SR result I_SR using Eq. 13.
```

B. Fine Stage Using Unsupervised Manner

Despite our method utilizes five bands to restore single band in coarse stage, it still lacks information exploration of a larger perspective in spectral dimension. As a result, the super-resolved hyperspectral image inevitably suffers serious spectral distortion by such single band SR. A natural way is to input all bands into the model to optimize the results by supervised manner. Nevertheless, this approach violates the original intention. Considering this limitation, we apply unsupervised manner to further refine initial results.

Back-projection [35] optimizes the reconstruction error through an efficient iterative strategy. For this algorithm, it usually utilizes multiple upsampling descriptors to upsample LR image and iteratively calculate the reconstruction error. Currently, back-projection is widely introduced in natural image SR [36]–[38], which has been proved to develop the quality of SR image. Yang *et al.* [36] employ back-projection to refine results produced according to coupled dictionaries through unsupervised strategy. However, the initialization which leads to an optimal solution remains unknown. The main reason is that the algorithm involves predefined hyperparameters, such as number of iteration and convolution kernel.

To extend this algorithm, we further develop back-projection without predefined hyperparameters, which is shown in Fig. 6. Specifically, given a LR image I_{LR}, the coarse result U is obtained by coarse model f according to scale factor s, which is expressed as

$$U = \mathcal{F}(s, I_{LR}). \tag{6}$$

Previous studies aim to resize SR image to the same size as LR image, and measure the reconstruction error between the two. As for large scale factor s, this approach discards too much useful information, resulting in very little performance gain for reconstruction. Therefore, in our paper, the coarse model is utilized to scale LR image I_{LR} again to obtain another initial SR result V for scale factor $\frac{s}{2}$, i.e.,

$$V = \mathcal{F}(\frac{s}{2}, I_{LR}), \tag{7}$$

Then, the coarse result U for scale factor s is downsampled using Bicubic. After generating LR image M, the reconstruction error is reckoned by

$$M = R_{bicu}(0.5, U), \tag{8}$$

$$Z = V - M, \tag{9}$$

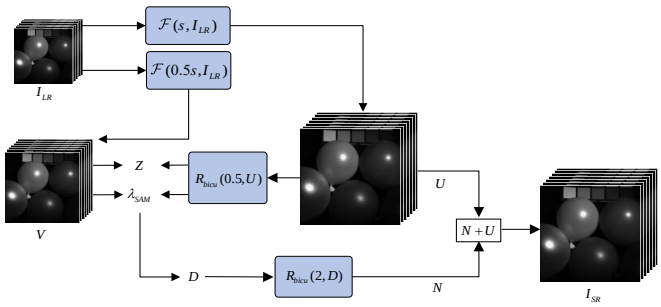


Fig. 6. Enhanced back-projection method via spectral angle constraint.

To yield high spectral fidelity, we introduce Spectral Angle Mapper (SAM) [39] to constrain error, which is represented as

$$\lambda_{SAM} = arccos\left(\frac{< M, V >}{||M||_2||V||_2}\right), \tag{10}$$

$$D = \begin{cases} \lambda_{SAM} \times Z, & \lambda_{SAM} < 1, \\ Z, & otherwise, \end{cases} \tag{11}$$

where $arccos(\cdot)$ is arccos function. $< \cdot, \cdot >$ is dot product. $|| \cdot ||_2$ denotes the L2 norm. The obtained error is upsampled using Bicubic and added to the coarse result U. Eventually, the fine result I_{SR} is obtained by

$$N = R_{bicu}(2, D), \tag{12}$$

$$I_{SR} = U + N. \tag{13}$$

Our method abandons the convolution operation executed in [36]. Importantly, the strategy does not require iteration. Since the refinement is conducted using unsupervised manner, it can be applied to any hyperspectral image SR algorithms to further optimize SR image, which indicates it belongs to a plug-and-play method. For the description of dual-stage SR algorithm, the procedures are summarized in **Algorithm 1**.

IV. Experiments

In this section, extensive experiments are conducted to evaluate the superiority of the proposed DualSR, and the comparative results are given from both quantitative and qualitative aspects.

A. Datasets

The CAVE dataset [40] was captured by a tunable filter and a cooled CCD camera. It contains 32 scenes for a wide variety of objects and materials, which is grouped into 5 parts. The range of full spectral resolution is from 400 nm to 700 nm at 10 nm steps, which means that each scene has 31 bands. The spatial resolution of scene is 512×512 pixels. Unlike CAVE dataset, Harvard dataset [41] is more challenging. It was collected under daylight, both outdoors and indoors. The dataset contains 71 hyperspectral images. Each scene consists of 1392×1040 pixels and 31 spectral bands in the range of 420 nm to 720 nm. In our work, we randomly select 80% of the data as the training set and the rest for testing. We augment the given training data by choosing 24 patches. With respect to each patch, its size is scaled 1, 0.75, and 0.5 times, respectively. We rotate these patches 90° and flip them horizontally. Through various blur kernels, we then subsample these patches into LR hyperspectral images with the size of $L \times 32 \times 32$.

TABLE I
Study of the effectiveness of adjacent bands for scale factor $\times 4$ on CAVE dataset.

Type	Scale	PSNR	SSIM	SAM
Adjacent bands	✗	39.617	0.9313	3.256
	✔	39.665	0.9328	3.152

B. Implementation Details

For our network, the convolution kernel after concatenation is set to 1×1, which reduces the number of channels. We adopt sub-pixel convolution layer to upscale the features into HR space in terms of upsampling operation. The kernel of other convolution operations involved in the network is fixed to 3×3, and the number of convolution kernels is defined as 64. In the training phase, our network is trained using $L1$ loss function. The mini-batch is set to 64. We optimize our network using ADAM optimizer with β_1 = 0.9 and β_2 = 0.999 and initial learning rate 10^{-4}. For learning rate, it is gradually updated by a half at every 30 epochs. Our model runs on PyTorch framework and is trained with NVIDIA GTX 1080 GPU.

C. Evaluation Metrics

To quantitatively evaluate the proposed method, we apply Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM), and Spectral Angle Mapper (SAM). Among these metrics, PSNR and SSIM are to evaluate the performance of super-resolved hyperspectral image in spatial domain. Generally, the higher their values are, the better the performance is. SAM is to analyze the performance of restored image in spectral domain. The smaller the value is, the less the spectral distortion is.

D. Effectiveness of Adjacent Bands

In coarse stage, adjacent bands are exploited to modulate the network to yield more complementary information, so as to enhance the performance of current band. To verify the study of the effectiveness of these adjacent bands, we replace adjacent band in other groups as current band, i.e., all other bands in each group been changed to the current band. It can avoid structural changes in the model. Table I reports the study of the effectiveness of adjacent bands. As seen from this table, the numerical results show that the adjacent bands do help to improve the overall performance of the model with adjacent bands. It indicates that relatively distant bands can also explicitly serve the current band, which reveals the effectiveness of adjacent bands.

TABLE II
ABLATION STUDY FOR SCALE FACTOR $\times 4$ ON CAVE DATASET. BICUBIC SR KERNEL IS USED.

Module	Different combinations of modules			
Inter-group fusion	✔	✗	✔	✔
Intra-group fusion	✗	✔	✔	✔
FCF	✗	✗	✗	✔
PSNR	38.720	39.521	39.635	39.665
SSIM	0.9295	0.9322	0.9327	0.9328
SAM	3.304	3.211	3.151	3.152

E. Ablation Study

To verify the effectiveness of key modules, we test variant versions of our model by removing each component, including inter-group fusion, intra-group fusion, and FCF.

Table II shows the ablation study for key modules. For clear show, we only provide the results evaluated by CoarSR. Specifically, with intra-group fusion, the network directly exploits 2D convolution to explore spatial information. In contrast with only the existence of inter-group fusion, the network exhibits better performance. It implies that the model should pay more attention to the exploration of spatial features and contribute to generating sharper results, which is consistent with the original intention of our design. When two kinds of group fusion are attached, the network tends to bring fine results. It indicates the combination of the two effectively integrates complementary information between different groups.

Finally, all three modules are involved in feature extraction. One can observe that this combination produces better performance than any other arbitrary module combination, except for SAM. It is concluded that these quantitative analyses revels the effectiveness and benefits of the proposed modules.

F. Study of Enhanced Back-projection

To further refine result obtained by CoarSR, we propose an enhanced back-projection under spectral angle constraint. In this section, we study the effectiveness of enhanced back-projection on bicubic SR kernel. Note that the proposed enhanced back-projection involves two different scale factors during fine stage. Therefore, the results are evaluated for scale factors $\times 4$ and $\times 8$ on two public datasets.

*1) **Performance Comparison of Back-projection**:* Since the proposed enhanced back-projection belongs to the post-processing method and is unsupervised, it is suitable for any hyperspectral image algorithms to further boost the performance. To investigate the performance gain for different SR algorithms, we apply it as post-processing of some methods to analyze and discuss. When the scale factor is set to 4, the designed back-projection method involves two different scale factors, i.e., $\times 2$ and $\times 4$. In SSPSR, the minimum upscale factor is set to $\times 4$ due to progressive upsampling. Therefore, the results of scale factor $\times 4$ cannot be provided when conducting our designed back-projection.

Table III shows the differences of two scale factors with or without post-processing. One can notice that all methods markedly reveal performance gains under action of enhanced back-projection, compared with without post-processing. Moreover, we also present the performance by setting back-projection from [36] as post-processing. Different from our enhanced back-projection, [36] only contains one kind of scale factor. Hence, we provide the results obtained by SSPSR for scale factor $\times 4$. It can be seen in Table IV that although the method can produce relatively good performance, it is not as obvious as the improvement in our work. Notably, the spectral distortion actually gets worse. Through analysis, this is enough to prove the effectiveness of our method.

*2) **Effect of Spectral Angle Constraint**:* In our work, we introduce Spectral Angle Mapper (SAM) to constrain reconstruction error, which can alleviate spectral distortion. To explore whether the constraint is effective, we delete it and see how the performance changes. Table V depicts the results using enhanced back-projection without spectral angle constraint. We combine the results in Table III for analysis. As one can see, some values do not change because the value calculated for SAM is greater than 1. As a result, the constraint does not work via Eq. 11. In addition to that, we observe that spectral angle constraint does enhance spectral fidelity in most cases. Meanwhile, the performance of other metrics has also been improved to a certain extent. It indicates our method can effectively integrate spatial information under constraint and make the spectral curves at different scales as consistent as possible.

G. Generalizability to Various Datasets

Using known bicubic downsampling condition, we compare our proposed DualSR with existing multiple approaches on CAVE and Harvard datasets, including 3D-FCNN [14], EDSR [42], SSPSR [25], MCNet [21], SFCSR [21], ERCSR [4]. Next, we investigate the performance by quantitative evaluation and qualitative comparison.

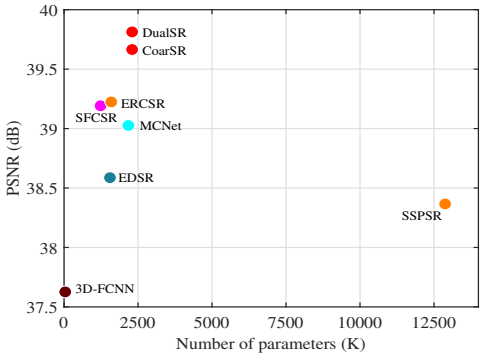


Fig. 7. PSNR performance versus number of parameters. The results between both are provided for scale factor $\times 4$ on CAVE dataset. Our CoarSR and DualSR obtain better trade off between performance and model size.

*1) **Quantitative Evaluation**:* Tables VI and VII show the quantitative evaluation of existing SR approaches for different scale factors on two datasets. We can notice that our method attains excellent results regardless of whether there is fine stage. Among the approaches only using 2D convolution, SSPSR attains superior performance across datasets. It first fuses features between groups for small scale factor, and then

TABLE III
PERFORMANCE GAIN BY SETTING ENHANCED BACK-PROJECTION METHOD AS POST-PROCESSING FOR SCALE FACTOR ×4 ON CAVE DATASET.

Post-processing	Scale	Metrics	CAVE				Harvard			
			SSPSR [25]	SFCSR [16]	ERCSR [4]	CoarSR	SSPSR [25]	SFCSR [16]	ERCSR [4]	CoarSR
✗	×4	PSNR	38.366	39.192	39.224	39.665	39.293	40.077	40.211	40.312
		SSIM	0.9272	0.9321	0.9322	0.9328	0.9333	0.9373	0.9374	0.9405
		SAM	3.484	3.221	3.243	3.152	2.448	2.407	2.384	2.384
	×8	PSNR	34.151	35.294	35.102	35.475	34.435	35.097	35.252	35.628
		SSIM	0.8718	0.8828	0.8824	0.8828	0.8676	0.8737	0.8733	0.8755
		SAM	4.831	4.378	4.445	4.269	2.899	2.911	2.888	2.794
✔	×4	PSNR	–	39.372	39.334	39.813	–	40.257	40.308	40.540
		SSIM	–	0.9335	0.9334	0.9339	–	0.9394	0.9401	0.9423
		SAM	–	3.201	3.196	3.130	–	2.375	2.367	2.372
	×8	PSNR	34.453	35.341	35.479	35.628	34.855	35.414	35.564	35.628
		SSIM	0.8771	0.8846	0.8855	0.8862	0.8719	0.8775	0.8793	0.8799
		SAM	4.661	4.267	4.264	4.141	2.837	2.803	2.829	2.774

TABLE IV
PERFORMANCE GAIN BY SETTING [36] AS POST-PROCESSING FOR SCALE FACTOR ×4 ON CAVE DATASET.

Scale	Metrics	CAVE				Harvard			
		SSPSR [25]	SFCSR [16]	ERCSR [4]	CoarSR	SSPSR [25]	SFCSR [16]	ERCSR [4]	CoarSR
×4	PSNR	38.426	39.225	39.244	39.675	39.350	40.089	40.224	40.327
	SSIM	0.9286	0.9326	0.9330	0.9332	0.9339	0.9380	0.9395	0.9406
	SAM	3.545	3.230	3.229	3.165	2.434	2.406	2.382	2.383
×8	PSNR	34.335	35.335	35.147	35.497	34.538	35.118	35.179	35.270
	SSIM	0.8730	0.8835	0.8830	0.8832	0.8680	0.8737	0.8735	0.8756
	SAM	4.922	4.398	4.467	4.283	2.905	2.925	2.858	2.814

TABLE V
PERFORMANCE GAIN BY SETTING ENHANCED BACK-PROJECTION WITHOUT SPECTRAL ANGLE CONSTRAINT AS POST-PROCESSING FOR SCALE FACTOR ×4 ON CAVE DATASET.

Scale	Metrics	CAVE				Harvard			
		SSPSR [25]	SFCSR [16]	ERCSR [4]	CoarSR	SSPSR [25]	SFCSR [16]	ERCSR [4]	CoarSR
×4	PSNR	–	39.221	39.317	39.774	–	40.246	40.295	40.541
	SSIM	–	0.9330	0.9333	0.9338	–	0.9394	0.9396	0.9422
	SAM	–	3.205	3.209	3.141	–	2.381	2.376	2.385
×8	PSNR	34.453	35.341	35.479	35.609	34.928	35.417	35.574	35.675
	SSIM	0.8771	0.8846	0.8855	0.8861	0.8724	0.8774	0.8796	0.8799
	SAM	4.661	4.267	4.264	4.145	2.839	2.805	2.837	2.786

TABLE VI
QUANTITATIVE EVALUATION OF EXISTING SR APPROACHES FOR DIFFERENT SCALE FACTORS ON CAVE DATASET. THE BEST RESULTS ARE RED FONT, AND THE SECOND BEST RESULTS ARE BLUE FONT.

Scale	Metric	Bicubic	3D-FCNN [14]	EDSR [42]	SSPSR [25]	MCNet [21]	SFCSR [16]	ERCSR [4]	**CoarSR (Ours)**	**DualSR (Ours)**
×4	PSNR ↑	35.755	37.626	38.587	38.366	39.026	39.192	39.224	39.665	39.813
	SSIM ↑	0.9071	0.9195	0.9292	0.9272	0.9319	0.9321	0.9322	0.9328	0.9339
	SAM ↓	3.944	3.360	3.804	3.484	3.292	3.221	3.243	3.152	3.130
×8	PSNR ↑	31.805	32.956	31.554	34.151	35.320	35.294	35.102	35.475	35.628
	SSIM ↑	0.8485	0.8600	0.8233	0.8718	0.8833	0.8828	0.8824	0.8828	0.8862
	SAM ↓	5.291	4.384	10.670	4.831	4.423	4.378	4.445	4.269	4.4107

TABLE VII
QUANTITATIVE EVALUATION OF EXISTING SR APPROACHES FOR DIFFERENT SCALE FACTORS ON HARVARD DATASET. THE BEST RESULTS ARE RED FONT, AND THE SECOND BEST RESULTS ARE BLUE FONT.

Scale	Metric	Bicubic	3D-FCNN [14]	EDSR [42]	SSPSR [25]	MCNet [21]	SFCSR [16]	ERCSR [4]	**CoarSR (Ours)**	**DualSR (Ours)**
×4	PSNR ↑	37.227	38.143	39.175	39.293	40.081	40.077	40.211	40.312	40.540
	SSIM ↑	0.9122	0.9188	0.9324	0.9333	0.9367	0.9373	0.9374	0.9405	0.9423
	SAM ↓	2.531	2.363	2.560	2.448	2.410	2.407	2.384	2.384	2.372
×8	PSNR ↑	33.275	33.363	34.068	34.435	34.927	35.097	35.138	35.252	35.628
	SSIM ↑	0.8518	0.8448	0.8598	0.8676	0.8732	0.8737	0.8733	0.8755	0.8799
	SAM ↓	2.884	2.726	3.051	2.899	2.957	2.911	2.888	2.794	2.774

performs SR for large scale factor. The progressive upsampling is more conducive to the generation of clear image. Thanks to the exploration of spectral information, in contrast, the model using 3D convolution has better performance than the model using 2D convolution on the whole, except for 3D-FCNN. With respect to 3D-FCNN, the main reasons for its low results are the shallow design and the lack of residual learning. With respect to recurrent networks, i.e., SFCSR and CoarSR, they exploit spectral structure from multiple neighboring bands to achieve complementary information. Their structures are more helpful to generate a single clear band. Although the number of parameters of CoarSR is more than that of SFCSR (see Fig. 7), the results obtained by our CoarSR are significantly higher than that of SFCSR, i.e., +0.473 dB, 0.0007, -0.069 on scale factor ×4 on CAVE dataset. In our opinion, the main reasons are in two aspects. The one is that more representation learning is added in spatial domain, when spectral information is available. The other is that our method fully exploit the complementary information of several neighboring bands, recovering missing details. Interestingly, without increasing the number of parameters, the values of proposed DualSR increase obviously in all three metrics, outperforming others by a non-negligible margin.

*2) **Qualitative Evaluation**:* We further analyze the proposed method by qualitative evaluation. To show the visual results in terms of spatial domain, in our work, two bands in single SR hyperspectral image are provided, i.e., 10-*th* band and 20-*th* band. Since the band is gray, it does not effectively observe the edge details of image. To clearly distinguish the difference between super-resolved and HR image, the absolute error map is given. Figs. 8 and 9 present visual example for scale factor ×4 on two datasets. One observe that our method produces low absolute errors. In particular, there are more shallow edges in some positions, which indicates that the proposed approach can generate sharper edges and finer details. It is consistent with the analysis in Tables VI and VII, which further demonstrates that our approach can simultaneously learn spectral and spatial knowledge while generating diverse textures.

We also exhibit the visual results in spectral domain. Concretely, one pixel in spatial domain is selected, and the spectral curve is plotted along spectral dimension at the same position. Fig. 10 displays the spectral distortion of super-resolved hyperspectral image by randomly choosing two pixels on two datasets, respectively. We can see that our DualSR maintains the same curve as the ground-truth in most cases. It validates that the proposed method can yield higher spectral fidelity against other approaches.

H. Generalizability to Various Blur Kernels

In this section, we show results on various blur kernel conditions to demonstrate generalizability. For fair comparison, each image is subsampled by different downsampling kernels using [43]. The LR images are input into the model to get the corresponding restored images, and the numerical results are obtained by calculating the average value. Note that the downscaling kernels in our work are *cubic*, *lanczos*, *box*, and *linear*.

*1) **Quantitative Evaluation**:* Table VIII compares our performance against other methods on CAVE dataset. As can be seen, these conventional SR networks obtain severe performance drops due to fixed bicubic SR kernel, which deviates from the non-bicubic SR kernels. It indicates these methods cannot deal with other non-bicubic SR kernels. On the contrary, we observe that the proposed CoarSR can address non-bicubic SR kernel degraded images well, and its performance exceeds that of SFCSR for scale factor ×4 by a very large margin: + 1.022 dB. This is due to the fact that our method uses various downscaling kernels in the process of training, while other methods do not. From this perspective, CoarSR has better generalization performance, compared to other methods. In other words, the proposed approach is more robust across different scenarios.

*2) **Qualitative Evaluation**:* Similarly, we take the same approach from Section IV-G2 to provide the visual results in terms of spatial and spectral domain. To make a simple comparison, the results subsampled by linear downscaling kernel are exhibited, which is shown in Figs. 11 and 12. As seen, our method still achieves better performance in two aspects, which reveals it can deal with LR image subsampled by non-bicubic kernel well. Through the above analysis, it proves the proposed model surpasses state-of-the-art networks over each dimension on various blur kernel conditions.

V. CONCLUSION

In this study, we propose a new structure for hyperspectral image SR, which contains coarse stage and fine stage. In coarse stage, the coarse SR image is obtained in band-by-band. Different from previous works, in this process, we add more adjacent bands to the model, and encourage the network by

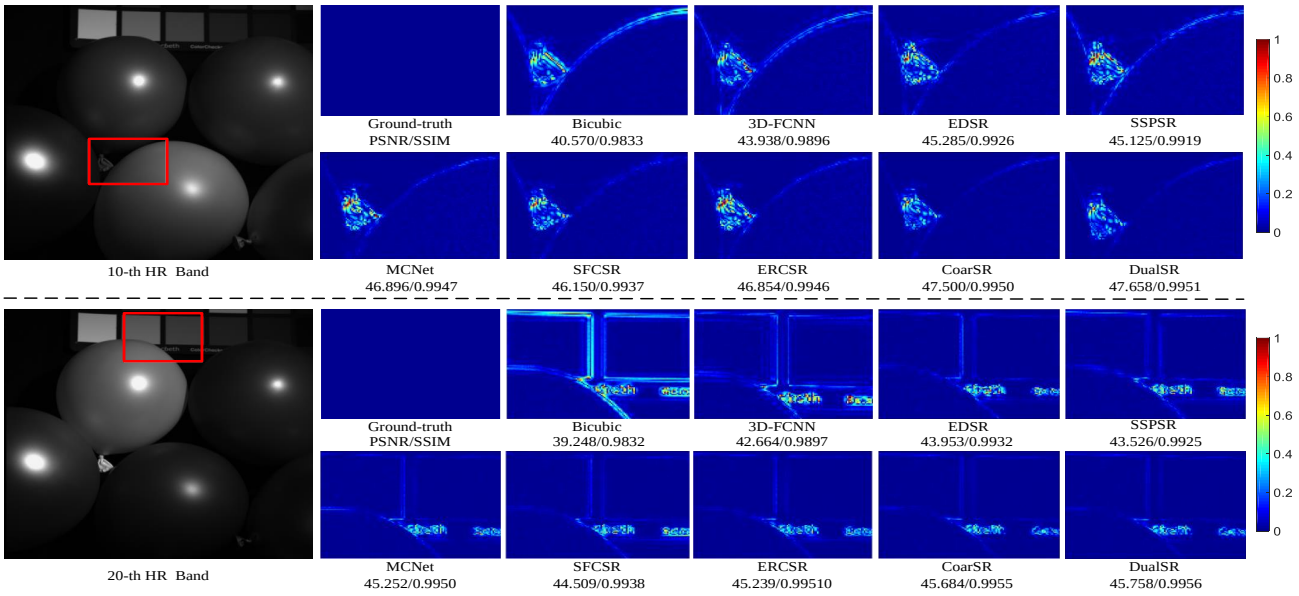


Fig. 8. Visual results in terms of spatial domain with existing SR methods on CAVE dataset. The results of *balloons* image are evaluated for scale factor $\times 4$. The first line denotes SR results of 10-*th* band, and the second line denotes SR results of 20-*th* band.

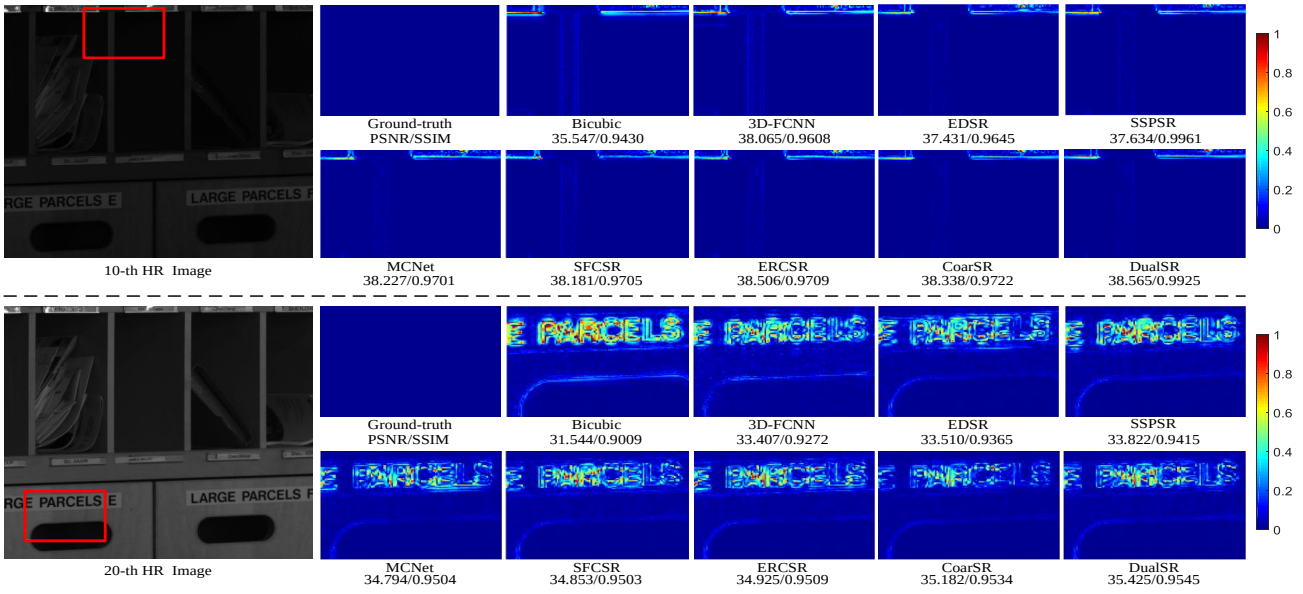


Fig. 9. Visual results in terms of spatial domain with existing SR methods on Harvard dataset. The results of *imgd5* image are evaluated for scale factor $\times 4$. The first line denotes SR results of 10-*th* band, and the second line denotes SR results of 20-*th* band.

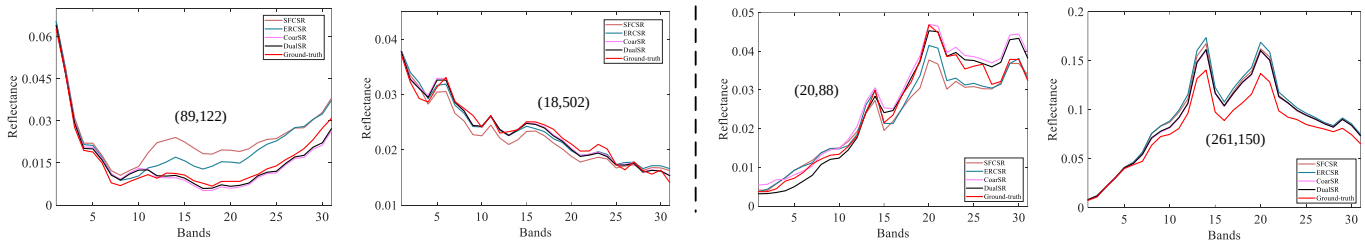


Fig. 10. Visual comparison in terms of spectral domain by randomly selecting two pixels for scale factor $\times 4$. The two on the left are the results of *balloons* image on CAVE dataset. The two on the right are the results of *imgd5* image on Harvard dataset. Note that to avoid confusion, only two representative algorithms are compared with our methods.

TABLE VIII
SR IN THE PRESENCE OF VARIOUS DOWNSCALING KERNELS. THE BEST RESULTS ARE RED FONT, AND THE SECOND BEST RESULTS ARE BLUE FONT.

Scale	Metric	Bicubic	3D-FCNN [14]	EDSR [42]	SSPSR [25]	MCNet [21]	SFCSR [16]	ERCSR [4]	**CoarSR (Ours)**	**DualSR (Ours)**
×4	PSNR ↑	35.652	37.263	38.204	37.932	38.317	38.410	38.390	39.296	39.432
	SSIM ↑	0.9067	0.9175	0.9263	0.9253	0.9281	0.9281	0.9286	0.9309	0.9312
	SAM ↓	3.871	3.241	3.601	3.578	3.410	3.390	3.403	3.242	3.238
×8	PSNR ↑	31.735	32.813	34.250	33.920	34.601	34.635	34.503	34.972	35.322
	SSIM ↑	0.8496	0.8584	0.8711	0.8699	0.8787	0.8828	0.8784	0.8802	0.8842
	SAM ↓	5.134	4.302	4.855	4.893	4.638	4.893	4.727	4.462	4.256

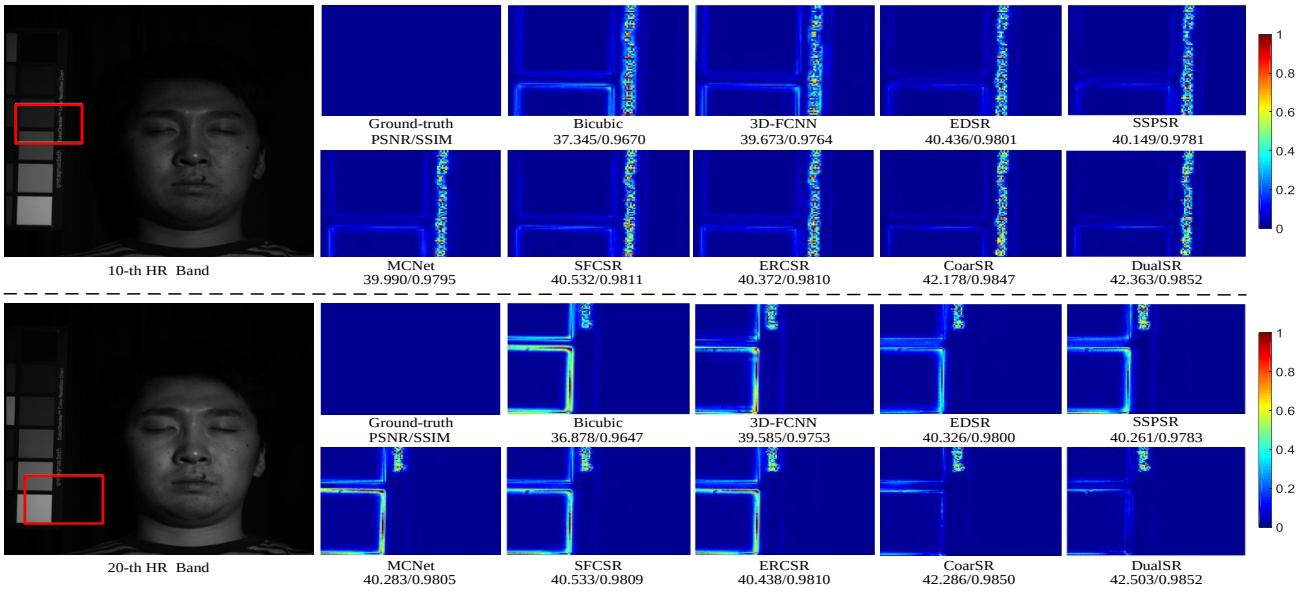


Fig. 11. Visual results subsampled by linear kernel in terms of spatial domain on CAVE dataset. The results of *face* image are evaluated for scale factor ×4. The first line denotes SR results of 10-*th* band, and the second line denotes SR results of 20-*th* band.

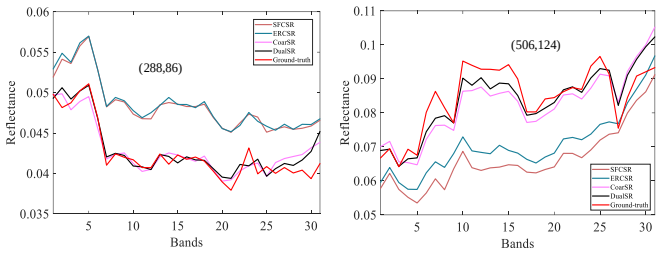


Fig. 12. Visual comparison in terms of spectral domain by randomly selecting two pixels for scale factor ×4. Note that to avoid confusion, only two representative algorithms are compared with our methods.

grouping them to learn the potential information of the current band. With the intra/inter-group fusion, the complementary information is borrowed from adjacent bands, achieving refined details. To learn the content of spectal-spatial consistency, an enhanced back-projection method is proposed. This method can be utilized as the post-processing for any hyperspectral image SR algorithms, dramatically boosting performance gain. Experiments demonstrate our proposed DualSR can produce state-of-the-art results over existing works. Compared with second best method, the designed approach in PSNR and SSIM increase by 0.589dB and 0.0017 for scale factor ×4, and its SAM decreased by 0.113.

In the future, we plan to extend our method by constructing more realistic training pairs. In the real scene, the image degradation process is complicated due to noise motion blur and other factors, and the performance of the traditional SR network which only uses Bicubic kernel training decreases obviously in the face of other kernels. Thus, our next focus is to build more realistic training data.

REFERENCES

[1] S. G. Bajwa, P. Bajcsy, P. Groves, and L. F. Tian, "Hyperspectral image data mining for band selection in agricultural applications," *Trans. ASAE*, vol. 47, no. 3, pp. 895, 2004.

[2] 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.*, 2017, pp. 3862–3871.

[3] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2329–2336.

[4] Q. Li, Q. Wang, and X. Li, "Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sensing*, 2020.

[5] R. Jiang, X. Li, A. Gao, L. Li, H. Meng, S. Yue, and L. Zhang, "Learning spectral and spatial features based on generative adversarial network for hyperspectral image super-resolution," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process Proc.*, 2019, pp. 3161–3164.

[6] 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 *IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3113–3116.

[7] D. Liu, J. Li, and Q. Yuan, "A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sensing*, 2021.

[8] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3070–3079.

[9] Z.-W. Pan and H.-L. Shen, "Multispectral image super-resolution via RGB image fusion and radiometric calibration," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1783–1797, 2019.

[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 Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1585–1594.

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

[12] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral image super-resolution via intrafusion network," *IEEE Trans. Geosci. Remote Sensing*, vol. 58, no. 10, pp. 7459–7471, 2020.

[13] J. Li, R. Cui, B. Li, R. Song, Y. Li, and Q. Du, "Hyperspectral image super-resolution with 1D–2D attentional convolutional neural network," *Remote Sens.*, vol. 11, no. 23, pp. 2859, 2019.

[14] 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, pp. 1139, 2017.

[15] J. Yang, L. Xiao, Y.-Q. Zhao, and J. C.-W. Chan, "Hybrid local and nonlocal 3-D attentive CNN for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, 2020.

[16] Q. Wang, Q. Li, and X. Li, "Hyperspectral image super-resolution using spectrum and feature context," *IEEE Trans. Ind. Electron.*, 2020.

[17] F. Fang, J. Li, and T. Zeng, "Soft-edge assisted network for single image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 4656–4668, 2020.

[18] K. Zhang, D. Tao, X. Gao, X. Li, and J. Li, "Coarse-to-fine learning for single-image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1109–1122, 2017.

[19] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sensing*, vol. 58, no. 7, pp. 4590–4603, 2020.

[20] M. Zolfaghari, K. Singh, and T. Brox, "ECO: Efficient convolutional network for online video understanding," in *Proc. European Conference on Computer Vision*, 2018, pp. 713–730.

[21] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution," *Remote Sens.*, vol. 12, no. 10, pp. 1660, 2020.

[22] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 7459–7471, 2020.

[23] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sensing*, vol. 54, no. 1, pp. 88–102, 2016.

[24] Q. Li, Q. Wang, and X. Li, "Hyperspectral image super-resolution via adjacent spectral fusion strategy," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 1645–1649.

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

[26] W. Wei, J. Nie, Y. Li, L. Zhang, and Y. Zhang, "Deep recursive network for hyperspectral image super-resolution," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1233–1244, 2020.

[27] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, 2017.

[28] X. Wang, J. Ma, and J. Jiang, "Hyperspectral image super-resolution via recurrent feedback embedding and spatial–spectral consistency regularization," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–13, 2021.

[29] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3073–3082.

[30] J. Li, R. Cui, B. Li, R. Song, Y. Li, Y. Dai, and Q. Du, "Hyperspectral image super-resolution by band attention through adversarial learning," *IEEE Trans. Geosci. Remote Sensing*, vol. 58, no. 6, pp. 4304–4318, 2020.

[31] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021.

[32] S. Anwar, S. Khan, and N. Barnes, "Hyperspectral image data mining for band selection in agricultural applications," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 60:1–60:34, 2020.

[33] Y. Hu, J. Li, and Y. Huang, "Channel-wise and spatial feature modulation network for single image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 59, no. 6, pp. 5028–5039, 2021.

[34] T. Mikolov, S. Kombrink, L. Burget, J. Černockỳ, and S. Khudanpur, "Extensions of recurrent neural network language model," in *IEEE Int. Conf. Acoust. Speech Signal Process Proc.*, 2011, pp. 5528–5531.

[35] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Model. Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.

[36] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.

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

[38] M. Haris, G. Shakhnarovich, and N. Ukita, "Recurrent back-projection network for video super-resolution," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 3897–3906.

[39] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H Goetz, "The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, no. 2-3, pp. 145–163, 1993.

[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, 2010.

[41] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 193–200.

[42] 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. 1132–1140.

[43] T. Michaeli and M. Irani, "Nonparametric blind super-resolution," in *IEEE Int. Conf. Comput. Vision*, 2013, pp. 945–952.

Qiang Li received the Ph.D. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China in 2022.

He is currently a postdoc with the School of Electronic Engineering, Xidian University, Xi'an. His research interests include remote sensing image processing and computer vision.

Yuan Yuan (M'05-SM'09) is currently a Full Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. She has authored or co-authored over 150 papers, including about 100 in reputable journals, such as the IEEE TRANSACTIONS AND PATTERN RECOGNITION, as well as the conference papers in CVPR, BMVC, ICIP, and ICASSP. Her current research interests include visual information processing and image/video content analysis.

Xiuping Jia (M'93-SM'03-F'21) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from the University of New South Wales, Canberra, ACT, Australia, in 1996.

Since 1988, she has been with the School of Information Technology and Electrical Engineering, University of New South Wales, where she is a Senior Lecturer. She is also a Guest Professor with Harbin Engineering University, Harbin, China, and an Adjunct Researcher with the China National Engineering Research Center for Information Technology in Agriculture, Beijing, China. She is a coauthor of the remote sensing textbook titled Remote Sensing Digital Image Analysis (Springer-Verlag, third edition, 1999, and fourth edition, 2006). Her research interests include remote sensing, image processing, and spatial data analysis.

Dr. Jia is also a Subject Editor of the Journal of Soils and Sediments and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Qi Wang (M'15-SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.