Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 32(7): 4517–4542.
DOI: 10.3934/era.2024205
Received: 29 March 2024
Revised: 17 June 2024
Accepted: 25 June 2024
Published: 23 July 2024

Research article

Single hyperspectral image super-resolution using a progressive upsampling deep prior network

Haijun Wang[1], **Wenli Zheng**[1,*], **Yaowei Wang**[1], **Tengfei Yang**[2], **Kaibing Zhang**[3] **and Youlin Shang**[1]

[1] School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang 471000, China

[2] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

[3] School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China

* **Correspondence:** Email: dip@haust.edu.cn.

Abstract: Hyperspectral image super-resolution (SR) aims to enhance the spectral and spatial resolution of remote sensing images, enabling more accurate and detailed analysis of ground objects. However, hyperspectral images have high dimensional characteristics and complex spectral patterns. As a result, it is critical to effectively leverage the spatial non-local self-similarity and spectral correlation within hyperspectral images. To address this, we have proposed a novel single hyperspectral image SR method based on a progressive upsampling deep prior network. Specifically, we introduced the spatial-spectral attention fusion unit (S^{2}AF) based on residual connections, in order to extract spatial and spectral information from hyperspectral images. Then we developed the group convolutional upsampling (GCU) to efficiently utilize the spatial and spectral prior information inherent in hyperspectral images. To address the challenges posed by the high dimensionality of hyperspectral images and limited training dataset, we implemented a parameter-sharing grouped convolutional upsampling framework within the GCU to ensure model stability and enhance performance. The experimental results on three benchmark datasets demonstrated that the proposed single hyperspectral image SR using a progressive upsampling deep prior network (PUDPN) method effectively improves the reconstruction quality of hyperspectral images and achieves promising performance.

Keywords: deep prior network; hyperspectral image super-resolution; progressive upsampling; deep learning; spatial-spectral attention

1. Introduction

With hundreds or even thousands of distinct discrete bands, hyperspectral sensors acquire rich spectral information, resulting in hyperspectral images that provide detailed and accurate ground object information. This enables a deeper understanding of object and scene characteristics than ever before, making hyperspectral imaging widely used in agriculture [1], pollution monitoring [2], military [3], remote sensing [4], and other fields. To meet the requirements of upstream application scenarios based on hyperspectral images, obtaining high-resolution (HR) hyperspectral images is essential. However, due to environmental factors, the arrangement density of the sensor array, and sensor size limitations, directly obtaining high spatial resolution and high spectral resolution images can be challenging. Therefore, hyperspectral image super-resolution (SR) technology has emerged as a solution to overcome these limitations [5–7].

Image SR is a technique that reconstructs an HR image from a single or multiple low-resolution (LR) images [8]. Since panchromatic images have higher spatial resolution and contain more detailed information, most existing hyperspectral image SR techniques utilize panchromatic, RGB, or multispectral images to assist in the reconstruction process, enhancing the details of LR hyperspectral images. Based on whether auxiliary images are used, hyperspectral image SR methods can be divided into two categories: fusion-based hyperspectral image SR and single hyperspectral image SR. Fusion-based hyperspectral image SR enhances spatial details by integrating the hyperspectral image with other HR images captured in the same scene. Bayesian inference [9], matrix decomposition [10], and sparse representation [11] have demonstrated good performance in recent years for these fusion-based methods. However, these methods often assume that the input LR hyperspectral image and the HR auxiliary image are well-aligned, which can be challenging to achieve in practice [12]. Single hyperspectral image SR methods aim to directly reconstruct HR hyperspectral images from LR counterparts without using auxiliary information. Successfully applying these methods relies on effectively leveraging the non-local self-similarity in space and the strong correlation between spectra. Single hyperspectral image SR can be divided into two main categories: traditional methods and methods based on deep learning. The former approaches often formulate SR as a constrained optimization problem, with regularization provided by priors like low-rank tensors [13], non-local similarity [14], or sparse representation [15]. Despite their widespread use, these methods encounter several limitations due to the high-dimensional nature of hyperspectral data, the complexity of spectral signatures, and the varied and uncertain degradation processes during imaging. Consequently, they struggle to extract representative features from the images effectively, often resulting in distorted reconstructions. Recent advancements in deep learning-based image SR have yielded impressive results [16–19]. For example, Long et al. [20] proposed a dual self-attention swin transformer SR network that utilizes the ability of the shifted windows (swin) transformer in the spatial representation of both global and local features and learns spectral sequence information from adjacent bands of hyperspectral imagery (HSI). Zhao et al. [21] proposed a novel method for HSI SR named attention-driven dual feature guidance net, which makes full use of the spatial–spectral information.

Hyperspectral images, unlike grayscale or RGB counterparts, possess a richer spectral content with more complex band interdependencies. Effectively leveraging the spatial non-local self-similarity and spectral correlation within hyperspectral images is crucial for reconstructing high-quality images. One of the primary challenges in hyperspectral image SR is the abundance of channels, which leads to an increase in model parameters and complexity. Moreover, the limited dataset often results in overfitting. Group convolution has been shown to play a crucial role in reducing computational load and decreasing the likelihood of model overfitting [22]. Inspired by the "spatial-spectral prior for super-resolution (SSPSR)" network structure and the utilization of the spatial-spectral attention module, we propose a new single hyperspectral image SR method based on a progressive upsampling deep prior network. In response to these insights, the present paper delineates the development of a group convolutional upsampling (GCU) module. At the heart of this module lies the spatial-spectral attention fusion (S^{2}AF) unit, which processes spatial and spectral image data in parallel. This unit is proficient in learning patterns of non-local self-similarity within the spatial domain and harnessing the strong correlations across spectral bands, enabling a comprehensive integration of these critical data streams. By implementing overlapping group convolution, our model achieves a reduction in parameter count, thereby alleviating the computational intensity of the training phase. The model further adopts a progressive upsampling scheme, effectively doubling the resolution at each iteration, which is particularly efficacious in the context of high upscaling factors for image reconstruction. The overall architecture of the proposed method is illustrated in Figure 1, and the contributions are enumerated below.

- This paper presents a novel approach for single hyperspectral image SR. Our method leverages the non-local self-similarity in the spatial domain and the significant correlation among spectral bands. By integrating information from different bands, we further enhance the spectral and spatial resolution of the reconstructed image. We have validated the effectiveness of the proposed model through comprehensive testing on multiple datasets.
- We introduce the S^{2}AF module, a novel component that integrates spatial and spectral attention mechanisms to significantly enhance both spatial detail and spectral accuracy. The S^{2}AF module effectively captures important features within the image and meticulously regulates information for each spectral band during reconstruction. By sequentially arranging multiple S^{2}AF modules, we have established the GCU network architecture, which excels at learning subtle textural and edge information, thereby greatly improving the overall quality and detail of the image. The architecture is further optimized by incorporating skip connections, which not only boost spectral fidelity but also simplify the learning process by reducing the complexity of feature extraction. This enables the network to be trained more efficiently and applied effectively to real-world data.
- Given the inherent challenges associated with high magnification factors in hyperspectral image SR, particularly in light of the limited size and high dimensionality of typical hyperspectral image datasets, we applied a progressive upsampling strategy to our network. This strategy incrementally increases image resolution, effectively avoiding the blurring and distortion issues commonly associated with single-step magnification. Ablation studies confirmed the practicality of this approach, demonstrating its ability to better preserve fine textural details compared to traditional single-step upsampling methods at elevated magnification rates. This not only enhances the visual quality of the images but also ensures consistency and accuracy across all spectral bands in the reconstruction results, providing a more robust foundation for the application of hyperspectral images.
- We have introduced a novel loss function that combines the L1 loss function with the GTV loss. Through ablation experiments, we have validated that the loss function we employ can significantly improve the quality of the reconstructed images by the model.

The remainder of this article is organized as follows: In Section 2, we delve into a review of prior works that have laid the groundwork for our study, establishing the relevance and novelty of our approach. Section 3 is dedicated to a detailed exposition of our proposed progressive upsampling deep prior network (PUDPN) model, highlighting its conceptual foundation and architectural specifics. In Section 4, the research narrative advances to the empirical validation of the PUDPN model, where its performance is rigorously evaluated against existing methods through a series of methodically designed experiments. The concluding Section 5 encapsulates the essence of our findings, reflecting on the contributions of our work to the field and proposing directions for future research that could further advance the state-of-the-art methods in hyperspectral image SR.

2. Related works

In this section, we briefly review some relevant works, including fusion-based hyperspectral image SR and single hyperspectral image SR.

2.1. Fusion-based hyperspectral image SR

In the realm of hyperspectral imaging, fusion-based SR techniques strive to elevate the spatial resolution of images by seamlessly blending hyperspectral data with complementary HR images depicting the same scene. These methods have attracted considerable attention due to their remarkable capability to reconstruct the intricate details inherent in HR hyperspectral images. The standard fusion-based approach involves extracting high-frequency spatial information from an HR auxiliary image and subsequently incorporating this information into the desired HR hyperspectral image. For instance, Wei et al. [23] utilized a variational approach to merge HR multispectral images with their corresponding LR hyperspectral counterparts. Yokoya et al. [10] introduced a method rooted in coupled non-negative matrix factorization (CNMF) to derive HR hyperspectral images from a combination of HR multispectral and LR hyperspectral imagery. Wan et al. [7] considered hyperspectral images as three-dimensional tensors and proposed a fusion technique leveraging non-local four-dimensional tensor dictionary learning. Additionally, various methods have been explored, leveraging concepts such as sparsity [6], non-local similarity [24], superpixel-guided self-similarity [25], clustering manifold structures [26], and tensor and low-rank constraints [27, 28], to fully exploit the spectral domain's redundancies and correlations. In recent years, the application of deep learning in fusion-based methods has gained significant momentum [29]. For example, Pan and Shen [30] fused LR multispectral images with HR RGB images, presenting a deep learning algorithm tailored for multispectral image SR. It should be noted that fusion-based hyperspectral image SR methods typically rely on the assumption that the HR auxiliary image aligns well with the LR hyperspectral image. However, obtaining suitable auxiliary images for alignment can pose a challenge in practical applications, thereby limiting the widespread adoption of these methods.

2.2. Single hyperspectral image SR

The challenge of reconstructing HR hyperspectral images from their LR versions, without any external assistance, holds immense practical significance. This problem has traditionally been approached as a constraint optimization task with regularization based on prior knowledge. For example, He et al. [13] used a combination of low-rank tensor modeling and total variation regularization. However, these methods often involve complex optimization procedures and rely on manually defined priors, limiting their application to specific scenarios or datasets.

In recent years, the remarkable progress of deep learning in various fields has led researchers to explore data-driven approaches for single hyperspectral image SR. A significant milestone was achieved by Dong et al. [31] with the introduction of the SRCNN algorithm, which marked the first application of convolutional neural networks in image SR. Subsequently, Liebel and Korner [32] extended this approach to enhance the resolution of individual remote sensing images. Yuan et al. [33] and Xie et al. [5] were pioneers in applying deep convolutional neural networks (DCNNs) to hyperspectral image SR, incorporating non-negative matrix factorization (NMF) to maintain spectral features in intermediate stages.

Advancements in this area have continued with the development of attention mechanisms like the convolutional block attention module (CBAM) by Woo et al. [34], which attends to both spatial and channel dimensions for effective feature extraction. The flexibility of CBAM allows it to be easily integrated into various CNN architectures. Alternatively, Mei et al. [18] proposed a three-dimensional SR network to extract relevant prior information, albeit with higher computational complexity due to three-dimensional convolution. To address this issue, Li et al. [22] introduced the grouped deep residual recursive neural network (GDRRN), which uses grouped convolution strategies to reduce the computational burden associated with high spectral dimensions.

Jiang et al. [19] further optimized this approach by incorporating grouped convolution with parameter sharing to capture spectral band relationships and spatial details while significantly reducing model parameters. In a unique approach, Lempitsky et al. [35] designed a network that leverages the neural network structure itself as prior information, eliminating the need for large datasets during training. Instead, it iteratively reconstructs HR images from LR inputs. Sidorov and Hardeberg [36] combined grouped convolution with a progressive upsampling framework to create a stable and efficient deep network for hyperspectral image SR. Dong et al. [37] utilized 2D convolution for extracting detailed image features at a higher level. Nevertheless, 2D convolution is primarily optimized for spatial feature extraction, making it less effective in handling the spectral dimension of hyperspectral images. Alternatively, 3D convolution [18] is capable of learning contextual information across adjacent bands in hyperspectral data but comes with increased computational demands and training difficulties. Consequently, a combined 2D-3D convolutional approach was introduced [16], demonstrating promising performance despite issues related to parameter explosion. Liu et al. [38] have presented a CNN-based hyperspectral image SR approach, christened the spectral grouping and attention-driven residual dense network (SGARDN), which aims to facilitate the modeling of all spectral bands and focus on the exploration of spatial-spectral features. Wang et al. [39] have introduced a novel group-based single hyperspectral image SR technique, referred to as a group-based embedding learning and integration network (GELIN), which reconstructs HR images in a group-wise manner, thus alleviating the complexity of feature extraction and reconstruction for hyperspectral images. Hou et al. [17] have developed a method, deep posterior distribution-based embedding for hyperspectral image SR (PDE-Net), which formalizes hyperspectral embedding as an approximation of the posterior distribution for a set of meticulously defined hyperspectral embedding events. Liu et al. [40] proposed a dual-domain network based on hybrid convolution (SRDNet) to fully exploit the spatial-spectral and frequency information among the hyperspectral data. Chen et al. [41] developed attention mechanisms that function across spatial and spectral domains to comprehensively model long-range spatial-spectral characteristics. Zhang et al. [42] sought to overcome the challenge of inadequate spectral information utilization, introducing the enhanced spectral self-attention former to mitigate the issue of artifacts post-upsampling. However, the quality of hyperspectral image reconstruction is still not at an ideal state. Despite the promising results achieved by deep learning-based methods, fully exploiting the spectral correlation between adjacent bands remains a challenge due to the high dimensionality of hyperspectral images and limited training data availability.

3. Proposed method

In this section, we will introduce the proposed PUDPN method in detail. First, we will provide the overall architecture of PUDPN. Then, we will introduce the GCU and the S^{2}AF. Finally, we will provide the structures of the spatial attention block (SAB) and the spectral attention block.

3.1. PUDPN network architecture

The proposed network architecture, depicted in Figure 1, aims to learn an end-to-end mapping from LR images to their corresponding HR counterparts. The architecture comprises two primary components: a residual connection-based group convolutional progressive upsampling module and a reconstruction layer. The upsampling module is further segmented into two distinct stages, with each stage incorporating a GCU network designed to enhance the input LR image by a factor of two in magnification. This staged approach to progressive image upsampling mitigates the learning complexity for the model, facilitating superior restoration of intricate image details and texture information.

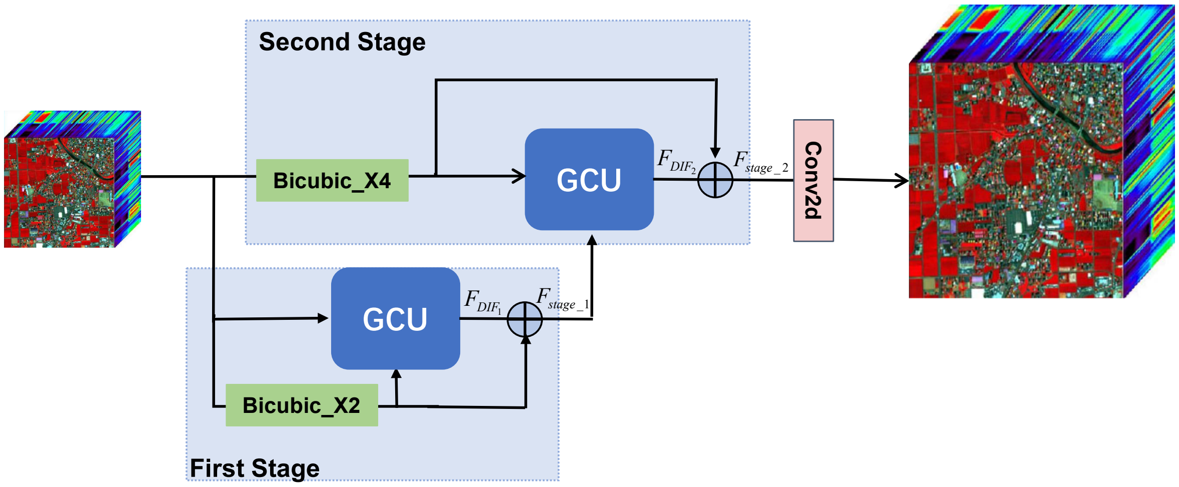

Figure 1. Overall framework of the proposed PUDPN.

In the following, we will use $I_{LR} \in \mathbb{R}^{h \times w \times c}$ to represent the input LR image, and $I_{SR} \in \mathbb{R}^{H \times W \times c}$ to represent the corresponding reconstructed HR image. Among them, $h(H)$ and $w(W)$ denote the height and width of the LR (HR) hyperspectral image, respectively, while c represents the number of spectral bands and r represents the amplification factor. The ground truth of the HR image is represented by $I_{HR} \in \mathbb{R}^{H\times W\times c}$. The degradation process from I_{HR} to I_{LR} can be expressed as

$$I_{LR} = \mathrm{D}(I_{HR}) + n^{*}, \tag{3.1}$$

where $\mathrm{D}(\cdot)$ represents the downsampling operator, and n^{*} represents noise. In this paper, we obtain the LR image by bicubic interpolation. Our goal is to predict the corresponding HR hyperspectral image I_{SR} for a given LR hyperspectral image I_{LR} through our proposed PUDPN method. This mapping function can be expressed by Eq (3.2),

$$I_{SR} = \mathrm{H}_{\mathrm{Net}}(I_{LR}), \tag{3.2}$$

where $\mathrm{H}_{\mathrm{Net}}(\cdot)$ represents our proposed PUDPN method. Our model does not require any auxiliary images, and takes $I_{LR} \in \mathbb{R}^{h\times w\times c}$ as the input to the network. We employ bicubic interpolation to upscale $I_{LR} \in \mathbb{R}^{h\times w\times c}$ by a factor of 2, and together they are fed into the first stage of the GCU modules input. This operation can be represented as

$$F_{DIF_1} = \mathrm{H}_{\mathrm{DIF}_1}(I_{LR}, \mathrm{bicubic}_{\times 2}(I_{LR})), \tag{3.3}$$

where $\mathrm{bicubic}_{\times 2}(\cdot)$ represents the bicubic interpolation with a factor of 2, $\mathrm{H}_{\mathrm{DIF}_1}(\cdot)$ represents the mapping function of the first stage GCU network, and F_{DIF_1} denotes the prior information of the first stage GCU network output. To preserve more details and semantic information from the original image, a residual connection is applied after the GCU network to obtain the output of the first stage, which can be represented as

$$F_{stage_1} = F_{DIF_1} \oplus \mathrm{bicubic}_{\times 2}(I_{LR}), \tag{3.4}$$

where $\oplus$ represents the pointwise addition operation, and F_{stage_1} is the output of the first stage. Subsequently, the output of the first stage and the result of upsampling I_{LR} with a fourfold magnification factor are fed into the second-stage GCU network, and the operation can be expressed as

$$F_{DIF_2} = \mathrm{H}_{\mathrm{DIF}_2}(F_{stage_1}, \mathrm{bicubic}_{\times 4}(I_{LR})), \tag{3.5}$$

where $\mathrm{bicubic}_{\times 4}(\cdot)$ represents the upsampling operation using bicubic interpolation with a factor of 4, $\mathrm{H}_{\mathrm{DIF}_2}(\cdot)$ represents the mapping function of the second-stage GCU network, and F_{DIF_2} represents the prior information of the second-stage GCU network output. Similarly to the first stage structure, a residual connection is applied after the GCU network in the second stage to obtain the output of the second stage, which can be represented as

$$F_{stage_2} = F_{DIF_2} \oplus \mathrm{bicubic}_{\times 4}(I_{LR}), \tag{3.6}$$

where F_{stage_2} is the output of the second stage, which is the output of the group convolutional progressive upsampling module based on residual connections.

The reconstruction layer of the overall structure of the PUDPN method consists of a 3 × 3 convolution. After the image undergoes the operation of the grouped convolution upsampling module, it passes through the reconstruction layer to ensure that the number of channels of the model's output I_{SR} is equal to that of the input I_{LR}. This operation can be represented as

$$I_{SR} = \mathrm{Conv}_{3\times 3}(F_{stage_2}), \tag{3.7}$$

where $\mathrm{Conv}_{3\times3}(\cdot)$ represents the convolution operation with a kernel size of 3×3.

We perform image feature extraction in two stages, where each stage involves upsampling operations with a magnification factor of 2. By incorporating residual connections that facilitate the integration of both deep and shallow features, we mitigate the challenges associated with feature learning for the model. This approach enables effective extraction of image features, enhances the model's expressive capabilities, and ultimately leads to the reconstruction of higher-quality images.

3.2. Grouped convolutional upsampling network (GCU)

In Figure 2, we outline the architecture of the proposed grouped convolutional upsampling network (GCU) for hyperspectral images. Addressing the computational challenges posed by their high dimensionality, we organize the spectral bands into groups, facilitating more efficient feature extraction.

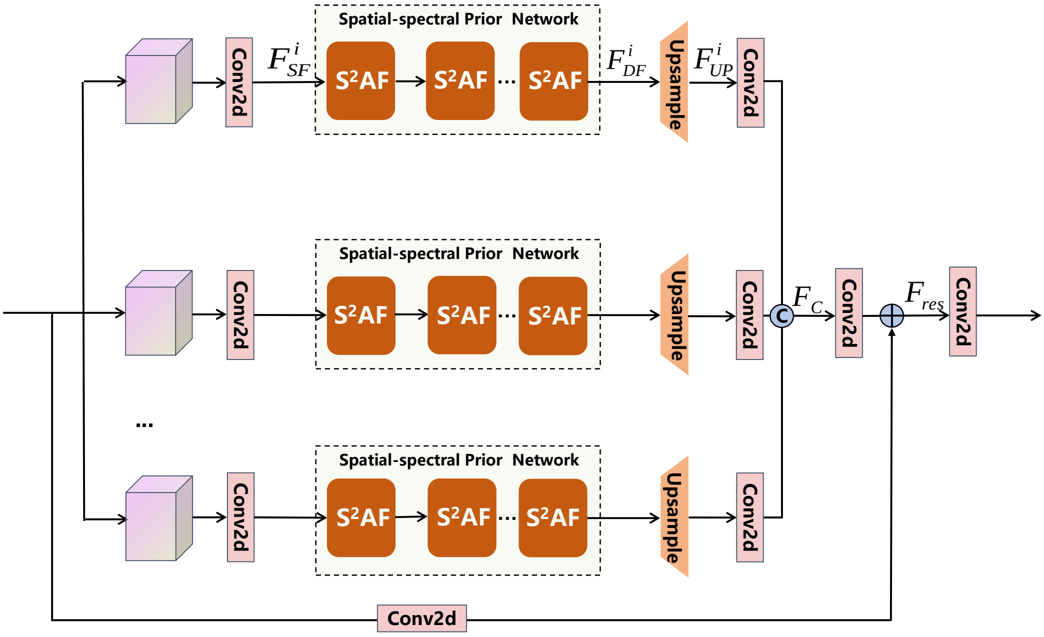
Spatial-spectral Prior Network
Spatial-spectral Prior Network
Spatial-spectral Prior Network
S²AF
Conv2d
Upsample
F^i_{SF}
F^i_{DF}
F^i_{UP}
F_C
F_{res}
C
...

Figure 2. The group convolutional upsampling network (GCU). C represents the concatenation operation along the channel dimension, and ⊕ represents the pixel-wise addition operation.

Grouped convolution can reduce model parameters, as illustrated in Figure 3. Suppose we have m feature maps that need to be expanded to n feature maps through convolution operations. As shown on the left side of the dashed line in Figure 3, without using grouped convolution, the number of convolution kernels required is $m \times n$. As shown on the right side of the dashed line in Figure 3, when using grouped convolution, if k adjacent feature maps are grouped together, resulting in m/k groups (assuming non-overlapping groups for simplicity), the number of convolution kernels required is $(m/k) \times (m/k) \times k$. Since $(m/k) \times (m/k) \times k = mn/k$ is less than $m \times n$, grouped convolution can effectively reduce the number of convolution kernels and thereby decrease the number of parameters and computational cost.

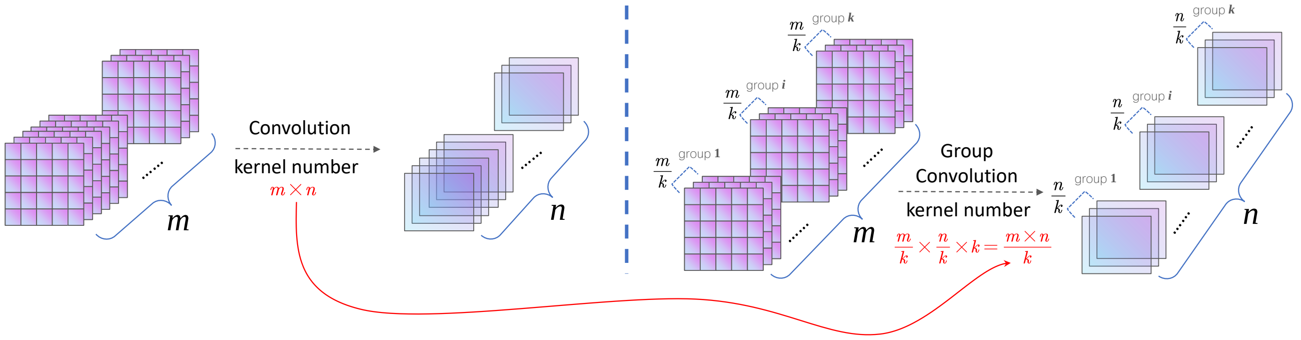

Figure 3. A visual representation of parameter reduction via grouped convolution.

Borrowing from the principles of grouped convolution and the GDRRN [22] model, we adopt a strategy of overlapping groups. When the final group exhibits fewer bands than its predecessors, we supplement it with the initial bands from the hyperspectral image. Each group is then processed using a unified set of operations, leveraging parameter sharing for efficiency. Within each group, a convolutional layer is tasked with extracting shallow features F^i_{SF}. These groups contain both effective N and overlapping u bands, enabling the capture of long-term spectral dependencies, the extraction of correlated band information, and a reduction in convolutional kernels via parameter sharing. The convolutional layer receives input channels commensurate with the group's band count and consistently outputs 256 channels.

$$F^i_{SF} = \mathrm{Conv}(I^i_{LR}), \tag{3.8}$$

where Conv(·) represents the convolution operation, and I^i_{LR} is the feature of the i-th group after grouping. Afterward, a spatial-spectral prior network, which is comprised of M consecutive S^{2}AF modules, is employed to comprehensively extract spatial non-local similarities and the relationship between distinct bands, thereby acquiring the deep feature F^i_{DF} of the hyperspectral image.

$$F^i_{DF} = \mathrm{H}_{\mathrm{S^2AF}}(\mathrm{H}_{\mathrm{S^2AF}}(\cdots \mathrm{H}_{\mathrm{S^2AF}}(F^i_{SF})\cdots)), \tag{3.9}$$

where $\mathrm{H}_{\mathrm{S^2AF}}(\cdot)$ is the S^{2}AF operation that we propose. In the subsequent sections, we will delve into the intricacies of the S^{2}AF and provide more detailed information. Once the features have been fully extracted, we employ the upsampling module to enhance the resolution of the image.

To optimize network performance and refine texture details, we adopt a staged upsampling technique that progressively doubles the image resolution at each step. The resulting upsampled feature of the i-th group, referred to as F^i_{UP}, is formally expressed as

$$F^i_{UP} = \mathrm{H}_{\mathrm{UP}}(F^i_{DF}), \tag{3.10}$$

where $\mathrm{H}_{\mathrm{UP}}(\cdot)$ represents the upsampling part feature mapping function. The architecture of the upsampling module employed in this study is illustrated in Figure 4. It comprises two convolutional layers, a PixelShuffle operation for upscaling, and a subsequent LeakyReLU activation function to boost the model's nonlinear representational capacity. Additionally, dropout is incorporated to mitigate the risk of overfitting. Prior to amalgamating the features from each group into F_C, a convolutional operation is executed to decrease the number of feature maps to the predefined effective spectral band count N, as determined by the grouping configuration.

$$F_C = \text{Concate}(\text{Conv}(F_{UP}^i)), \tag{3.11}$$

where Concate(·) represents the concatenation along the channel dimension. Following the concatenation of distinct feature map groups, they are subjected to a convolutional layer for further computation, thereby augmenting the feature map count to 256. Subsequently, a residual connection is introduced to tackle the problem of channel mismatch. Within this residual connection framework, a convolutional layer is employed to expand the channel dimension to 256.

$$F_{res} = \text{Conv}(F_C) + \text{Conv}(lms), \tag{3.12}$$

where F_{res} represents the feature after the residual connection, lms is the result of the LR image upsampled by bicubic interpolation, and + denotes the pixel-wise addition. Finally, a convolutional layer is utilized to revert the number of feature maps to that of the initial input, yielding the output of the GCU network.

$$F_{GCU} = \text{Conv}(F_{res}), \tag{3.13}$$

where F_{GCU} represents the output result of the GCU network.

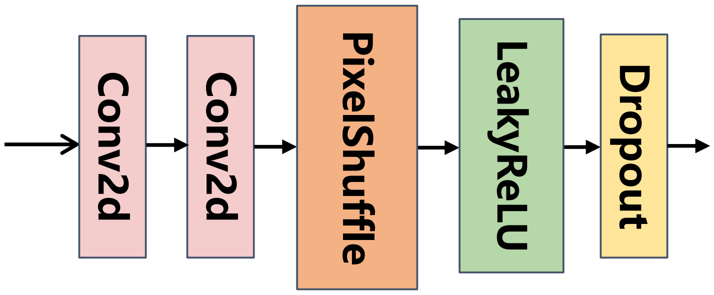

Figure 4. Upsampling module.

3.3. *Spatial-spectral attention fusion module (S²AF)*

A crucial challenge in hyperspectral image reconstruction lies in the efficient extraction of correlation information across diverse bands. Motivated by the performance improvements observed in residual networks, channel attention, and spatial attention within the domain of image SR, we designed the S^{2}AF module based on the spatial-spectral block (SSB) in the SSPSR model, as depicted in Figure 5. The S^{2}AF module is primarily composed of a spatial attention block and a channel attention block (CAB). Notably, as the attention between distinct channels in hyperspectral imagery pertains to spectral attention, we interchangeably refer to the CAB as the spectral attention block. the SAB first guides the model to focus on the important feature regions in the data, ignoring redundant information. This allows the model to concentrate more on the most critical areas. The spectral attention block then guides the model to focus on the essential features in specific spectral bands of the input data. Hyperspectral images contain multiple consecutive bands of data, each corresponding to different spectral information. Spectral attention helps the model better select and utilize the spectral information most useful for the task. To synergistically leverage both spectral and spatial information, we concatenate the results of spatial attention and spectral attention, enabling better handling of hyperspectral image data that contains complex multidimensional features. To allow the spatial and spectral attention modules to fully realize their respective functions, we separately concatenate spatial attention and spectral attention modules afterwards. The design of the S^{2}AF module captures rich feature information while maintaining the modularity of the model.

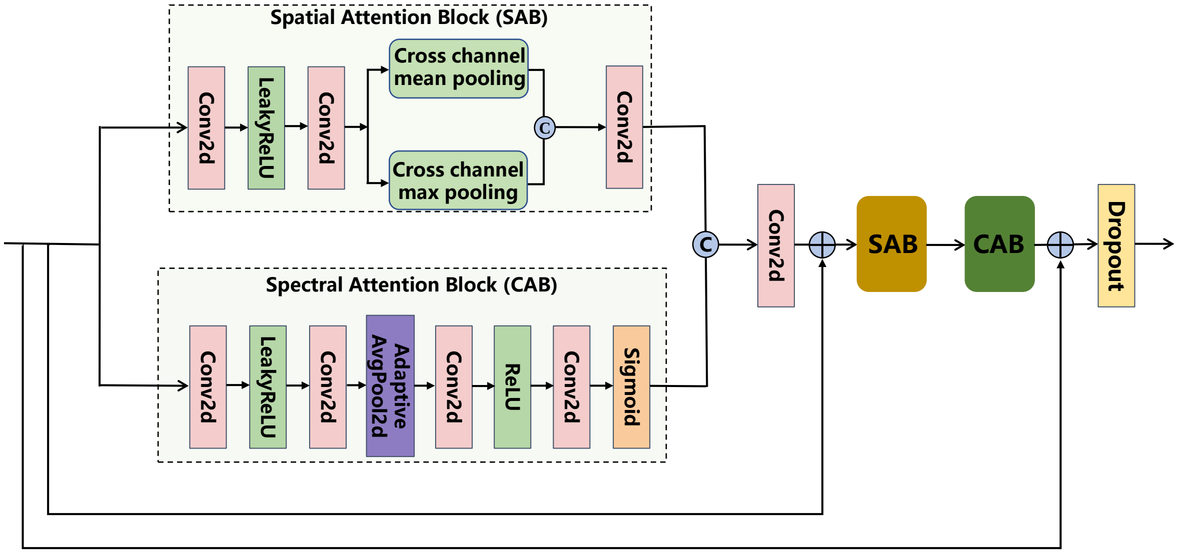
Spatial Attention Block (SAB)
Conv2d
LeakyReLU
Conv2d
Cross channel mean pooling
Cross channel max pooling
C
Conv2d
Spectral Attention Block (CAB)
Conv2d
LeakyReLU
Conv2d
Adaptive AvgPool2d
Conv2d
ReLU
Conv2d
Sigmoid
C
Conv2d
SAB
CAB
Dropout

Figure 5. Spatial-spectral attention fusion unit.

The spatial attention module plays a vital role in SR tasks, as it allows for the adaptive processing of different image regions based on their significance. This selective approach enables a more accurate representation of image details, leading to improved precision and accuracy in the final super-resolved image. In our spatial attention block design, we first extract spatial features using a 3 × 3 convolution kernel. These features are then passed through a LeakyReLU activation function, enhancing the nonlinearity of the network's mapping. Additionally, we employ cross channel mean pooling and cross channel max pooling operations to obtain average and maximum feature maps, respectively. These maps are concatenated along the channel dimension to effectively capture the spatial characteristics of the image. Finally, a convolutional layer with a single output channel is utilized to compute the spatial attention score, which is represented as

$$F_{Spa}^{j} = \mathrm{H}_{\mathrm{Spa}}(F_0), \tag{3.14}$$

where F_{Spa}^{j} represents the output of the first SAB in the j-th S^{2}AF block, $\mathrm{H}_{\mathrm{Spa}}(\cdot)$ represents the mapping function of the SAB, and F_0 is the input feature of the j-th S^{2}AF block. In order to effectively harness the inter-band correlations in hyperspectral images and embed this valuable information into the newly formed spectral bands, the endeavor is akin to learning a set of weight parameters that accurately represent the hyperspectral data. These parameters can be determined by performing a 1 × 1 convolutional operation along the channel dimension, as described in reference [19]. Additionally, we propose the integration of a channel attention mechanism, visually represented in Figure 5. This mechanism commences with adaptive average pooling to condense the input feature maps and extract global channel-wise information. Subsequently, a 1 × 1 convolution layer is employed to reduce dimensionality while simultaneously capturing the correlational patterns between adjacent spectral bands. The introduction of the ReLU activation function serves to introduce nonlinearity, followed by a second 1 × 1 convolution layer aimed at restoring the original input dimension. Conclusively, a sigmoid activation function layer computes a weighted vector for each channel, facilitating channel-wise attention. The CAB is diagrammatically presented in Figure 5 and mathematically formalized as follows,

$$F_{Spe}^{j} = \mathrm{H}_{\mathrm{Spe}}(F_0), \tag{3.15}$$

where F_{Spe}^{j} represents the output of the first CAB module in the j-th S^{2}AF block, and $\mathrm{H}_{\mathrm{Spe}}(\cdot)$ represents the mapping function of the CAB module. Following the channel attention module's processing, the resultant features acquire channel attention scores. These are fused with spatial attention scores, and the amalgamated data is processed through convolutional layers to derive spatial-spectral attention fusion scores. By leveraging residual connections, these fusion scores are added back to the S^{2}AF block's input data, enabling the execution of spatial-spectral attention operations on the feature map. To preclude the attenuation of either spatial or spectral attention, we sequentially introduce SAB and CAB via residual connections. Finally, we use dropout to prevent overfitting. The overall operation of the S^{2}AF module can be represented as

$$F_{S^2AF}^{j} = \mathrm{H}_{\mathrm{S^2AF}}(F_0), \tag{3.16}$$

where $F_{S^2AF}^{j}$ represents the output feature of the j-th S^{2}AF block, and $\mathrm{H}_{\mathrm{S^2AF}}(\cdot)$ is the mapping function of the S^{2}AF block.

Through the implementation of the S^{2}AF block, our methodology achieves a harmonious fusion of spatial and channel attentions, effectively preserving feature robustness. The block is specifically engineered to extract non-local self-similarities across spatial dimensions and to harness the inherent correlations among spectral channels in hyperspectral images. This dual extraction capability ensures an enriched feature representation, pivotal for the advanced processing and analysis of hyperspectral imagery.

3.4. Loss function

The efficacy of an SR model is significantly influenced by the choice of loss function, which not only evaluates model quality but also dictates the optimization trajectory. Various loss functions can lead the model to prioritize different aspects of the data. The L1 loss is particularly effective in penalizing slight inaccuracies, a feature crucial for the precision required in SR tasks. It also excels in preserving image edges, a key factor for maintaining visual quality. Its lower computational complexity and memory footprint, coupled with notable training stability, make it an optimal choice for SR. Consequently, this paper adopts the L1 loss for the proposed SR method, highlighting its contributions to improving model efficiency and training stability. The definition of L1 loss is below:

$$L_1(\theta) = \frac{1}{N}\sum_{n=1}^{N} ||I_{HR} - I_{SR}||_1, \tag{3.17}$$

where θ is the parameter of the network, N is the batch size, and n is the index of the reconstructed image.

Drawing inspiration from TVLoss, we introduce GTVLoss, a metric that assesses the gradient differences between the predicted SR image and the ground truth in both height and width directions. By computing the mean of the squared sum of these differences, GTVLoss excels in conserving fine image details and texture information. Moreover, its computational efficiency is notable, requiring only one squaring and one division operation, thus offering an advantage over MSE loss. The specific formula for calculating GTVLoss is detailed as follows:

$$GTV = \frac{1}{N}(\frac{\sum_{h=1}^{H}(\nabla I_{SR_H} - \nabla I_{HR_H})^2}{H} + \frac{\sum_{w=1}^{W}(\nabla I_{SR_W} - \nabla I_{HR_W})^2}{W}), \tag{3.18}$$

where ∇I_{SR_H} represents the gradient of each pixel in the height direction predicted by the model, ∇I_{HR_H} represents the gradient of each pixel in the height direction in the ground truth, ∇I_{SR_W} represents the gradient of each pixel in the width direction predicted by the model, and ∇I_{HR_W} represents the gradient of each pixel in the width direction in the ground truth. H , W and N represent the height, width, and batch size of the HR image, respectively.

In summary, the total loss function in this paper is defined as

$$L = L_1 + \alpha L_{GTV}, \tag{3.19}$$

where α represents the parameter used to balance the contributions of different losses, and we set the value of α to 1e-3.

4. Experimental results and analysis

In this section, we commence by presenting an overview of the three benchmark datasets employed in our study. Subsequently, we delve into the specifics of the experimental configurations. Ultimately, we offer a comprehensive assessment of our model's efficacy through both quantitative and qualitative comparisons with five other competing methodologies. For the quantitative analysis, we have adopted the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as evaluation metrics. The PSNR values mentioned later are all measured in dB. Notably, for PSNR and SSIM, we have computed their respective average values across all spectral bands.

4.1. Dataset

In this article, we employ three publicly available datasets: Pavia_Center [43], CAVE [44], and Chikusei [45]. Specifically, both the Chikusei and Pavia_Center datasets pertain to remote sensing hyperspectral imagery, whereas the CAVE dataset falls under the category of natural hyperspectral images.

- Pavia_Center: The Pavia_Center dataset is a hyperspectral image dataset obtained by the ROSIS sensor. It is often used for hyperspectral image SR, remote sensing image classification, and target detection tasks. The dataset contains one image with a raw size of 1096 × 1096 pixels. After removing the noise bands, the number of spectral bands is 102. After discarding the central part of the image that does not contain information, the image size becomes 1096 × 715 pixels. Therefore, after processing, the image size is 1096 × 1096 × 102.
- CAVE: The CAVE dataset was captured using a cooled CCD camera in the 400–700 nm wavelength range with a 10nm interval, and consists of 32 images. One of the images is an RGB image with a spatial size of 512 × 512 pixels, and the remaining 31 images are hyperspectral images. The dataset mainly contains images of toys, clothes, paints, food, and faces. Since the CAVE dataset is a publicly available dataset, ethical approval and informed consent were obtained by the original creators of the dataset, ensuring compliance with relevant ethical guidelines and standards.
- Chikusei: The Chikusei dataset was acquired using the Hyperspec-VNIR-CIRIS spectrometer, and is a hyperspectral image of the Chikusei area in Ibaraki Prefecture, Japan. The ground sampling distance is 2.5 meters, and the spectral range is from 363 nm to 1018 nm. A total of 128 bands were collected, and the image size is 2517 × 2335.

4.2. Experimental details

For the Pavia_Center dataset, we cropped the left portion (384 × 715 × 102) to serve as the test and validation sets. Specifically, the leftmost section (256 × 715 × 102) was used to extract test image patches of the size 256 × 256, while the remaining area was divided into 64 × 64 validation image patches. It is important to note that the validation and test sets were non-overlapping during this partitioning. The remaining data constituted the training set, with all training samples being further divided into overlapping blocks of size 64 × 64. In cases where the upsampling factor was set to 4, a 32-pixel overlap was introduced.

In the case of the CAVE dataset, a random selection process was utilized to allocate 80% of the samples to form the training set. The remaining 20% of the dataset was then equally divided, with 10% allocated to the test set and another 10% to the validation set, ensuring a balanced distribution for comprehensive model evaluation.

For the Chikusei dataset, the original image size was 2517 × 2335 × 128. We first cropped the central region of the image, resulting in a size of 2304 × 2048 × 128. We selected an area of 2304 × 384 × 218 pixels on the left side of the image and divided it into the test set and validation set. Half of this area was generally used to create 256 × 256 test samples, and the other half was used to create 64 × 64 validation samples. The validation and test sets did not overlap when cropped. The remaining data was used as the training set and was divided into 64 × 64 overlapping patches as training samples. When the upsampling factor was set to 4, there was an overlap of 32 pixels.

Subsequent to the preparation of HR training samples, a bicubic interpolation technique was employed to generate their LR equivalents, each with dimensions of 32 × 32 pixels. To augment the dataset and ensure the development of a robust model, data augmentation was performed. This involved the application of rotation and mirroring transformations to the training samples, effectively quadrupling the size of the dataset. Such augmentation is critical for improving the model's generalization capabilities across varied imaging conditions. In the methodology employed for the grouping process, we specified several key parameters to optimize performance. The number of effective spectral bands N was set to 17, with the overlapping band channel u determined to be 4. To mitigate overfitting, a *dropout rate* of 0.3 was implemented, alongside a residual *weight res_scale* of 0.1 to control the contribution of residual learning. The Pavia_Center dataset served as the basis for both training and evaluation. The cascading of different numbers of S^{2}AF blocks affects the model's performance. Here, we denote the number of cascading S^{2}AF blocks as M. As detailed in Table 1, an increase in M correlates with a rise in the parameter count. Optimal performance was observed when $M = 3$, yielding a PSNR of 32.2196 dB, despite a slight decrease in the SSIM by 0.0074 compared to $M = 1$. This balance between parameter efficiency and model efficacy led to the selection of $M = 3$ for our configuration. We use ADAM as the optimizer for our network, with parameters $\beta_1 = 0.9$, $\beta_1 = 0.99$, and *weight_decay* set to 1e-5. In the proposed model, a strategic warm-up training approach is employed to facilitate optimal parameter initialization and enhance convergence speed. Initially, the model undergoes a preliminary training phase for two epochs with a reduced *learning rate* of 1e-5. This warm-up phase is designed to guide the model toward a favorable parameter space, effectively circumventing potential local minima that could impede learning efficiency. Subsequently, the *learning rate* is elevated to 1e-4 for the formal training process. To ensure sustained learning progress and adaptability, the *learning rate* is further adjusted by a factor of 0.1 every 30 epochs. The implementation of our experiments is conducted within the PyTorch framework, leveraging the computational capabilities of an NVIDIA RTX A6000 GPU, which boasts a memory capacity of 47.5 GB.

Table 1. Metrics under different numbers of S^{2}AF modules.

MODELS	d	PSNR ↑	SSIM ↑	PARAMETER
Ours-M = 1	4	32.1058	0.8556	16,200,702
Ours-M = 2	4	31.9897	0.8482	17,101,080
Ours-M = 3	4	32.2196	0.8542	21,201,124
Ours-M = 4	4	32.0452	0.8530	31,201,968

Notes: The red font indicates the best metric. ↑ represents that a larger value indicates better performance.

4.3. Ablation experiments

The proposed PUDPN method is characterized by the integration of a S^{2}AF unit alongside a progressive upsampling strategy. To empirically substantiate the effectiveness of these innovations, ablation studies were conducted. These experiments were designed to compare models equipped with these modules against their counterparts lacking such components. The objective was to quantitatively demonstrate the enhancements these modules impart on the model's performance. The dataset selected for training comprised images from the Pavia_Center dataset, while the evaluation of model performance was based on test images measuring 256 × 256 pixels, also derived from the Pavia_Center dataset.

1) To substantiate the utility of the S^{2}AF, we introduced a model variant in which the S^{2}AF was replaced by three 3 × 3 convolutional layers. The performance evaluation, summarized in Table 2, compares our original approach ("Ours") with its S^{2}AF-absent counterpart ("Ours-w/o S^{2}AF") on test images scaled up by a factor of 4. The results reveal that the absence of the S^{2}AF module leads to a

noticeable deterioration in performance metrics, including a reduction in SSIM and PSNR scores across different test image sizes relative to the "Ours-w/o S^2AF" method. By incorporating the S^2AF module, our method effectively fuses spatial and spectral information from multiple dimensions, leveraging the inherent richness of hyperspectral images. This leads to consistently better reconstruction results across varying spatial sizes of test images, thereby validating the efficacy of the S^2AF module.

2) Table 3 delineates the comparative performance analysis between our proposed method ("Ours") and its variant devoid of the progressive upsampling strategy ("Ours-w/o PU"), across test images of varying dimensions, all subjected to a magnification factor of four. Notably, for test images sized 64 × 64, our method demonstrates superior performance in terms of the PSNR and SSIM. In addition, for images sized 128 × 128, our methodology unequivocally surpasses the variant across all metrics. While the PSNR for images sized 256 × 256 slightly favors the variant method by 0.0272 dB, our approach maintains an edge in SSIM metrics. These findings collectively validate that the progressive upsampling strategy we employed can enhance the model's performance to a certain extent.

3) In Table 4, we report the reconstruction results of ablation experiments on test samples with different target sizes from the Pavia_Center dataset, at a 4x magnification factor, using different loss functions. It is evident that incorporating GTV loss on top of L1 loss can lead to better outcomes.

Table 2. Ablation experiment results for the spatial-spectral attention fusion unit (S^2AF).

MODELS	*d*	HR_image_size	PSNR ↑	SSIM ↑
Ours	4	64 × 64	33.9523	0.8438
Ours-w/o S^2AF	4	64 × 64	33.8413	0.8381
Ours	4	128 × 128	33.2196	0.8542
Ours-w/o S^2AF	4	128 × 128	31.9682	0.8462
Ours	4	256 × 256	30.2815	0.8365
Ours-w/o S^2AF	4	256 × 256	30.1522	0.8277

Table 3. Ablation experiment results for the progressive upsampling strategy.

MODELS	*d*	HR_image_size	PSNR ↑	SSIM ↑
Ours	4	64 × 64	33.9523	0.8438
Ours-w/o PU	4	64 × 64	33.8657	0.8432
Ours	4	128 × 128	33.2196	0.8542
Ours-w/o PU	4	128 × 128	32.0811	0.8509
Ours	4	256 × 256	30.2815	0.8365
Ours-w/o PU	4	256 × 256	30.3087	0.8355

Table 4. Ablation experiment results for the loss functions.

LOSSES	d	HR_image_size	PSNR ↑	SSIM ↑
L1	4	64 × 64	33.8401	0.8114
GTVLoss	4	64 × 64	28.8261	0.6841
Ours	4	64 × 64	33.9523	0.8438
L1	4	128 × 128	32.1175	0.8150
GTVLoss	4	128 × 128	28.8782	0.6150
Ours	4	128 × 128	32.2196	0.8542
L1	4	256 × 256	30.1552	0.8233
GTVLoss	4	256 × 256	27.8743	0.6521
Ours	4	256 × 256	30.2815	0.8366

4.4. Comparison with five other representative methods

To validate the performance of our proposed PUDPN method, we conducted experiments on three public datasets and compared it with five other representative methods: PDE-Net [17], SSPSR [19], GDRRN [22], Deep_hs_Prior [36], and Bicubic [46].

1) On the Pavia_Center dataset, as shown in Table 5, with a test image size of 256 × 256 and a magnification factor of 4, our method outperformed the others in both metrics, with the best results emphasized in red. The second-best outcomes are highlighted in blue. PDE-Net formulates hyperspectral embedding as an approximation of the posterior distribution of a set of carefully defined hyperspectral embedding events. The approximated posterior distribution helps the model capture the underlying data distribution more effectively, thus PDE-Net also achieves good performance. Diverging from SSPSR, our approach not only integrates spatial and spectral attention blocks but also optimizes the application of these attentions to mitigate the potential dilution of focus on either spatial or spectral details. This refined extraction and application of spatial and spectral attention information underpin the superior performance of our method. Furthermore, for a more illustrative comparison, we visualized the SR results of different comparison methods on three datasets and presented the corresponding absolute error maps.

Table 5. Quantitative comparison of different methods on the Pavia_Center dataset. The best results are highlighted in red font, while the second-best results are highlighted in blue font.

MODELS	d	HR_image_size	PSNR ↑	SSIM ↑
Bicubic	4	256 × 256	29.3110	0.7948
GDRRN	4	256 × 256	29.6292	0.8018
Deep_hs_Prior	4	256 × 256	29.4749	0.8164
SSPSR	4	256 × 256	30.1413	0.8306
PDE-Net	4	256 × 256	29.2435	0.8057
PUDPN (Ours)	4	256 × 256	30.2815	0.8366

Figure 6 elucidates the comparative performance of various SR models on the Pavia_Center dataset, including ground truth, bicubic interpolation, Deep_hs_Prior, GDRRN, SSPSR, PDE-Net, and our proposed PUDPN model. The sequence of images in the first row of each figure delineates the SR outcomes for a specific spectral band, arranged in the order mentioned. Subsequently, the second row in each figure reveals the absolute error maps corresponding to the SR results displayed above. These error maps, serving as a visual gauge, quantify the pixel-level reconstruction accuracy, thereby offering a nuanced view of the spatial fidelity achieved by each model. Notably, regions of lower error are depicted in cooler hues, facilitating an intuitive assessment of spatial reconstruction quality.

A detailed visual analysis demonstrates the exceptional reconstruction capability of our PUDPN model. This is highlighted in the region outlined by the red rectangle in the figure. The yellow rectangle, in turn, shows an enlarged view of this area, the performance of the Deep_hs_Prior model, which is predicated on leveraging the inherent prior knowledge embedded within the network's structure for image reconstruction, was found wanting. This shortfall is primarily attributed to the indiscriminate application of uniform iteration parameters across disparate data samples, a practice that likely undermines the adaptability and, consequently, the efficacy of the model.

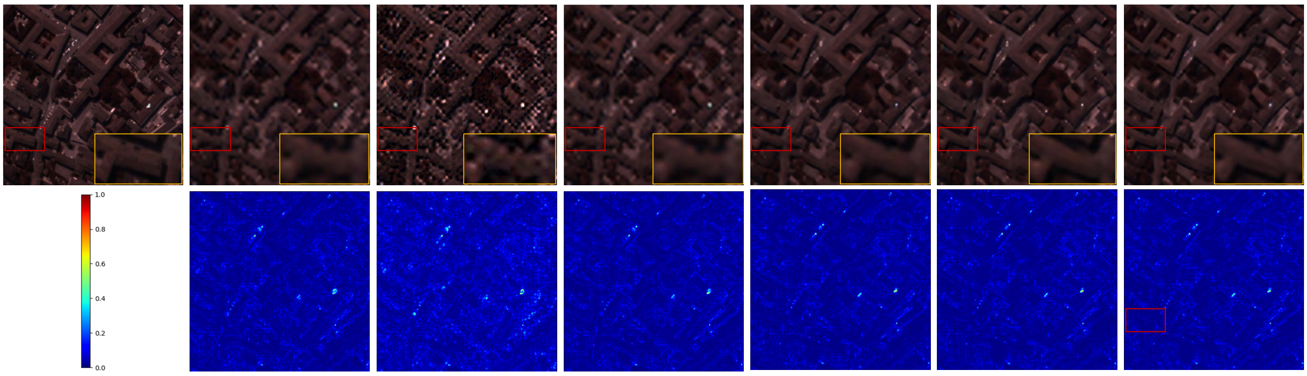

Figure 6. In the Pavia_Center dataset, with an upsampling factor of 4 and selecting the 32-21-11 bands as R-G-B, visualizations of pseudo-colored images of different models' reconstructed images (first row) and their corresponding error maps (second row) are shown. From left to right, they are: ground truth, bicubic [46], Deep_hs_Prior [36], GDRRN [22], SSPSR [19], PDE-Net [17], and PUDPN (ours).

2) Table 6 presents a comparative analysis of the quantitative results achieved by six SR methods applied to the CAVE dataset, utilizing an upsampling factor of $d = 4$ and evaluating performance on images resized to 256 × 256 pixels. PDE-Net exhibits the best overall performance among the evaluated methods, while our proposed PUDPN method achieves the second-best performance. Although our PUDPN model slightly lags behind PDE-Net in terms of PSNR and SSIM metrics, the visualizations in Figures 7–9 and the error maps reveal that the reconstructed images still outperform PDE-Net in certain regions. Particularly in the regions highlighted by the red rectangles, the superior reconstruction accuracy of PUDPN is evident. Furthermore, although PDE-Net leads by 0.4 dB in PSNR and by 0.005 in SSIM on the CAVE dataset, our proposed PUDPN method surpasses PDE-Net by 1 dB in PSNR and by 0.03 in SSIM on the Pavia_Center dataset. Furthermore, on the subsequent Chikusei dataset, PDE-Net continues to lag behind our PUDPN method in both metrics. Overall, our method remains superior to PDE-Net.

Table 6. Quantitative comparison of different methods on the CAVE dataset.

MODELS	d	HR_image_size	PSNR ↑	SSIM ↑
Bicubic	4	256 × 256	33.4488	0.9355
GDRRN	4	256 × 256	31.5601	0.8832
Deep_hs_Prior	4	256 × 256	31.0288	0.9064
SSPSR	4	256 × 256	35.5034	0.9512
PDE-Net	4	256 × 256	36.2361	0.9571
PUDPN (Ours)	4	256 × 256	35.8857	0.9525

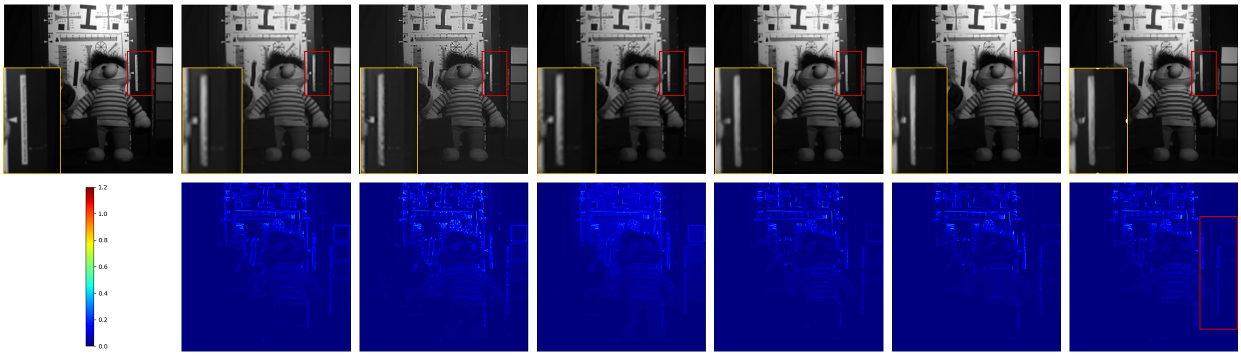

Figure 7. In the CAVE dataset with an upsampling factor of 4, a test image with content of toys was selected. The 30th spectral band of the reconstructed images by different models was visualized (first row), along with the corresponding reconstruction error maps (second row). The models are arranged from left to right as follows: ground truth, bicubic [46], Deep_hs_Prior [36], GDRRN [22], SSPSR [19], PDE-Net [17], and PUDPN (ours).

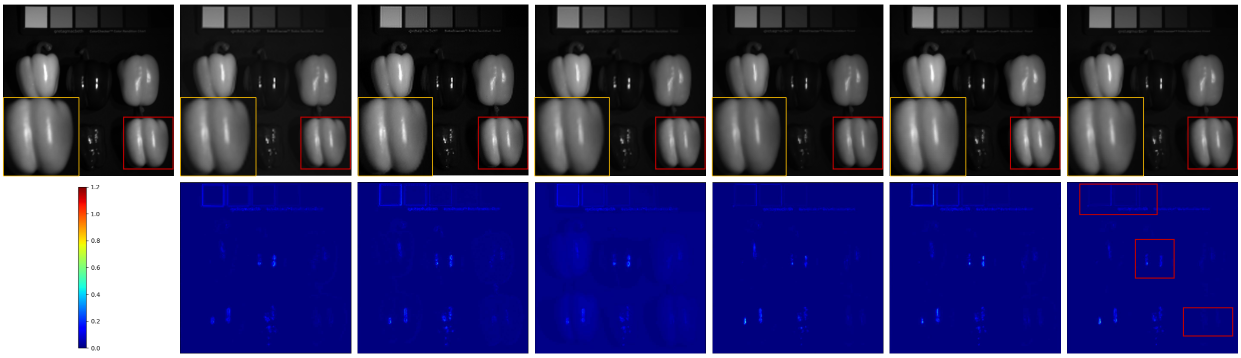

Figure 8. In the CAVE dataset with an upsampling factor of 4, a test image with content of peppers was selected. The 30th spectral band of the reconstructed images by different models was visualized (first row), along with the corresponding reconstruction error maps (second row). The models are arranged from left to right as follows: ground truth, bicubic [46], Deep_hs_Prior [36], GDRRN [22], SSPSR [19], PDE-Net [17], and PUDPN (ours).

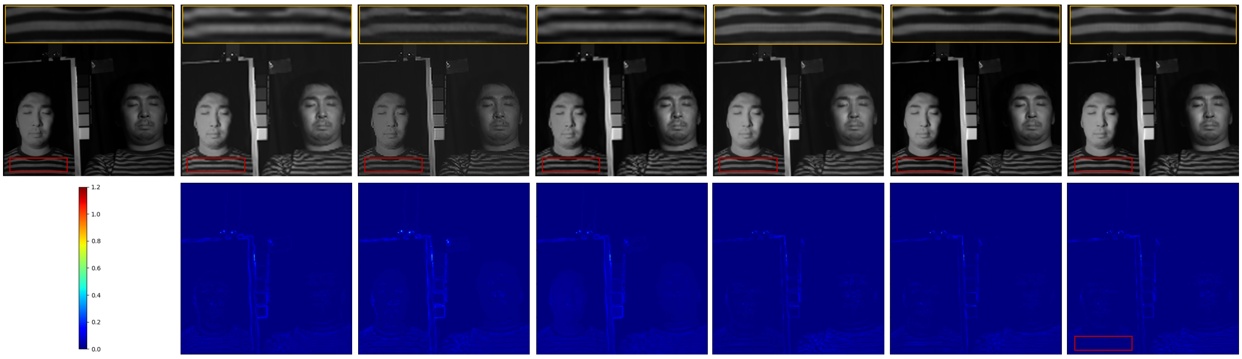

Figure 9. In the CAVE dataset with an upsampling factor of 4, a test image with content of faces was selected. The 30th spectral band of the reconstructed images by different models was visualized (first row), along with the corresponding reconstruction error maps (second row). The models are arranged from left to right as follows: ground truth, bicubic [46], Deep_hs_Prior [36], GDRRN [22], SSPSR [19], PDE-Net [17], and PUDPN (ours).

3) Table 7 presents a comparative study of five SR techniques on the Chikusei dataset, with an emphasis on an upsampling factor of $d = 4$ and an evaluation image size of 256 × 256. Our PUDPN method emerges as the frontrunner, achieving the highest PSNR value, 0.1184 dB above the SSPSR, the next best method. The reconstructed images and their error maps, depicted in Figure 10, facilitate a direct comparison between the methods. Notably, the PUDPN, PDE-Net, and SSPSR methods, while not capturing significant contour details, excel in reconstructing specific areas, indicating their superior performance over alternative models. A comparison between the Chikusei and Pavia_Center datasets suggests a superior reconstruction performance on the former. This is likely due to the Chikusei dataset's larger size and its abundance of uniform areas, which enhances model training efficiency. The PUDPN method, in particular, showcases the best recovery results in areas marked by red rectangles, affirming its leading position in SR technology.

Table 7. A Quantitative comparison of different methods on the Chikusei dataset, with the best results highlighted in red font.

MODELS	d	HR_image_size	PSNR ↑	SSIM ↑
Bicubic	4	256 × 256	37.4098	0.9131
GDRRN	4	256 × 256	37.9635	0.9123
Deep_hs_Prior	4	256 × 256	38.1485	0.9233
SSPSR	4	256 × 256	38.3621	0.9232
PDE-Net	4	256 × 256	37.6435	0.9242
PUDPN (Ours)	4	256 × 256	38.4805	0.9246

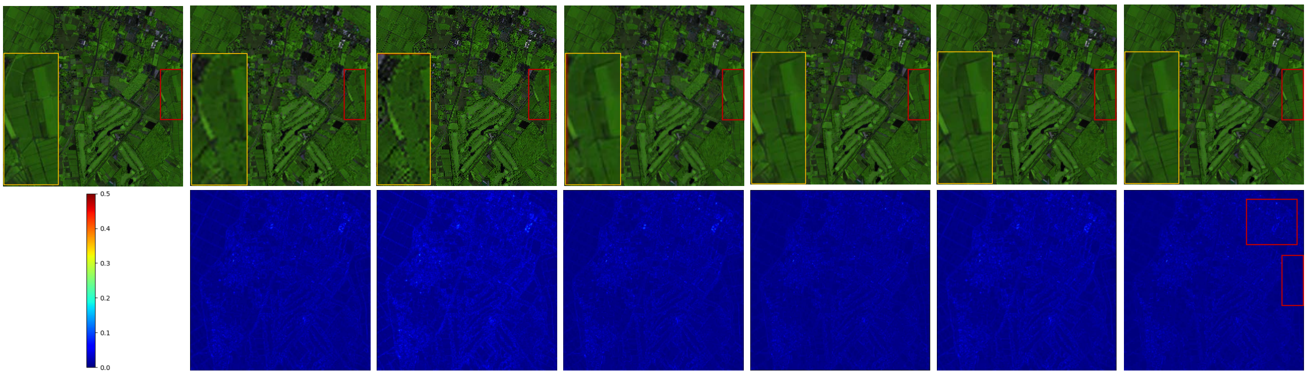

Figure 10. In the Chikusei dataset, with an upsampling factor of 4 and selecting the 70-100-36 bands as R-G-B, visualizations of pseudo-colored images of different models' reconstructed images (first row) and their corresponding error maps (second row) are shown. From left to right, they are: ground truth, bicubic [46], Deep_hs_Prior [36], GDRRN [22], SSPSR [19], PDE-Net [17], and PUDPN (ours).

5. Conclusions

In this paper, we propose a novel framework for hyperspectral image SR, termed the progressive upsampling deep prior network (PUDPN). Central to our approach is the integration of a S^2AF module, aimed at exploiting the rich spatial and spectral information embedded within hyperspectral images. The PUDPN framework addresses critical challenges in the field, including high-dimensionality, limited availability of training samples, and the need for significant upsampling ratios. To this end, we have developed a group convolutional upsampling (GCU) network that incorporates a parameter-sharing, progressive upsampling strategy, enhanced by residual connections to preserve both high- and low-level image details. Comparative analyses on three widely recognized datasets demonstrate the superiority of PUDPN over five benchmark methods in terms of both quantitative and qualitative metrics. Additionally, ablation studies provide further insights into the efficacy of the proposed framework.

Despite the excellent performance of the proposed PUDPN on specific datasets, it faces challenges in cross-dataset applications due to inconsistencies in the number of bands and significant differences in image features. Additionally, due to the limitations of convolution operations in capturing global information, the model's ability to capture global features needs to be further enhanced.

Given the unique encoder-decoder structure of the U-Net architecture, which facilitates multi-scale feature extraction and integration, enabling better capture of global and local information in images and enhancing the quality of super-resolved images, we plan to explore integrating the sliding window mechanism from the swin transformer [47] into the U-Net architecture. This integration aims to enhance the model's ability to extract global features, more effectively utilize the spatial and spectral characteristics of deep images, and further improve image reconstruction and model generalization through network structure optimization and feature learning enhancement. Additionally, we will investigate effective transfer learning strategies to overcome dataset discrepancies and enhance the model's versatility and stability.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part the National Natural Science Foundation of China under Grant 12071112, in part by the Open Project of National Earth Observation Data Center under Grant NODAOP2023005, in part by the Training Plan for Young Key Teachers in Colleges and Universities of Henan Province under Grant 2020GGJS247, in part by the Major Science and Technology Projects of Longmen Laboratory under Grant 231100220400, in part by the Science and Technology Research Project of Henan Province under Grants 222102210053, 222102210127, 232102210057, and 232102210111, in part by the Key Scientific Research Projects of Colleges and Universities in Henan Province under Grant 22A120006, in part by the Ministry of Education Industry-Education Cooperation Collaborative Education Project under Grant 202101289003, and in part by the Henan University of Science and Technology Higher Education Teaching Reform Research and Practice Project under Grant 2024SZ05.

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. B. Lu, P. D. Dao, J. Liu, Y. He, J. Shang, Recent advances of hyperspectral imaging technology and applications in agriculture, *Remote Sens.*, **12** (2020), 2659. https://doi.org/10.3390/rs12162659

2. B. P. Banerjee, S. Raval, P. J. Cullen, UAV-hyperspectral imaging of spectrally complex environments, *Int. J. Remote Sens.*, **41** (2020), 4136–4159. https://doi.org/10.1080/01431161.2020.1714771

3. M. Shimoni, R. Haelterman, C. Perneel, Hyperspectral imaging for military and security applications: combining myriad processing and sensing techniques, *IEEE Geosci. Remote Sens. Mag.*, **7** (2019), 101–117. https://doi.org/10.1109/MGRS.2019.2902525

4. J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, *IEEE Geosci. Remote Sens. Mag.*, **1** (2013), 6–36. https://doi.org/10.1109/MGRS.2013.2244672

5. W. Xie, X. Jia, Y. Li, J. Lei, Hyperspectral image super-resolution using deep feature matrix factorization, *IEEE Trans. Image Process.*, **57** (2019), 6055–6067. https://doi.org/10.1109/TGRS.2019.2904108

6. W. Dong, F. Fu, G. Shi, X. Gao, J. Wu, G. Li, et al., Hyperspectral image super-resolution via non-negative structured sparse representation, *IEEE Trans. Image Process.*, **25** (2016), 2337–2352. https://doi.org/10.1109/TIP.2016.2542360

7. W. Wan, W. Guo, H. Huang, J. Liu, Nonnegative and nonlocal sparse tensor factorization-based hyperspectral image super-resolution, *IEEE Trans. Geosci. Remote Sensing*, **58** (2020), 8384–8394. https://doi.org/10.1109/TGRS.2020.2987530

8. S. C. Park, M. K. Park, M. G. Kang, Super-resolution image reconstruction: a technical overview, *IEEE Signal Process. Mag.*, **20** (2003), 21–36. https://doi.org/10.1109/MSP.2003.1203207

9. Q. Wei, N. Dobigeon, J. Y. Tourneret, Bayesian fusion of multiband images, *IEEE J. Sel. Top. Signal Process.*, **9** (2015), 1117–1127. https://doi.org/10.1109/JSTSP.2015.2407855

10. N. Yokoya, T. Yairi, A. Iwasaki, Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion, *IEEE Trans. Geosci. Remote Sensing*, **50** (2011), 528–537. https://doi.org/10.1109/TGRS.2011.2161320

11. N. Akhtar, F. Shafait, A. Mian, Sparse spatio-spectral representation for hyperspectral image super-resolution, in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, (2014), 63–78. https://doi.org/10.1007/978-3-319-10584-0_5

12. Y. Zhou, A. Rangarajan, P. D. Gader, An integrated approach to registration and fusion of hyperspectral and multispectral images, *IEEE Trans. Geosci. Remote Sensing*, **58** (2020), 3020–3033. https://doi.org/10.1109/TGRS.2019.2946803

13. S. He, H. Zhou, Y. Wang, W. Cao, Z. Han, Super-resolution reconstruction of hyperspectral images via low rank tensor modeling and total variation regularization, in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, (2016), 6962–6965. https://doi.org/10.1109/IGARSS.2016.7730816

14. R. Dian, L. Fang, S. Li, Hyperspectral image super-resolution via non-local sparse tensor factorization, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2017), 3862–3871. https://doi.org/10.1109/CVPR.2017.411

15. H. Huang, J. Yu, W. Sun, Super-resolution mapping via multi-dictionary based sparse representation, in *2014 IEEE International Conference on Acoustics, Speech Signal Processing (ICASSP)*, IEEE, (2014), 3523–3527. https://doi.org/10.1109/ICASSP.2014.6854256

16. Q. Li, Q. Wang, X. Li, Mixed 2D/3D convolutional network for hyperspectral image super-resolution, *Remote Sens.*, **12** (2020), 1660. https://doi.org/10.3390/rs12101660

17. J. Hou, Z. Zhu, J. Hou, H. Zeng, J. Wu, J. Zhou, Deep posterior distribution-based embedding for hyperspectral image super-resolution, *IEEE Trans. Image Process.*, **31** (2022), 5720–5732. https://doi.org/10.1109/TIP.2022.3201478

18. S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, Q. Du, Hyperspectral image spatial super-resolution via 3d full convolutional neural network, *Remote Sens.*, **9** (2017), 1139. https://doi.org/10.3390/rs9111139

19. J. Jiang, H. Sun, X. Liu, J. Ma, Learning spatial-spectral prior for super-resolution of hyperspectral imagery, *IEEE Trans. Comput. Imaging*, **6** (2020), 1082–1096. https://doi.org/10.1109/TCI.2020.2996075

20. Y. Long, X. Wang, M. Xu, S. Zhang, S. Jiang, S. Jia, Dual self-attention swin transformer for hyperspectral image super-resolution, *IEEE Trans. Geosci. Remote Sensing*, **61** (2023), 5512012. https://doi.org/10.1109/TGRS.2023.3275146

21. M. Zhao, J. Ning, J. Hu, T. Li, Attention-driven dual feature guidance for hyperspectral super-resolution, *IEEE Trans. Geosci. Remote Sensing*, **61** (2023), 5525116. https://doi.org/10.1109/TGRS.2023.3318013

22. Y. Li, L. Zhang, C. Dingl, W. Wei, Y. Zhang, Single hyperspectral Image super-resolution with grouped deep recursive residual network, in *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, IEEE, (2018), 1–4. https://doi.org/10.1109/BigMM.2018.8499097

23. Q. Wei, J. Bioucas-Dias, N. Dobigeon, J. Y. Tourneret, Hyperspectral and multispectral image fusion based on a sparse representation, *IEEE Trans. Geosci. Remote Sensing*, **53** (2015), 3658–3668. https://doi.org/10.1109/TGRS.2014.2381272

24. Y. Xu, Z. Wu, J. Chanussot, Z. Wei, Nonlocal patch tensor sparse representation for hyperspectral image super-resolution, *IEEE Trans. Image Process.*, **28** (2019), 3034–3047. https://doi.org/10.1109/TIP.2019.2893530

25. X. H. Han, B. Shi, Y. Zheng, Self-similarity constrained sparse representation for hyperspectral image super-resolution, *IEEE Trans. Image Process.*, **27** (2018), 5625–5637. https://doi.org/10.1109/TIP.2018.2855418

26. L. Zhang, W. Wei, C. Bai, Y. Gao, Y. Zhang, Exploiting clustering manifold structure for hyperspectral imagery super-resolution, *IEEE Trans. Image Process.*, **27** (2018), 5969–5982. https://doi.org/10.1109/TIP.2018.2862629

27. M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. BioucasDias, J. Chanussot, Hyperspectral super-resolution of locally low rank images from complementary multisource data, *IEEE Trans. Image Process.*, **25** (2015), 274–288. https://doi.org/10.1109/TIP.2015.2496263

28. R. Dian, S. Li, Hyperspectral image super-resolution via subspacebased low tensor multi-rank regularization, *IEEE Trans. Image Process.*, **28** (2019), 5135–5146. https://doi.org/10.1109/TIP.2019.2916734

29. Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, Z. Xu, Multispectral and hyperspectral image fusion by ms/hs fusion net, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, (2019), 1585–1594. https://doi.org/10.1109/CVPR.2019.00168

30. Z. W. Pan, H. L. Shen, Multispectral image super-resolution via RGB image fusion and radiometric calibration, *IEEE Trans. Image Process.*, **28** (2019), 1783–1797. https://doi.org/10.1109/TIP.2018.2881911

31. C. Dong, C. C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in *Computer Vision-ECCV 2014*, Springer, (2014), 184–199. https://doi.org/10.1007/978-3-319-10593-2_13

32. L. Liebel, M. Körner, Single-image super resolution for multispectral remote sensing data using convolutional neural networks, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **41** (2016), 883–890. https://doi.org/10.5194/isprs-archives-XLI-B3-883-2016

33. Y. Yuan, X. Zheng, X. Lu, Hyperspectral image superresolution by transfer learning, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **10** (2017), 1963–1974. https://doi.org/10.1109/JSTARS.2017.2655112

34. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, CBAM: convolutional block attention module, in *Proceedings of the European conference on computer vision (ECCV)*, IEEE, (2018), 3–19.

35. V. Lempitsky, A. Vedaldi, D. Ulyanov, Deep image prior, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2018), 9446–9454. https://doi.org/10.1109/CVPR.2018.00984

36. O. Sidorov, J. Y. Hardeberg, Deep hyperspectral prior: single-image denoising, inpainting, super-resolution, in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, IEEE, (2019), 3844–3851. https://doi.org/10.1109/ICCVW.2019.00477

37. C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **38** (2016), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281

38. D. Liu, J. Li, Q. Yuan, A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution, *IEEE Trans. Geosci. Remote Sensing*, **59** (2021), 7711–7725. https://doi.org/10.1109/TGRS.2021.3049875

39. X. Wang, Q. Hu, J. Jiang, J. Ma, A group-based embedding learning and integration network for hyperspectral image super-resolution, *IEEE Trans. Geosci. Remote Sensing*, **60** (2022), 5541416. https://doi.org/10.1109/TGRS.2022.3217406

40. T. Liu, Y. Liu, C. Zhang, L. Yuan, X. Sui, Q. Chen, Hyperspectral image super-resolution via dual-domain network based on hybrid convolution, *IEEE Trans. Geosci. Remote Sensing*, **62** (2024), 5512518. https://doi.org/10.1109/TGRS.2024.3370107

41. S. Chen, L. Zhang, L. Zhang, Cross-scope spatial-spectral information aggregation for hyperspectral image super-resolution, preprint, arXiv:2311.17340.

42. M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao, J. Zhang, Essaformer: efficient transformer for hyperspectral image super-resolution, in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, (2023), 23016–23027. https://doi.org/10.1109/ICCV51070.2023.02109

43. X. Huang, L. Zhang, A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, *Int. J. Remote Sens.*, **30** (2009), 3205–3221. https://doi.org/10.1080/01431160802559046

44. F. Yasuma, T. Mitsunaga, D. Iso, S. K. Nayar, Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum, *IEEE Trans. Image Process.*, **19** (2010), 2241–2253. https://doi.org/10.1109/TIP.2010.2046811

45. N. Yokoya, A. Iwasaki, Airborne hyperspectral data over chikusei, *Space Appl. Lab., Univ. Tokyo, Tokyo*, **5** (2016), 1–6. https://doi.org/10.1109/TIP.2010.2046811

46. H. Hou, H. Andrews, Cubic splines for image interpolation and digital filtering, *IEEE Trans. Acoust. Speech Signal Process.*, **26** (1978), 508–517. https://doi.org/10.1109/TASSP.1978.1163154

47. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al., Swin transformer: hierarchical vision transformer using shifted windows, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, (2021), 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

AIMS Press