Multi-Scale Factor Joint Learning for Hyperspectral Image Super-Resolution

Qiang Li, *Member, IEEE,* Yuan Yuan, *Senior Member, IEEE,* and Qi Wang, *Senior Member, IEEE*

***Abstract*—Hyperspectral image super-resolution (SR) using auxiliary RGB image has obtained great success. Currently, most methods respectively train single model to handle different scale factors, which may lead to the inconsistency of spatial and spectral contents when converted to the same size. In fact, the manner ignores the exploration of potential interdependence among different scale factors in single model. To this end, we propose a multi-scale factor joint learning for hyperspectral image super-resolution (MulSR). Specifically, to take advantage of the inherent priors of spatial and spectral information, a deep architecture using single scale factor is designed by terms of symmetrical guided encoder (SGE) to explore the hyperspectral image and RGB image. Considering that there are obvious differences in texture details at various scale factors, another architecture is proposed which is basically the same as above, except that its scale factor is larger. On this basis, a multi-scale information interaction (MII) unit is modeled between two architectures by a direction-aware spatial context aggregation (DSCA) module. Besides, the contents generated by the model with multi-scale factor are combined to build a learnable feedback compensation correction (LFCC). The difference is fed back to the architecture with large scale factor, forming an interactive feedback joint optimization pattern. This calibrates the representation of spatial and spectral contents in the reconstruction process. Experiments on synthetic and real datasets demonstrate that our MulSR shows superior performance in terms of qualitative and quantitative aspects. Our code is publicly available at https://github.com/qianngli/MulSR.**

***Index Terms*—Hyperspectrla image, super-resolution, contextual aggregation, compensation correction, joint optimization.**

I. Introduction

As an important branch of optical images, hyperspectral image has been widely applied in some fields due to its noticeable advantage with high spectral resolution, such as image classification [1], anomaly detection [2], etc. Although the perception ability of spectral imaging systems has been continuously improved in recent years, it still does not fundamentally change the current dilemma, i.e., the imbalance between geometric state and attribute acquisition ability. Moreover, the external natural conditions are complicated and uncertain during observation. These situations can easily lead to image degradation. As a result, it cannot accurately describe the real information of the objects, which brings great challenges to the accurate analysis for hyperspectral image.

Qiang Li is with the School of Electronic Engineering, Xidian University, Xi'an 710071, P.R. China (e-mail: liqmges@gmail.com).

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

As one of the effective means of image sharpening [3], hyperspectral image super-resolution (SR) aims to restore low-resolution (LR) image to high-resolution (HR) image, so as to improve the image quality [4], [5], [6], [7], [8]. Since RGB image contains rich textures in space, many SR models are designed by combining LR hyperspectral image with RGB image generated by spectral response function [9]. According to the input pattern, existing methods are roughly classified into two categories, namely stack input pattern [10], [11], [12] and parallel input pattern [13] [14], [15]. Here, the purpose of stack input pattern is to stack upsampled LR hyperspectral image and RGB image, forming a new image cube. Then, the cube is viewed as a whole to establish SR model. This manner actually destroys the rich textures of RGB image and abundant spectra of hyperspectral image. It cannot utilize significant priors to improve the feature representation. Therefore, the performance of those approaches is relatively poor.

To make better use of these priors, the researchers replace the stack input pattern with parallel input pattern, and study each modality through two stream networks [16], [17], [18], [19], [20], where each image is regarded as one modality. The techniques involved in these methods are extremely similar to multi-modal fusion models. Concretely, previous methods mainly integrate two modalities in the initial or intermediate stage, where most of them are the former. Compared with approaches via stack input pattern, these methods obtain satisfying performance. However, they ignore a crucial issue, i.e., the textures are obviously diverse between RGB image and hyperspectral image. While some methods [9], [13], [19] fuse both features from two modalities in the middle stage, they do not take advantage of the inherent priors mentioned above. To address this drawback, a symmetrical feature propagation approach is developed [21]. The network adopts the symmetrical multi-step propagation mechanism to explore two inherent priors. In contrast to previous works, the approach attains high performance. Inspired by this framework, we also exploit the strategy to construct our deep model in this paper.

As for the SR task, current hyperspectral image SR methods usually construct corresponding models according to different scale factors, respectively. Without loss of generality, the super-resolved images with different scale factors contain obvious texture differences. Additionally, the greater the scale factor is, the more serious the detail loss in the reconstructed images is. If this texture difference can be used effectively, it really enhances performance in scenarios. For example, the features with low detail loss can be utilized to guide the model learning with high detail loss. Nevertheless, the above hyperspectral image SR approaches do not build models from

this perspective. Moreover, SR images obtained under different scale factors are converted to the images with same size. Ideally, the results should be consistent in terms of spatial and spectral contents. This can actually constrain the generated SR images by this manner. At present, existing hyperspectral image SR methods do not consider the consistency of information representation. Theoretically, the texture details of the image can be refined by adding more scale factors for inconsistent analysis. Although current natural SR methods [22]–[24] utilize some scale factors to build models, they do not fully explore the interdependence by more branches. Therefore, how to utilize the model with small scale factor to assist the model with large scale factor for joint SR, so as to establish the interdependence among different scale factors in single model needs further study.

Motivated by these analyses, we propose a multi-scale factor joint learning for hyperspectral image SR (MulSR) in this paper, and combine multi-scale factor to design the SR model instead of single scale factor. Specifically, we first propose a deep architecture using single scale factor by means of symmetrical guided encoder (SGE) to study the hyperspectral image and RGB image, so as to yield the corresponding spatial and spectral features. Similar to the above architecture, we add and construct the same network with large scale factor. Its aim is to establish the interdependence between two architectures. In this process, a multi-scale information interaction (MII) unit is designed to assemble feature sequences with different sizes. Under its effect, the clear contents with small scale factor are transferred to the network with large scale factor. By doing so, the model can be guided to achieve the detail enhancement. To obtain consistent contents in spatial and spectral aspects when converted to the same size, a learnable feedback compensation correction (LFCC) is developed. It computes the difference to explicitly correct the error data. In summary, the contributions of this paper are three-fold:

- We propose a MulSR to study the spatial and spectral contents under two types of scale factors. Unlike previous approaches, MulSR explores the inherent knowledge with diverse resolutions by SGEs, and establishes the relationship between architectures, which improves the SR performance for large scale factor.
- We design a MII unit to encode multi-scale feature sequences by direction-aware spatial context aggregation (DSCA) module. The Unit aggregates context information through four perspectives to mine interdependence and realize cross-scale information fusion interaction.
- We verify that the proposed model with multi-scale factor has better performance than the network with single scale factor by model analysis, which indicates that this framework is effective. Furthermore, extensive experiments on several datasets demonstrate that our MulSR can deal with degraded LR hyperspectral images well.

The rest of this paper is organized as follows: Section II reviews deep models with different scale factors. Section III introduces the designed approach in detail. We analyze and discuss the experiments in Section IV. Finally, Section V summarizes this work.

II. Realted Work

While there are many methods for hyperspectral image SR, we mainly introduce deep models using single scale factor and multi-scale factor in this section.

A. Deep Models Using Single Scale Factor

As for hyperspectral image SR, existing methods almost adopt single scale factor to design models. Here, we only introduce the methods using parallel input pattern. Dong *et al.* [20] first upsample LR hyperspectral image, and the features of HR RGB image are fed to different depths to analyze two modalities. To learn a general prior, Zhang *et al.* [19] design three parallel sub-networks to investigate the hyperspectral image, RGB image, and combined image of the two, respectively. In this process, a mutual-guiding fusion module is proposed to fuse those deep features. Zhou *et al.* [3] model pan-sharpening and its corresponding degradation process to achieve image SR. The network adopts an invertible neural network to conduct bidirectional closed-loop operation. Zhu *et al.* [25] extract LR bands with different numbers along spectral dimension, and propose a progressive residual network to analyze the zero-centric residual information from the two images. To exploit the spatial and spectral properties of the two images, a spatial-spectral fusion framework via double paths is developed [16]. This manner naturally preserves the specificity of each image. In contrast, the performance of the approaches using parallel input pattern is significantly higher than that of using stack input pattern. Nevertheless, these methods neglect the use of two inherent priors, namely the rich textures of RGB image and the abundant spectra of hyperspectral image. For that reason, the symmetrical propagation mechanism is adopted [21] to study two inherent priors. Inspired by this strategy, we also use similar architecture to construct SR model in our paper.

As mentioned in Section I, the textures contained in the images generated by different SR scale factors are remarkably distinctive. Taking advantage of these noticeable differences can actually improve performance. At present, current hyperspectral image SR approaches all ignore this point, which leaves room for further improvement. Importantly, these methods lack the exploration of the relationship among different scale factors in single model.

B. Deep Models Using Multi-Scale Factor

Existing hyperspectral image SR methods do not typically construct the models by combining multiple scale factors. Unlike this study for hyperspectral image SR, there are many ways to model natural image SR using different scale factors [22], [26]–[28]. For instance, the researchers propose progressive upsampling frameworks, such as LapSRN [23], MS-LapSRN [29], and ProSR [30]. These methods gradually achieve image SR by cascading multiple stages, where each stage contains convolutional neural network and upsampling layers. This framework extremely alleviates the learning difficulty, particularly for large scale factor. Importantly, it also does not introduce too many parameters. To better establish

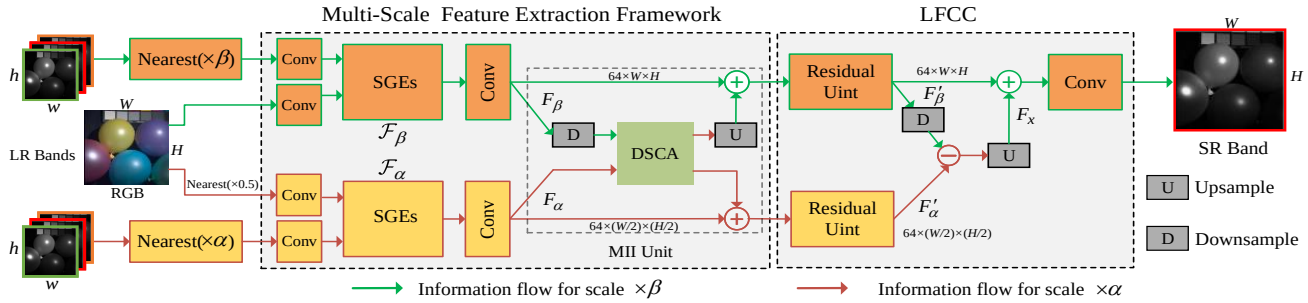


Fig. 1. The architecture of proposed MulSR. The network mainly contains multi-scale feature extraction and learnable feedback compensation correction (LFCC) module. In multi-scale feature extraction part, two branches with identical architecture are designed to extract corresponding features. In this process, a MII unit is proposed to achieve information interaction. Considering the inconsistency of information representation across scale factors, a LFCC module based on deep features is proposed to correct the spectral distortion and inconsistent details. Finally, the SR band is obtained.

the relationship between LR-HR image pairs, an iterative up-and-downsampling is incorporated into the SR task. For example, Haris *et al.* [22] apply back-projection to calculate the reconstruction error by upsampling and downsampling layers alternately, and then fuse it to adjust the super-resolved image. Other approaches such as SRFBN [31] and RBPN [24] also exploit this framework to discuss the relationship between LR-HR image pairs and attain relatively high performance. These networks can provide an error feedback to constraint generated SR image well.

Through the investigation of these methods, it can be found that almost all of these methods are based on single model with multiple scale factors. Although iterative up-and-downsampling can capture the interdependence of LR-HR image pairs, it does not jointly analyze the interdependence among cross-scale images in the form of multiple branches. Therefore, the research in this aspect is not enough and further exploration is needed.

III. The Proposed Method

Given a LR hyperspectral image $Z \in \mathbb{R}^{w\times h\times L}$ and a corresponding RGB image $Y \in \mathbb{R}^{W\times H\times 3}$, where w and h are width and height for image Z. W and H denote width and height for image Y. Hyperspectral image SR network is designed to produce estimated hyperspectral image $X \in \mathbb{R}^{W\times H\times L}$. The relationship among three images can be denoted as

$$Y = XT, \tag{1}$$

$$Z = RX, \tag{2}$$

where T is the intrinsic parameter of the imaging system, and R is the degradation function.

A. Motivation and Overview

Unlike the natural image, hyperspectral image has rich spectral information, while its spatial resolution is low. For hyperspectral image SR, its design principle is to produce an image with high spatial resolution in space. Currently, almost all methods focus on the single scale factor while ignoring the exploration of potential interdependence among different scale factors in single model. Moreover, these methods do not consider the consistent representation in the spatial and spectral contents generated by different scale factors. To establish the interdependence among different scale factors in single model, we propose a MulSR by means of auxiliary RGB image, as shown in Fig. 1. Since hyperspectral image with dozens or even hundreds of bands requires more memory footprint than RGB image, we refer to [32], and only employ several bands instead of the whole hyperspectral image and a HR RGB image to achieve SR. Overall, the proposed network mainly includes two architectures and two interaction modules. In the architecture $\mathcal{F}_\alpha$ with scale factor $\times\alpha$ ($\alpha > 1$), to fully explore rich texture of RGB image and abundant spectra of hyperspectral image, we exploit the symmetrical pattern to study the information in two modalities over a long range. Specifically, we analyze each modality separately through the residual unit, and design some SGEs to realize the information interaction between them. To maintain the specificity of each modality in a long range, the fusion contents are fed back to the two modalities, respectively, which allows for more accurate coding. Similar to [21], we adopt multi-step propagation mechanism to further extract the deep information in each modality.

Generally, the larger the scale factor is, the more serious the detail loss of the reconstructed image is. Different from small scale factor used in natural image SR, hyperspectral image SR using auxiliary RGB image usually requires larger scale factors, such as $\times 16$. The details of the image that can be restored for large scale factors are significantly reduced during reconstruction. For larger scale factors, if we use the clear details of the model with small scale factor, it can recover the details of larger scale factor in a certain sense. Motivated by this discovery, we add and design another architecture $\mathcal{F}_\beta$ which is basically the same as above architecture, except that its scale factor is larger ($\alpha < \beta$). Its aim is to utilize the model with small scale factor to assist the model with large scale factor for joint SR. To learn the inter-scale correlation, the MII unit is designed to encode feature sequences by multiple directions. It aggregates contexts via four perspectives to mine the interdependence and realize cross-scale information fusion interaction. Considering the inconsistency of information representation, a LFCC module based on deep features is proposed for two scale factors. This analyzes the inconsistent contents and feeds it back to the architecture $\mathcal{F}_\beta$. The manner

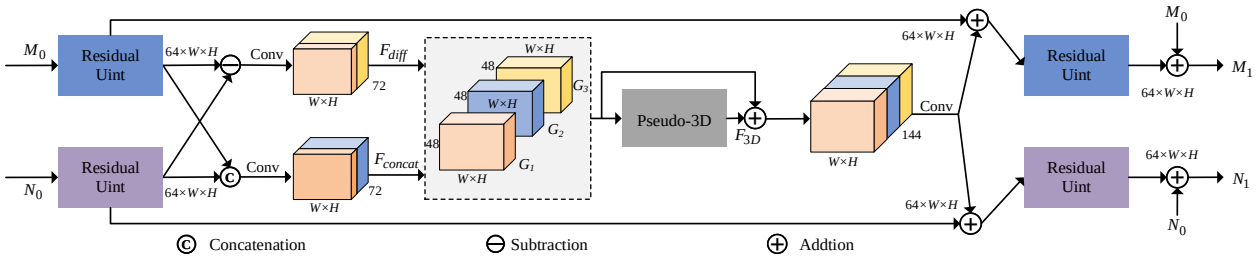


Fig. 2. The architecture of proposed Symmetrical Guided Encoder (SGE).

corrects the spectral distortion and inconsistent details, forming an interactive feedback joint optimization pattern. Thus, it achieves image detail refinement and structure-preserving for large scale factor. Through the above procedure, single super-resolved band is generated. Finally, all bands are obtained by recursion.

B. Single Scale Network Using Symmetrical Guided Encoder

In this section, we describe only one of two architectures. Considering that the adjacent bands of hyperpsectral image are highly correlated, as for a LR hyperspectral image Z, three adjacent bands are first selected to form a new combined image $I(i)$, i.e.,

$$I(i) = \begin{cases} [Z_1, Z_2, Z_3], & i = 1 \\ [Z_{i-1}, Z_i, Z_{i+1}], & 2 \leq i \leq L-1. \\ [Z_{L-2}, Z_{L-1}, Z_L], & i = L \end{cases} \tag{3}$$

By doing so, it is designed to reduce the memory footprint and make effective use of similar adjacent spectral information. To easily fuse RGB image and combined image $I(i)$ while avoiding unnecessary up-and-downsampling operations, the image $I(i)$ is interpolated by *Nearest* interpolation to be the same size as corresponding RGB image Z. Since RGB image has rich textures and hyperspectral image exhibits abundant spectral information, the symmetrical pattern is designed to study the information of two modalities over a long range, forming two symmetrical branches, which can fully make use of two inherent priors. As for two branches, the corresponding shallow features are extracted by a convolution layer, respectively. Then, the SGE is proposed to propagate and fuse information between the two forms. Fig. 2 displays the architecture of the module. Supposing that the shallow features for hyperspectral image are defined as F_0, it is first input into the residual unit to extract its information. Here, the residual unit consists of two convolution layers and a Rectified Linear Unit (ReLU), where skip connection is set. To realize the information complementarity between the two modalities, the features extracted from the two branches are aggregated and the model is allowed to modulate the contents.

Let M_0 and $N_0 \in \mathbb{R}^{64\times W\times H}$ denote the features after performing residual units for two branches. Their features are organized by concatenation and subtraction operations, which compute the fusion and difference of two forms, respectively. It enables the model to learn more detailed textures by different perspectives. Then, the corresponding features are analyzed by means of a convolution layer, respectively. The process is formulated as

$$F_{diff} = f_{conv}(M_1 - N_1), \tag{4}$$

$$F_{concat} = f_{conv}([M_1, N_1]), \tag{5}$$

where $f_{conv}(\cdot)$ represents convolution function. Adjacent bands in hyperspectral image have high similarity [33]. To utilize this peculiarity, the number of the features F_{diff} and F_{concat} is first increased from 64 to 72 during convolution. It can easily conduct 3D convolution operation along spectral dimension by generating more features. In this process, these features are split three groups with same size. We exploit a pseudo-3D convolution [34] to mine the features in spatial and spectral domains, i.e.,

$$F_{3D} = f_{spectral}(f_{spatial}([G_1, G_2, G_3])), \tag{6}$$

where G_1, G_2, and $G_3 \in \mathbb{R}^{48\times W\times H}$. $f_{spatial}(\cdot)$ and $f_{spectral}(\cdot)$ denote spatial and spectral extraction functions. After generating multi-domain information, we decompose these features and concatenate them, which is followed by a convolution layer. Deep fusion features are embedded into two branches to obtain stronger feature representation. It can be coded more accurately, so as to maintain the specificity of each modality. Similar to the above steps, the residual unit is utilized again to mine the additional features. To encode each modality in a long range, we adopt multi-step propagation mechanism to perform feature learning in the same way. It can calibrate the feature representation. Importantly, the manner restrains the generation and spread of misleading information and forms more robust feature representation. Finally, we unify the two modalities during deep feature extraction, and a convolution layer is conducted to produce coarse SR results.

C. Multi-Scale Network Using Information Interaction

To exploit the model with small scale factor to assist the model with large scale factor to jointly execute SR, we add and design another architecture $\mathcal{F}_\beta$, which is basically the same as the above architecture, except that its scale factor is larger. Since there exists texture differences in SR results with different scales, the relationship of the two architectures is modeled to modulate the whole network. The process mainly involves a MII unit and a LFCC module.

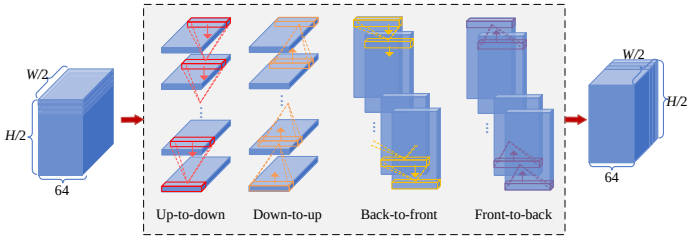


Fig. 3. The architecture of proposed Direction-Aware Spatial Context Aggregation (DSCA) module.

1) Multi-Scale Information Interaction: To establish the connection between two architectures, the MII unit is designed to assemble feature sequences with different sizes. According to the obtained deep features F_α and F_β, we first downsample features F_β to the same size as features F_α. These features are then handled by concatenation operation, followed by a convolution layer to reduce channels. It achieves cross-scale information fusion. Considering that the rich spectral information can promote the feature learning in spatial domain, inspired by the work [35], we propose a DSCA module to learn the spectral and spatial contexts by joint analysis in a multi-view way. Fig. 3 illustrates the detailed architecture of DSCA module. The module encodes feature sequences by multi-direction, which aggregates contexts through four perspectives to mine the interdependence and realize cross-scale information fusion interaction. Taking up-down aggregation as an example, its formula is defined as

$$F_c(k) = \begin{cases} f_{conv}(F_d^k), & k = 1 \\ f_{conv}(F_d^k + f_{conv}(F_d^{k-1})), & k = 2, 3, ..., W \end{cases}, \tag{7}$$

where F_d is intermediate features. The DSCA module can maintain the same size of the input. Meanwhile, the global spatial information is introduced into the whole feature map to learn. According to DSCA mechanism, the features of specific region contain the global information with different extents. The region actually improves feature representation by learning similar patterns. Thus, the representation is able to implicitly generate reliable contents. Similarly, the fusion features are incorporated into two architectures to further enhance feature learning.

2) Learnable Feedback Compensation Correction: SR images obtained with different scale factors are converted into the images of the same size. Ideally, SR results should be consistent in terms of spatial and spectral contents. At present, existing hyperspectral image SR methods almost ignore the consistency of information representation. Notably, the constructed multi-scale network has the natural advantage of multi-scale information. To purposely handle this characteristic, a LFCC module is proposed based on coarse SR results for two scale factors. The architecture of this module is shown in Fig. 1. Let $F_\alpha^{'} \in \mathbb{R}^{64\times W/2\times H/2}$ and $F_\beta^{'} \in \mathbb{R}^{64\times W\times H}$ be the deep features after residual unit. To generate consistent information, the deep features $F_\beta^{'}$ are downsampled into the same size as features $F_\alpha^{'}$ by a learnable way. The overall process of LFCC is conducted by

$$F_x = U(F_\alpha^{'} - D(F_\beta^{'})), \tag{8}$$

where $U(\cdot)$ and $D(\cdot)$ represent the learnable upsampling and downsampling functions. The module explicitly computes inconsistent contents and feeds it back to the architecture $\mathcal{F}_\beta$. The manner corrects the spectral distortion and inconsistent details in space, forming an interactive feedback joint optimization pattern. It ensures the content consistent in spatial and spectral domains. Thus, the module achieves image detail refinement and structure-preserving for large scale factor.

IV. Experiments

This section conducts extensive experiments in various aspects. First, we introduce datasets, implementation details, etc. Then, the effectiveness of key modules is analyzed and discussed. Finally, the performance comparison with mainstream methods is performed on multiple datasets.

A. Datasets

1) CAVE: The CAVE dataset was captured by Apogee Alta U260 camera [36]. As typical dataset for hyperspectral image SR, it has 32 images and mainly 5 sections. Here, the size of each image is $512 \times 512 \times 31$. Here, the spectral response function of the dataset is known.

2) Harvard: The Harvard dataset was collected by Nuance FX, CRI Inc camera in indoor and outdoor [37]. In contrast to CAVE dataset, the dataset has more images, where the size of each image is $1392 \times 1040 \times 31$. Similar to existing works, we select 50 outdoor images to participate in experiment. Likewise, the spectral response function of the dataset is known.

3) Sample of Roman Colosseum: The real dataset was obtained by WorldView-2 over Roman Colosseum area. It has one hyperspectral remote sensing image. The size of the image is $419 \times 658 \times 128$. Accordingly, the size of corresponding HR RGB image is 1676×2632 pixels. To train the model well, the part of hyperspectral image ($209 \times 658 \times 8$) and RGB image ($836 \times 2632 \times 3$) in top left are cropped, and the remaining part is chosen to test. Note that the spectral response function of the dataset is unknown.

B. Comparison Methods and Evaluation Metrics

To demonstrate the superiority of the proposed method, four approaches are compared with our method in each dimension. They are MoG-DCN [20], UAL [19], PZRes-Net [25], and CoarseNet [21]. Among these competitors, UAL and CoarseNet contain two steps. The first step is to learn an general model by supervised manner. The second step is to optimize the model in specific image by unsupervised manner. Note that two methods need spectral response function in second step. For fair comparison, we remove the second step. The remaining works are supervised approaches.

To evaluate the performance, we apply Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM), Spectral Angle Mapper (SAM), and Root Mean Squared Error (RMSE). Here,

the higher values of PSNR and SSIM indicate better quality of reconstructed image. Besides, the obtained image is better in the aspects of edge and texture, when the values of SAM and RMES are small.

C. Implementation Details

Since current datasets have fewer hyperspectral images even one image, we augment them to obtain more training samples. As for CAVE and Harvard datasets, we select 80% samples to train. Then, these samples are randomly flipped, rotated, and rolled. With respect to Sample of Roman Colosseum dataset, the image in the training set is randomly cropped to obtain 64 patches with the size $12\beta \times 12\beta$. Similarly, these patches are augmented by above way. In test stage, we refer to [21] to obtain test set. Concretely, anisotropic Gaussian is first applied to blur the HR hyperspectral images. Then, we downsample the blur images according scale factor and add Gaussian noise to obtain test images. Here, the mean and variance of parameters are set to 0 and 0.001, respectively.

With respect to experiment setup, we select the size of convolution kernels to be 3×3, except for the kernels mentioned above. Moreover, the number of these kernels is set to 64. Following previous works, we fix the learning rate at 10^{-4}, and its value is halved every 30 epoch. To optimize our model, the ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$ is chosen. Moreover, we set $2\alpha = \beta$ in our paper. All experiments are carried out on the PyTorch framework using NVIDIA GeForce GTX 1080 GPU.

D. Model Analysis

In this section, we conduct the experiment for scale factor $\beta = 8$ on CAVE dataset to study the effectiveness of different modules in model. For simply analysis, we adopt two steps propagation mechanism to extract the deep information of each modality in SGE.

TABLE I
EFFECT OF DIFFERENT PARTS FOR MII UNIT ON THE PERFORMANCE. THE UNDERLINE AND BOLD REPRESENT THE WAY USED IN THIS PAPER AND BEST PERFORMANCE, RESPECTIVELY.

Module	PNSR	SSIM	SAM	RMSE	Param.
w/o MII	43.686	0.9910	3.904	0.0069	**2427K**
w/ MII	**44.745**	**0.9923**	**3.694**	**0.0061**	2509K
w/o DSCA	43.487	0.9901	4.337	0.0071	2452K

1) Study of Multi-Scale Information Interaction: To establish the connection between two architectures, the multi-scale information interaction (MII) unit is designed to assemble feature sequences with different sizes. In this section, we demonstrate this unit to analyze the effectiveness. Table I depicts the study of different parts for MII. We remove the MII unit and DSCA module in MII unit respectively, and compare them with our method. As seen this table, the best results are obtained by adding MII unit when the parameters do not change much, which reflects that the module is effective. Without MII unit, the results of each metric decrease significantly. In fact, the module and unit are the key bridges between the two architectures with different scale factors. It plays the role of information flow transmission and promotes the network optimization. Therefore, it demonstrates that the unit and module are effective for feature representation.

TABLE II
EFFECT OF LFCC ON THE PERFORMANCE. THE UNDERLINE AND BOLD REPRESENT THE WAY USED IN THIS PAPER AND BEST PERFORMANCE, RESPECTIVELY.

Module	PNSR	SSIM	SAM	RMSE	Param.
w/o LFCC	44.556	**0.9923**	3.752	0.0063	**2381K**
w/ LFCC	**44.745**	**0.9923**	**3.694**	**0.0061**	2509K

TABLE III
EFFECT OF MULTI-SCALE STRATEGY ON THE PERFORMANCE. THE UNDERLINE AND BOLD REPRESENT THE WAY USED IN THIS PAPER AND BEST PERFORMANCE, RESPECTIVELY.

Type	PNSR	SSIM	SAM	RMSE	Param.
Single scale	42.346	0.9866	5.196	0.0080	**1187K**
Multi-scale	**44.745**	**0.9923**	**3.694**	**0.0061**	2509K

2) Study of Learnable Feedback Compensation Correction: Considering the inconsistency of information representation after images with different resolutions are converted to the same scale, a learnable feedback compensation correction (LFCC) module is adopted to optimize it. To analyze its effectiveness in this section, we observe if it improves the performance by adding it. Table II reports the effect of LFCC module on the performance. Its main purpose is to correct the contents converted from different scales to the same scale, and autonomously calibrate the content representation through learnable means. Therefore, the module achieves image detail refinement and structure-preserving for large scale factor. Meanwhile, it also serves as an information interaction between the two architectures. As a result, the module modulates the entire network and achieves performance improvement, especially for PNSR, which verifies that the module is helpful to learn model parameters.

3) Study of Multi-Scale Framework: The larger the scale factor is, the more serious the detail loss of the reconstructed image is. To address more detail loss in large scale factor, a multi-scale framework is proposed, which utilizes the model with small scale factor to assist the model with large scale factor. This section mainly investigates the effectiveness of multi-scale network by introducing single scale framework. Table III displays the performance of multi-scale strategy. When two scale factors are involved with the model, their parameters are approximately doubled relative to a single scale. Intuitively, there is no way to avoid this in our approach, and this is one of the major weaknesses for multi-scale network. Through the numbers in the table, we notice that the difference between the two is large in terms of evaluation metrics. There are two main reasons for this. One is that information interaction improves the information flow transmission of the model and further promotes the model optimization. The other is to transfer the clear details with small scale factor to the content learning of the model with large scale factor, which is helpful to compensate for the information loss. These two

TABLE IV
EFFECT OF DIFFERENT NUMBERS OF CROSS-MODAL FUSION MODULE ON THE PERFORMANCE. THE BOLD REPRESENTS THE BEST PERFORMANCE.

Metrics	1	2	3	4
PSNR	44.199	44.745	44.835	**45.438**
SSIM	0.9911	0.9923	0.9925	**0.9932**
SAM	3.904	3.694	3.534	**3.486**
RMSE	0.0065	0.0061	0.0061	**0.0057**
Param.	**1446K**	2509K	3572K	4636K

TABLE V
PERFORMANCE OF COMPETITORS ON CAVE DATASET. THE BEST RESULT AND SECOND RESULT ARE DENOTED AS THE BOLD AND UNDERLINE, RESPECTIVELY.

Scale	Method	PSNR	SSIM	SAM	RMSE
×8	PZRes-Net	41.07	0.9513	12.57	0.0089
	MoG-DCN	43.63	0.9873	5.76	0.0069
	UAL	43.73	0.9889	6.10	0.0068
	CoarseNet	<u>43.95</u>	<u>0.9931</u>	**3.46**	<u>0.0070</u>
	MulSR	**45.44**	**0.9933**	<u>3.49</u>	**0.0057**
×16	PZRes-Net	39.91	0.9443	12.80	0.0103
	MoG-DCN	33.76	0.8815	15.77	0.0215
	UAL	**42.50**	<u>0.9887</u>	<u>6.18</u>	<u>0.0083</u>
	CoarseNet	32.36	0.9140	9.66	0.0250
	MulSR	<u>41.99</u>	**0.9889**	**4.51**	**0.0082**

actions together result in a significant increase in evaluation metrics. Thus, it is concluded that multi-scale network is indeed effective for feature learning.

E. Study of Number of Symmetrical Guided Encoder

In single scale architecture, we adopt multi-steps propagation mechanism to extract the deep information in each modality. To investigate the number of propagation mechanism in this section, we set the steps from 1 to 4. Table IV reports the performance of different numbers of SGE. As seen this table, different numbers exhibit great distinction in evaluation metrics and parameter. In particular, the difference is most noticeable when the number is set to 1 or 4. From the point of view of parameters, it can be found that the parameters increase by almost 1000K after each single step. Since our model is constructed by multi-scale network, it means that this module has 500K. This is also the main source of parameters for the proposed approach. The manner requires computational resources and memory to run, compared with single architecture. Based on the above considerations, we select four steps propagation pattern to implement the following experiments.

F. Performance Comparison with Existing Approaches

To show the superiority of the proposed method, we compare the proposed method with four existing approaches on different scale factors and datasets, including PZRes-Net [25], MoG-DCN [20], UAL [19], and CoarseNet [21]. Table V shows the performance comparison of competitors on CAVE dataset. It can be found that the proposed MulSR attains the better performance. Specifically, among these methods, the shallow RGB image features of PZRes-Net and MoG-DCN

TABLE VI
PERFORMANCE OF COMPETITORS ON HARVARD DATASET. THE BEST RESULT AND SECOND RESULT ARE DENOTED AS THE BOLD AND UNDERLINE, RESPECTIVELY.

Scale	Method	PSNR	SSIM	SAM	RMSE
×8	PZRes-Net	39.01	0.9455	8.84	0.0128
	MoG-DCN	42.78	0.9765	3.68	0.0098
	UAL	41.81	0.9735	4.80	0.0105
	CoarseNet	<u>42.88</u>	<u>0.9768</u>	**3.39**	<u>0.0098</u>
	MulSR	**43.29**	**0.9770**	<u>3.43</u>	**0.0096**
×16	PZRes-Net	38.71	0.9468	8.88	0.0134
	MoG-DCN	36.53	0.9315	10.08	0.0169
	UAL	41.42	0.9747	4.66	0.0108
	CoarseNet	<u>41.68</u>	<u>0.9752</u>	**3.60**	<u>0.0107</u>
	MulSR	**42.15**	**0.9754**	<u>3.77</u>	**0.0105**

are embedded into hyperspectral image branch with different depths to extract features by fusion. Obviously, the feature extraction of RGB image is regarded as an auxiliary branch to realize hyperspectral image SR. In fact, this manner does not take full advantage of the rich spatial information of RGB image to guide model learning. Importantly, the difference of texture detail between the bands in hyperspectral image and RGB image has not been fully explored, which leads to unsatisfactory results. The performance in this table reveals this point well. MoG-DCN and CoarseNet show large changes in results, especially for scale ×16. It indicates that these approaches have poor robustness across scales. In our opinion, the main reason for this situation is that the models constructed by these methods can not generate the image details well when using little training data. In contrast, MulSR and UAL can obtain relatively stable performance. Fig. 5 displays the relationship between PSNR performance and the number of parameters/FLOPs. According to this figure, we can find that our MulSR in parameters has more advantage than UAL. However, our MulSR involves two branches with different scale factors and performs image SR through the interaction between them, which inevitably produces the large amount of FLOPs. Unlike this result, MulSR obtains the absolute performance in terms of PSNR and RMSE. It can effectively bridge the gap in FLOPs. Similarly, Table VI also depicts that our method achieves better performance on Harvard dataset. Unlike the results on CAVE dataset, the values of CoarseNet and MoG-DCN do not fluctuate as much when more training samples are added. Through comprehensive analysis, it can be concluded that our MulSR overall has better performance in terms of evaluation metrics and parameter, except for FLOPs.

The visual comparisons are also shown in spatial reconstruction and spectral distortion aspects. As for spatial reconstruction, one band in reconstructed hyperspectral image is selected. Since the grayscale band is inconvenient to show the details, we compute the absolute values of the difference between reconstructed image and ground-truth to clearly describe through heat map. Generally, the color is bluer, which indicates that the details of the image are sharper. Fig. 4 shows the visual comparison of spatial reconstruction. One can obverse that our method obtains more bluer in the enlarged area. In particular, the contents around the edges is very light in this

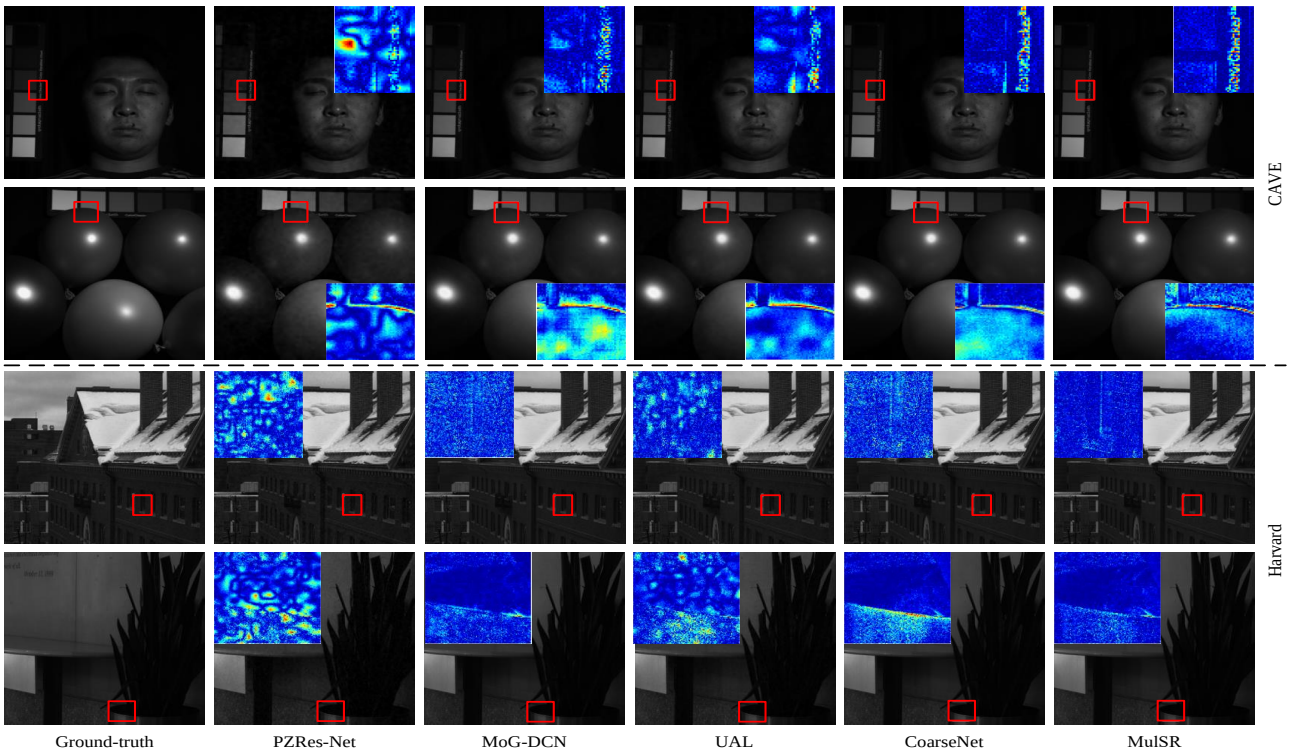


Fig. 4. Visual comparison in spatial reconstruction on two dataset. The first and second lines are the visual images of the 10-th band and 20-th band, respectively.

area. Besides, the visual comparison of spectral distortion is displayed in Fig. 6. Likewise, the red curves of our MulSR are closer to ground-truth. In summary, our method can produce more desirable results in both spatial and spectral aspects.

G. Performance Comparison Under Various Degradations

The spectral data tends to suffer from various degradation in the process of imaging. In general, the degradation mainly contains noise, blur, and downsampling. To demonstrate the robustness of the proposed method and other competitors, we set three ways to generate LR hyperspectral image under various conditions, including 1) add Gaussian noise with mean 0 and variance of three values. 2) utilize three blur kernels with size 7×7, 11×11, and 15×15. 3) employ anisotropic and isotropic Gaussian kernels.

Table VII reports the performance comparison under various degradation on CAVE dataset. The performance of all methods fluctuates in terms of the three types of degradation. As for noise level, PZRes-Net does not take into account the effect of noise in the process of training. Therefore, the method achieves ideal results in the absence of noise. However, its performance degrades very badly in the presence of noise. This also shows that this method can not effectively deal with the situation with noise. Different from this method, MoG-DCN and UAL adds noise to train the models. Although they can address and analyze well in the case of low noise level, the performance degradation is more obvious under strong noise when the variance is set to 6. In contrast, other competitors show reasonable results, and the performance fluctuations are not so great, especially for our MulSR. With respect to kernel size and kernel type, these approaches show relatively stable results, which also shows that these degenerative factors have little influence on the model. Overall, our method can well handle images under different degradation conditions, which verifies that the proposed method has very good generalization performance.

H. Application on Real Hyperspectral Image

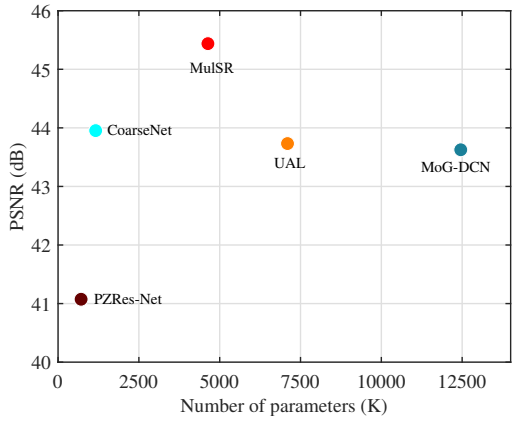
To examine the processing ability of the proposed method on the real data, Sample of Roman Colosseum dataset is exploited to analyze. Considering that the HR image is not available, the approach in [38] is utilized to deal with this trouble. First, the patches are randomly cropped in hyperspectral image and corresponding RGB image from training set, respectively. As a result, the sizes of hyperspectral image and RGB image are $36 \times 36 \times 8$ and $144 \times 144 \times 3$. Similarly, we downsample the obtained patches according to scale factor $\times 1/4$. The training LR-HR samples are constructed by downsampled patch and original patch. Fig. 7 provides the visual results on this dataset. The figure illustrates that our method yields good visual effect on the details, especially the edges. It reveals that our MulSR can address real degraded images well.

V. Conclusion

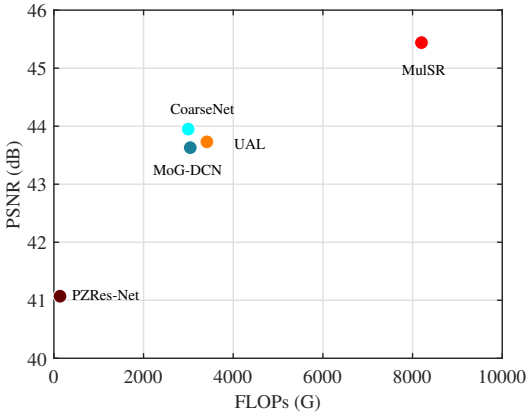
This paper develops a multi-scale factor joint learning for hyperspectral image SR (MulSR). The main contributions are as follows: 1) We explore the spatial and spectral knowledge

TABLE VII
PERFORMANCE COMPARISON UNDER VARIOUS DEGRADATION ON CAVE DATASET.

Method	Metric	Noise Level			Kernel Size			Kernel Type	
		0	3	6	7×7	11×11	15×15	anisotropic	isotropic
PZRes-Net	PSNR	47.17	41.07	36.53	41.23	41.13	41.07	41.07	40.967
	SSIM	0.994	0.9513	0.8904	0.9528	0.9518	0.9513	0.9513	0.9507
	SAM	2.98	12.57	18.68	12.38	12.52	12.57	12.57	13.01
	RMSE	0.0049	0.0089	0.015	0.0088	0.0089	0.0089	0.0089	0.0091
MoG-DCN	PSNR	44.58	43.63	41.79	43.82	43.68	43.63	43.63	43.8
	SSIM	0.9909	0.9873	0.9785	0.9874	0.9873	0.9873	0.9873	0.9873
	SAM	4.20	5.76	8.01	5.7	5.73	5.76	5.76	5.78
	RMSE	0.0063	0.0069	0.0084	0.0068	0.0069	0.0069	0.0069	0.0068
UAL	PSNR	45.51	43.73	41.37	43.99	43.81	43.73	43.73	44.29
	SSIM	0.9941	0.9889	0.9777	0.9888	0.9886	0.9889	0.9889	0.9891
	SAM	3.02	6.10	9.05	6.02	6.06	6.10	6.10	5.98
	RMSE	0.0058	0.0068	0.0087	0.0066	0.0067	0.0068	0.0068	0.0063
CoarseNet	PSNR	44.62	43.95	42.66	44.05	44.12	43.95	43.95	44.29
	SSIM	0.9937	0.9931	0.9897	0.9924	0.9924	0.9931	0.9931	0.9921
	SAM	2.89	3.46	5.03	3.85	3.84	3.46	3.46	3.93
	RMSE	0.0061	0.007	0.0076	0.0065	0.0065	0.007	0.007	0.0065
MulSR	PSNR	46.19	45.44	44.02	45.13	45.28	45.44	45.44	45.74
	SSIM	0.9943	0.9933	0.9896	0.993	0.9931	0.9933	0.9933	0.9928
	SAM	2.71	3.49	4.63	3.57	3.53	3.49	3.49	3.60
	RMSE	0.0053	0.0057	0.0066	0.0059	0.0058	0.0057	0.0057	0.0057



(a) PSNR versus Parameters



(b) PSNR versus FLOPs

Fig. 5. PSNR performance versus number of parameters and FLOPs for scale factor ×8 on CAVE dataset.

through two scale factors, which utilizes the model with small scale factor to assist the model with large scale factor for joint SR. 2) To establish the interdependence between scale factors, a multi-scale information interaction (MII) unit is designed to assemble feature sequences by direction-aware spatial context aggregation (DSCA) module. It can be guided to achieve detail enhancement. 3) Extensive experiments are conducted by various aspects, and different module analyses indicate that the proposed multi-scale framework is effective. Moreover, the results on three datasets verify that MulSR can handle degraded images well in both spatial and spectral aspects. In the future, we extend our approach in parameter reduction to quickly process LR images.

REFERENCES

[1] D. Hong, R. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2020.

[2] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sensing*, 2023.

[3] M. Zhou, J. Huang, D. Hong, F. Zhao, C. Li, and J. Chanussot, "Rethinking pan-sharpening in closed-loop regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, 2023.

[4] Q. Li, Y. Yuan, X. Jia, and Q. Wang, "Dual-stage approach toward hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 7252–7263, 2022.

[5] Y. Liu, J. Hu, X. Kang, J. Luo, and S. Fan, "Interactformer: Interactive Transformer and CNN for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–15, 2022.

[6] 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.

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

[8] 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, 2019.

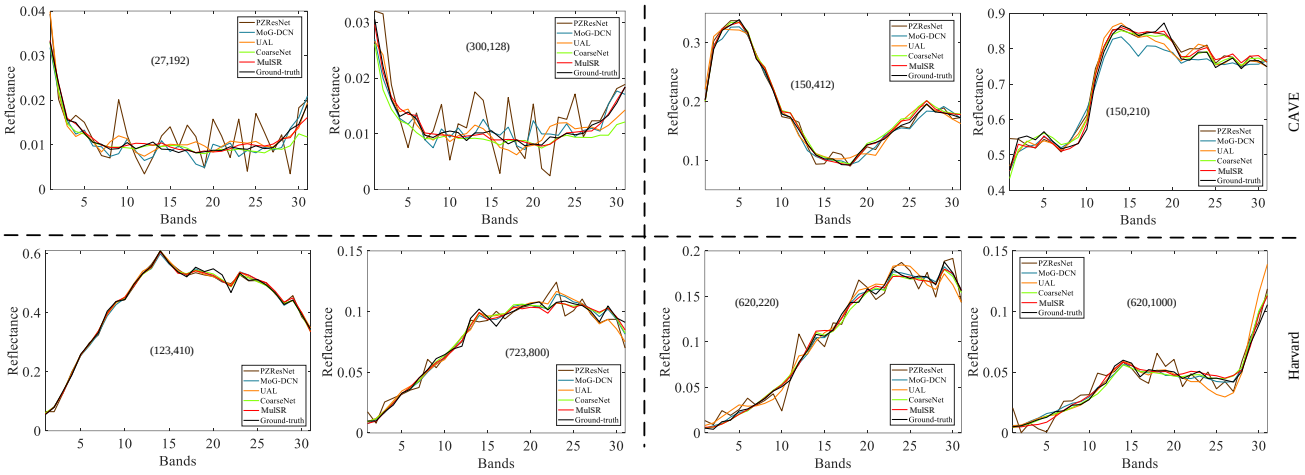


Fig. 6. Visual comparison in spectral distortion by selecting two pixels. The left to right columns are the visual results of above images.

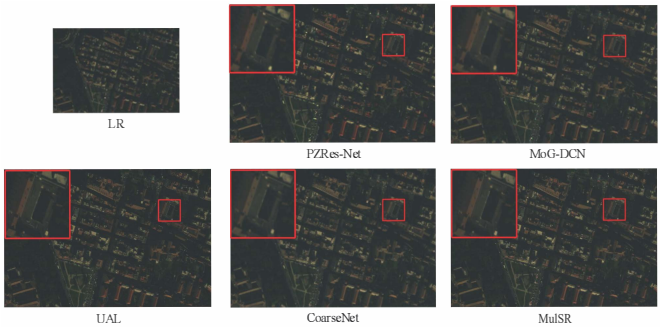


Fig. 7. Visual comparison on real hyperspectral image dataset. We choose the 2-3-5 bands after SR to synthesize the pseudo-color image.

[9] J.-F. Hu, T.-Z. Huang, L.-J. Deng, H.-X. Dou, D. Hong, and G. Vivone, "Fusformer: A transformer-based fusion network for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[10] X.-H. Han, B. Shi, and Y. Zheng, "SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution," in *Proc. Int. Conf. Image Process.*, 2018, pp. 2506–2510.

[11] X. Zhang, W. Huang, Q. Wang, and X. Li, "SSR-NET: Spatial-spectral reconstruction network for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 7, pp. 5953–5965, 2021.

[12] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.

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

[14] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Process.*, vol. 30, pp. 1423–1438, 2021.

[15] S.-Q. Deng, L.-J. Deng, X. Wu, R. Ran, D. Hong, and G. Vivone, "PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion," *IEEE Trans. Geosci. Remote Sensing*, vol. 61, pp. 1–15, 2023.

[16] X.-H. Han, Y. Zheng, and Y.-W. Chen, "Multi-level and multi-scale spatial and spectral fusion cnn for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2019, pp. 4330–4339.

[17] R. Ran, L.-J. Deng, T.-X. Jiang, J.-F. Hu, J. Chanussot, and G. Vivone, "Guidednet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution," *IEEE Trans. Cybern.*, 2023.

[18] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "RGB-induced feature modulation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sensing*, vol. 61, pp. 1–11, 2023.

[19] 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. Vision Pattern Recognit.*, 2020, pp. 3073–3082.

[20] 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, 2021.

[21] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "Symmetrical feature propagation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–12, 2022.

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

[23] 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. Vision Pattern Recognit.*, 2017, pp. 624–632.

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

[25] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Process.*, vol. 30, pp. 1423–1438, 2021.

[26] 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.

[27] P. Behjati, P. Rodriguez, A. Mehri, I. Hupont, C. F. Tena, and J. Gonzalez, "Overnet: Lightweight multi-scale super-resolution with overscaling network," in *IEEE Winter Conf. Appl. Comput. Vis.*, 2021, pp. 2694–2703.

[28] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5791–5800.

[29] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, 2018.

[30] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2018, pp. 864–873.

[31] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 3867–3876.

[32] Q. Wang, Q. Li, and X. Li, "Hyperspectral image super-resolution using spectrum and feature context," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11276–11285, 2021.

[33] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 6, pp. 5028–5039, 2020.

[34] S. Xie, C. Sun, J. Huang, Z. Tu, and K. Murphy, "Rethinking spatiotem-

poral feature learning: Speed-accuracy trade-offs in video classification," in *Eur. Conf. Comput. Vis.*, 2018, pp. 305–321.
[35] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial CNN for traffic scene understanding," in *AAAI Conf. Artif. Intell.*, 2018, vol. 32.
[36] 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.
[37] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 193–200.
[38] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sensing*, vol. 56, no. 9, pp. 5443–5457, 2018.

Qiang Li (Member, IEEE) 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.

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, pattern recognition and remote sensing.