Multiple Frame Splicing and Degradation Learning for Hyperspectral Imagery Super-Resolution

Chenwei Deng, Xingshi Luo, and Wenzheng Wang

***Abstract*—Hyperspectral imagery (HSI) is an emerging remote sensing technology to discriminate different remote sensing objects. However, the HSI spatial resolution is relatively low due to the trade-off in restricted physical hardware and various imaging conditions, restricting the subsequent object detection applications. At present, the single hyperspectral super resolution (SHSR) strategy has encountered the bottleneck on more precise details extraction, and the fusion hyperspectral image super resolution (FHSR) strategy must need extra RGB/multispectral information, which is not suitable for general HSI usage. Also, both types of current strategies focus less on the multiple degradation causes of low spatial resolution. In this article, a step forward in designing a novel framework of multiple frame splicing strategy to greatly improve the SHSR quality, and applying multiple HSI degradation models to better fit the real degradation circumstance. Specifically, the framework is an end-to-end super resolution (SR) network that supersedes a single up-sampling module and removes complex attention residual model due to the same size of multiple splicing low-resolution input samples with high-resolution outputs. The effective framework will alleviate the vague at higher multiples, and accelerate the training convergence. Based on this framework, multiple degradation low-resolution samples can be simultaneously combined to fit better for the blind SR result. Concretely, the degradation focus on the blur, noise, compression, and their combinations to simulate the real degradation. Experimental results on three different hyperspectral datasets demonstrate that the proposed multiple frame splicing and degradation model (MFSDM) algorithm can significantly enhance the details in the recovered high-resolution hyperspectral images, and outperforms the state-of-the-art SHSR methods.**

***Index Terms*—Hyperspectral remote sensing, hyperspectral super resolution, image super resolution, multiple degradation super resolution, multiple frame super resolution.**

I. Introduction

HYPERSPECTRAL imagery (HSI) is a challenging and comprehensive interdisciplinary subject that arose in the 1980s [1]. Compared with the traditional panchromatic and multispectral (MS) remote sensing, the most prominent advantage of hyperspectral imaging is the 3-D data cube can record both 2-D spatial information and 1-D richer reflectance or radiance information of a scene at nanometer spectral resolution in the range of ultraviolet, visible, near-infrared to long-wave infrared [2].

Manuscript received 1 July 2022; revised 5 September 2022; accepted 6 September 2022. Date of publication 19 September 2022; date of current version 6 October 2022. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62171040. *(Corresponding author: Wenzheng Wang.)*

The authors are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: cwdeng@bit.edu.cn; xingshi_luo@bit.edu.cn; wang_wenzheng@bit.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3207777

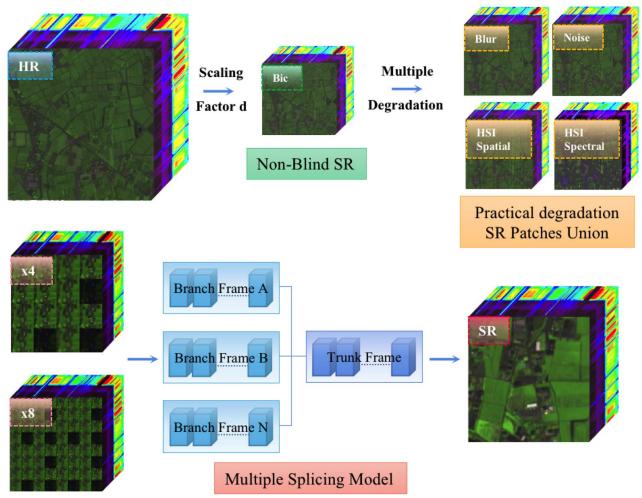


Fig. 1. General idea of MFSDM, which contains multiple degradation (upper) and multiple splicing network (lower).

Therefore, many problems that can not be solved in traditional remote sensing imaging methods can be solved by hyperspectral technology. Nowadays, the HSI technology has been widely applied in many remote sensing fields, such as object detection [3], [4], [5], ground classification [6], [7], environmental monitoring [8], anomaly detection [9], etc [10]. Recently, the multidimensional and multimodal HSIs have expanded the possibility of more applications, such as visual recognition and tracking [11], [12], [13].

For these practical applications, the spatial resolution of acquired HSIs limits the subsequent tasks. Under the influence of complex imaging conditions, hardware limitations, and other blind degradation factors, the current HSIs fail to obtain high-resolution spatial with comparably high-resolution spectral information. The spatial features are hard to utilize and difficult to identify ground objects with acceptable accuracy.

Therefore, reconstructing the HSIs into HR images with economical super-resolution (SR) algorithms is a valuable and challenging task. As illustrated in Fig. 1, the general idea and structure of the multiframe slicing and degradation model are to further improve the low spatial resolution of HSIs and increase the practicality of current methods.

Hyperspectral image super-resolution (HSISR) technology is a booming research topic that aims at improving the spatial resolution of the input HSIs through signal processing technologies without further improving the hardware of the imaging system.

It can solve the contradiction between the decrease of spectral resolution, increase of hardware quantity and weight caused by the improvement of spatial resolution, and provide a better foundation for follow-up applications.

Although the HSISR is of important research and application value, the reconstruction is a theoretically ill-posed inverse problem, which suffers from great accuracy loss due to the absence of prior image degradation information and the limited blind reconstruction model [14], [15].

To better solve the HSISR ill-posed inverse problem, there are two main types of solution: Make the best use of the limited available information, which is called the single hyperspectral image super-resolution (SHSR) method

$$\theta^* = \arg\min E\left(f_\theta(\mathrm{HSI}_0), x_{\mathrm{ori}}\right). \quad (1)$$

The current SHSR algorithm is gradually replacing the traditional interpolation reconstruction methods, developed into dictionary learning [16], sparse representation, and low-rank approximation methods [17], and stepping forward on more advanced deep neural network (DNN) learning methods. The typical deep learning HSISR algorithms include 2-D/3-DCNN SR learning algorithms [18], [19], [20], transfer natural image SR learning algorithms [21], generative adversarial network SR learning algorithms [22], spatial spectrum joint prior learning algorithms [23], and other frontier algorithms [24], [25], [26].

In general, these deep learning methods show better performance in contrast to the traditional methods. However, there are limited remote sensing HSI samples that are better than 2.0 m spatial-resolution, and even fewer samples focus on the same area when photographed at a different time and even harder photographed by more different HSI devices. Meanwhile, the high-dimensional bands increase the difficulty in fully exploiting the spatial domain features and emulating the relationship in the spectral domain of the HSI data.

Make the most use of relevant and complementary information, which is called the fusion hyperspectral image super resolution (FHSR) method

$$\theta^* = \arg\min E\left(f_\theta(\mathrm{HSI}_0, \mathrm{MSI}_0, \mathrm{RGB}_0, ...), x_{\mathrm{ori}}\right). \quad (2)$$

As opposed to SHSR, the FHSR methods have developed rapidly in recent years and has made substantial SR performance improvements. Since the MS imaging systems (e.g., RGB cameras, space-borne or air-borne MS sensors) can provide supplementary information [27], the finer spatial details can be preserved at the expense of reducing spectral resolution. These fusion methods based on Bayesian inference [28], matrix decomposition [29], sparse representation [30], manifold structure [31], or the advanced deep learning fusion techniques [32], [33], [34] have gained more attention and obtained extremely better SR result than SHSR.

However, it is very hard or even impossible to simultaneously obtain such a well-co-registered auxiliary image in practice. Therefore, most of the research on those fusion methods can only be tested on limited datasets, which cannot be widely applied in actual airborne/spaceborne data.

In general natural image SR algorithms, the multiframe super-resolution (MFSR) can reconstruct the original HR image by using multiple LR images instead of a single image super-resolution (SISR). Therefore, MFSR can learn more image information or at least the same amount of image information by learning multiple down-sampling LR images, which leads to reconstruction quality improvement.

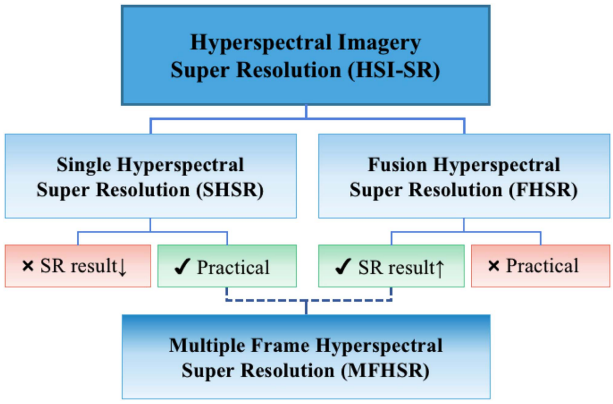


Fig. 2. HSISR methods category and their main pros and cons, MFHSR relationship with current methods.

The multiframe learning central idea is helpful to improve the HSISR. As illustrated in Fig. 2, multiframe learning for hyperspectral SR is a novel type of method besides the SHSR and the FHSR. The advantage is obvious, since it can further improve the performance of SHSR without obtaining additional RGB/MS information, so it can be more practical in HSI applications.

However, the difficulty is also prominent. Most of the existing MFSR methods require two stages to perform SR reconstruction: 1) Fusion or share multiple LR image information and 2) Depend on single-step LR image upsampling module to perform end-to-end learning.

The first stage will increase the computation parameters by at least N (multiple sample number) times and make the HSISR even harder to converge. The second stage often requires prior degradation knowledge such as blur kernel, and noise level, otherwise, this step will produce edge blur or edge jag caused by a single-step up-sampling module.

In addition, the direct application of channel attention and spatial attention, which proved to be useful in SISR and SHSR cannot achieve excellent results in the MFSR task, thus limiting the further improvement of the representational ability of the MFSR network.

Moreover, despite the rapid development of deep learning SR algorithms, there are still few methods that could realize joint reconstructing of MFSR in a unified end-to-end network and even fewer for MFHSR.

To complement the above limitations, a novel strategy and an improved model, named multiple frame splicing and degradation model (MFSDM), is proposed in this article. This is a MFHSR model, which uses multiple splicing unions and multiple degradation combinations as the training inputs to reconstruct high-resolution outputs. The detail of MFSDM is illustrated in Fig. 3.

In summary, our main contributions are as follows:

1) We propose a novel strategy of utilizing multiple splicing unions as same-size LR inputs with HR outputs. Based on this strategy, multiframe HSIs information is naturally shared in one input, simplifying the fusion process.

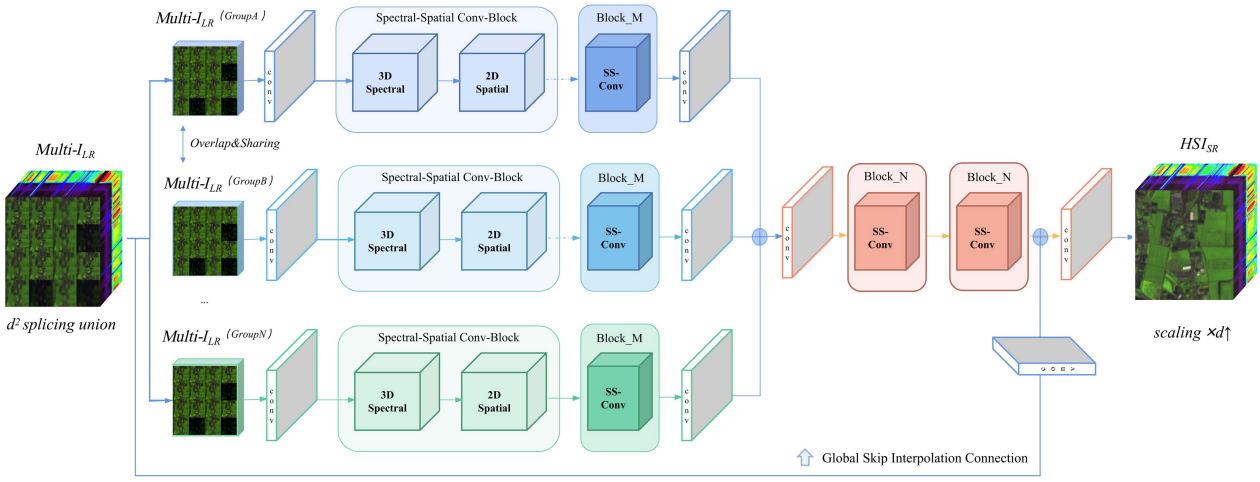


Fig. 3. Overall architecture of MFSDM. The multi-LR splicing unions are learned by N overlapping branch frame with M spectral-spatial convolution block (SSCB) and concatenate at trunk frame with global skip interpolation connection.

2) Based on the multiple splicing union, we propose multiple degradation augmentation unions to better fit blind degradation models. Based on this strategy, the degradation situation is enriched and the particular degradation models of HSIs are combined to better fit real applications.
3) Based on the multiple splicing and degradation strategies, we proposed a more concise end-to-end framework that eliminates the attention block, and improve the upsampling process, which proves good convergence performance.
4) Compared with the latest SHSR and FHSR methods, the proposed MFSDM method achieves better results than the state-of-the-art SHSR method and achieves a similar result to state-of-the-art FHSR with less amount of network parameters.

II. Related Work

The most relevant methods are further viewed in this section, which include the most representative SHSR methods and some preeminent FHSR methods. In addition, the typical framework of the latest single/multiple general nature image SRs are reviewed to further explain the basis of the proposed approach and the focused question to solve.

A. Single Hyperspectral Image Super-Resolution

With the development and superior performance of deep learning technology, the SHSR algorithm is gradually developing from the traditional interpolation or representation method to the deep learning based SR method.

Traditional SR methods include various basic interpolation, such as nearest-neighbor interpolation, bilinear interpolation, bicubic linear interpolation, and local adaptive nonlinear interpolation, etc. Those interpolation methods are simple in computation and still useful as SR baseline. Even though in the DNN learning process, Most SHSR methods will down-sample the HR hyperspectral images applying the bicubic interpolation to obtain the corresponding LR hyperspectral images or use bicubic up-scaled hyperspectral images for residual learning.

However, only relying on the surrounding pixels and a single interpolation core will cause the aliasing and edge blur. Therefore, some frequency domain approximation methods are given, such as convex set projection method [35] and Bayesian analysis method [36], etc; other multiresolution analysis methods, such as wavelet transform, pyramid algorithm [37], etc; and some subpixel location methods, such as pixel replication method [38], etc. However, those hand-crafted priors can only reflect the portion spatial features and finite spectral characteristics of the hyperspectral data, the complex details recovery is hard to realize based on those methods.

Besides those interpolation and pixel combination methods, the dictionary learning-based approaches are widely proposed [39], [40], which also suffer several defects. First, complex optimization in the testing phase is time-consuming or even not soluble. Second, the image priors are always based on the internal limited example, and then the external information from external samples is invalid.

To combine the advantage of interpolation and representation methods, DNN, especially convolutional neural networks (CNN) are widely used in SR problems and have been introduced into the SHSR task. Recently, the graph convolutional networks (GCN) show prior learning ability for hyperspectral data structure and is possible to further improve the interpretation and learning ability [41].

According to the characteristics of hyperspectral data, the framework should concern both the spatial and spectral information. The current CNN SHSR methods focus on the balance between 2-DCNN and 3-DCNN structures and explore more reasonable spatial-spectrum joint network structures.

For example, Mei's 3-D full convolution neural network (3-D-FCNN) [8] improved the SR effect and preserved spectral information. Compared with the 3-D convolution, 2-D convolution can not retain the spectral feature information well, which is a serious shortage for the subsequent hyperspectral processing task.

However, compared with the 2-D convolution, the parameters are increased by an order of magnitude, resulting in a large amount of calculation and complex training. In addition,

according to the DHPSR method designed by Oleksii [42], [43], 3-D convolution is not always better than 2-D convolution, which is because of insufficient hyperspectral dataset and high spatial dimension.

Therefore, the spatial or spatial-spectral information extraction block will greatly influence the SR result since the single LR information is limited. The 2-DCNN can extract the spatial features with less computation and 3-DCNN is more suitable for the joint acquisition of spatial-spectral features. Thus, Li proposed a convolution method MCnet combining the 2-DCNN and 3-DCNN [19], which can solve the problem of large parameters partially.

The above methods focus more on improving the spatial block of hyperspectral data, but there is still a large space for development in the retention of spectral correlation characteristics. Jiang [23] introduced a spatial-spectral prior network (SSPSR) to fully utilize the correlation relationship between spatial information and spectral characteristic.

SSPSR proposed a group convolution strategy with shared network parameters and a progressive up-sampling structure, which can reduce the difficulty of high-dimensional feature extraction of HSI data and better stabilize the training. Also, the spatial-spectral block (SSB) is set to integrate spatial and spectral information.

This work achieve state-of-the-art performance, which has better model structure, superior SR quality indices, and visualization result. However, the improvement is still limited compared with FHSR methods, and the module is more computationally intensive.

Above all, the SHSR have achieved much progress, but the further improvement is limited. On the one hand, it is because the single LR samples have limited information to learn, on the other hand, the limited type of degradation is incapable to fully explore more blind SR circumstance.

B. Fusion Hyperspectral Image Super-Resolution

FHSR is a challenging but effective SR strategy. To exploit the redundancy in the spectral domain and correlation in the spatial domain, some algorithms have been proposed by utilizing the sparsity [30], self-similarity [44], low-rank prior [45], clustering manifold structure [46], and tensor constraints [47].

Last, deep learning based methods become more prevalent in FHSR, since they can achieve superior performance in SR consequence and suppose fewer assumptions about the image degradation and knowledge prior.

Since Xie et al. [32] proposed the MHFnet unfolding network, first proved the feasibility of the fusion model, more approaches start to focus on the deep learning-based FHSR. For example, Wang [48] put forward a DBIN model and optimized the estimation of the fusion problem. Recently, Dong [49] proposed an MoG-DCN method based on the deep hyperspectral denoiser. Basically, the above approaches focus on casting the fusion optimization problem of MS/HSI into joint learning of a deep denoiser prior and a particular observation model.

However, these learning-based fusion methods require either large data for supervision training or the knowledge of degradation prior, which are both unrealistic in a real HSISR scenario.

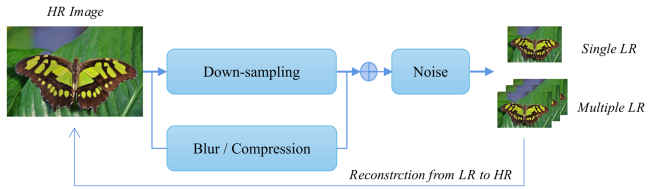


Fig. 4. General process of downsampling and upsampling in SR of SISR and MFSR.

Thus, some unsupervised blind FHSR methods were proposed. For example, Zheng et al. [50] proposed a two-stage network based on unsupervised adaption learning, while estimating the unknown spatial degradation. And lately, Ma et al. [51] optimized the HSISR model by applying the proximal gradient algorithm and introduced a transformer prior to unfolding network for iterative solution simulation.

The above methods partly solve the fusion problem between MS and hyperspectral. Also, providing some unsupervised strategies to solve the blind HSISR problems. However, the fusion-based method relies on supplementary information, which is not fully accessible in most cases. This shortcoming limits further practical applications.

C. Multiframe Image Super-Resolution

The general process of the single/multiple deep learning image SR is shown in Fig. 4.

First, a HR image is down-sampled, and then the noise is added to the simulated real degraded image. Next, the end-to-end mapping relationship between the single LR image or multiple LR images and the HR is trained to obtain the SR reconstruction model [52].

Especially, multiframe image super-resolution (MFSR) aims at restoring the original HR image f_H by extracting and merging the information from multiple LR images f_n.

Mathematically, the original HR image f_H can be reconstructed by minimizing the error between f_n and estimated $\hat{f}_n = \mathbf{H}(\mathbf{S}_n(\hat{f}_H))$ with a regularization. The general estimate $\hat{f}_H$ can be obtained as following:

$$\hat{f}_H = \arg\min_{f_H} \frac{1}{2\sigma^2 N} \sum_{f=1}^{N} ||f_n - \mathbf{H}(\mathbf{S_n}(f_H))||_2^2 + \lambda \mathcal{R}(f_H) \tag{3}$$

where $\mathcal{R}(f_H)$ represents the regularization term related to the original HR image f_H and λ is the penalty parameter.

Although there are more methods proposed for MFSR in nature image processing [53], as of now, there is barely no network focus on multiple LR samples learning for HSIs, and the reconstruction is concentrated upon a limited degradation model only.

III. Proposed MFSDM Framework

A. Problem Formulation

In the MFSDM, we regard the inverse task of single HSI restoration, such as inpainting, noise removal, and SR, can be formulated as an energy minimization problem as follows, which

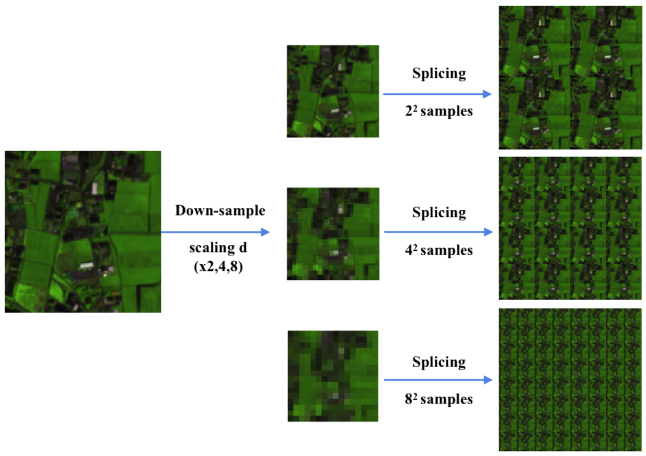


Fig. 5. General process of splicing strategy of scaling parameter x2,x4,x8. The spliced unions are in the same size of the HR image.

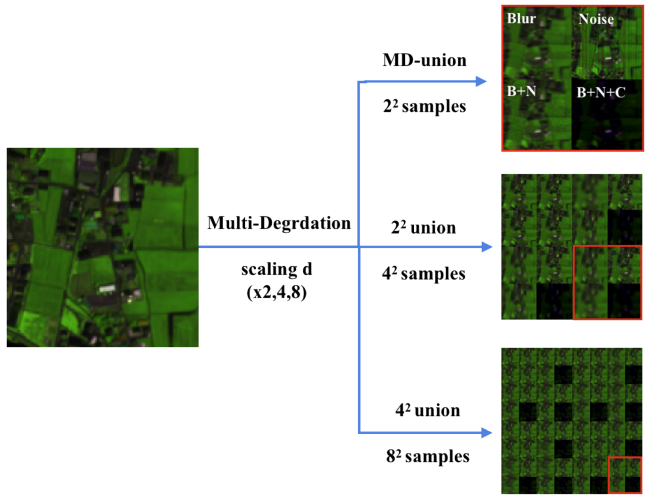


Fig. 6. General process of multiple degradation strategy of scaling parameter x2,x4,x8.

concerns more down-sampling models:

$$\mathrm{p}^* = \min \mathrm{E}(\mathrm{x}(\text{Multi-Degrade(N,R,B,C,...)}), \mathrm{x0}) + \mathrm{R}(\mathrm{x}). \quad (4)$$

In the reality, the multiple degradation model can be very complex, such as multinoise, multireshape, multiblur kernel, and multicompression loss. We use multiple combinations of image and HSI particular degradation in our method.

Based on the single HSI restoration, the multiple-frame slicing can be formulated as an energy minimization problem as follows, which concerns combination of multi-LR samples

$$\mathrm{p}^* = \min \mathrm{E}(\mathrm{x}(\text{Multi-Splice}(S^1_{d_1}, ..., S^n_{d_m})), \mathrm{x0}) + \mathrm{R}(\mathrm{x}). \quad (5)$$

In the reality, multiframe HSI image is hard to acquire, since the available HSI video and same location dataset photoed at a different time is scarce. The satellite-borne HSI can provide such information, however, the original spatial resolution is limited, usually lower than 15 m. Airborne or UAV-borne HSI continuous data can be more helpful in the research of HSISR. However, the available open dataset can not fully meet the multiframe data source requirement.

Since the multiframe SR is useful to improve SR quality, we get multiframe HSI images using the same image copies with the same or different degradation model to simulate multiframe HSI data.

The details of the multiple slicing strategy and multiple degradation strategy are elaborated as follows.

B. Multiple Slicing

The slicing strategy is a simple way to combine the multiframe information and eliminate the influence of using single step upsampling module. The multiframe slicing strategy is illustrated as Fig. 5, and the x4, x8 scaling results are compared in the later experiments.

1) Multiple Slicing Motivation: The process of up-sampling and the up-sampling module design is very critical in defining the performance of the SR method, since it is an ill-posed problem. Based on the upsampling methods, there are four primary frameworks for SR with different learning strategies, and various network types: preupsampling, postupsampling, iterative up-and-down sampling, and progressive upsampling [52].

All the implementation of the main SR process in the network framework is based on the single or progressive up-sampling step, which makes the LR image size reach the original HR image size. However, the current up-sampling step is only based on the results of the previous layer, which causes the loss of the learning features, and the up-sampling step is generally single-step interpolation, which also limits the performance of existing SR algorithms. The up-sampling module is improved in recent works but still can not eliminate all of the ill effects brought by the up-sampling module.

2) Multiple Slicing Strategy: In our method, the slicing strategy is to first down-sampling the HR hyperspectral by bicubic interpolation or later multiple degradation method, then combining the d^2 low-resolution samples together as a splicing training union.

The advantages of multiple slicing are it combines the up-sampling module for each LR sample, and increase the information richness of training samples, which is more useful when the scaling factor is larger. In the latter experiment, we compare the bicubic slicing union SR result with the SHSR result, the converge rate and SR result are improved significantly.

C. Multiple Degradation

The multiple degradation consider different types of degradation effects including down-sampling, blur, noise, and compression [15]. Except for the typical blur types (generalized Gaussian blur), down-sampling types (bilinear, and bicubic), random noises types (Poisson noise and camera sensor noise), the particular degradation types of HSI are also concerned. In our method, we consider the band noise and PCA compression degradation, which heavily influence the spatial information and spectral information.

1) Multiple Degradation Strategy: Specifically, in the slicing of d^2 patches, each 2*2 splicing union uses a different type of degradation strategy. As illustrated in Fig. 6, the upper left uses

the multiple blur model, the upper right uses the multiple noise model, the bottom left uses the multiple blur and noise model, and the bottom right uses multiple combinations of the blur, noise, and compression model.

2) Multiple Degradation Parameters: In our method, the different parameters are randomly chosen in the following range with each union and different samples. The degradation blur model includes isotropic Gaussian blur [0.1, 1.0], anisotropic Gaussian blur [0.1, 1.0], and 3-D Gaussian blur [0.1, 1.0]. The parameters are randomly chosen in that parameter range, and other degradations are similar.

The degradation noise model includes Gaussian additive noise $[0.01\%, 0.1\%]$, Gaussian multiplicative noise $[0.01\%, 0.1\%]$, and striping noise $[1\%, 10\%]$. The striping noise here is added by row following the energy proportion of each row in the Gaussian distribution. The band or striping noise is a normal type of spatial noise in HSI, since the imaging process of a large area will leave threadiness and bad pixel noise in the final imaging result. Therefore, this degradation should also be concerned in the HSISR.

The degradation blur and noise model includes a random combination of the above blur and noise type. The other parameters are the same with striping noise decrease in the range of $[1\%, 5\%]$. Since the higher noise level will make spatial information unavailable.

The degradation mixing model includes a random combination of the above blur, noise, and compression types. Besides the spatial degradation, the compression of HSI influence both the spatial and spectral information. The normal type of compression is using principal component analysis (PCA) to select the main spectral band information. The reconstruction result by the principal part of the original hyperspectral will reduce the data amount and simulate the real spectral shifting in our degradation model. For example, although the variance of the top 15 principal bands of the Chikusei dataset is usually larger than 99%, the reconstructed HSI spectral information is often shifted or inaccurate. The spatial information is also influenced by wrong color space mapping and high-frequency details loss. In the mixing model, the other parameter is the same as the principal number in the range of [15, 30]. The less principal selection will harm the effectiveness of reconstruction. For different datasets, the principal number selection varies and the variability comparison is shown in the experiment part.

D. MFSDM Learning Model

1) MFSDM Network Architecture: Fig. 3 shows the framework of the proposed MFSDM, which is improved from the SSPSR idea [23]. The method mainly includes three parts: shallow feature extraction combined with upscale, spatial-spectral deep feature extraction branch, and reconstruction trunk. The structure mainly omitting the single upscale module and condensing the SSB, which reduces the spectral attention module but uses a 3-D convolution block as a spectral feature module.

The input low-resolution HSI image is the splicing union $I_{\mathrm{L\,R\ union}} \in \mathbb{R}^{H\times W\times C}$, which is d^2 union of degradation $I_{\mathrm{L\,R}} \in \mathbb{R}^{h\times w\times C}$. The h size of the degradation sample is H size divided by 2^n scaling factor. And the corresponding high-resolution output HSI image is $I_{S\mathrm{R}} \in \mathbb{R}^{H\times W\times C}$. The original high-resolution HSI image $I_{\mathrm{H\,R}} \in \mathbb{R}^{H\times W\times C}$ is the ground truth of the input low-resolution union.

The framework of MFSDM is to better utilize the information from the multiple sample unions of $I_{\mathrm{L\,R}}$ and better minimize the error between $I_{S\mathrm{R}}$ and $I_{\mathrm{H\,R}}$.

$$I_{S\mathrm{R}} = H_{\mathrm{MFSDM}}\left(I_{\mathrm{L\,R\ union}}\right). \tag{6}$$

Similar to SSPSR, we keep the branch networks and the global trunk network, since the coarse-to-fine manner is useful and the branch networks share the parameters in the learning process, which reduce the learning parameters greatly.

2) The Branch and Trunk Network: We first divide the $I_{\mathrm{L\,R}}$ into N groups to better exploit the correlations among neighboring spectral bands, also reduce the feature dimensions of of each group. $I_{\mathrm{L\,R}} = \{I_{\mathrm{L\,R}}^{(1)}, I_{\mathrm{L\,R}}^{(1)}, \cdots, I_{\mathrm{L\,R}}^{(N)}\}$. For each group $I_{\mathrm{L\,R}}^{(N)}$, one convolutional layer was directly applied to obtain its shallow features and upscale each samples

$$F_0^{(N)} = H_{\mathrm{F\,E_{conv-up}}}\left(I_{\mathrm{L\,R}}^{(N)}\right). \tag{7}$$

$I_{\mathrm{L\,R}}^{(N)}$ then used for extracting the deep features with the improved spectral-spatial convolution block (SSCB).

$$F_{\mathrm{SSCB}}^{(N)} = H_{\mathrm{SSCB}}\left(F_0^{(N)}\right). \tag{8}$$

The SSCB contains 3-D-spectral block and 2-D-spatial block. After the M series of SSCB, one convolutional layer is added after each branch. The purpose is to concatenate feature channels and reconstruct them to the same band channels of the input group HSI

$$F_{\mathrm{rec}}^{(N)} = H_{\mathrm{rec}}\left(F_{\mathrm{SSCB}}^{(N)}\right). \tag{9}$$

The branch and trunk frame are almost the same, the features from all branches are concatenated together, and the average feature values with overlapped bands are computed according to their original spectral band positions. After the trunk frame, the SR HSI is obtained by feeding the up-scaled random low-resolution samples as input HSIs with the learned concatenate features

$$I_{\mathrm{S\,R}} = H_{\mathrm{GlobalREC}}\left(F_{\mathrm{Trunk}} + H_{\mathrm{GlobalFE}}\left(I_{\mathrm{L\,R}}\uparrow\right)\right). \tag{10}$$

3) The Spatial-Spectral Conv Block: Similar to SSPSR, the SSCB exploits the spatial-spectral information from the hyperspectral data. Each SSCB contains two modules, a spectral 3-DCNN module, and a spatial 2-DCNN module. The architecture of SSCB is given in Fig. 7. For the spectral module, we leverage the 3-D convolutions block with $3 \times 3 \times 3$ convolutions to extract the joint features, and for the spatial module, the 3×3 2-D convolutions to extract the spatial features instead of 1×1 convolutions. The channel attention mechanism used in the SSPSR is also reduced since the channel-wise feature calculation is complex and time-consuming. The corresponding SSCB parameters comparison test is elaborate in the experiment part.

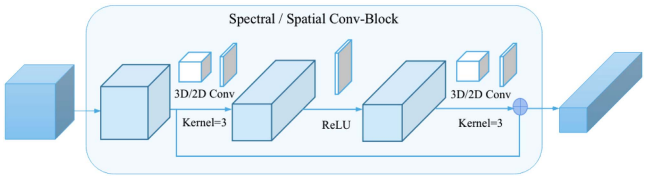


Fig. 7. General spectral spatial convolution block which using 2-D/3-DCNN and skip connection.

4) Loss Function: The loss function selected in MFSDM also adapts the L_1 loss and the spatial-spectral total variation (SSTV) loss weighted sum as the final reconstruction loss.

The L_1 loss is as follows:

$$\mathcal{L}_1(\Theta) = \frac{1}{N}\sum_{n=1}^{N} \left\| I_{\mathrm{H\,R}}^n - H_{\mathrm{MFSDM}}\left(I_{\mathrm{L\,R}}^n\right)\right\|_1. \tag{11}$$

The L_{SSTV} loss is as follows:

$$\mathcal{L}_{\mathrm{SSTV}}(\Theta) = \frac{1}{N}\sum_{n=1}^{N} \left(\|\nabla_h I_{\mathrm{S\,R}}^n\|_1 + \|\nabla_w I_{\mathrm{S\,R}}^n\|_1 + \|\nabla_c I_{\mathrm{S\,R}}^n\|_1\right) \tag{12}$$

where ∇_h, ∇_w, ∇_c are functions to measure the gradient of horizontal, vertical, and spectral, respectively.

The final reconstruction loss is as follows:

$$\mathcal{L}_{\mathrm{total}}(\Theta) = \mathcal{L}_1 + \alpha\mathcal{L}_{\mathrm{S\,S\,T\,V}}. \tag{13}$$

In summary, the MFSDM is a more simplified model than the state-of-the-art SHSR model and preserves the most advantages. The structure is fit for dealing with multiple HSI SR tasks.

IV. Experiments

The MFSDM network is compared with the latest SHSR methods on two remote sensing datasets: Chikusei [54] and Pavia Center [55]. Based on the Chikusei dataset, the ablation studies are performed to test each component's effectiveness and explore the proper parameter settings. It confirms the superiority of the MFS and MFSD strategy with different scales and further validated it through comparison with the FHSR method on the CAVE [56] dataset.

The proposed method is compared with the most representative SHSR methods on Chikusei and Pavia Center dataset, including three state-of-the-art deep single gray/RGB image SR methods, EDSR [57], RCAN [58], and SAN [59], and five most relevant and representative deep SHSR methods, TLCNN [21], 3DCNN [18], GDRRN [60], DeepPrior [43], and SSPSR [23]. The Bicubic interpolation is undertaken as a baseline method.

The proposed method also compared with most representative FHSR methods, LTTR [61], uSDN [62], MHFnet [32], CU-CaNet [33], MoG-DCN [49], 3DT-Net [51] on CAVE dataset. The SR results and the PQIs are shown as follows.

A. Datasets and Training Details

1) *Chikusei Dataset*: An urban area in Chikusei, Ibaraki, Japan, taken on July 29, 2014 from 363 nm to 1018 nm by Headwall imaging sensor. It contains 128 spectral bands in the spectral domain and 2517 × 2335 pixels in spatial domain. The ground sampling distance was 2.5 m.
2) *Pavia Center dataset*: A central area of Pavia, northern Italy, taken in a flight campaign in 2001 by ROSIS sensor. It contains 102 available spectral bands leaving 1096 × 715 pixels meaningful region in the spatial domain.
3) *CAVE dataset*: Widely used in many MS and HSI fusion-based restoration tasks. This dataset covers 32 scenes of real and fake daily objects with 512 × 512 spatial pixels, including 31 spectral bands from 400 nm to 700 nm with the RGB images in corresponding scenes.

Training Details. For different dataset, the HSI preprocessing and learning processes are similar. Take the procedure of the Chikusei dataset as an example, the image center is first cropped into a subimage with 2048 × 2048 × 128 pixels, the upper left 25% region is further divided into four none overlapping 512 × 512 pixels to form the testing data.

Then extracted 1875 overlapping patches from the remaining 75% region to form HR HSI training reference, and each training sample is 128 × 128 pixels. Random 10% of the training data are excluded forming the validation set. The LR HSI are downsampled by bicubic with scaling factor 4 and 8. Further, both bicubic and multiple down-sampling samples are generated to evaluate the splicing validation and the effectiveness of blind recovery of the MFSDM model.

Evaluation Measures. The following typical quantitative PQIs are calculated for MFSDM performance evaluation. Except for the most used peak signal-to-noise ratio (PSNR) [63], and structure similarity (SSIM) [44] index, the spectral angle mapper (SAM) [64], cross correlation (CC) [65] show the spectral recovery quality, and the root mean squared error (RMSE), erreur relative globale adimensionnelle de synthese (ERGAS) [44] evaluate the global reconstruction quality. The following tables show comparison results of those indicators between different methods.

Implementation Details. We use MATLAB to generate training and testing samples and rely on Pytorch libraries to implement the proposed network. We train the corresponding MFSDM multislicing (MFSDM_MS) and MFSDM multidegradation (MFSDM_MD) model for scaling factors 4 and 8 with random initialization and test each strategy on different dataset. In the training process, we adopt the ADAM optimizer [66] with 1e-4 initial learning rate. The learning rate decays by a factor of 10 after each 10 epochs and the batch size is 32. The model takes about 40 epochs to achieve stable performance. We use four parallel Nvidia GA100 GPU to train the model and the average training time is 280 min for 40 epochs to achieve the best SR results.

B. Ablation Studies and Quantitative Evaluation

In order to better explain the role of each structural module, and the computational complexity, convergence, and reasoning time between the comparison methods, we conducted quantitative comparative analysis of representative algorithms, and finally selected the compromise optimal module design scheme.

TABLE I
TYPICAL GROUPING PARAMETERS AND SHARING STRATEGY PARAMETERS COMPARISON OF THE PROPOSED MFSDM METHOD

B	O	G	p×10^6	F×10^9	PSNR	SSIM	SAM	ERGAS
128	0	1	12.28	15.12	41.8915	0.9622	2.1612	4.6821
1	0	128	11.61	189.28	41.9012	0.9681	2.0927	4.5910
8	0	16	11.72	29.36	42.9209	0.9792	1.9332	3.9232
16	0	8	11.89	37.12	43.2372	0.9823	1.9028	3.8623
8	2	21	11.72	32.17	43.7321	0.9834	1.7828	3.1922
8	**4**	**31**	**11.72**	**38.48**	**43.7843**	**0.9852**	**1.7579**	**2.8956**
8	6	61	11.72	89.37	43.7913	0.9889	1.7549	2.8729

TABLE II
TYPICAL SPECTRAL SPATIAL CONVOLUTION BLOCK PARAMETERS AND STRUCTURE COMPARISON OF THE PROPOSED MFSDM METHOD

Models	Num	p×10^6	F×10^9	PSNR	SSIM	SAM	ERGAS
SSCB_2D	2	6.38	15.09	37.7815	0.9392	5.0642	5.9811
SSCB_2D	3	8.41	30.38	40.9483	0.9691	2.0947	4.6387
SSCB_2D	4	10.52	45.36	43.0219	0.9803	1.8732	3.6232
SSCB_3D	2	9.27	18.26	38.2382	0.9415	2.3028	4.9028
SSCB_3D	**3**	**11.72**	**38.48**	**43.7321**	**0.9834**	**1.7828**	**3.1922**
SSCB_3D	4	13.22	59.27	43.8843	0.9874	1.7769	3.0926

Grouping Parameters and Sharing Strategy. The grouping parameters include spectral band number (p) and the overlap band number (o) between the neighboring groups, based on which the groups are determined

$$\text{Groups} = \text{ceil}\left(\frac{\text{Bands} - o}{p - o}\right). \tag{14}$$

The parameter sharing strategy refers to sharing the training parameters of each branch network, which can greatly reduce the training parameters and complexity, accelerating the convergence of the model.

In the SSPSR, these two strategies proved to be useful. Here, we also compare the bands and overlaps with corresponding parameters and flops. We compare the PQIs on chikusei test images of scaling factor 4, the result is shown in Table I. The results show that the group sharing strategy is useful especially when there are more overlap bands and more bands in a single group, however, the computation flops also increases a lot. For better performance and less computation complexity, eight bands for each group and four overlaps between each group for the chikusei dataset are selected as the appropriate parameter selection. For Pavia Center dataset, 16 bands for each group and 4 overlaps are choose. For CAVE dataset, 10 bands for each group and 3 overlaps are choose.

SSCB Parameters and Structure Comparison. Further, the parameter M blocks of SSCB and the validity of choosing 3-DCNN in each block have been experimented. Here, we compare the usage of 2-DCNN and 3-DCNN as the conv-block in SSCB and the number of SSCB blocks. The result is shown in the Table II with corresponding computation complexity. The results show that the 3-DCNN achieves better reconstruction results although the 2-DCNN achieves equivalent effect by increasing the number of blocks. Here, we choose the relatively better structure of 3-DCNN as spectral-spatial extraction blocks with the number of blocks as three. It seems that more number of blocks setting may achieve better results but the computation complexity will also increase at exponential growth. This parameter setting is adopted on the other two datasets to achieve a balance between the complexity and the effect.

TABLE III
TYPICAL COMPUTATION COMPLEXITY AND CONVERGE PARAMETERS COMPARISON

Models	P(10^6)	F(10^9)	C_T(h)	R_T(ms)	PSNR	SSIM
RCAN	15.37	50.56	8.5	402.35	39.9041	0.9359
3-DCNN	8.81	25.28	3.3	204.85	38.6091	0.9127
MCNet	12.53	29.78	4.0	380.35	38.7812	0.9103
SSPSR	13.56	43.51	7.5	320.50	40.3613	0.9412
MFSDM_MS	11.72	38.48	4.6	342.40	43.7843	0.9852
MFSDM_MD	11.72	38.48	5.2	358.20	42.4274	0.9734

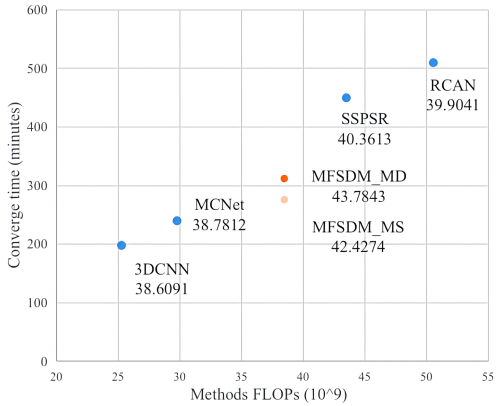


Fig. 8. FLOPs versus Converge time on Chikusei dataset in ×4 scale factor.

Computation Complexity and Converge Comparison. In order to better evaluate the performance of the algorithm, we record the number of parameters, the FLOPs, the converge training time, and the running time among representative SHSR methods to further compare the computation complexity with other methods. The result is showed in Table III and the corresponding visualization Fig. 8 can show the compare results more intuitive. Compared with the current optimal method, it has advantages in parameter quantity and SR effect. Through comparison, it is proved that the MFSDM method is easy to converge when the parameters at moderate level, and the quality of SR is improved at the same time. The inference running time of the algorithm is also within a reasonable range.

Spectral and Various Variability Comparison. Considering the spectral variability, the remote sensing data usually tend to suffer from various degradation, noise effects, or variability in the process of imaging [67]. The MFSDM mainly consider this problem as the PCA selection and reconstruction process. It is a general process of HSI compression and the principle component choose can cause various variability in the spectral domain. Since the blind SR can not fully acquire the spectral degradation reasons, we compare the variance caused by principal component selection. The result is shown in the Table IX and the spectral various after the reconstruction is shown in Fig. 13. The experiments show that the spectral changes caused by the principal component compression are one of the factors that have a great impact on the hyperspectral degradation and reconstruction, and further analysis is required for the characteristics of compressed information loss.

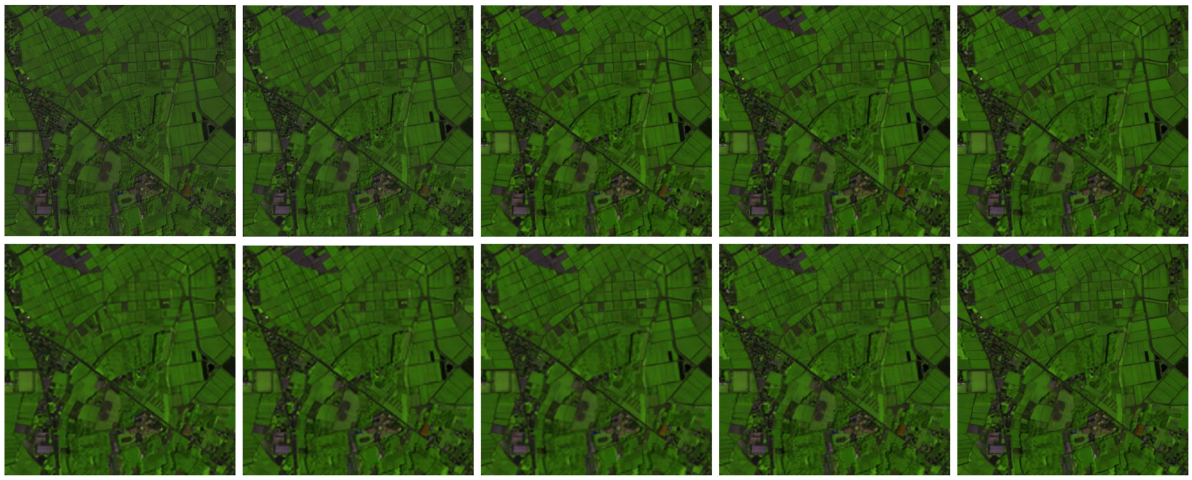


Fig. 9. Reconstructed images ($d = 4$) of one test Chikusei HSI with spectral bands [70-100-36]. The images from left to right, top to down, show the ground truth, results of MFSDM, EDSR [57], RCAN [58], SAN [59], TLCNN [21], 3DCNN [18], GDRRN [60], DeepPrior [43], SSPSR [23] method, respectively.

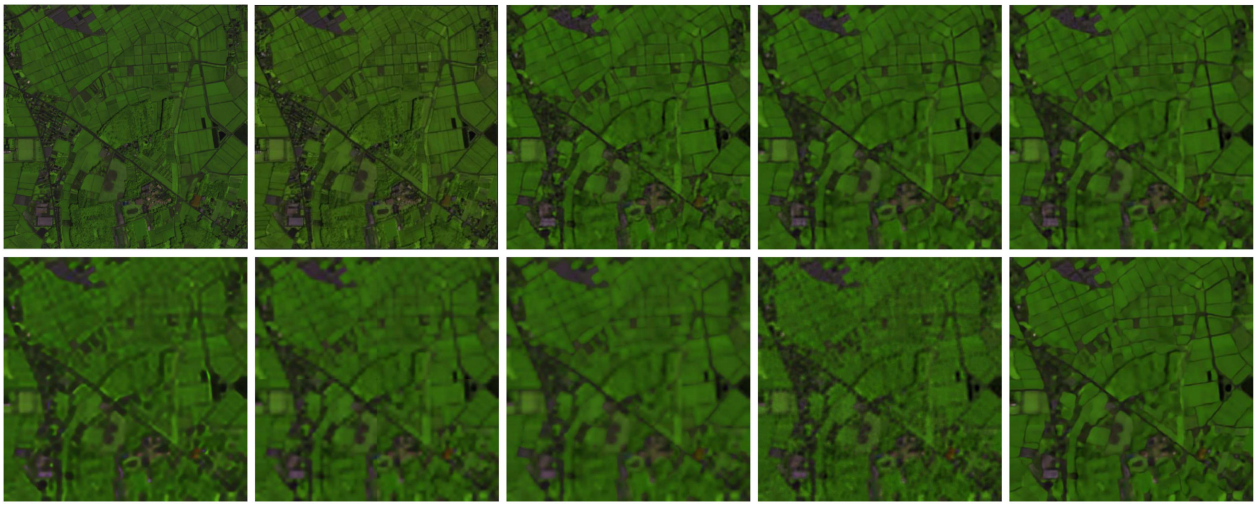


Fig. 10. Reconstructed images ($d = 8$) of one test Chikusei HSI with spectral bands [70-100-36]. The images from left to right, top to down, show the ground truth, results of MFSDM, EDSR [57], RCAN [58], SAN [59], TLCNN [21], 3DCNN [18], GDRRN [60], DeepPrior [43], SSPSR [23] method, respectively.

TABLE IV
SIX PQIS COMPARISONS OF NINE APPROACHES OVER TESTING IMAGES FROM CHIKUSEI DATASET (SCALING 4)

***d*=4**	Bicubic	EDSR	RCAN	SAN	TLCNN	3DCNN	GDRRN	DeepPrior	SSPSR	**MFSDM_MS**	**MFSDM_MD**
CC	0.9212	0.9517	0.9519	0.9514	0.9196	0.9355	0.9369	0.9293	0.9565	**0.9893**	**0.9842**
SAM	3.4040	2.5548	2.5396	2.5547	3.8573	3.1174	2.5021	3.5584	2.3527	**1.7579**	**1.7956**
RMSE	0.0156	0.0122	0.0119	0.0120	0.0150	0.0140	0.0137	0.0147	0.0114	**0.0051**	**0.0082**
ERGAS	6.7564	5.3709	5.3205	5.3349	6.7522	6.0026	5.9540	6.2095	4.9892	**2.8956**	**3.5805**
PSNR	37.6377	39.8389	39.9041	39.8671	37.7251	38.6091	38.7198	38.1923	40.3613	**43.7843**	**42.4274**
SSIM	0.8949	0.9351	0.9359	0.9357	0.9007	0.9127	0.9193	0.9011	0.9412	**0.9852**	**0.9734**

C. Results on Chikusei Dataset

Figs. 9 and 10 display the reconstructed images of MFSDM in Chikusei dataset and the comparing visual result with other methods. It can be observed that the proposed MFSDM method shows better performance compared with other SHSR algorithms, better recovery of both fine grain textures and coarse grain structures in visualization. At Tables IV and V, we report the metrics of the MFSDM reconstructed images with other methods. On most PQIs, our method has considerable advantages. The average PSNR value is 0.6 dB higher than the second best SSPSR method.

The MFSDM of multiple splicing is the single bicubic downsampling union, which achieves the best SR result. Obviously, when the degradation is simple, the recovery will have a better result. The MFSDM of multiple degradations is the combination of degrade union and achieving the second best SR result. Taking the scaling $d = 8$ MFSDM average as an example, the CC index increased 0.16 db, the SAM index decreased 0.68 db, the RMSE

TABLE V
SIX PQIS COMPARISONS OF NINE APPROACHES OVER TESTING IMAGES FROM CHIKUSEI DATASET (SCALING 8)

d=8	Bicubic	EDSR	RCAN	SAN	TLCNN	3DCNN	GDRRN	DeepPrior	SSPSR	MFSDM_MS	MFSDM_MD
CC	0.8314	0.8636	0.8665	0.8664	0.8249	0.8428	0.8421	0.8366	0.8766	**0.9289**	**0.9096**
SAM	5.0436	4.4205	4.3757	4.3922	5.3041	4.8432	4.3160	5.3386	4.0127	**3.0889**	**3.7891**
RMSE	0.0224	0.0201	0.0198	0.0197	0.0224	0.0215	0.0214	0.0219	0.0191	**0.0107**	**0.0169**
ERGAS	4.8488	4.5091	4.5229	4.5170	4.8843	4.5964	4.5879	4.6789	4.3120	**3.1791**	**3.4379**
PSNR	34.5049	35.4217	35.5044	35.5018	34.3488	34.8375	34.8153	34.6692	35.8368	**39.88028**	**37.9238**
SSIM	0.8228	0.8501	0.8531	0.8527	0.8215	0.8313	0.8357	0.8126	0.8624	**0.9139**	**0.9046**

TABLE VI
SIX PQIS COMPARISONS OF NINE APPROACHES OVER TESTING IMAGES FROM PAVIA CENTER DATASET (SCALING 4)

d=4	Bicubic	EDSR	RCAN	SAN	TLCNN	3DCNN	GDRRN	DeepPrior	SSPSR	MFSDM_MS	MFSDM_MD
CC	0.8594	0.8922	0.8917	0.8927	0.8563	0.8813	0.8829	0.8723	0.9003	**0.9484**	**0.9290**
SAM	6.1399	5.8657	5.9785	5.9590	6.9013	5.8669	5.4750	6.2665	5.4612	**4.6318**	**5.0588**
RMSE	0.0437	0.0379	0.0376	0.0374	0.0431	0.0396	0.0393	0.0410	0.0362	**0.0143**	**0.0281**
ERGAS	6.8814	6.0199	6.0485	5.9903	6.9139	6.2665	6.2264	6.4845	5.8014	**3.3126**	**4.6197**
PSNR	27.5874	28.7981	28.8165	28.8554	27.6682	28.4114	28.4726	28.1061	29.1581	**33.2441**	**32.0786**
SSIM	0.6961	0.7722	0.7719	0.7740	0.7141	0.7501	0.7530	0.7365	0.7903	**0.8711**	**0.8574**

TABLE VII
SIX PQIS COMPARISONS OF NINE APPROACHES OVER TESTING IMAGES FROM PAVIA CENTER DATASET (SCALING 8)

d=8	Bicubic	EDSR	RCAN	SAN	TLCNN	3DCNN	GDRRN	DeepPrior	SSPSR	MFSDM_MS	MFSDM_MD
CC	0.6969	0.7215	0.7152	0.7104	0.6880	0.7163	0.7111	0.7007	0.7359	**0.8132**	**0.7981**
SAM	7.8478	7.8594	7.9992	8.0371	8.3843	7.6878	7.3531	7.9281	7.3312	**7.7573**	**7.8923**
RMSE	0.0630	0.05983	0.0604	0.0609	0.0633	0.0605	0.0607	0.0618	0.0586	**0.0328**	**0.0302**
ERGAS	4.8280	4.6359	4.6930	4.7646	4.9143	4.6469	4.6220	4.7366	4.5266	**2.7994**	**3.0173**
PSNR	24.5972	25.0041	24.9183	24.8485	24.5215	24.9336	24.8648	24.7252	25.1985	**30.8733**	**28.2837**
SSIM	0.4725	0.5130	0.5086	0.5054	0.4790	0.5038	0.5014	0.4963	0.5365	**0.7717**	**0.6827**

TABLE VIII
FOUR PQIS COMPARISONS OF SIX APPROACHES OVER TESTING IMAGES FROM CAVE DATASET (SCALING 8)

d=8	LTTR	USDN	MHF-Net	CUCA-Net	MOG-DCN	3DT-Net	MFSDM_MS	MFSDM_MD
SAM	2.97	3.31	3.39	3.32	2.04	2.01	**2.31**	**2.21**
ERGAS	0.667	0.663	0.645	0.631	0.45	0.45	**0.479**	**0.463**
PSNR	45.89	46.01	46.31	46.21	48.89	50.93	**49.57**	**49.37**
SSIM	0.994	0.993	0.994	0.995	0.996	0.996	**0.995**	**0.994**

Fig. 11. Reconstructed images ($d = 8$) of one test Pavia Center HSI with spectral bands [32-21-11]. The images from left to right show the ground truth, results of MFSDM, RCAN [58], SAN [59], 3DCNN [18], GDRRN [60], SSPSR [23] method, respectively.

index decreased 1.41 db, the ERGAS index decreased 1.22 db, the PSNR index increased by 0.45 db, the SSIM index increased by 0.25 db. The average index improvement is 0.695 db, which is 1.173 times.

D. Results on Pavia Center Dataset

As the visualization result shown in Fig. 11 and the super-resolution result in Tables VI and VII, the average performance of MFSDM outperforms on six PQIs of testing images between nine competing approaches. The average index improvement is 0.765 db, which is 1.192 times better. Especially, the MFSDM performance is better compared to the existing methods on the higher scaling $d = 8$, since the sampling union has more slicing union patch together as training samples, the details recover much better and the objective, which is a blur in the other method can be more clear in MFSDM method. To a certain extent, it also proves that multiframe SR is a more effective strategy when the lack of training data.

E. Results on CAVE Dataset

In addition, the HSIs of natural scenes are compared on the CAVE dataset. The average values of SAM, ERGAS, PSNR, and SSIM of most competing methods are shown in Table VIII. Generally, the PSNR and SSIM values of the proposed MFSDM method at $d = 8$ are slightly lower but also very close to other methods. As shown in the visualize reconstruction Fig. 12, even compared with the state-of-the-art FHSR method, by only using the HSI information and using the multiframe slicing and

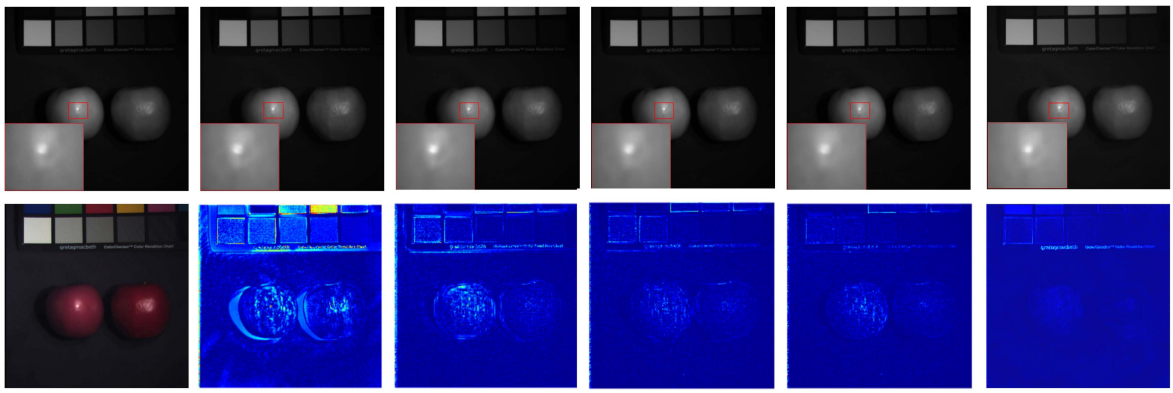


Fig. 12. Reconstructed images and error maps ($d = 8$) of real and fake apples test CAVE HSI (spectral bands 31). From left to right shows the ground truth and real RGB images, results of Hysure [35], MHF-net [32], MoG-DCN [49], 3DT-Net [51], and MFSDM methods.

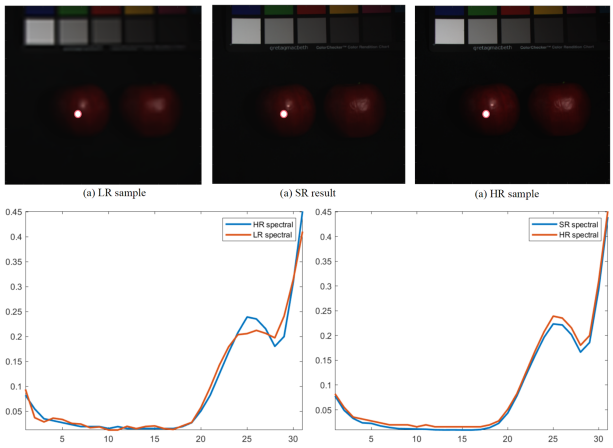


Fig. 13. Comparison of degradation spectra before and after reconstruction. Within the allowable range of variance error, the reconstructed spectral characteristics remained stable.

TABLE IX
INFLUENCE OF SPECTRAL CHANGES CAUSED BY SPECTRAL PRINCIPAL COMPONENTS ON MFSDM

PCA_selection	Variance_error	PSNR	SSIM	SAM	ERGAS
[15,30]	<0.01%	43.7843	0.9852	1.7579	2.8956
[10,15]	0.01%-0.05%	40.2729	0.9138	2.7838	5.9283
[5,10]	>0.05%	38.1023	0.9082	3.7832	6.5392

multidegradation strategy, the single HSI SR MFSDM can also have a competitive or even local part better SR result.

V. CONCLUSION

In this article, a novel multi-frame slicing and multidegradation model (MFSDM) is proposed to address the HSI SR problem. In particular, in order to further improve the SHSR performance and solve the problem of multiframe fusion tasks, we designed a multiple-slicing strategy to overcome the shortage of a single-step up-sampling module. It can better exploit the low-resolution union information and correlation between the union features. In addition, to deal with the problem that the single degradation is unpractical and the HSI real degradation is more intricate, a multiple-degradation union strategy and a corresponding simplified framework is proposed. In this way, we can balance the parameters of the model and fit hostile degradation conditions under limited hyperspectral data. Most evaluations on three public hyperspectral datasets demonstrate that our model not only gains a higher SHSR evaluation index, but also generates more clear visualization SR output. The SR performance is almost the same as the FHSR, which proves that the multislicing and multidegradation strategy is effective. In the future, we will consider making improvements on higher spatial-resolution datasets, further explore better network structure on the basis of multislicing, and further analyze the generalization of its method to other types of 3-D data structures, such as MRI and video structure. Further, maintaining the stability of spectral information in SR reconstruction needs detailed consideration, and subsequent research need to improve the model for this challenging problem.

REFERENCES

[1] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.
[2] B. Rasti et al., "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.
[3] L. Tang, W. Tang, X. Qu, Y. Han, W. Wang, and B. Zhao, "A scale-aware pyramid network for multi-scale object detection in SAR images," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 973.
[4] C. Deng, D. Jing, Y. Han, S. Wang, and H. Wang, "FAR-Net: Fast anchor refining for arbitrary-oriented object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Jan. 2022.
[5] H. Bi, C. Zhang, K. Wang, J. Tong, and F. Zheng, "Rethinking camouflaged object detection: Models and datasets," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5708–5724, Sep. 2022.
[6] D. Hong et al., "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
[7] W. Wang, Y. Han, C. Deng, and Z. Li, "Hyperspectral image classification via deep structure dictionary learning," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2266.
[8] Z. Zhao, Y. Han, T. Xu, X. Li, H. Song, and J. Luo, "A reliable and real-time tracking method with color distribution," *Sensors*, vol. 17, no. 10, 2017, Art. no. 2303.
[9] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.
[10] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *J. Imag.*, vol. 5, no. 5, 2019, Art. no. 52.
[11] Y. Gu et al., "Multimodal hyperspectral remote sensing: An overview and perspective," *Sci. China Inf. Sci.*, vol. 64, no. 2, pp. 1–24, 2021.

[12] Y. Han, C. Deng, B. Zhao, and B. Zhao, "Spatial-temporal context-aware tracking," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 500–504, Mar. 2019.

[13] Y. Han, C. Deng, B. Zhao, and D. Tao, "State-aware anti-drift object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4075–4086, Aug. 2019.

[14] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2388–2400, Jun. 2021.

[15] W. Zhang, G. Shi, Y. Liu, C. Dong, and X.-M. Wu, "A closer look at blind super-resolution: Degradation models, baselines, and performance upper bounds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 527–536.

[16] 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.*, 2014, pp. 3523–3527.

[17] 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, 2017, Art. no. 1286.

[18] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1139.

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

[20] 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.*, vol. 18, no. 7, pp. 1274–1278, Jul. 2021.

[21] 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, May 2017.

[22] X. Dou, C. Li, Q. Shi, and M. Liu, "Super-resolution for hyperspectral remote sensing images based on the 3D attention-srgan network," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1204.

[23] 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, May 2020.

[24] Y. Fu, Z. Liang, and S. You, "Bidirectional 3D quasi-recurrent neural network for hyperspectral image super-resolution," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2674–2688, Feb. 2021.

[25] H. Liu et al., "Video super-resolution based on deep learning: A comprehensive survey," *Artif. Intell. Rev.*, pp. 1–55, 2022.

[26] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8628–8638.

[27] D. Hong et al., "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.

[28] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.

[29] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[30] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2014, pp. 63–78.

[31] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5969–5982, Dec. 2018.

[32] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, and D. Meng, "MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1457–1473, Mar. 2022.

[33] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 208–224.

[34] J.-F. Hu, T.-Z. Huang, and L.-J. Deng, "Fusformer: A transformer-based fusion approach for hyperspectral image super-resolution," 2021, *arXiv:2109.02079*.

[35] 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.

[36] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3631–3640.

[37] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 624–632.

[38] X. Xu et al., "Hyperspectral image super resolution reconstruction with a joint spectral-spatial sub-pixel mapping model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 6129–6132.

[39] 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.

[40] W. Huang, L. Xiao, H. Liu, and Z. Wei, "Hyperspectral imagery super-resolution by compressive sensing inspired dictionary learning and spatial-spectral regularization," *Sensors*, vol. 15, no. 1, pp. 2041–2058, 2015.

[41] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[42] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9446–9454.

[43] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019.

[44] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[45] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2269–2280, Mar. 2021.

[46] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LEMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019.

[47] X.-H. Han, B. Shi, and Y. Zheng, "Self-similarity constrained sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5625–5637, Nov. 2018.

[48] W. Wang, W. Zeng, Y. Huang, X. Ding, and J. Paisley, "Deep blind hyperspectral image fusion," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4150–4159.

[49] W. Dong, C. Zhou, F. Wu, J. Wu, G. Shi, and X. Li, "Model-guided deep hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 5754–5768, May 2021.

[50] K. Zheng et al., "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, Mar. 2021.

[51] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Learning a 3D-CNN and transformer prior for hyperspectral image super-resolution," 2021, *arXiv:2111.13923*.

[52] S. M. A. Bashir, Y. Wang, M. Khan, and Y. Niu, "A comprehensive review of deep learning-based single image super-resolution," *PeerJ Comput. Sci.*, vol. 7, 2021, Art. no. e621.

[53] M. M. Khattab, A. M. Zeki, A. A. Alwan, A. S. Badawy, and L. S. Thota, "Multi-frame super-resolution: A survey," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res.*, 2018, pp. 1–8.

[54] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over chikusei," Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27, vol. 5, 2016.

[55] P. Gamba, "A collection of data for urban area characterization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2004, vol. 1, paper no. 72.

[56] 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.

[57] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–144.

[58] 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.

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

[60] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *Proc. IEEE 4th Int. Conf. Multimedia Big Data*, 2018, pp. 1–4.

[61] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

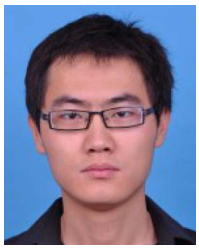
[62] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2511–2520.
[63] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. IEEE 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369.
[64] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. JPL, Summaries 3rd Annu. JPL Airborne Geosci. Workshop Volume 1: AVIRIS Workshop*, 1992, p. 147.
[65] J.-C. Yoo and T. H. Han, "Fast normalized cross-correlation," *Circuits, Syst. Signal Process.*, vol. 28, no. 6, pp. 819–843, 2009.
[66] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[67] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

Chenwei Deng received the Ph.D. degree in signal and information processing from the Beijing Institute of Technology, Beijing, China, in 2009.

Since 2012, he has been an Associate Professor and then a Full Professor with the School of Information and Electronics, Beijing Institute of Technology. Prior to this, he was a Postdoctoral Research Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore. He has authored or coauthored more than 50 technical papers in refereed international journals and conferences. He has co-edited one book. His current research interests include video coding, quality assessment, perceptual modeling, feature representation, object recognition, and tracking.

Xingshi Luo received the bachelor's degree in electronic science and technology from the Beijing Institute of Technology, Beijing, China, in 2021. She is currently working toward the master's degree in information and communication engineering from the Beijing Institute of Technology, Beijing, China.

Her research interests include computer vision, deep learning, and their applications on remote sensing, such as multimodal hyperspectral image processing.



Wenzheng Wang received the Ph.D. degree in information and communication engineering from the School of Electrical and Information Engineering, Beijing Institute of Technology, Beijing, China, in 2019.

He was a Postdoctoral Research Fellow with the School of Electronics Engineering and Computer Science, Peking University, Beijing, China. He is currently a tenure-track Assistant Professor with the Beijing Institute of Technology, Beijing, China. His research interests include hyperspectral/optical imagery target detection and image analysis.