Information Fusion 94 (2023) 92–111

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

An efficient unfolding network with disentangled spatial-spectral representation for hyperspectral image super-resolution

Denghong Liu [a,b], Jie Li [b], Qiangqiang Yuan [b,*], Li Zheng [b], Jiang He [b], Shuheng Zhao [a], Yi Xiao [b]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[b] School of Geodesy and Geomatics, Wuhan University, Wuhan, China

ARTICLE INFO

Keywords:
Hyperspectral image
Super-resolution
Unfolding network
Disentangled spatial-spectral representation

ABSTRACT

Hyperspectral image super-resolution (HSI SR) is dramatically impacted by high spectral dimensionality, insufficient spatial resolution, and limited availability of training samples. Current approaches mainly rely on complex data-driven models to address some of these challenges, and the characteristics of HSI are not fully considered in the model design. In this paper, we propose an efficient unfolding network with disentangled spatial-spectral representation (EUNet) for HSI SR by combining domain knowledge (i.e., spectral correlation, degradation model, and structure prior) with deep learning. Specifically, the optimization process of the super-resolution prior-driven Maximum A Posterior (MAP) framework is unfolded into an interpretable multi-stage network, which inherits the advantages of deep learning-based image super-resolution (e.g., feature extraction in low-resolution space) and explicitly imposes the degradation model constraint. To well incorporate the structure prior of HSI, spatial and spectral feature extraction is disentangled by a variant of depthwise separable convolution, and spectral correlation is embedded by a lightweight spectral attention mechanism, so that the difficulty and computational complexity of feature learning are greatly reduced. Experiments on benchmark datasets with different degradation models demonstrate the feasibility and superiority of the proposed EUNet over other state-of-the-art methods in terms of evaluation metrics and computational complexity. The source code is available at https://github.com/denghong-liu/EUNet.

1. Introduction

Hyperspectral sensors sample the reflective portion of the electromagnetic spectrum in hundreds of narrow contiguous spectral channels to form three-dimensional (3D) hyperspectral images (HSIs) [1] shown in Fig. 1. In the spatial dimension, a 3D HSI consists of a set of gray-scale images (spectral bands) to provide structural information. From a spectral perspective, each pixel is a spectral vector corresponding to the reflected radiation from a specific region of the Earth. Such detailed spectral sampling facilitates the accurate identification of different materials and subtle changes [2,3]. However, due to the limitations of imaging technology, there is a tradeoff between spatial resolution, spectral resolution, and signal-to-noise ratio (SNR) [4], that is, precise spectral information generally comes at the cost of low spatial resolution such that it cannot meet the needs of unmixing, classification and identification, hindering practical applications of HSI.

Hyperspectral image super-resolution (HSI SR) is an economical and effective signal post-processing technique to enhance the spatial resolution of given low-resolution (LR) images. Depending on whether or not auxiliary information is used, HSI SR can be divided into two categories, fusion-based methods and single HSI SR [5]. Fusion-based methods aim to find missing high-frequency details in high-resolution (HR) panchromatic images (PAN) or multispectral images (MSI). These auxiliary co-registered images are, however, difficult to obtain in real scenarios. In this regard, single HSI SR has increasingly attracted attention in reality. Given the limited information provided by a single LR HSI, such method explores image priors from a large external image dataset to constrain the solution space of the ill-posed SR and is therefore called leaning-based single HSI SR, early represented by sparse representation [6,7]. Inspired by the remarkable advances in natural image super-resolution over recent years, using deep learning methods to achieve the single HSI SR has come into vogue. Deep learning, with a powerful ability to learn realistic image priors, as manifested by convolutional neural networks (CNNs), is capable of modeling complex mappings between LR and HR images [8–10]. A straightforward way to accomplish the single HSI SR is to apply natural image super-resolution

* Corresponding author.
E-mail address: yqiang86@gmail.com (Q. Yuan).

https://doi.org/10.1016/j.inffus.2023.01.018
Received 16 May 2022; Received in revised form 26 December 2022; Accepted 17 January 2023
Available online 20 January 2023

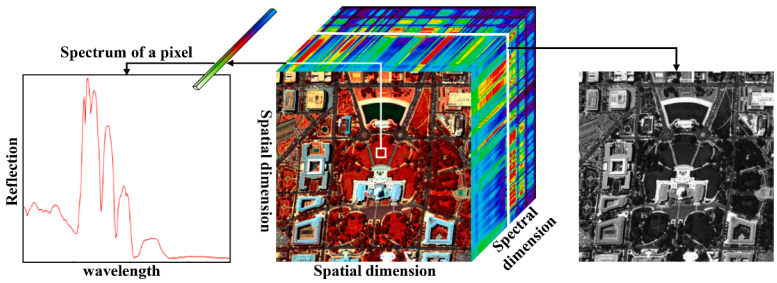


Fig. 1. An example of a 3D HSI.

methods to each spectral band independently [11].

Nevertheless, single HSI SR is more challenging than its natural image counterpart in view of the intrinsic characteristics of HSIs. First, HSIs typically have hundreds of bands with complex spectral modes. Strong similarities exist between adjacent bands, known as spectral correlation, and the SNR varies widely by band. The high spectral dimensionality increases the difficulty of feature extraction and requires spectral consistency during reconstruction. Second, hyperspectral images usually cover a wide area with very limited spatial resolution relative to the size of the object being detected, struggling to provide sufficient detailed information and consequently deteriorating the ill-posedness of SR. Third, since the collection of HSIs is either expensive or time demanding, the number of such samples currently available is relatively small, making it difficult to build high-precision data-driven models.

Existing literature focus on the utilization of spectral correlation, introducing network modules and learning strategies such as spectral difference learning [12], 3D convolution [13], spectral angle mapper (SAM) loss [14], and attention mechanisms [15]. However, they almost all follow the framework of directly learning an end-to-end mapping, which not only lacks interpretability but also overlooks the role of the degradation model that characterizes the relationship between LR HSI and HR HSI. Moreover, the general trend is to design more complicated networks to improve performance. For example, the number of parameters of 2D CNNs for HSI SR grows from hundreds of thousands (e.g., GDRRN [14]) to more than ten million (e.g., SSPSR [16]), and similarly, that of 3D CNNs rises from tens of thousands (e.g., 3DFCNN [13]) to more than a million (e.g., ERCSR [17]). The high computational complexity of the models limits their deployment in mobile and embedded applications (e.g., satellite and airborne platforms).

To tackle the challenges of HSI SR and compensate for the deficiencies of current approaches, we incorporate the domain knowledge of degeneration model (Section 3.1.1), 3D structure prior (Section 3.3.1), and spectral correlation (Section 3.3.2) in the network design to build an efficient unfolding network with disentangled spatial-spectral representation (EUNet). In summary, the main contributions of this work are three-fold.

(1) The proposed EUNet is an interpretable multi-stage network under the super-resolution prior-driven Maximum A Posterior (MAP) framework, which can encompass the popular design of deep learning-based super-resolution and explicitly impose the degradation model constraint in searching for high-quality solutions.
(2) A hyperspectral image super-resolution prior network (HSRPN) is designed for the benchmark bicubic degradation, where a variant of depthwise separable convolution disentangles the HSI from different dimensions and a lightweight spectral attention mechanism embeds spectral correlation. In this way, the difficulty and computational complexity of feature learning are greatly reduced, achieving efficient spatial-spectral representation.
(3) An in-depth analysis of different unfolding strategies, spatial-spectral convolutions, and ablation experiments is presented to validate the effectiveness of different parts of the proposed method. Experiments on benchmark datasets with different degradation models testify that the proposed approach can achieve state-of-the-art performance with lightweight structures.

The rest of this paper is organized as follows. Section 2 briefly reviews the deep learning-based single HSI SR, lightweight SR networks, and the deep unfolding strategies in the field of SR. Section 3 describes in depth the insights and details of the proposed efficient unfolding network. Comprehensive performance comparisons and analyses are provided in Section 4. A detailed discussion of the structure and pros of our method is presented in Section 5, and finally this paper is concluded in Section 6.

2. Related work

2.1. Deep learning-based single HSI SR

Since SRCNN [8] has striking success in the SR field, deep learning techniques, most prominently CNNs have become prevalent for single HSI SR due to their strong capabilities. In general, HSIs are either super-resolved band-by-band or all spectral bands are modeled simultaneously. In the first way, Liebel and Körne [11] extended SRCNN from natural images to single-band images of SENTINEL-2 with promising results by training on a domain-specific dataset. Taking into account the spectral information, Yuan et al. [18] and Xie et al. [19] proposed to first super-resolve all or key bands independently by CNNs, and then apply nonnegative matrix factorization (NMF) to enforce collaborations between low- and high-resolution hyperspectral images. Li et al. [12] introduced a SDCNN for spectral difference mapping learning to preserve spectral information, combined with a spatial constraint to improve spatial details. The aforementioned multi-step methods are inflexible and unstable, suffering from error accumulation and time-consuming inference. Hu et al. [20] further developed a generic spectral difference module to enable end-to-end prediction per band. However, the band-wise manner makes it difficult to effectively explore spectral relationships.

For the other way, Li et al. [14] proposed a grouped deep recursive residual network (GDRRN) with grouped recursive modules enabling the compact network, and SAM loss reducing spectral distortion, which learns to directly map the input LR HSI to the corresponding HR HSI. Unfortunately, the output of normal 2D convolution is generated by a summation over all channels, that is, channel dependencies are implicitly embedded, which would lead to spectral disorder and the non-discrimination of channel importance. Zheng et al. [21] designed separable-spectral convolution to independently extract features and then fused them together through inception-residual blocks. Jiang et al. [16] proposed a group convolution and progressive upsampling framework to alleviate the difficulty of feature extraction, within which a spatial-spectral block consisting of residual blocks, convolutions and a channel attention module was devised to exploit spatial and spectral correlations. Liu et al. [22] introduced traditional spectral grouping scheme and second-order spectral attention into residual dense network

to facilitate the modeling of all spectral bands.

Another popular route is to use 3D convolution for exploring both neighboring spatial context and spectral correlation. Mei et al. [13] first constructed a four-layer 3DFCNN to mitigate spectral distortion. Yang et al. [23] further integrated wavelet transform into 3D CNNs to enhance detailed structure reconstruction. Li et al. [17,24] successively proposed MCNet and ERCSR with hybrid 2D/3D convolutions to strengthen the spatial learning ability while considering spectral information. Although separable 3D convolutions and 3D attention mechanisms have been applied in HSI SR [25–27], 3D CNNs still have expensive computational costs and memory demands, especially for extremely high spectral dimensionality, and the global features of HSI are difficult to be extracted by 3D convolution.

2.2. Lightweight SR networks

There has been rising interest in studying lightweight SR networks to boost practical real-world applications such as mobile systems and video streaming services. Earlier works [28,29] relied on recursive structures to reduce the number of parameters but ignored the fact that SR performance was guaranteed at the cost of the increased number of operations. Ahn et al. [30] proposed a cascading residual network (CARN) with the efficient residual block and recursive scheme that is lightweight in both size and computation. Hui et al. [31] designed an information distillation network (IDN) to gather and distill more useful information despite having low model complexity. IMDN further improved IDN through progressive feature refinement and contrast-aware channel attention [32]. In addition to manually designed network architectures, techniques such as neural architecture search (NAS) [33], pruning [34], and knowledge distillation [35] are implemented to explore more efficient lightweight SR networks.

However, research on lightweight networks is scarce in the field of remote sensing SR [36]. Remote sensing images are generally quite large, especially HSIs with hundreds of spectral bands. Fast processing remote sensing big data can support emergency response (e.g., earthquakes, typhoons, and coastal monitoring). In addition, lightweight models can alleviate the training difficulties and overfitting problems with the limited availability of HSI samples. In recent years, there has been a trend to deploy various missions on space devices (e.g., satellites and airborne equipment) to reduce operational costs [37]. As the communication channel between space devices and ground stations is often limited, onboard data processing (e.g., joint SR and detection) can avoid transferring useless images. Due to the low computing power and small memory of space devices, it is essential to study lightweight HSI SR networks.

2.3. Deep unfolding strategy for SR

The unfolding strategy is a mainstream interpretable SR framework that can leverage both model-based and learning-based methods. Based on the degradation model characterizing the relationship between LR and HR images, SR can be transformed into a Bayesian optimization problem, consisting of a data fidelity term that guarantees the solution conforms to the degradation process and a prior regularizer that enforces desired properties of the output [38]. With the help of variable splitting techniques, such as the half-quadratic splitting (HQS) algorithm [39] and alternating direction method of multipliers (ADMM) algorithm [40], the traditional handcrafted images priors can be replaced by implicit priors defined by off-the-shelf denoisers [41], generally utilizing deep learning-based denoisers, which are called deep plug-and-play (PnP) methods or deep denoising-based methods [42]. Although deep learning plays a role to some extent, such methods are essentially traditional iterative optimization, subject to manual parameter selection and a high computational burden.

On this basis, the deep unfolding strategy has been developed [38,43,44]. Specifically, the iterative process of the SR optimization problem is unfolded into a feed-forward neural network that is jointly optimized by end-to-end training. The methods based on deep unfolding strategy are not only interpretable, that is, similar to the deep PnP methods integrating model- and learning-based methods, but also can produce better results with fewer parameters and iterations. Regarding the reconstruction of HSIs, recent works have introduced the deep unfolding strategy into hyperspectral image fusion [45, 46], spectral super-resolution [47], and spatiospectral super-resolution [48] with promising results. To the best of our knowledge, there is very little research work on interpretable networks for single HSI SR. Furthermore, the existing unfolding strategy of SR mostly corresponds to two separated subproblems, one for dealing with the data fidelity term and the other is a pure denoising problem. It is non-trivial to efficiently handle the data fidelity term involving deblurring and upsampling, and generic super-resolution networks cannot be adopted in this strategy.

Table 1
Notations and abbreviations.

Abbr.	Description
HSI	hyperspectral image
SR	super-resolution
LR	low-resolution
HR	high-resolution
MAP	Maximum A Posterior
y	LR HSI $\in R^{h\times w\times B}$
X	HR HSI $\in R^{H\times W\times B}$
H	blurring matrix
D	bicubic downsampling
EUNet	efficient unfolding network
RP	regularization parameter
DC	data consistency
SRP	super-resolution prior
HSRPN	hyperspectral image super-resolution prior network
ESSG	efficient spatial-spectral group
ESSB	efficient spatial-spectral block
DSSM	disentangled spatial-spectral module
RLSAM	residual local spectral attention module

3. Methodology

The proposed EUNet aims to learn an end-to-end mapping between the observed LR HSI and the corresponding HR HSI in a recursive way and apply the learned mapping to reconstruct the HR HSI from a given LR HSI. By incorporating the domain knowledge of HSI, EUNet can better deal with the ill-posedness of SR and achieve efficient spatial-spectral representations. In this section, we first introduce the super-resolution prior-driven unfolding framework, which converts the MAP optimization problem into two independent subproblems to be solved alternately. Then, EUNet is designed based on the above unfolding framework with three interconnected modules, regularization parameter (RP), data consistency (DC), and super-resolution prior (SRP) module to jointly optimize these two subproblems. Finally, the neural network form of the SRP module is illustrated in detail. The notations used in this paper are listed in Table 1.

3.1. Super-resolution prior-driven unfolding framework

Determining how to characterize the degradation relationship between LR HSI and HR HSI is the key to the success of SR. The widely-used degradation models include: (1) bicubic degradation, which assumes the LR HSI is a bicubicly downsampled version of HR HSI [49]; (2) a general degradation model, which assumes the LR HSI is a blurred, bicubicly downsampled and noisy version of HR HSI [10]. The existing unfolding framework is usually based on the latter degradation model, where the original optimization problem can be decoupled into a data fidelity term and a regularization term corresponding to the pure denoising problem, so that various denoising prior, such as excellent CNN denoisers, can be

plugged in as a modular part [42, 50, 51]. However, dealing with the data fidelity term in this framework is a challenge since it involves both deblurring and upsampling. In addition, it cannot take advantage of state-of-the-art super-resolution methods based on deep learning. Therefore, we employ another simple yet effective degradation model to define the unfolding framework, which corresponds to a deblurring problem and an SR problem with bicubic degradation [52]. It can encourage data consistency while inheriting the merits of existing SR networks.

3.1.1. Degradation model

Denote the observed LR HSI as y in $R^{h\times w\times B}$ contains B bands with spatial size of $h \times w$ and the corresponding HR HSI as $X \in R^{H\times W\times B}$, in which $H = s \times h$, $W = s \times w$ and s represents the scale factor. In general, the blur convolution is assumed to be separable over spectral bands [53]. The degradation model is given by:

$$y_b = (X_b\downarrow_s \otimes K_b) + N_b \tag{1}$$

where for bth band of HSI, $X_b\downarrow_s \otimes K_b$ represents two-dimensional convolution between blur kerner K_b and downsampled image X_b, N_b is the additive independent and identically distributed (i.i.d) Gaussian noise with standard deviation (noise level) σ, and $\downarrow_s$ is a s-fold bicubic downsampler. Two types of downsampler, direct downsampler and bicubic downsampler, are widely considered in the existing literature. We consider the bicubic downsampler since when K is delta kernel and the noise level is zero, the problem turns into the widely-used bicubic degradation model.

By unfolding 3D data cubes into 2D matrices (i.e., $y \in R^{hw\times B}$ and $X \in R^{HW\times B}$), the hyperspectral degradation model can be written as:

$$y = HDX + N \tag{2}$$

where $H \in R^{hw\times hw}$ and $D \in R^{hw\times HW}$ denotes the blurring matrix and bicubic downsampling operation, respectively. The degradation model conveys that the LR HSI is a bicubicly downsampled, blurred and noisy version of the HR HSI.

3.1.2. Unfolding optimization

According to the Maximum A Posteriori (MAP) framework [38], once the degradation model is defined, the estimated HR HSI $\widehat{X} \in R^{H\times W\times B}$ could be obtained by minimizing the following objective function:

$$\widehat{X} = \arg\min_X \frac{1}{2\sigma^2}\| y - HDX \|_2^2 + \lambda\Phi(X) \tag{3}$$

where the first squared-error term is the data fidelity term, $\Phi(X)$ is the regularization (prior) term, and λ is the trade-off parameter. This objective function forces $\widehat{X}$ to accord with degradation process and have the desired properties.

With the variable splitting technique, such as the half quadratic splitting (HQS) algorithm and alternating direction method of multipliers (ADMM) algorithm, the unfolding inference for Eq. (3) can be achieved. By introducing an auxiliary variable Z, Eq. (3) can be reformulated as:

$$\widehat{X} = \arg\min_X \frac{1}{2\sigma^2}\| y - HZ \|_2^2 + \lambda\Phi(X), s.t. Z = DX \tag{4}$$

For simplicity, HQS is used to convert the above constrained optimization problem into an equivalent non-constrained optimization problem as follows:

$$(X, Z) = \arg\min_{X,Z} \frac{1}{2\sigma^2}\| y - HZ \|_2^2 + \lambda\Phi(X) + \frac{\mu}{2}\| Z - DX \|_2^2 \tag{5}$$

where μ is the variable penalty parameter that constrains the similarity between Z and X. Such a problem can be resolved by alternately solving two independent subproblems associated with the fidelity and regularization terms, respectively.

$$\begin{aligned} Z_{t+1} &= \arg\min_Z \| y - HZ_t \|_2^2 + \mu_{t+1}\sigma^2\| Z_t - DX_t \|_2^2 \\ X_{t+1} &= \arg\min_X \frac{\mu_{t+1}}{2}\| Z_{t+1} - DX_t \|_2^2 + \lambda\Phi(X_t) \end{aligned} \tag{6}$$

The Z-subproblem is a quadratic optimization problem that has a closed-form solution, but involves matrix inversion, which is computationally expensive especially for large-scale matrices. Therefore, we follow [38] to compute an approximate solution with a single step of gradient descent, as

$$\begin{aligned} Z_{t+1} &= Z_t - \eta_{t+1}\left[H^T(HZ_t - y) + \alpha_{t+1}(Z_t - DX_t)\right] \\ &= \overline{H}Z_t + \eta_{t+1}H^T y + \eta_{t+1}\alpha_{t+1}DX_t \\ &= DC(Z_t, X_t, y, H, D, \alpha_{t+1}, \eta_{t+1}) \end{aligned} \tag{7}$$

where $\alpha_{t+1} \triangleq \mu_{t+1}\sigma^2$, $\overline{H} = [(1 - \eta\alpha)\mathrm{I} - \eta H^T H]$, and η is the parameter controlling the step size. $DC(\cdot)$ represents a gradient step that promotes data consistency.

The X-subproblem can be rewritten as

$$X_{t+1} = \arg\min_X \frac{1}{2\left(\sqrt{\lambda/\mu_{t+1}}\right)^2}\| Z_{t+1} - DX_t \|_2^2 + \Phi(X_t) \tag{8}$$

From a Bayesian perspective, Eq. (8) corresponds to super-resolving the image Z_{t+1} which is bicubicly downsampled and corrupted by Gaussian noise with noise level $\beta_{t+1} \triangleq \sqrt{\lambda/\mu_{t+1}}$. In another word, the X-subproblem is a classic super-resolution problem with bicubic degradation model as follows:

$$Y = DX + N \tag{9}$$

As a result, any super-resolution methods can be used to solve Eq. (8), as

$$X_{t+1} = SRP(Z_{t+1}, \beta_{t+1}) \tag{10}$$

where $SRP(\cdot)$ denotes a super-resolver.

As stated above, the two subproblems can be relatively easy to solve by Eqs. (7) and (10). Since the prior term $\Phi(X)$ is implicitly encoded in $SRP(\cdot)$, we refer to it as super-resolution prior-driven unfolding framework, which is summarized below in Algorithm 1. This framework has a clear interpretation, as it iteratively addresses the distortion of blur and bicubic degradation.

Algorithm 1
Super-resolution prior-driven unfolding framework.

Input: y

Initialization:
- Degradation matrices D, H, and H^T;
- Regularization parameters α, β, and η;
- Auxiliary variable $Z = H^T y$.

for T *steps* **do**
- Compute X_t via Eq. (10);
- Compute Z_{t+1} via Eq. (7);

end

Output: X

3.2. Efficient unfolding network

In the above framework, the explicit image prior term can be unknown in solving Eq. (3). Various plug-and-play (PnP) methods have been proposed to plug deep implicitly priors encoded by pre-trained SR networks into the iterative solution, which explores the advantages of powerful modeling capacity, efficient inference, and promising performance of deep learning. However, they are essentially model-based methods, which not only involve manual selection of regularization parameters and time-consuming iterative inference, but also fail to

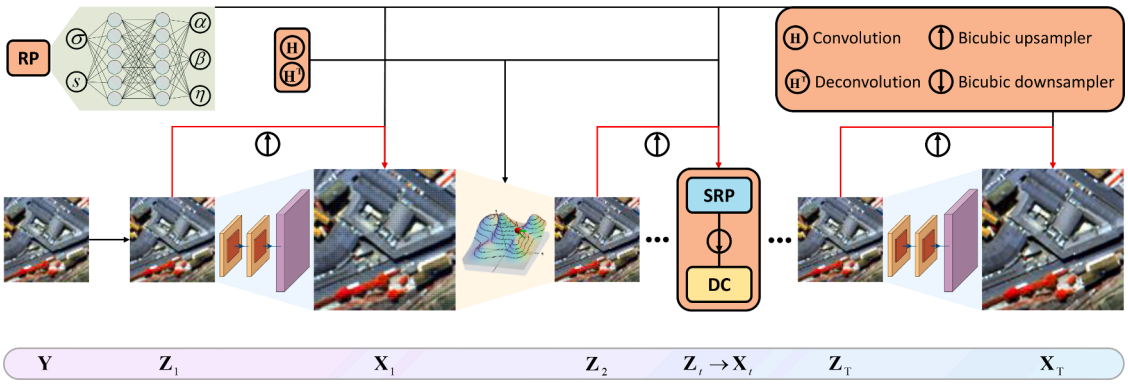


Fig. 2. The overall architecture of our proposed EUNet for HSI SR.

jointly optimize the two subproblems.

Considering the two subproblems can be solved with neural network modules, an efficient unfolding network for HSI SR (EUNet) is designed by unfolding the iterative process into a multistage network. It is an end-to-end trainable network, which can jointly deal with data fidelity and prior terms. Fig. 2 illustrates the overall architecture of EUNet with three interconnected modules, regularization parameter (RP), data consistency (DC), and super-resolution prior (SRP) module. The RP module adaptively sets the parameters that balance the contribution of two subproblems. The DC module is responsible for the updating of auxiliary variable Z, which serves as the input of the SRP module to obtain the reconstruction result. EUNet includes T stages, corresponding to T iterations in Algorithm 1. The architectures and parameters are shared at different stages. In this way, we can easily control the trade-off between performance and computational complexity, just like the recursive neural network, increasing depth without introducing new parameters for additional convolutions [54].

3.2.1. Regularization parameter module

The regularization parameter of the unfolding framework is generically composed of a series of internal parameters, including penalty parameters, denoising strengths, and step size of gradient descent during iterations. Due to the complexity of hyperspectral images and the inexact solution of Z-subproblem, these parameters do not necessarily change monotonously and are difficult to be initialized empirically. Therefore, a regularization parameter module similar to [43], is employed to control the outputs of the data fidelity and prior terms in an adaptive manner. It is a fully connected neural network with two hidden layers, which maps the scale factor and noise level that affect the degree of ill-posedness to regularization parameters as follows:

$$[\alpha, \beta, \eta] = RP(\sigma, s) \tag{11}$$

where $\alpha = [\alpha_1, \cdots, \alpha_t, \cdots, \alpha_T]$, $\beta = [\beta_1, \cdots, \beta_t, \cdots, \beta_T]$, and $\eta = [\eta_1, \cdots, \eta_t, \cdots, \eta_T]$. This subnetwork enables regularization parameters to be compatible with other parts of the network and vary with two key elements.

3.2.2. Data consistency module

The data consistency module aims to reduce the distortion of blur by minimizing a weighted combination of the data term $\| Y - HZ_t \|_2^2$ and the quadratic regularization term $\| Z_t - DX_t \|_2^2$. It is a parameter-free module that performs a gradient step on the objective function of Z-subproblem to obtain an approximated solution, as shown in Eq. (12).

$$Z_{t+1} = \mathrm{DC}(Z_t, X_t, y, H, D, \alpha_{t+1}, \eta_{t+1}) \tag{12}$$

By taking as input the blur matrix and downsampling matrix, it imposes an explicit degradation constraint on the solution (i.e., the reconstructed HR HSI should accord with the degradation process), thus promoting data consistency in the iterative process. The wide coverage and relatively large ground sampling distance of HSIs lead to the insufficient detailed texture features, which further deteriorates the ill-posedness of super-resolution. This module can provide an additional constraint in searching for HR solutions to improve both accuracy and robustness.

Note that the blur matrix H can be simply implemented with a convolutional layer. Z_0 is initialized as $H^T y$, which can be implemented as the transposed convolution of y. As mentioned in Section 3.1.1, the convolution is separable over spectral bands, so we use depthwise convolution with small computation cost, where each band corresponds to an independent convolutional kernel. In this study, two types of degradation operators are considered: Gaussian downsampling and bicubic downsampling. For bicubic downsampling, H and H^T are replaced by the identity matrix. For Gaussian downsampling, synthesize the corresponding LR HSI via Eq. (2) with bicubic downsampling and Gaussian kernel.

3.2.3. Super-resolution prior module

The super-resolution prior module acts as the regularization term to facilitate the search for high-quality solutions by solving a super-resolution problem with bicubic degradation model. Since the expressive capacity of hand-crafted priors and shallow heuristic models is limited to handle the high-dimensional hyperspectral data, powerful deep learning methods are employed to bring strong exterior image priors. In the above unfolding framework, X-subproblem corresponds to a super-resolution problem with bicubic degradation model, which has become the benchmark setting in deep learning-based SR methods. Therefore, we can extend these methods to facilitate implicit learning of spectral-spatial priors while keeping their merits, such as feature extraction in LR space. Based on this, the hyperspectral image super-resolution prior network (HSRPN) is proposed, which concatenates the learned noise level map and the output of DC module Z as input to handle various noise levels for efficiency.

$$X_t = SRP(Z_t, \beta_t) \tag{13}$$

In addition, considering the high dimensionality and limited training samples of HSI, we incorporate domain knowledge to design a lightweight disentangled spatial-spectral module (DSSM) and residual local spectral attention module (RLSAM), which can achieve complex spectral pattern modeling and full utilization of spectral correlation with low computational complexity.

3.2.4. Spectral preservation

In addition to enhancing spatial detail, HSI SR also aims to preserve spectral information. Traditional model-based optimization methods usually define the spectral penalty as

$$L_{spectral} = \sum_{b=1}^{B} \| X_b \otimes K_b - y_b \uparrow_s \| = \| HX - D^T y \| \tag{14}$$

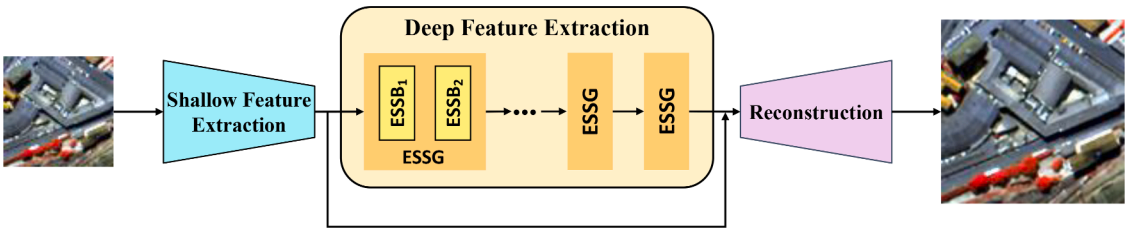


Fig. 3. The architecture of HSRPN for super-resolution prior module.

where $\uparrow_s$, D^T indicates the s-fold upsampling operation. This penalty can encourage the estimated solution to share the same spectral information in the original LR HSI.

The HSRPN at each stage learns an end-to-end mapping between Z and X to obtain a clearer image, by minimizing

$$L = \| SRP(Z, \beta; \Theta) - X \| \tag{15}$$

where Θ denotes learnable network parameter.

To promote spectral consistency, a skip connection (the red arrow in Fig. 2) is employed to pass the upsampled LR HSI with abundant spectral information to the end of HSRPN, and thus Eq. (15) can be rewritten as:

$$L = \| SRP(Z, \beta; \Theta) + D^T Z - X \| \tag{16}$$

This term is similar as Eq. (14), enforcing the spectral information sharing. However, unlike the blur kernel used in Eq. (14), the deep network can automatically modify the HR HSI [55].

3.2.5. Loss function

For HSI SR, both spatial reconstruction and spectral consistency should be considered. Therefore, l_1 loss and SAM loss function are used to jointly optimize the network.

$$L = L_{spatial} + \delta L_{spectral} \tag{17}$$

where δ is the balance parameter. Denote $\{(y_i, X_i)\}_{i=1}^N$ as N LR-HR patch pairs. The spatial loss is constructed from l_1 norm, which has been shown to be effective in SR, as follows:

$$L_{spatial} = \frac{1}{N} \sum_{i=1}^{N} \| X_i - \widehat{X}_i \|_1 \tag{18}$$

SAM loss [14] measures the spectral similarity by calculating the angle between two corresponding spectral vectors pixel by pixel:

$$L_{spectral} = \frac{1}{N} \sum_{i=1}^{N} \arccos\left(\frac{X_i \cdot \widehat{X}_i}{\| X_i \|_2 \| \widehat{X}_i \|_2} \right) \tag{19}$$

We empirically set δ to 0.1 so that the two loss values are in the same order of magnitude. Above-mentioned hybrid loss encourages the network to improve both spatial and spectral reconstruction accuracy.

3.3. Hyperspectral image super-resolution prior network

An overview of HSRPN is shown in Fig. 3. Following the design of the classic SR network, it consists of three parts: shallow feature extraction, deep hierarchical feature extraction, and reconstruction [56]. Specifically, a convolutional layer is first used to represent the LR input as feature maps because images gradients are more informative than the

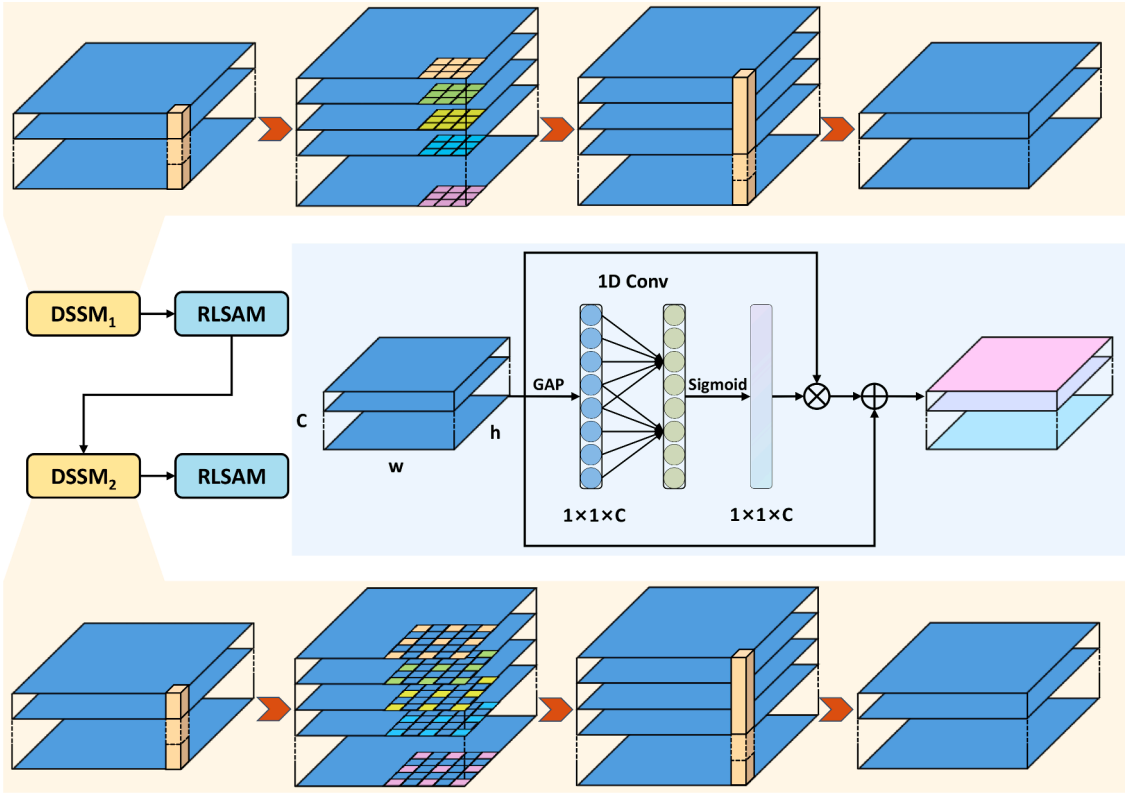


Fig. 4. The structure of ESSG in HSRPN.

Table 2
Dataset statistics.

Dataset	Sensor	Spatial resolution (m)	Sepctral range (nm)	Spatial size	Spectral band
Pavia center	ROSIS	1.3	430–860	1096 × 715	102
Pavia University	ROSIS	1.3	430–860	610 × 340	103
Chikusei	Hyperspec-VNIR-C	2.5	363–1018	2517 × 2335	128
Houston	ITRES CASI 1500	1.0	380–1050	4172 × 1202	48

raw intensities for SR [54]. Then, deep spatial-spectral features are extracted by a cascade of efficient spatial-spectral groups (ESSG), which contains two successive efficient spatial-spectral blocks (ESSB) with different receptive fields. The ESSB including DSSM and RLSAM is designed based on the fact that the hyperspectral image is a 3D cube integrating 1D spectrum and 2D imaging. It first decouples the spatial-spectral feature extraction by DSSM and then uses RLSAM to embed spectral correlation, achieving effective spatial-spectral representation learning with limited computation complexity. Finally, shallow and deep features are aggregated with a global skip connection to reconstruct the HR image. The deconvolution layers are utilized in the reconstruction part to upsample features, followed by a convolutional layer.

3.3.1. Disentangled spatial-spectral module

Existing HSI SR methods usually utilize 2D convolution or 3D convolution for feature learning. The output of 2D convolution is produced by the summation of all spectral bands, and thus spectral dependencies are implicitly embedded and entangled with local spatial correlation, which hinders the learning of spatial and spectral features, respectively. Although 3D convolution can preserve spectral information, it is difficult to extract global features and requires too many computational resources. To address these problems, we design the disentangled spatial-spectral module built primarily from depthwise separable convolutions [57], to decouple the cross-spectral and spatial information, as shown in Fig. 4.

This module employs the $_{3\times3}$ depthwise convolution and $_{1\times1}$ convolution to process features in spatial dimension and spectral dimension, respectively, which can well incorporate the structure prior of HSI and effectively handle high spectral dimensionality, making representation learning easier. Specifically, the $_{1\times1}$ convolution enables cross-channel interaction, that is, each spectral band can be reconstructed by a linear combination of all spectral bands [16]. Therefore, it can focus on the global correlation along with spectra. The $_{1\times1}$ depthwise convolution applies a single filter per each input spectral band for spatial context encoding in a bandwise manner. Before spatial feature extraction, the input feature dimension F is expanded using an additional $_{1\times1}$ convolution (the expansion rate is denoted as e) to perform spatial transformation at a higher dimension, thus enhancing the network expressiveness. The experiments in Section 5.5 demonstrate that the disentangled spatial-spectral module not only saves a lot of computation but also improves the performance of HSI SR.

Considering the wide range of scales within the ground objects in HSIs, using a uniform scale for feature extraction cannot guarantee optimal spatial features. Therefore, dilated convolution [58] is further adopted in the depthwise convolution for spatial feature extraction to adjust the receptive fields, which enables the network to capture multiscale representations without increasing the number of parameters and network complexity. As shown in Fig. 4, dilated convolution exponentially expands the receptive fields by padding zeros between two pixels in convolutional kernels. With different dilation rates r, a fixed convolutional kernel can achieve various receptive fields, e.g., the receptive field of 3×3 convolution with $r=2$ is 5. In each ESSG, we alternately employ one ESSB with standard depthwise convolution and one ESSB with dilated depthwise convolution, both to alleviate the "gridding issue" [59] caused by the standard dilated convolution and to reduce structural redundancy.

3.3.2. Residual local spectral attention module

Hyperspectral images usually have hundreds of spectral bands with strong correlations between adjacent bands. However, normal convolution treats all bands equally and thus cannot distinguish the differences among spectral bands and fully exploit the spectral correlation. To this end, a residual local spectral attention module is proposed, as shown in Fig. 4, which only involves a few parameters while bringing performance gain. This module applies a gating mechanism [60] to explicitly recalibrate features $x \in R^{h\times w\times C}$ as:

$$\hat{x} = x \odot (1 + F(x)) \tag{20}$$

where $\odot$ is channel-wise multiplication operation and $F(\cdot)$ represents a transformation consisting of channel-wise global average pooling (GAP), 1D convolution and Sigmoid function, to generate a set of spectral gate $g \in R^C$. It first performs average pooling along spatial dimensions to embed global spatial information, and then local cross-spectral interactions are captured by fast 1D convolution of size k. Since it considers every spectral band and its k neighbors, spectral correlation is exploited to enhance the network representation ability. Note that the spectral gate ranges from [0,1], which would degrade the value of features in deep layers [61]. Therefore, we introduce a residual connection to keep the good properties of original features.

The benefits of this module can be clearly seen in the backward propagation of the network. The gradient of this module can be written as

$$\nabla\hat{x} = \nabla x + \nabla x \odot F(x) + x \odot \nabla F(x) \tag{21}$$

On the one hand, the existence of ∇x ensures the gradient information can be backpropagated directly, which is a key to avoiding the gradient vanishing problem. On the other hand, ∇x is weighted by the spectral gates $F(x)$, which work as feature selectors to enhance informative features.

4. Experiments

4.1. Experimental settings

4.1.1. Data sets

We evaluate the proposed method on publicly-available aerial hyperspectral image datasets from three different sensors, as presented in Table 2, including Pavia Center and University,[1] Chikusei,[2] and Houston.[3] The Pavia Center and Pavia University datasets are two hyperspectral images of different scenes acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) during a flight campaign in the Pavia region of northern Italy in 2001. The ground sampling distance is 1.3 m. The spectral range being sensed is from 430 nm to 860 nm, with a total of 115 bands, some of which are removed due to water vapor absorption and noise. Specifically, Pavia Center has 102 bands with a spatial size of 1096×1096, but the middle part contains no information and has been discarded, leaving 1096×715 valid pixels. A sub-image of size 128×715×102 at the bottom is used for the test, and to

[1] https://ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes
[2] https://www.sal.t.u-tokyo.ac.jp/hyperdata/
[3] https://hyperspectral.ee.uh.edu/?page_id=1075

Table 3
Quantitative comparisons of different methods on Pavia center, Chikusei and Houston with bicubic downsampling for scale factors 2, 3 and 4.

Dataset	Scale factor	Bicubic	IFN	3DFCNN	GDRRN	SAGRDN	ERCSR	EUNet	
Pavia center	2	PSNR	30.89	33.47	33.13	33.40	34.71	34.17	**35.03**
		SSIM	0.9544	0.9749	0.9729	0.9742	0.9802	0.9781	**0.9813**
		FSIM	0.9356	0.9697	0.9662	0.9680	0.9795	0.9749	**0.9804**
		SAM	3.87	3.43	3.50	3.60	3.09	3.26	**3.01**
	3	PSNR	27.82	29.28	28.97	29.44	30.17	29.70	**30.33**
		SSIM	0.9049	0.9362	0.9312	0.9386	0.9485	0.9420	**0.9499**
		FSIM	0.8619	0.9074	0.8999	0.9101	0.9291	0.9152	**0.9301**
		SAM	4.86	4.58	4.63	4.60	4.14	4.53	**4.07**
	4	PSNR	26.19	27.14	27.00	27.35	27.63	27.40	**27.63**
		SSIM	0.8563	0.8931	0.8897	0.8993	0.9062	0.8996	**0.9064**
		FSIM	0.7997	0.8482	0.8435	0.8577	**0.8713**	0.8553	0.8710
		SAM	5.53	5.35	5.34	5.22	4.96	5.36	**4.84**
Chikusei	2	PSNR	35.40	38.45	37.92	38.25	39.86	39.42	**39.99**
		SSIM	0.9843	0.9922	0.9913	0.9921	0.9942	0.9936	**0.9944**
		FSIM	0.9711	0.9880	0.9864	0.9879	0.9918	0.9904	**0.9920**
		SAM	1.79	1.42	1.50	1.50	1.22	1.34	**1.14**
	3	PSNR	31.56	33.53	33.03	33.67	34.67	34.24	**34.82**
		SSIM	0.9627	0.9769	0.9743	0.9781	0.9819	0.9799	**0.9825**
		FSIM	0.9285	0.9596	0.9545	0.9623	0.9703	0.9651	**0.9712**
		SAM	2.69	2.26	2.39	2.23	1.89	2.09	**1.80**
	4	PSNR	29.84	31.29	30.96	31.68	32.17	31.90	**32.30**
		SSIM	0.9401	0.9584	0.9553	0.9627	0.9658	0.9632	**0.9668**
		FSIM	0.8875	0.9269	0.9212	0.9365	0.9460	0.9358	**0.9462**
		SAM	3.41	2.97	3.07	2.77	2.51	2.75	**2.41**
Houston	2	PSNR	35.29	37.80	37.30	36.29	38.41	38.45	**38.49**
		SSIM	0.9937	0.9966	0.9962	0.9956	0.9971	0.9970	**0.9971**
		FSIM	0.9911	0.9967	0.9961	0.9944	0.9975	0.9974	**0.9975**
		SAM	1.32	1.09	1.15	1.52	1.02	1.02	**0.98**
	3	PSNR	30.93	33.20	32.64	32.33	34.34	34.07	**34.41**
		SSIM	0.9821	0.9898	0.9885	0.9885	0.9924	0.9916	**0.9925**
		FSIM	0.9699	0.9853	0.9833	0.9821	0.9897	0.9881	**0.9898**
		SAM	2.06	1.76	1.84	2.20	1.49	1.64	**1.45**
	4	PSNR	28.12	29.88	29.43	29.50	31.10	30.68	**31.11**
		SSIM	0.9648	0.9776	0.9756	0.9774	0.9836	0.9812	**0.9837**
		FSIM	0.9412	0.9659	0.9630	0.9654	**0.9759**	0.9707	0.9758
		SAM	2.81	2.45	2.57	2.75	2.05	2.35	**1.98**

avoid the interference of discontinuity between the left and right parts, center cropping is performed on the left and right sides of the test area, respectively, to obtain four non-overlap images of size 128×128×102. Pavia University has 103 bands with a spatial size of 610×340. An original sub-region with rich details of size 200×200×103 is selected for the real experiment.

The Chikuesi dataset was taken by the Headwall Hyperspec-VNIR-C imaging sensor over agricultural and urban areas in Chikusei, Ibaraki, Japan, with a ground sampling distance of 2.5 m. It contains 2517×2335 pixels with 128 bands covering the spectral range from 363 nm to 1018 nm. Following the experimental setup of [16], center cropping is first performed for missing edge information to obtain a sub-image of size 2304×2048×128. The top region of this image is further extracted as the test set containing four non-overlap images of size 512×512×128. As part of the 2018 IEEE GRSS Data Fusion Contest, the Houston dataset was captured by the ITRES CASI 1500 spectral imager, covering the University of Houston campus and its surrounding urban area, with a ground sampling distance of 1 m. It contains a total of 48 bands in the 380 nm to 1050 nm range and spatial size of 4172×1202. The bottom region of size 128×1202×48 is chosen as the test data, and eight non-overlap images of size 128×128×48 are further obtained by center cropping.

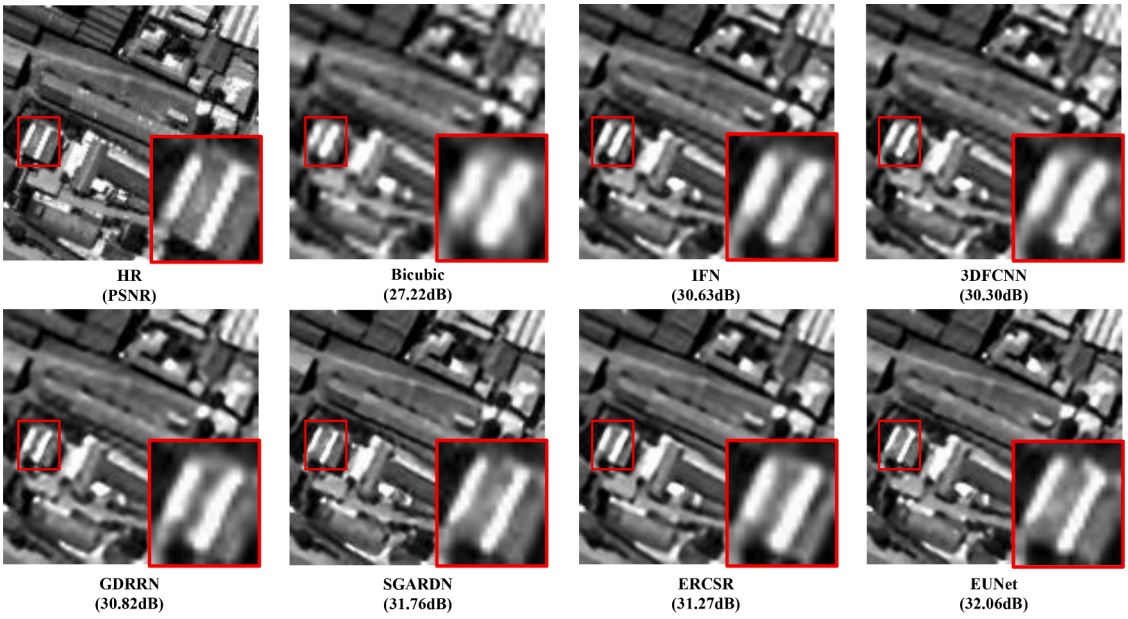
Overlap patches are extracted from the remaining regions of each dataset as reference HR images for training, of which 10% is used for validation. When the scale factor is 2 and 4, the size of each patch is 64×64 pixels, or 63×63 pixels when the scale factor is 3. To fully demonstrate the effectiveness of our proposed method, we use two degradation models to simulate LR images. The first one is bicubic downsampling by adopting the Matlab function imresize with the option bicubic. We use the first degradation model to simulate LR images with scale factors 2, 3, and 4. The second one is Gaussian downsampling according to Eq. (2), where HR images are first bicubicly downsampled with the scale factor 3, and then blurred by Gaussian kernel of size 7×7 with a standard deviation 1.6.

4.1.2. Evaluation metrics

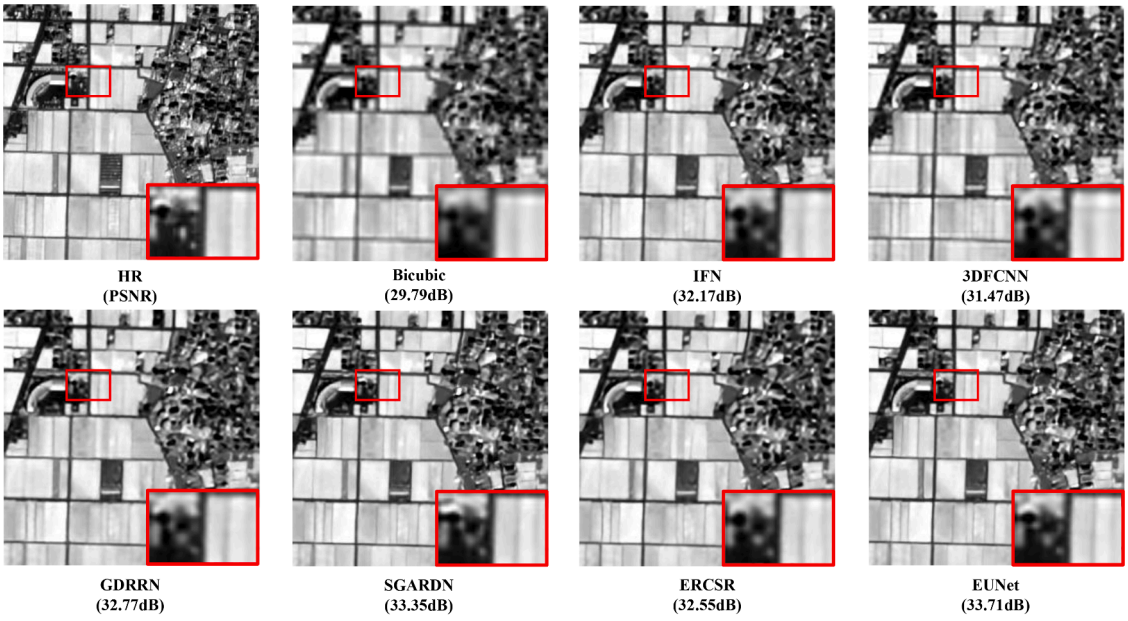
As for objective comparisons, we use four widely-used evaluation metrics to comprehensively assess the performance of different methods in terms of spatial, perceptual, and spectral aspects, including Peak Signal-to-Noise Ratio (PSNR), Structure Similarity Index (SSIM) [62], Feature Similarity Index (FSIM) [63], and Spectral Angle Mapper (SAM). For each 2D band image, PSNR, SSIM, and FSIM are calculated to measure the similarity between the reconstructed HSI and the ground truth based on the mean square error (MSE), structural consistency, and phase congruency and gradient magnitude, respectively. Their individual averages over all spectral bands are used as indicators of the spatial quality of the whole HSI. SAM calculates the angle between two vectors of the recovered and reference spectra to quantify the spectral consistence at each pixel. Its average over all pixels is used as the spectral quality indicator of the whole HSI. Larger PSNR, SSIM and FSIM values mean better results, while smaller values of SAM indicate less spectral distortion.

4.1.3. Competing methods

We compare the proposed EUNet with six representative HSI SR methods, including Bicubic, IFN [20], 3DFCNN [13], GDRRN [14], SGARDN [22], and ERCSR [17]. To be specific, Bicubic is the classic interpolation method; IFN is the recent advanced band-wise method; 3DFCNN and GDRRN are classic 3D and 2D convolution-based methods, respectively; SGARDN and ERCSR are state-of-the-art 2D convolution-based and 2D-3D hybrid convolution-based methods, respectively. We reproduce IFN and SGARDN in the Pytorch framework,



(a) A sub-image in the 50th band of Pavia Centre



(b) A sub-image in the 85th band of Chikusei

Fig. 5. Visual comparison of spatial SR with bicubic downsampling for the scale factor 3.

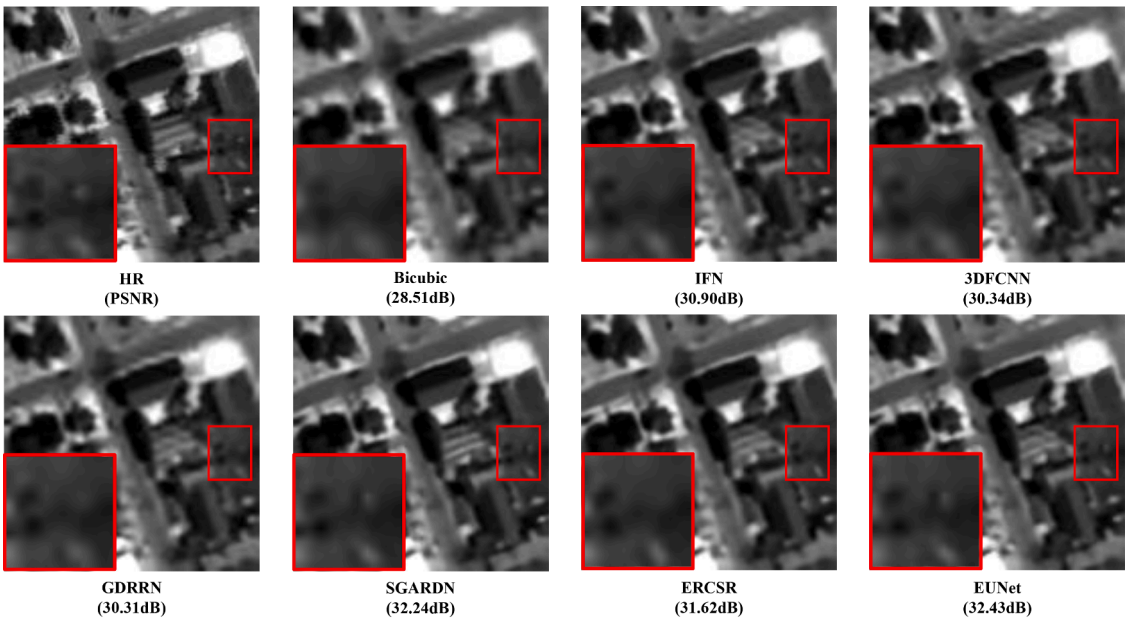
and the rest methods use open-source codes, all following the parameter settings in corresponding papers and trained until convergence using the same dataset.

4.1.4. Implementation details

Unless otherwise specified, we set the iteration T to 4 and the number of ESSG K to 3 for each iteration. Within each ESSB, the number of feature maps F is 128, the expansion rate e is 2, and the dilation rate r is 2. We implement our EUNet with the PyTorch framework on two Nvidia RTX 2080Ti GPUs. The Adam [64] solver with the default setting is adopted to optimize the parameters of EUNet. The learning rate is initialized to $1e-4$ for bicubic downsampling and $1e-3$ for Gaussian downsampling, and decreases by half for every 100 epochs. Considering the scale of the network, 32 LR patches are randomly cropped from training images as input for each training batch and are augmented by randomly flipping vertically and rotating 90°, 180°, and 270° to alleviate the overfitting and improve the generalization ability of the network.

4.2. Experimental results

To comprehensively evaluate the effectiveness of the proposed method, bicubic downsampling and Gaussian downsampling simulation experiments are carried out on three hyperspectral image datasets with different scenes, as well as a real experiment. Pavia Center and Houston datasets were acquired with smaller ground sampling distances over urban areas where detailed textures are relatively complex and adequate, while Chikusei dataset is a mixed urban and rural scene with



(c) A sub-image in the 35th band of Houston

Fig. 5. (*continued*).

large areas of structurally simple farmland, but the structural features of densely built-up areas are not sufficiently clear due to the lower spatial resolution. Compared to the Pavia Center and Chikusei datasets with hundreds of bands, the Houston dataset has only 48 bands so that the spectral reconstruction is easier.

4.2.1. Experiment1: Bicubic downsampling

Quantitative results. Table 3 shows quantitative comparisons for $\times 2$, $\times 3$, and $\times 4$ SR on three hyperspectral image datasets, where bold indicates the best result. Our EUNet performs the best on all datasets with all scale factors. This is because it alleviates the difficulty of feature learning and overfitting by disentangling spatial and spectral feature extraction while enhancing spectral consistency by RLSAM to embed spectral correlation, and SAM loss to guide spectral reconstruction. IFN, 3DFCNN, and GDRRN can obtain good results for considering spectral correlation but are still limited by the model capacity. The 2D unit in ERCSR improves the spatial learning ability but fails to exploit the complementary information between bands since it treats each band separately; in addition, the 3D unit has difficulty in exploring the global spectra due to the limitation of the receptive field, resulting in unsatisfactory spectral reconstruction results. SGARDN employs a residual dense structure to fully utilize hierarchical features and introduces a second-order spectral attention module for global spectral relationship modeling, thus enhancing spatial-spectral feature representation to achieve the second-best performance.

Visual results. Some test regions in each dataset for the scale factor 3 are displayed to qualitatively analyze different methods. Fig. 5 shows the spatial SR results and enlarges the same area marked by the red rectangle for better visualization. In comparison, our EUNet generates the most visually pleasing results. Concretely, our method recovers sharper edges and outlines of the building, while the reconstruction results of other methods are blurry. All the field ridges on Chikusei dataset reconstructed by methods except SGARDN and EUNet suffer from distortion. In addition, the majority of methods fail to capture the target in the upper right of the red box on Houston dataset, and only SGARDN and EUNet can successfully recover it.

Fig. 6 presents the SAM visualizations, which reflect the spectral reconstruction quality of each pixel so that it is clear where spectral distortions occur. In contrast, our EUNet yields fewer spectral anomalies, while other methods are prone to noticeable spectral distortions at the edges and shadows. The example spectral curves are shown in Fig. 7. The spectral pattern of Houston is relatively simple such that all methods preserve the features of the reference spectral curve. For Pavia center and Chikusei, the spectral curves reconstructed by EUNet are closer to the complex reference ones. It is demonstrated that our method has a great advantage in spectral consistency.

4.2.2. Experiment2: Gaussian downsampling

Quantitative results. Table 4 shows quantitative comparisons for SR on three hyperspectral image datasets, where bold indicates the best result. From the results of Bicubic, the Gaussian downsampling degradation model is more challenging, as reflected by the decrease in PSNR of 5.17 dB, 5.74 dB and 8.95 dB and the increase in SAM of 3.18, 2.97 and 4.12 on the three data sets, respectively. Evidently, our EUNet achieves the best spatial reconstruction results on all datasets except Pavia center, and all the best spectral reconstruction results, suggesting that our method can better deal with serious ill-posed cases by fully exploiting spatial-spectral priors and imposing degradation constraints. IFN and 3DFCNN give rise to poor performance attributed to the limited capability of the shallow model. Although GDRRN increases depth by recursive structure, the small number of parameters still restricts model ability. By comparison, the high-complexity models exhibit a clear superiority in that the quantitative results are all substantially improved, which reveals the importance of network structure and learning strategy for HSI SR. ERCSR obtains the second-best result, indicating that 2D and 3D hybrid convolution can enhance discriminative spatial-spectral prior learning ability to handle challenging scenarios.

Visual results. Part of test regions for the scale factor 3 are shown for qualitative comparison. Fig. 8 depicts the spatial SR results and highlights the same area marked by the red rectangle for better observation. It is clear that the proposed method performs best in recovering the details of the original HSIs. Specifically, LR images lose most of the high-frequency information due to complex Gaussian downsampling, but Bicubic cannot recover any details. Low-complexity models (i.e., IFN, 3DFCNN and GDRRN) using interpolated LR images as input produce very unpleasant artifacts and are unable to alleviate the blurriness. The reconstructions of SGARDN, ESCSR, and EUNet have some similarity, and all of them can recover some correct details. For example, on Pavia

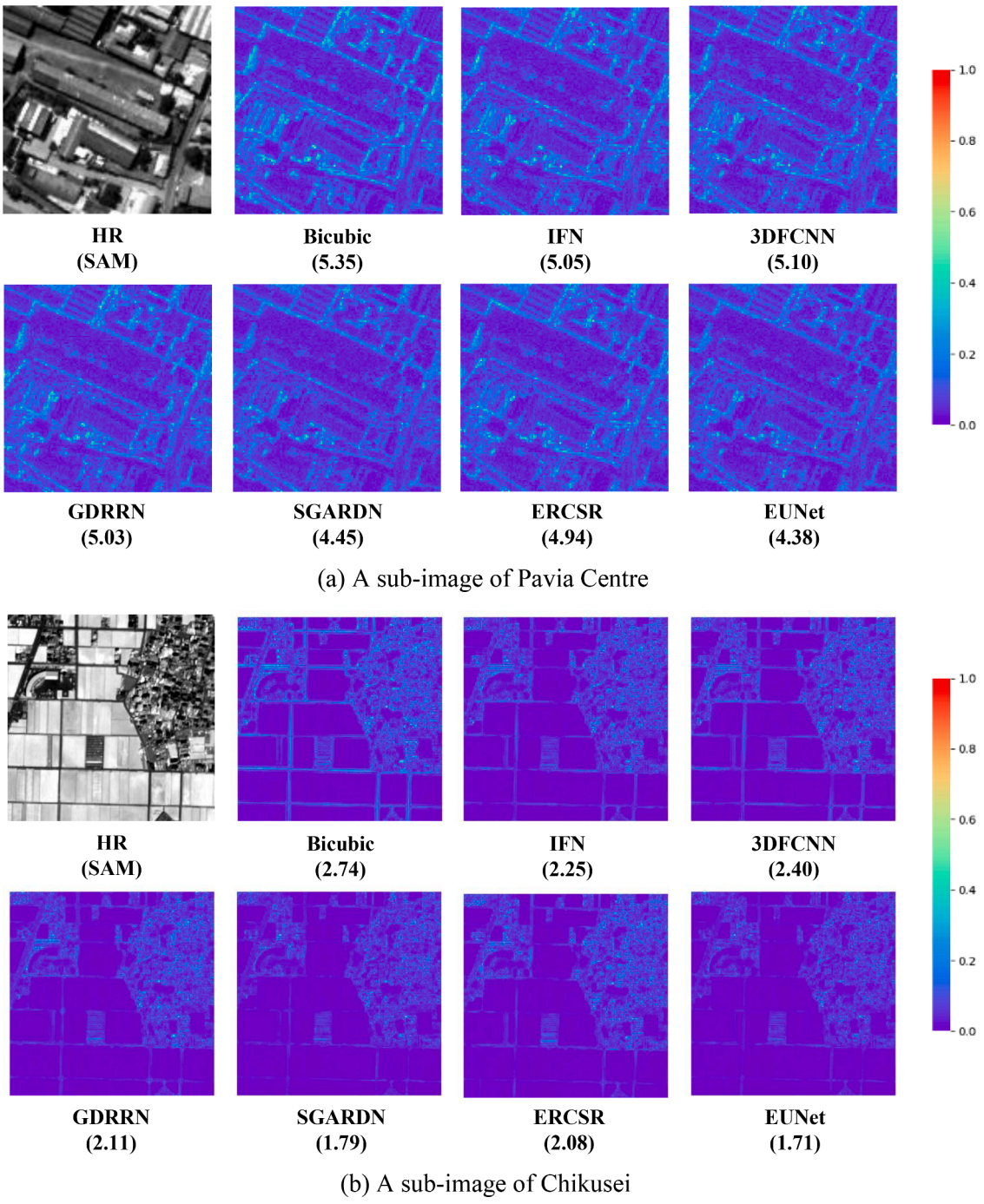


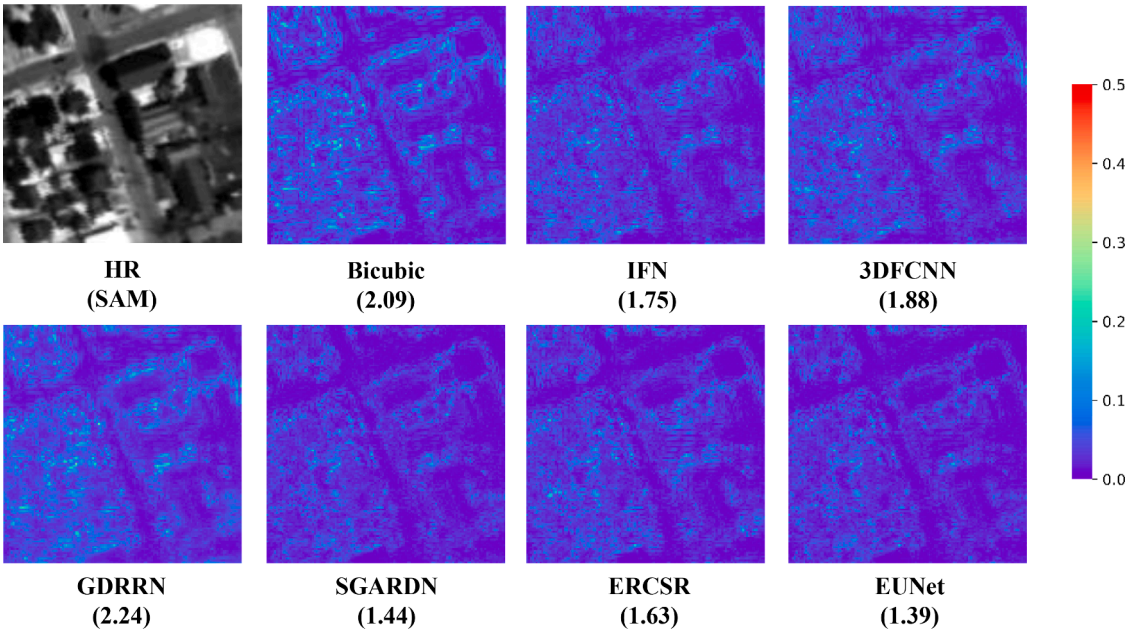
Fig. 6. SAM visualizations with bicubic downsampling for the scale factor 3.

Center, ERCSR results are closer to the ground-truth in overall shape, but different targets are connected, while EUNet results are consistent with the target separation phenomenon. EUNet ensures the continuity of the intermediate river on Chikusei dataset, while the rivers reconstructed by other methods all appear distorted and disconnected. In addition, EUNet generates sharper edges on Houston.

The SAM visualizations are shown in Fig. 9. We notice that Bicubic, IFN and 3DFCNN appear severe spectral distortion in a large area, and GDRRN can alleviate spectral distortion to some extent by SAM loss. In contrast, sophisticated models (i.e., SGARDN, ERCSR and EUNet) that combine hyperspectral image characteristics with network structures, greatly facilitate spectral preservation. Fig. 10 depicts the example spectral curves, which indicates that the spectral curves reconstructed by EUNet can better fit various reference spectral curves. The above results verify that our method can fully explore spectral dependencies to maintain the important high-dimensional spectral information.

4.2.3. *Real experiment*

We further carry out the real experiment on Pavia University to verify the effectiveness of our method under unknown degradations. Pavia University without downsampling is fed into the $\times 2$ SR model trained on Pavia Center to improve original resolution. Note that we discard the last band of Pavia University for unity. Since the HR image is unavailable in this case, we perform a visual inspection of the SR results,



(c) A sub-image of Houston

Fig. 6. (*continued*).

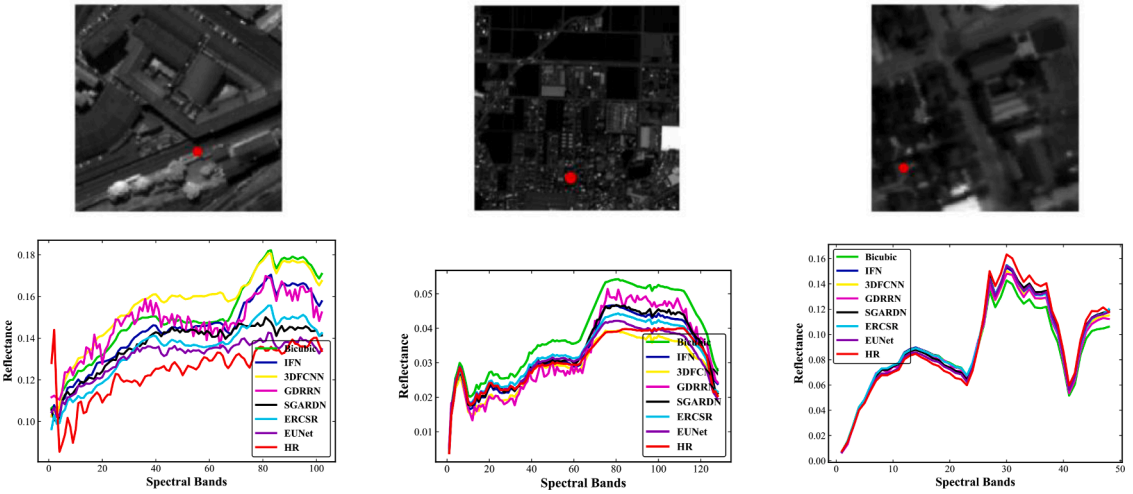
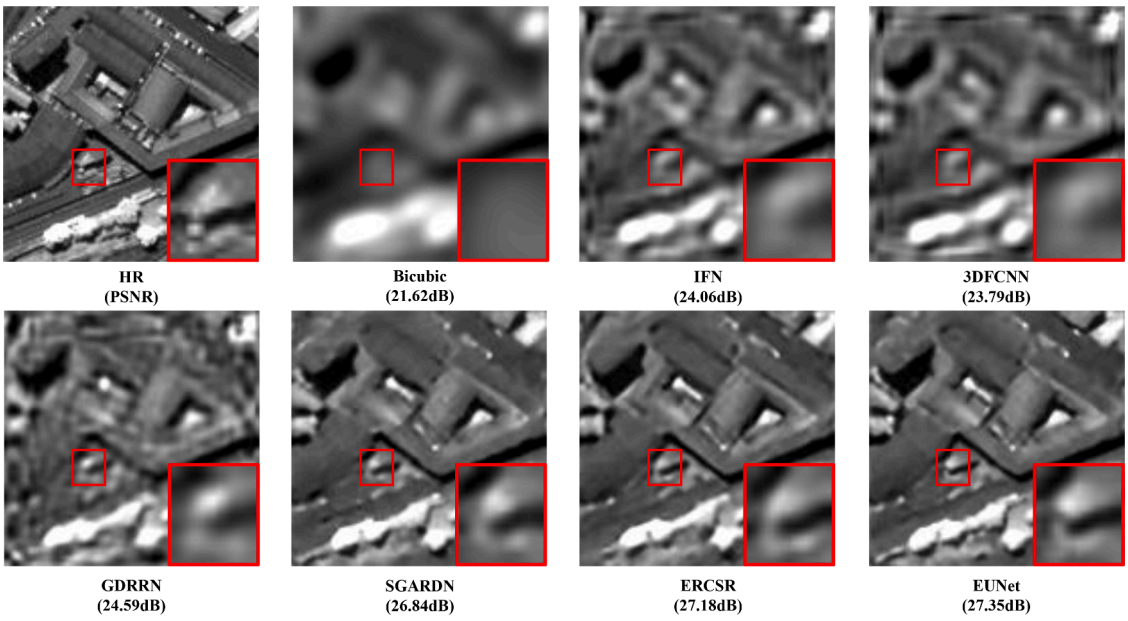


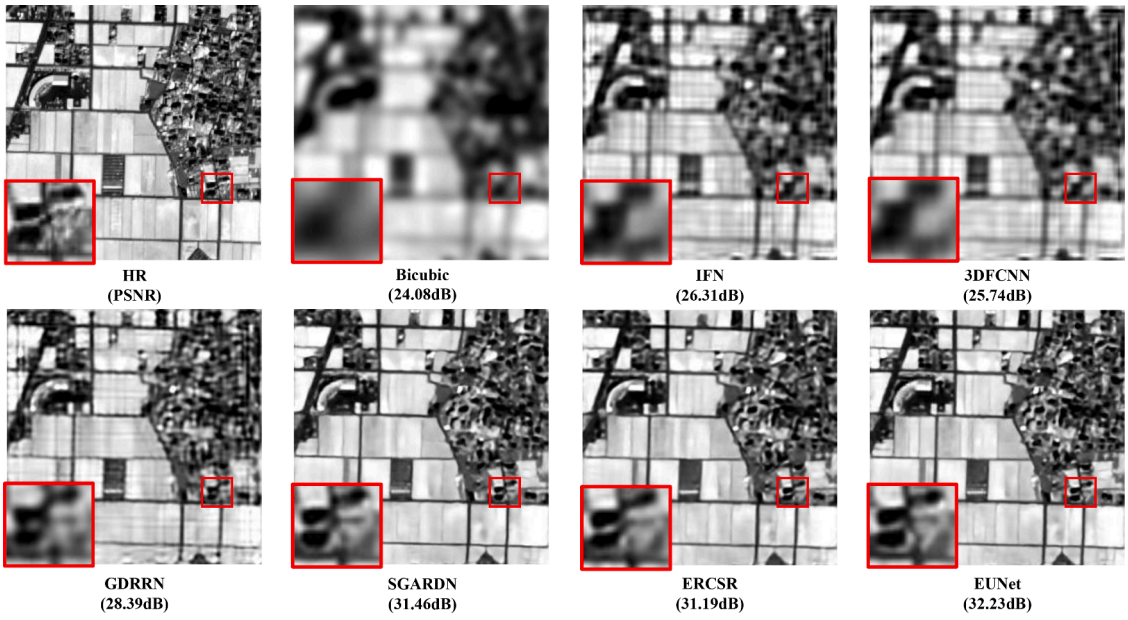
Fig. 7. Example spectral curves of selected locations on Pavia Center, Chikusei, and Houston sequentially, with bicubic downsampling for the scale factor 3.

Table 4
Quantitative comparisons of different methods on Pavia center, Chikusei and Houston with Guassian downsampling for the scale factor 3.

Dataset	Scale factor	Bicubic	IFN	3DFCNN	GDRRN	SAGRDN	ERCSR	EUNet	
Pavia center	3	PSNR	22.65	24.64	24.30	25.33	27.15	**27.80**	27.55
		SSIM	0.5950	0.7975	0.7818	0.8392	0.8933	**0.9090**	0.9042
		FSIM	0.5854	0.7627	0.7465	0.8154	0.8588	0.8690	**0.8705**
		SAM	8.04	7.03	7.36	6.81	5.50	5.61	**5.26**
Chikusei	3	PSNR	25.82	28.11	27.58	29.24	32.33	32.71	**32.92**
		SSIM	0.8344	0.9156	0.9046	0.9391	0.9690	0.9716	**0.9734**
		FSIM	0.7353	0.8636	0.8469	0.9107	0.9473	0.9505	**0.9552**
		SAM	5.66	4.27	4.54	3.97	2.69	2.64	**2.40**
Houston	3	PSNR	21.98	24.92	24.22	25.93	29.72	30.53	**30.75**
		SSIM	0.8299	0.9283	0.9163	0.9480	0.9778	0.9809	**0.9828**
		FSIM	0.7962	0.9055	0.8852	0.9355	0.9693	0.9723	**0.9760**
		SAM	6.18	4.17	4.50	4.27	2.58	2.64	**2.23**



(a) A sub-image in the 85th band of Pavia Centre



(b) A sub-image in the 80th band of Chikusei

Fig. 8. Visual comparison of spatial SR with Gaussian downsampling for the scale factor 3.

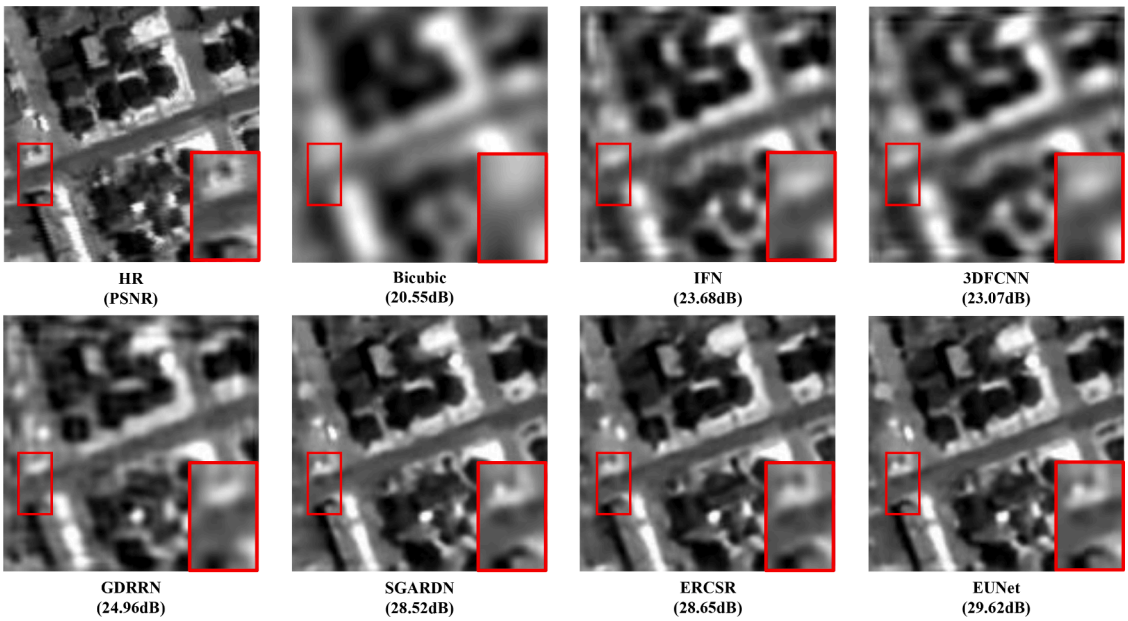
as shown in Fig. 11, which provides the reconstructed composite images with bands 28–13–5 as R-G-B. It can be seen that our method recovers sharper edges and finer details, while other methods show varying degrees of blurring and artifacts. We also provide the no-reference image quality score (BRISQUE) for spatial quantitative evaluation [65]. The lower value indicates the better perceptual quality. Our method still shows competitive performance. According to the above investigations, the spatial-spectral priors learned by EUNet perform robustly in real scenarios.

5. Discussion

The proposed EUNet incorporates the domain knowledge of spectral correlation, degeneration model, and structure prior. In this section, we perform a series of experimental analyses for the scale factor 2 on Pavia Center with bicubic downsampling to verify the effectiveness of the network components.

5.1. Parameter analysis

We investigate the basic network parameters, including the ESSG number K, the number of iterations T, and the number of feature maps F. As shown in Table 5, we report the number of parameters and floating point operations (FLOPs), as well as evaluation metrics. We first explore various combinations of K and T to construct different EUNet structures with different depths, and see how these two parameters affect the performance by variable control methods within an appropriate range. When more ESSGs are stacked, the parameter number linearly increases.



(c) A sub-image in the 35th band of Houston

Fig. 8. (*continued*).

When more iterations are performed, the parameter number keeps the same but the FLOPs linearly increases. Larger K can enhance the feature representation contributing to better performance, but the performance will decrease when $K > 3$, which may be attributed to overfitting caused by limited training samples. The performance improves as the number of iterations increases, especially in spatial reconstruction, because DC and SRP modules can facilitate each other during iterations by alternately imposing degradation constraints and reusing network parameters to recover details. However, the performance decreases slightly when $T > 4$, which indicates that our method can quickly converge with a few iterations while on the other hand, it is non-trivial to train deep and complex networks. In addition, we construct a very lightweight model by changing the number of feature maps from 128 to 64 for the sake of real-time processing. It can be observed that the width of the network has an apparent effect on performance. A narrower network is not sufficient to characterize complex HSIs but can significantly reduce the model complexity. Thanks to the dimension expansion before spatial feature extraction, this very lightweight network can still achieve promising reconstruction results.

5.2. Ablation study

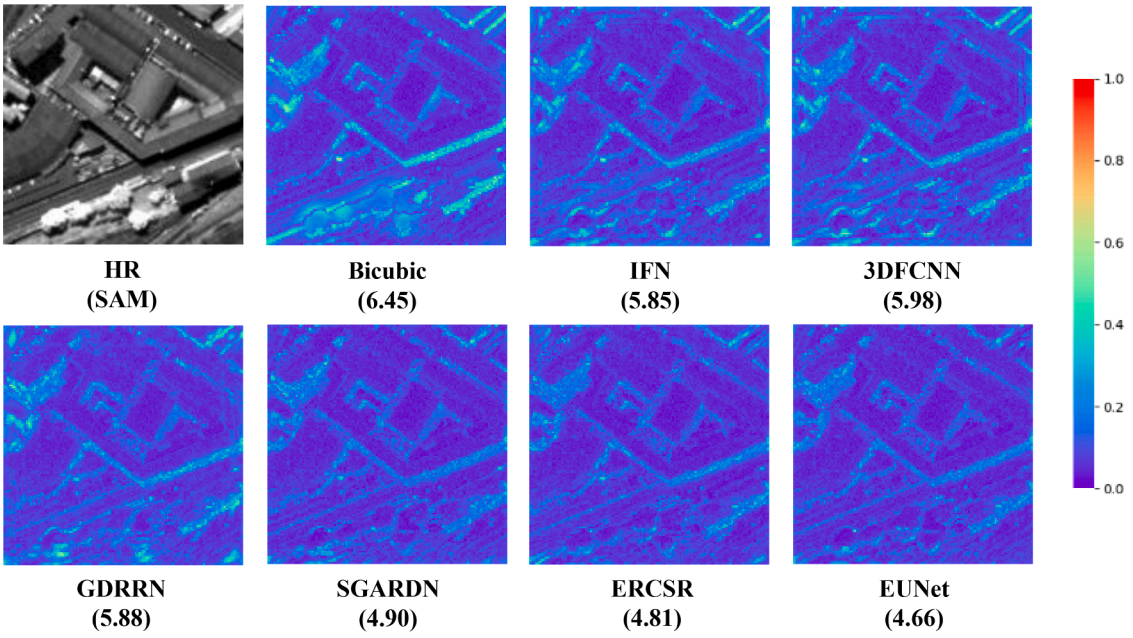
Table 6 shows the ablation study on the effects of spectral preservation, multi-scale feature, and RLSAM. These four networks have the same basic structure and similar complexity. The network in the first row removes the residual connections at each stage (i.e., the red arrows in Fig. 2) so that the end of HSRPN cannot directly access to LR images. We find that spectral preservation has a pronounced effect on the results of both spatial and spectral reconstruction. This is because residual connections can facilitate reconstructed images to share abundant spectral information of LR images. If we set the dilation rate to 1, the network cannot explicitly extract multi-scale features. The comparison between the second and last rows can validate that multi-scale features help in spatial reconstruction. Since analyzing more neighboring pixels provides more clues to the ill-posed SR problem, the network benefits from dilated convolutions with larger receptive fields. RLSAM is capable of mining complementary information in adjacent bands by embedding spectral correlation in a lightweight manner, and therefore the performance is slightly degraded when RLSAM is removed (third row). As discussed above, all three strategies contribute to HSI SR and are complementary in terms of spatial-spectral representation. With respect to the number of parameters and FLOPs, they can be seamlessly integrated into the network while introducing negligible computational cost. Apart from this, we also analyze the role of residual connection within RLSAM. Removing this connection, we observe a decrease in PSNR/SSIM/SAM metrics (35.00/0.9813/3.02), which is consistent with the analysis in Section 3.3.2 and indicates that the residual connection can maintain the good properties of original features.

5.3. Investigation of encoder-decoder architecture

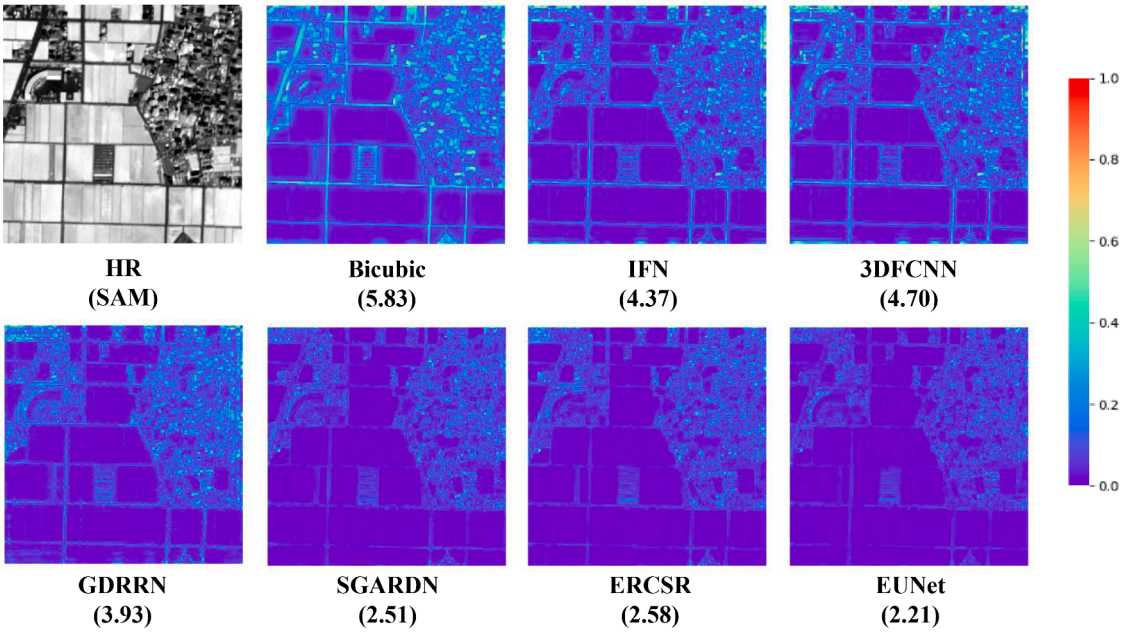
As an alternative, we investigate the popular encoder-decoder architecture for multi-scale feature extraction and fusion. UNet [66] is applied to hyperspectral image super-resolution by modifying the input and output channels. However, we observe a significant decrease in the quantitative results from the first row of Table 7. This is presumably due to the following reasons: first, the large number of model parameters leads to training difficulties and overfitting; second, the details of the image content are lost during multiple downsampling operations. Since the ground sampling distance of HSI is large (i.e., a pixel covers a wide area), the downsampling operation causes more serious information loss, which is difficult for the decoder to recover. Considering the model complexity, we use a two-stage UNet (i.e., containing only two downsampling operations) as the deep feature extractor of HSPRN (middle part of Fig. 3) and refer to this network as the efficient deep unfolding UNet (EUUNet). As shown in Table 7, EUUNet improves the results of UNet under our proposed super-resolution prior-driven unfolding framework but is still not comparable to EUNet restricted by the above-mentioned reasons. In contrast, the proposed EUNet incorporates the domain knowledge of HSI (e.g., structure prior and spectral correlation) in the feature extraction process to facilitate the learning of informative representations.

5.4. Data consistency module analysis

The proposed EUNet is designed in accordance with the interpretable MAP optimization framework, consisting of an almost parameter-free data consistency module and a deep learning-based super-resolution



(a) A sub-image of Pavia Centre



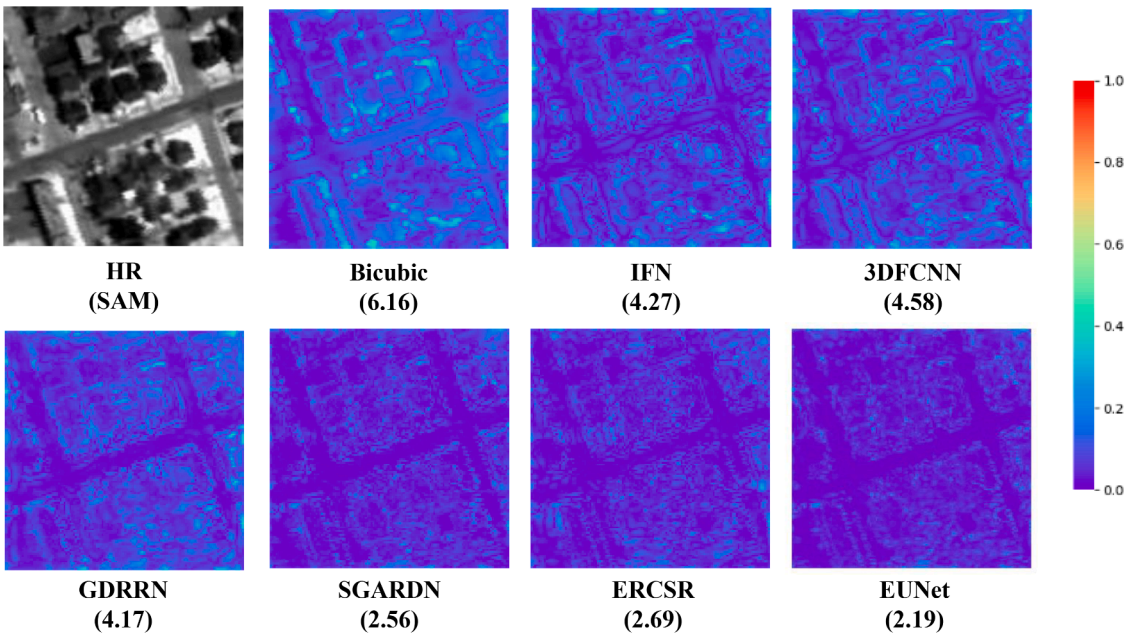
(b) A sub-image of Chikusei

Fig. 9. SAM visualizations with Gaussian downsampling for the scale factor 3.

prior module. In this subsection, we analysis this architecture that combines model-based optimization and deep learning. By directly removing the data consistency (DC) module, the traditional MAP optimization process is broken. Table 8 shows the comparison results under different degradation models. As we can see, the performance of the model without the data consistency module noticeably degrades, indicating that the two subproblems under the MAP framework are complementary to each other and both need to be considered. The performance degradation is more severe in the case of challenging Gaussian downsampling, which demonstrates that the data consistency module can effectively alleviate the ill-posedness of SR by imposing degradation model constraints on the results of super-resolution prior network.

5.5. *Disentangled spatial-spectral representation*

We propose the DSSM based on depthwise separable convolution for disentangled spatial-spectral representation. To verify its validity, we construct two other models by replacing DSSM with the commonly used residual block (RB) and make one of them similar to EUNet in terms of number of parameters by reducing the number of modules. Table 9 reflects that the disentangled spatial-spectral representation plays an important part in HSI SR. The model in the first row has limited performance due to insufficient depth, and the second row indicates that



(c) A sub-image of Houston

Fig. 9. (*continued*).

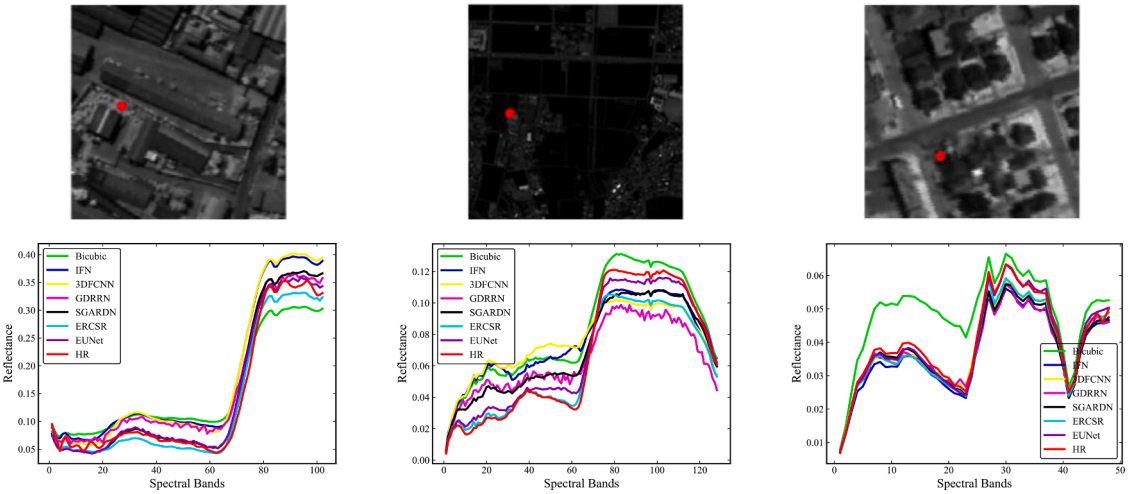


Fig. 10. Example spectral curves of selected locations on Pavia Center, Chikusei, and Houston sequentially, with Guassian downsampling for the scale factor 3.

the model with a large number of parameters is prone to overfitting, especially in the case of limited training samples. The proposed DSSM achieves an efficient representation of HSI, since it not only disentangles the 3D HSI into spatial and spectral subspaces to reduce the difficulty of feature learning, but also decreases the complexity of 2D convolution while increasing the network depth to enhance the feature representation.

5.6. Denoising prior vs. super-resolution prior

To demonstrate the superiority of the super-resolution prior-driven MAP framework, Table 10 provides a comparison between the denoising prior commonly used in the unfolding strategy and the super-resolution prior employed in this paper. Specifically, the denoising prior-driven MAP framework decouples SR problem into a data term and a prior term encoded implicitly by the denoiser. We modify the data consistency module of EUNet according to [38] and replace the upsampling operator in HSRPN with a convolutional layer so that HSRPN works as a deep denoiser. In terms of reconstruction performance, although they are similar in the bicubic downsampling case, the super-resolution prior has a remarkable advantage in the Gaussian downsampling case, primarily because it is non-trivial to efficiently handle the data term in the denoising prior framework that includes both deblurring and upsampling, while the data item in the super-resolution prior framework involves only deblurring, which means that it is possible to more fully use the power of deep learning. In addition, we note that the denoising prior substantially increases the FLOPs, which is attributed to the fact that deep feature extraction is performed in the high-resolution space. That is, the super-resolution prior framework can inherit the speed advantage of existing SR networks.

5.7. Model complexity

Fig. 12 shows the comparison in terms of performance, network

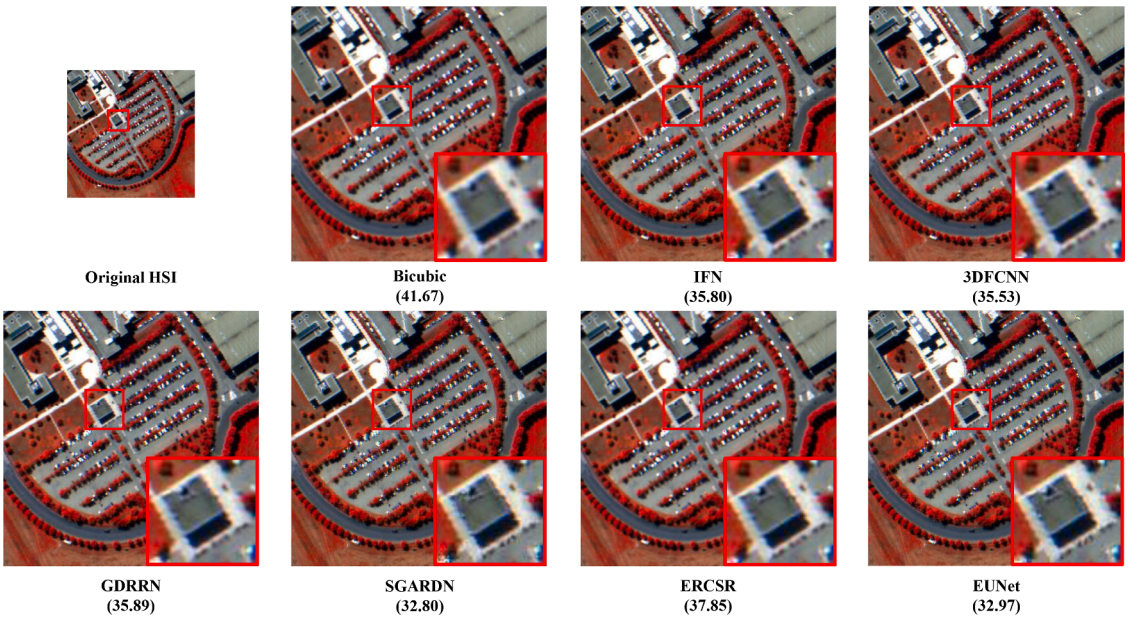


Fig. 11. Real experiment results on Pavia University.

Table 5
Parameter analysis of EUNet with different values of K, T and F.

Parameter type	Model	PSNR	SSIM	SAM	#params	FLOPs
K	EUNet ($T=1, K=1, F=128$)	34.64	0.9799	3.09	**442,291**	**16.20G**
	EUNet ($T=1, K=3, F=128$)	**34.79**	**0.9806**	**3.04**	713,671	20.68G
	EUNet ($T=1, K=5, F=128$)	34.70	0.9800	3.06	985,051	25.16G
T	EUNet ($T=1, K=3, F=128$)	34.79	0.9806	3.04	**713,671**	**20.68G**
	EUNet ($T=2, K=3, F=128$)	34.90	0.9809	3.02	713,866	41.35G
	EUNet ($T=4, K=3, F=128$)	**35.03**	**0.9813**	**3.01**	714,256	82.70G
	EUNet ($T=6, K=3, F=128$)	34.98	0.9812	3.02	714,646	124.06G
F	EUNet ($T=4, K=3, F=64$)	34.72	0.9801	3.08	**245,072**	**30.63G**
	EUNet ($T=4, K=3, F=128$)	**35.03**	**0.9813**	**3.01**	714,256	82.70G

Table 6
Ablation study of spectral preservation, multi-scale feature and RLSAM.

Model	PSNR	SSIM	SAM	#params	FLOPs
EUNet w/o Spectral Perservation	34.95	0.9811	3.05	714,256	82.70G
EUNet w/o Multi-scale Feature	34.96	0.9812	3.01	714,256	82.70G
EUNet w/o RLSAM	34.99	0.9812	3.02	**714,226**	**82.50G**
EUNet	**35.03**	**0.9813**	**3.01**	714,256	82.70G

Table 7
Investigation of encoder-decoder architecture for feature extraction.

Model	PSNR	SSIM	SAM	#params	FLOPs
UNet	33.97	0.9763	3.47	34,590,630	**69.68G**
EUUNet	34.10	0.9776	3.15	2,458,034	95.10G
EUNEt	**35.03**	**0.9813**	**3.01**	**714,256**	82.70G

Table 8
Analysis of data consistency module.

Degradation model	Model	PSNR	SSIM	SAM	#params	FLOPs
Bicubic downsampling	EUNet w/o DC	34.87	0.9808	3.03	**713,736**	**82.64G**
	EUNet	**35.03**	**0.9813**	**3.01**	714,256	82.70G
Gaussian downsampling	EUNet w/o DC	29.94	0.9441	4.51	**714,654**	**82.65G**
	EUNet	**31.99**	**0.9647**	**3.91**	717,928	83.50G

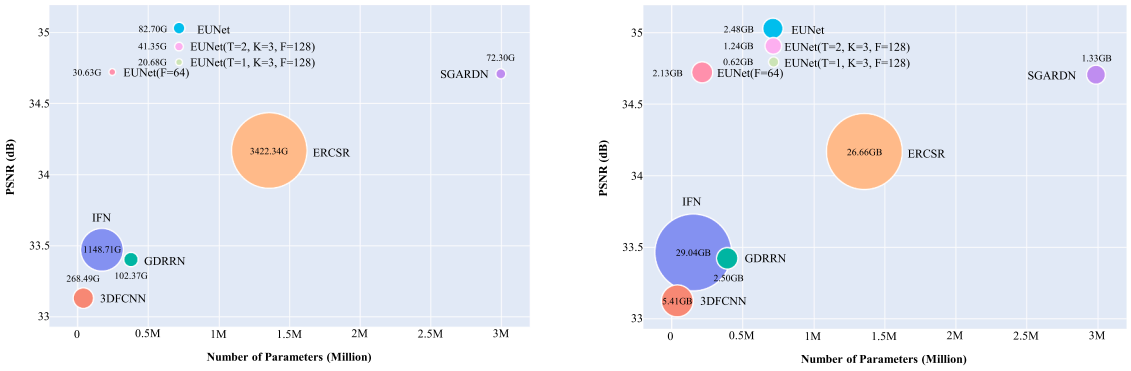
Table 9
Study on the effects of disentangled spatial-spectral representation.

Model	PSNR	SSIM	SAM	#params	FLOPs
EUNet-RB ($T=4, K=1, F=128$)	34.79	0.9804	3.07	796,176	94.55G
EUNet-RB ($T=4, K=3, F=128$)	34.61	0.9798	3.10	2,076,688	172.00G
EUNet ($T=4, K=3, F=128$)	**35.03**	**0.9813**	**3.01**	**714,256**	**82.70G**

parameters, FLOPs indicated by radiuses of circles in (a), and memory indicated by radiuses of circles in (b). Note that FLOPs and memory are calculated from all test images on Pavia center with a scale factor 2. The proposed EUNet achieves the best quality while having low model complexity. Specifically, IFN, 3DFCNN, and GDRRN all take the interpolated LR images as input, which increases the computation complexity quadratically. In addition, IFN is inherently a band-by-band approach that cannot perform fast processing of hyperspectral images and requires a large amount of memory. 3DFCNN demands large computational

Table 10
Comparison of denoising prior and super-resolution prior.

Degradation model	Prior type	PSNR	SSIM	SAM	#params	FLOPs
Bicubic downsampling	denoising	35.02	0.9813	**3.00**	796,176	277.42G
	super-resolution	**35.03**	**0.9813**	3.01	**714,256**	**82.70G**
Gaussian downsampling	denoising	31.19	0.9583	4.08	799,848	280.55G
	super-resolution	**31.99**	**0.9647**	**3.91**	**717,928**	**83.50G**



(a) PSNR vs. Parameters and FLOPs (b) PSNR vs. Parameters and Memory

Fig. 12. Model complexity comparison.

resources due to 3D convolution and spectral preservation. ERCSR further deepens the 3D CNN, leading to an increase in model complexity, particularly significant in FLOPs and memory demand. Although SGARDN has a relatively large number of parameters due to the deeper layers and hierarchical feature fusion, the amount of computational operations and memory is well controlled as a result of the small growth rate in residual dense blocks. Our EUNet develops disentangled spatial-spectral representations on top of the HSI structure prior, which greatly reduces the computational complexity of 2D convolution. Meanwhile, the complexity of EUNet can be easily scaled to achieve a good trade-off between performance and running time on different devices by controlling the number of iterations and features. As shown in Fig. 12, FLOPs can be reduced from 82.70 G to 20.68 G and memory from 2.48GB to 0.62GB without noticeable performance degradation.

6. Conclusion

In this study, we develop a novel HSI SR approach (EUNet) by combining the domain knowledge of spectral correlation, degradation model, and structure prior with deep learning to tackle the challenges posed by high spectral dimensionality, insufficient spatial resolution, and limited availability of training samples. The proposed EUNet, an interpretable multi-stage network, builds upon the super-resolution prior-driven MAP framework with an appropriate degradation assumption, which can extend existing super-resolution networks and explicitly impose the degradation model constraint. In addition, we design lightweight generic modules for spatial-spectral prior learning. Specifically, a disentangled spatial-spectral module is devised based on depthwise separable convolution to respect the inherent structural information of HSI and greatly reduce the difficulty and computational complexity of feature learning. Meanwhile, a residual local spectral attention module is proposed to embed spectral correlation with only a handful of parameters. Sufficient discussion confirms the effectiveness of the unfolding strategy and disentangled spatial-spectral representation for expressing HSIs and constraining the ill-posed SR. Extensive experimental results on public HSI datasets have demonstrated that the proposed EUNet has excellent performance under different scenarios, and its lightweight structure can satisfy the applications with limited computational resources. In the future, we will explore a unified deep framework using promising techniques (e.g., self-supervised learning, Transformers, and knowledge distillation) for HSIs with different imaging conditions and spectral band numbers to facilitate their real-world applications.

CRediT authorship contribution statement

Denghong Liu: Conceptualization, Methodology, Software, Data curation, Investigation, Validation, Writing – original draft. **Jie Li:** Conceptualization, Methodology, Writing – review & editing. **Qiangqiang Yuan:** Conceptualization, Methodology, Funding acquisition, Supervision. **Li Zheng:** Resources, Supervision. **Jiang He:** Resources, Writing – review & editing. **Shuheng Zhao:** Visualization, Writing – review & editing. **Yi Xiao:** Data curation, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 41922008, Grant 61971319, and Grant 62071341.

References

[1] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, A. Plaza, Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art, IEEE Geosci. Remote Sens. Mag. 5 (4) (2017) 37–78, https://doi.org/10.1109/MGRS.2017.2762087.

[2] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, X. Zhu, Interpretable hyperspectral artificial intelligence: when nonconvex modeling meets hyperspectral remote sensing, IEEE Geosci. Remote Sens. Mag. 9 (2) (2021) 52–87, https://doi.org/10.1109/MGRS.2021.3064051.

[3] S. Zhao, Q. Yuan, J. Li, Y. Hu, X. Liu, L. Zhang, A fast and effective irregular stripe removal method for moon mineralogy mapper, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–19, https://doi.org/10.1109/TGRS.2021.3054661.

[4] N. Yokoya, C. Grohnfeldt, J. Chanussot, Hyperspectral and multispectral data fusion: a comparative review of the recent literature, IEEE Geosci. Remote Sens. Mag. 5 (2) (2017) 29–56, https://doi.org/10.1109/MGRS.2016.2637824.

[5] M. Zhang, X. Sun, Q. Zhu, G. Zheng, A survey of hyperspectral image super-resolution technology, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 4476–4479, https://doi.org/10.1109/IGARSS47720.2021.9554409.

[6] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873, https://doi.org/10.1109/TIP.2010.2050625.

[7] Y. Zhao, C. Yi, J. Yang, J.C.-W. Chan, Coupled hyperspectral super-resolution and unmixing, in: Proceedings of the IEEE Geoscience and Remote Sensing Symposium, 2014, pp. 2641–2644, https://doi.org/10.1109/IGARSS.2014.6947016.

[8] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 295–307, https://doi.org/10.1109/TPAMI.2015.2439281.

[9] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 286–301.

[10] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image restoration, IEEE Trans. Pattern Anal. Mach. Intell. 43 (7) (2021) 2480–2495, https://doi.org/10.1109/TPAMI.2020.2968521.

[11] L. Liebel, M. Körner, Single-image super resolution for multispectral remote sensing data using convolutional neural networks, ISPRS-international archives of the photogrammetry, Remote Sens. Spat. Inf. Sci. 41 (2016) 883–890.

[12] Y. Li, J. Hu, X. Zhao, W. Xie, J. Li, Hyperspectral image super-resolution using deep convolutional neural network, Neurocomputing 266 (2017) 29–41.

[13] 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 (11) (2017) 1139.

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

[15] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-.J. Zha, F. Wu, Deep residual attention network for spectral image super-resolution, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, pp. 214–229.

[16] 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.

[17] Q. Li, Q. Wang, X. Li, Exploring the relationship between 2d/3d convolution for hyperspectral image super-resolution, IEEE Trans. Geosci. Remote Sens. 59 (10) (2021) 8693–8703, https://doi.org/10.1109/TGRS.2020.3047363.

[18] Y. Yuan, X. Zheng, X. Lu, Hyperspectral image superresolution by transfer learning, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 10 (5) (2017) 1963–1974, https://doi.org/10.1109/JSTARS.2017.2655112.

[19] W. Xie, X. Jia, Y. Li, J. Lei, Hyperspectral image super-resolution using deep feature matrix factorization, IEEE Trans. Geosci. Remote Sens. 57 (8) (2019) 6055–6067, https://doi.org/10.1109/TGRS.2019.2904108.

[20] J. Hu, X. Jia, Y. Li, G. He, M. Zhao, Hyperspectral image super-resolution via intrafusion network, IEEE Trans. Geosci. Remote Sens. 58 (10) (2020) 7459–7471, https://doi.org/10.1109/TGRS.2020.2982940.

[21] K. Zheng, L. Gao, Q. Ran, X. Cui, B. Zhang, W. Liao, S. Jia, Separable-spectral convolution and inception network for hyperspectral image super-resolution, Int. J. Mach. Learn. Cybern. 10 (10) (2019) 2593–2607.

[22] D. Liu, J. Li, Q. Yuan, A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution, IEEE Trans. Geosci. Remote Sens. 59 (9) (2021) 7711–7725, https://doi.org/10.1109/TGRS.2021.3049875.

[23] J. Yang, Y.Q. Zhao, J.C.-W. Chan, L. Xiao, A multi-scale wavelet 3d-cnn for hyperspectral image super-resolution, Remote Sens. 11 (13) (2019) 1557.

[24] Q. Li, Q. Wang, X. Li, Mixed 2d/3d convolutional network for hyperspectral image super-resolution, Remote Sens. 12 (10) (2020) 1660.

[25] J. Li, R. Cui, B. Li, R. Song, Y. Li, Y. Dai, Q. Du, Hyperspectral image super-resolution by band attention through adversarial learning, IEEE Trans. Geosci. Remote Sens. 58 (6) (2020) 4304–4318, https://doi.org/10.1109/TGRS.2019.2962713.

[26] J. Yang, L. Xiao, Y.Q. Zhao, J.C.W. Chan, Hybrid local and nonlocal 3-d attentive cnn for hyperspectral image super-resolution, IEEE Geosci. Remote Sens. Lett. 18 (7) (2021) 1274–1278, https://doi.org/10.1109/LGRS.2020.2997092.

[27] D. Liu, J. Li, Q. Yuan, Enhanced 3d convolution for hyperspectral image super-resolution, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 2452–2455, https://doi.org/10.1109/IGARSS47720.2021.9553962.

[28] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645.

[29] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3147–3155.

[30] N. Ahn, B. Kang, K.-.A. Sohn, Fast, accurate, and lightweight super-resolution with cascading residual network, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 252–268.

[31] Z. Hui, X. Wang, X. Gao, Fast and accurate single image super-resolution via information distillation network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 723–731.

[32] Z. Hui, X. Gao, Y. Yang, X. Wang, Lightweight image super-resolution with information multi-distillation network, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 2024–2032.

[33] X. Chu, B. Zhang, H. Ma, R. Xu, Q. Li, Fast, accurate and lightweight super-resolution with neural architecture search, in: Proceedings of the 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 59–64.

[34] X. Jiang, N. Wang, J. Xin, X. Xia, X. Yang, X. Gao, Learning lightweight super-resolution networks with weight pruning, Neural Netw. 144 (2021) 21–32.

[35] X. Luo, Q. Liang, D. Liu, Y. Qu, Boosting lightweight single image super-resolution via joint-distillation, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 1535–1543.

[36] Z. Wang, L. Li, Y. Xue, C. Jiang, J. Wang, K. Sun, H. Ma, Fenet: feature enhancement network for lightweight remote-sensing image super-resolution, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–12, https://doi.org/10.1109/TGRS.2022.3168787.

[37] V. Kothari, E. Liberis, N.D. Lane, The final frontier: deep learning in space, in: Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications, 2020, pp. 45–49.

[38] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, X. Lu, Denoising prior driven deep neural network for image restoration, IEEE Trans. Pattern Anal. Mach. Intell. 41 (10) (2019) 2305–2318, https://doi.org/10.1109/TPAMI.2018.2873610.

[39] D. Zoran, Y. Weiss, From learning models of natural image patches to whole image restoration, in: Proceedings of the International Conference on Computer Vision, 2011, pp. 479–486, https://doi.org/10.1109/ICCV.2011.6126278.

[40] S.H. Chan, X. Wang, O.A. Elgendy, Plug-and-play admm for image restoration: fixed-point convergence and applications, IEEE Trans. Comput. Imaging 3 (1) (2017) 84–98, https://doi.org/10.1109/TCI.2016.2629286.

[41] S.V. Venkatakrishnan, C.A. Bouman, B. Wohlberg, Plug-and-play priors for model based reconstruction, in: Proceedings of the IEEE Global Conference on Signal and Information Processing, 2013, pp. 945–948, https://doi.org/10.1109/GlobalSIP.2013.6737048.

[42] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep cnn denoiser prior for image restoration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2808–2817, https://doi.org/10.1109/CVPR.2017.300.

[43] K. Zhang, L. Van Gool, R. Timofte, Deep unfolding network for image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3214–3223, https://doi.org/10.1109/CVPR42600.2020.00328.

[44] Q. Ning, W. Dong, G. Shi, L. Li, X. Li, Accurate and lightweight image super-resolution with model-guided deep unfolding network, IEEE J. Sel. Top. Signal Process. 15 (2) (2021) 240–252, https://doi.org/10.1109/JSTSP.2020.3037516.

[45] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, D. Meng, Mhf-net: an interpretable deep network for multispectral and hyperspectral image fusion, IEEE Trans. Pattern Anal. Mach. Intell. 44 (3) (2022) 1457–1473, https://doi.org/10.1109/TPAMI.2020.3015691.

[46] W. Dong, C. Zhou, F. Wu, J. Wu, G. Shi, X. Li, Model-guided deep hyperspectral image super-resolution, IEEE Trans. Image Process. 30 (2021) 5754–5768, https://doi.org/10.1109/TIP.2021.3078058.

[47] J. He, Q. Yuan, J. Li, L. Zhang, Ponet: a universal physical optimization-based spectral super-resolution network for arbitrary multispectral images, Inf. Fusion 80 (2022) 205–225.

[48] Q. Ma, J. Jiang, X. Liu, J. Ma, Deep unfolding network for spatiospectral image super-resolution, IEEE Trans. Comput. Imaging 8 (2022) 28–40, https://doi.org/10.1109/TCI.2021.3136759.

[49] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, L. Zhang, Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–19, https://doi.org/10.1109/TGRS.2021.3107352.

[50] S. Gu, R. Timofte, L. Van Gool, Integrating local and non-local denoiser priors for image restoration, in: Proceedings of the 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2923–2928, https://doi.org/10.1109/ICPR.2018.8545043.

[51] A. Brifman, Y. Romano, M. Elad, Unified single-image and video super-resolution via denoising algorithms, IEEE Trans. Image Process. 28 (12) (2019) 6063–6076, https://doi.org/10.1109/TIP.2019.2924173.

[52] K. Zhang, W. Zuo, L. Zhang, Deep plug-and-play super-resolution for arbitrary blur kernels, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1671–1681, https://doi.org/10.1109/CVPR.2019.00177.

[53] S. Henrot, C. Soussen, D. Brie, Fast positive deconvolution of hyperspectral images, IEEE Trans. Image Process. 22 (2) (2013) 828–833, https://doi.org/10.1109/TIP.2012.2216280.

[54] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1637–1645, https://doi.org/10.1109/CVPR.2016.181.

[55] X. Fu, W. Wang, Y. Huang, X. Ding, J. Paisley, Deep multiscale detail networks for multiband spectral image sharpening, IEEE Trans. Neural Netw. Learn. Syst. 32 (5) (2021) 2090–2104, https://doi.org/10.1109/TNNLS.2020.2996498.

[56] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: image restoration using swin transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.

[57] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-.C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520, https://doi.org/10.1109/CVPR.2018.00474.
[58] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122 (2015).
[59] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, Understanding convolution for semantic segmentation, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 1451–1460, https://doi.org/10.1109/WACV.2018.00163.
[60] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, Eca-net: efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11531–11539, https://doi.org/10.1109/CVPR42600.2020.01155.
[61] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6450–6458, https://doi.org/10.1109/CVPR.2017.683.
[62] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[63] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[64] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[65] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708, https://doi.org/10.1109/TIP.2012.2214050.
[66] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.