SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2023, Vol. 66 140303:1–140303:15
https://doi.org/10.1007/s11432-022-3610-5

Special Topic: Artificial Intelligence Innovation in Remote Sensing

Learning the external and internal priors for multispectral and hyperspectral image fusion

Shutao LI[1], Renwei DIAN[2*] & Haibo LIU[2]

[1]*College of Electrical and Information Engineering, Hunan University, Changsha* 410082*, China;*
[2]*School of Robotics, Hunan University, Changsha* 410082*, China*

Received 28 April 2022/Revised 27 July 2022/Accepted 28 October 2022/Published online 27 March 2023

Abstract Recently, multispectral image (MSI) and hyperspectral image (HSI) fusion has been a popular topic in high-resolution HSI acquisition. This fusion leads to a challenging underdetermined problem, which image priors are used to regularize, aiming at improving fusion accuracy. To fully exploit HSI priors, this paper proposes two kinds of priors, i.e., external priors and internal priors, to regularize the fusion problem. An external prior represents the general image characteristics and is learned from abundant training data by using a Gaussian denoising convolutional neural network (CNN) trained in the additional gray images. An internal prior represents the unique characteristics of the HSI and MSI to be fused. To learn the external prior, we first segment the MSI into several superpixels and then enforce a low-rank constraint for each superpixel, which can well model local similarities in the HSI. In addition, to model a low-rank property in the spectral mode, the high-resolution HSI is decomposed into a low-rank spectral basis and abundances. Finally, we formulate the fusion as an external and internal prior-regularized optimization problem, which is efficiently tackled through the alternating direction method of multipliers. Experiments on simulated and real datasets demonstrate the superiority of the proposed method.

Keywords multispectral and hyperspectral image fusion, external and internal prior learning, high-resolution hyperspectral imaging

Citation Li S T, Dian R W, Liu H B. Learning the external and internal priors for multispectral and hyperspectral image fusion. Sci China Inf Sci, 2023, 66(4): 140303, https://doi.org/10.1007/s11432-022-3610-5

1 Introduction

Hyperspectral imaging techniques can simultaneously collect hundreds of spectral bands corresponding to different spectral wavelengths for a scene. Hyperspectral images (HSIs) with high spectral resolution ensure an accurate identification of materials, and therefore they find broad application in remote sensing [1–6], medical diagnosis [7], and face recognition [8]. However, because of the limitations of the imaging sensors, an inevitable tradeoff emerges between the spectral and spatial resolutions [9–11]. Therefore, an HSI of high spectral resolution can be obtained with low spatial resolution, which limits HSI application. Meanwhile, multispectral images (MSIs) with fewer spectral bands can be obtained with higher spatial resolution. Recently, MSI and HSI fusion has become an emerging technique to acquire high spatial resolution HSI. HSI and MSI fusion mainly has two applications. On the one hand, it can improve HSI spatial resolution, which can be used to design high-resolution HSI cameras. On the other hand, this fusion improves HSI performance in many high-level computer vision tasks, such as anomaly detection [12], object classification [13], and change detection [14].

MSI and HSI fusion is a challenging underdetermined problem. To tackle this problem, it can be formulated as an image prior-regularized optimization problem based on the maximum posterior probability (MAP) estimation rule. Therefore, image priors play a crucial role in improving fusion accuracy. We categorize the various image priors into two types, i.e., internal and external priors. An external prior represents the general image characteristics, while an internal prior represents the unique prior in the

* Corresponding author (email: drw@hnu.edu.cn)

HSI and MSI to be fused. However, the existing methods often only exploit one type of prior, hindering fusion performance.

To solve this problem, we propose a novel external and internal prior learning-based method (ExInL) for MSI and HSI fusion. First, to exploit the low-rank property in the spectral mode, the high-resolution HSI is decomposed into a low-rank spectral basis and abundances, and then the spectral basis is learned from the HSI. Next, we formulate the estimation of abundances as an external and internal prior-regularized optimization problem. To fully exploit the external prior, we incorporate a Gaussian denoising convolutional neural network (CNN), trained in the additional gray images, into the optimization. To model the internal prior, we first segment the MSI into several superpixels and then enforce the low-rank constraint for abundances in each superpixel, which fully exploits the local similarities of high-resolution HSI. The optimization problem is tackled efficiently through the alternating direction method of multipliers (ADMMs). The main contributions of this paper are summarized as follows.

(1) Based on the MAP, we formulate the HSI and MSI fusion as an internal and external prior-regularized optimization problem, which can effectively model HSI image priors.

(2) We propose a novel method for HSI and MSI fusion, which models the external prior by learning a Gaussian denoizer from the additional training data and learns the internal priors by imposing spatially local low-rank priors in the MSI and HSI.

(3) The experimental results on two simulated datasets and one real dataset demonstrate the effectiveness of the proposed method.

The remaining sections of this paper are organized as follows. Section 2 introduces the representative MSI and HSI fusion approaches. We give a detailed introduction to the proposed ExInL approach in Section 3. The experiments are presented in Section 4. The conclusion is given in Section 5.

2 Related work

In this section, the representative works in MSI and HSI fusion are reviewed. Ref. [15] gives a comprehensive review of HSI and MSI fusion. On the basis of the characteristics of these methods, we classify the fusion approaches into three types: pan-sharpening methods, model-based methods, and deep learning-based methods.

2.1 Pan-sharpening methods

Pan-sharpening methods [16–20] are originally designed for merging a high-resolution panchromatic (PAN) image with a low-resolution MSI. Representative pan-sharpening approaches include component substitution approaches and multi-resolution analysis approaches. Because MSI and HSI fusion has a similar form to pan-sharpening, some studies adapt pan-sharpening approaches to MSI and HSI fusion. For example, Ref. [21] first partitions the spectral bands of HSI as several groups according to the correspondence between the spectral bands of HSI and MSI and then uses pan-sharpening methods to fuse each spectral band of MSI with counterpart bands in HSI. Moreover, a high-resolution image is generated by linear regression, and then each band of the HSI is fused with the generated image [22]. MSI and HSI fusion methods derived from pan-sharpening often have a low computational burden. However, a fused HSI often has substantial spectral distortions.

2.2 Model-based methods

The model-based fusion approaches assume that HSIs and MSIs can be considered the spatially and spectrally degraded versions of fused HSIs [23], respectively. On the basis of this assumption, methods in this category establish the degradation model for HSIs and MSIs, where the degradation process depends on the spectral response function and point spread function of the sensors. These methods combine an imaging model with handcrafted priors to address the fusion problem. For example, Dong et al. [24] decomposed the desired high-resolution HSI as a spectral basis and abundances and estimate the spectral basis and abundances based on an imaging model and a nonlocal sparse regularization. To exploit the local and nonlocal similarities of an HSI, Han et al. [25] presented a nonlocal similarity-constrained sparse representation approach to obtain the spectral basis and abundances. Except for spare priors, some studies also exploited a low-rank prior to estimate a high-resolution HSI. For example, Simoẽs et al. [26] first obtained a low-dimensional spectral basis by using vertex component analysis [27] and then obtained the abundances by using a spatially smooth prior. Zhou et al. [28] argued that a high-resolution HSI has strong local self-similarities and used a locally low-rank prior to regularize the fusion procedure. Liu et al. [29] proposed a truncated matrix factorization method for MSI and HSI fusion, which exploits the superpixel-level low-rank properties of an HSI. Furthermore, a global and local low-rank-regularized method [30] is proposed for the fusion. Ren et al. [31] considered the spectral variability of subclass objects and formulate the fusion problem as a local prior-regularized optimization problem. In contrast to the matrix factorization, many studies apply the tensor decomposition technique to accurately represent a high-resolution HSI and solve the fusion problem, which includes Tucker decomposition, Canonical polyadic, tensor train decomposition, tenor singular value decomposition (SVD), and tensor ring decomposition. On the basis of Tucker decomposition, Dian et al. [32, 33] presented a sparse tensor factorization method for the fusion, which decomposes the desired high-resolution HSI as a core tensor and factor matrices of three modes and estimates them with a tensor sparse prior. Furthermore, Chang et al. [34] presented a weighted low-rank tensor recovery model for the fusion, which imposes different sparsity regularization weights on different elements of the core tensor. Xu et al. [35] presented a nonlocal CP decomposition approach to fusing an MSI and HSI, which fully exploits the nonlocal spatial-spectral similarities of the HSI. To exploit the spatial low-rank of the HSI, Liu et al. [36] proposed a low tensor trace norm regularization-based method. He et al. [37, 38] proposed the tensor ring decomposition-based methods, which can effectively learn the low-rank property of an HSI.

2.3 Deep learning-based methods

Deep CNNs have advantages in learning image features and have achieved considerable progress in image processing. Many efforts have been made to apply CNNs to MSI and HSI fusion. Methods in this category are supervised, which first learn the mapping function from the MSI and HSI to the high-resolution HSI from available training data. A CNN with various structures has been proposed for MSI and HSI fusion. For example, Fu et al. [39] proposed a deep detail network for the fusion, which uses grouped multiscale dilated convolutions to effectively preserve contextual features. Ref. [40] presented a mature Gaussian-Laplacian pyramid network for the fusion, which comprises several Laplacian pyramid dense modules. To extract the spatial and spectral features, Hu et al. [41] used the attention and pixelShuffle modules to construct an efficient network, which obtains outstanding performance. Wang et al. [42] proposed a variational probabilistic autoencoder-based CNN for the fusion, which exploits the local spectral structures and spatial correlation. To tackle the blind MSI and HSI fusion problem, Wang et al. [43, 44] iteratively and alternatively optimized an imaging model and the fusion procedure by using an iterative refinement unit. Ref. [45] first initialized a high-resolution HSI based on an imaging model and then learnt the residual features between the desired HSI and unused HSI. Zheng et al. [46] proposed an edge-conditioned feature transform network to exploit an edge map prior. To fully exploit imaging models, Dong et al. [47] presented an imaging model-guided CNN to fuse an HSI and MSI. Furthermore, a deep unfolding CNN was proposed by Xie et al. [48, 49] for the fusion, which combines deep proximal gradient descent with a deep CNN. Yang et al. [50] exploited a CNN with two paths to obtain the spatial features and spectral features. To solve the problem of insufficient training data, Ref. [51] used the CNN denoizer for the HSI sharpening and combined imaging models with image priors learned by the CNN, which can flexibly cope with data of different types. To exploit the imaging model in the CNN, Ref. [52] proposed a variational regularization network, which uses gradient descent to obey the imaging model and attention scheme to learn an HSI prior. Given that an HSI and MSI are often not well aligned, Zheng et al. [53] proposed a spectral unmixing and image deformation correction network to jointly align and fuse an HSI and MSI.

3 Proposed method

3.1 Imaging model

Let $\boldsymbol{X} \in \mathbb{R}^{S\times n}$ $(n = w \times h)$, $\boldsymbol{Y} \in \mathbb{R}^{s\times N}$ $(N = W \times H)$, and $\boldsymbol{Z} \in \mathbb{R}^{S\times N}$ represent an HSI, MSI, and high-resolution HSI, respectively, where (w, h) and (W, H) are the spatial dimensions of the HSI and MSI, respectively. Here, each row and column of the matrices denotes a band and a pixel of the image, respectively. An imaging model describes the inherent relationship among the desired high-resolution HSI, HSI, and MSI. Fusion imaging models have been studied and introduced in many previous studies [53–55].

These studies assume that an HSI and MSI are spatially and spectrally downsampled versions of a high-resolution HSI, which are expressed as

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{ZBD} + \boldsymbol{N}_x, \\ \boldsymbol{Y} &= \boldsymbol{RZ} + \boldsymbol{N}_y, \end{aligned} \tag{1}$$

where $\boldsymbol{B} \in \mathbb{R}^{N\times N}$ denotes a spatially blurring matrix and is often assumed to be identical in all spectral bands. The function of the matrix $\boldsymbol{B}$ can be seen as conducting circular convolution with convolutional kernel $\boldsymbol{K} \in \mathbb{R}^{N\times N}$, where $\boldsymbol{K} \in \mathbb{R}^{N\times N}$ is a diagonal matrix and preserves elements of the convolutional kernel in its diagonal line. Therefore, $\boldsymbol{B} \in \mathbb{R}^{N\times N}$ can be decomposed as

$$\boldsymbol{B} = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\mathrm{H}}, \tag{2}$$

where $\boldsymbol{F}$ denotes a 2-dimensional discrete Fourier transform, and $\boldsymbol{F}^{\mathrm{H}}$ denotes its conjugate transpose. The convolution kernel depends on the point spread function of the sensor. $\boldsymbol{D} \in \mathbb{R}^{N\times n}$ denotes a spatially downsampling matrix, which conducts uniform downsampling with stride $d = \sqrt{\frac{N}{n}}$, and satisfies

$$\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D} = \boldsymbol{I}, \tag{3}$$

where $\boldsymbol{I}$ is the identity matrix. The matrix $\boldsymbol{R}$ is a spectrally downsampling matrix, which is up to the spectral response function of the imaging sensor. $\boldsymbol{N}_x$ and $\boldsymbol{N}_y$ denote the additive noise in the HSI and MSI, respectively.

Zhuang et al. [56,57] used the low-rank property of an HSI in the spectral domain for HSI reconstruction, which has shown state-of-the-art performance. In this way, a high-resolution HSI can be written as

$$\boldsymbol{Z} = \boldsymbol{SA}, \tag{4}$$

where $\boldsymbol{S} \in \mathbb{R}^{S\times L}$ and $\boldsymbol{A} \in \mathbb{R}^{L\times N}$ are the subspace and abundances, respectively. On the basis of the subspace representation, the imaging model (8) can be rewritten as

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{SABD} + \boldsymbol{N}_x, \\ \boldsymbol{Y} &= \boldsymbol{RSA} + \boldsymbol{N}_y. \end{aligned} \tag{5}$$

3.2 Subspace estimation

According to the subspace representation, MSI and HSI fusion is formulated as an estimate of $\boldsymbol{S}$ and $\boldsymbol{A}$. A low-resolution HSI preserves the most spectral information, and therefore high- and low-resolution HSIs are assumed to share the same spectral subspace. In this way, we estimate the spectral subspace from the low-resolution HSI. We first conduct SVD on the low-resolution HSI, that is

$$\boldsymbol{X} = \boldsymbol{U}\hat{\boldsymbol{\Sigma}}\boldsymbol{V}^{\mathrm{T}}. \tag{6}$$

Here, $\boldsymbol{U}$ and $\boldsymbol{V}$ are semi-unitary, and a diagonal matrix $\hat{\boldsymbol{\Sigma}}$ contains the singular values, which are arranged in non-increasing order. The large singular values represent the key information of the HSI, and therefore we reserve the L largest singular values and ignore the rest. In this way, the subspace $\boldsymbol{S}$ can be computed as

$$\boldsymbol{S} = \boldsymbol{U}(:, 1:L). \tag{7}$$

3.3 External and internal prior learning for the fusion

Because the observed HSI and MSI are the downsampled versions of the underlying high-resolution HSI, MSI and HSI fusion is a severally underdetermined problem. To tackle this challenging problem, various priors have been proposed to regularize it, such as sparse priors [24], low-rank priors [55], spatial smoothness [26], nonlocal similarities [25], and image priors learned by a deep CNN [51]. Image priors reflect HSI characteristics and play an important role in MSI and HSI fusion. In general, priors can be categorized as internal and external priors. The internal priors are derived from the images to be fused, which differ between image types. The internal priors model the unique characteristics of the HSI and MSI to be fused. The external priors model the general priors for images, which are common in the images and are often learned from the additional training data by a deep CNN. However, the existing approaches often only use one type of prior to solve the fusion problem, which hinders fusion performance improvement. To address this issue, we propose a novel method that can simultaneously exploit the external and internal HSI priors. After acquiring the subspace $\boldsymbol{S}$, the abundances $\boldsymbol{A}$ should be estimated to acquire the high-resolution HSI. On the basis of the MAP, the estimate of the abundances can be formulated as

$$\begin{aligned}\boldsymbol{A} &= \underset{\boldsymbol{A}}{\operatorname{argmax}}\, P(\boldsymbol{A}|\boldsymbol{X},\boldsymbol{Y}) = \underset{\boldsymbol{A}}{\operatorname{argmax}}\, \frac{P(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{A})P(\boldsymbol{A})}{P(\boldsymbol{X},\boldsymbol{Y})} \\ &= \underset{\boldsymbol{A}}{\operatorname{argmin}} - \log\{P(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{A})\} - \log\{P(\boldsymbol{A})\}.\end{aligned} \tag{8}$$

In (8), the log-likelihood term $-\log\{P(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{A})\}$ depicts the imaging model between $\boldsymbol{A}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$. On the basis of the imaging model (5), $-\log\{P(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{A})\}$ is equivalent to

$$-\log\{P(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{A})\} = \frac{1}{2\sigma_x^2}\|\boldsymbol{SABD}-\boldsymbol{X}\|_{\mathrm{F}}^2 + \frac{1}{2\sigma_y^2}\|\boldsymbol{RSA}-\boldsymbol{Y}\|_{\mathrm{F}}^2, \tag{9}$$

where σ_x^2 and σ_y^2 denote the variance of white Gaussian noise in the observation model of the HSI and MSI, respectively. The term $-\log\{P(\boldsymbol{A})\}$ in (8) denotes the prior information of $\boldsymbol{A}$. Here, we use the internal and external prior information to regularize the estimate of $\boldsymbol{A}$, and therefore we write the prior information term as

$$-\log\{P(\boldsymbol{A})\} = \lambda_1\phi_1(\boldsymbol{A}) + \lambda_2\phi_2(\boldsymbol{A}), \tag{10}$$

where λ_1 and λ_2 denote the regularization parameters, and $\phi_1(\cdot)$ and $\phi_2(\cdot)$ denote the internal and external priors of $\boldsymbol{A}$, respectively. By combing the imaging model and priors, the estimate of $\boldsymbol{A}$ is written as

$$\min_{\boldsymbol{A}} ||\boldsymbol{X}-\boldsymbol{SABD}||_{\mathrm{F}}^2 + ||\boldsymbol{Y}-\boldsymbol{RSA}||_{\mathrm{F}}^2 + \lambda_1\phi_1(\boldsymbol{A}) + \lambda_2\phi_2(\boldsymbol{A}). \tag{11}$$

The optimization (11) effectively combines the external and internal priors for MSI and HSI fusion.

The optimization problem (11) is difficult to solve directly because of the existence of regularization. We adopt the ADMMs [58] to address this problem iteratively. The idea of the ADMMs is to decompose the original problem into iteratively solving several treatable subproblems. By introducing new variables $\widehat{\boldsymbol{A}}$ and $\overline{\boldsymbol{A}}$, the optimization (11) is transferred into minimizing the following augmented Lagrange function:

$$\begin{aligned}\mathcal{L}(\boldsymbol{A},\widehat{\boldsymbol{A}},\overline{\boldsymbol{A}},\widehat{\boldsymbol{G}},\overline{\boldsymbol{G}}) = {} & ||\boldsymbol{X}-\boldsymbol{SABD}||_{\mathrm{F}}^2 + ||\boldsymbol{Y}-\boldsymbol{RSA}||_{\mathrm{F}}^2 + \lambda_1\phi_1(\widehat{\boldsymbol{A}}) + \lambda_2\phi_2(\overline{\boldsymbol{A}}) \\ & + \mu\left\|\boldsymbol{A}-\widehat{\boldsymbol{A}}+\frac{\widehat{\boldsymbol{G}}}{2\mu}\right\|_{\mathrm{F}}^2 + \mu\left\|\boldsymbol{A}-\overline{\boldsymbol{A}}+\frac{\overline{\boldsymbol{G}}}{2\mu}\right\|_{\mathrm{F}}^2.\end{aligned} \tag{12}$$

The function (12) can be minimized by iteratively addressing the following subproblems:

$$\begin{aligned}\boldsymbol{A} &= \underset{\boldsymbol{A}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{A},\widehat{\boldsymbol{A}},\overline{\boldsymbol{A}},\widehat{\boldsymbol{G}},\overline{\boldsymbol{G}}), \\ \widehat{\boldsymbol{A}} &= \underset{\widehat{\boldsymbol{A}}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{A},\widehat{\boldsymbol{A}},\overline{\boldsymbol{A}},\widehat{\boldsymbol{G}},\overline{\boldsymbol{G}}), \\ \overline{\boldsymbol{A}} &= \underset{\overline{\boldsymbol{A}}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{A},\widehat{\boldsymbol{A}},\overline{\boldsymbol{A}},\widehat{\boldsymbol{G}},\overline{\boldsymbol{G}}).\end{aligned} \tag{13}$$

3.3.1 *Optimization of* $\boldsymbol{A}$

The optimization of $\boldsymbol{A}$ can be written as

$$\begin{aligned}\boldsymbol{A} &= \underset{\boldsymbol{A}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{A},\widehat{\boldsymbol{A}},\overline{\boldsymbol{A}},\widehat{\boldsymbol{G}},\overline{\boldsymbol{G}}) \\ &= \underset{\boldsymbol{A}}{\operatorname{argmin}}\, ||\boldsymbol{X}-\boldsymbol{SABD}||_{\mathrm{F}}^2 + ||\boldsymbol{Y}-\boldsymbol{RSA}||_{\mathrm{F}}^2 + \mu\left\|\boldsymbol{A}-\widehat{\boldsymbol{A}}+\frac{\widehat{\boldsymbol{G}}}{2\mu}\right\|_{\mathrm{F}}^2 + \mu\left\|\boldsymbol{A}-\overline{\boldsymbol{A}}+\frac{\overline{\boldsymbol{G}}}{2\mu}\right\|_{\mathrm{F}}^2.\end{aligned} \tag{14}$$

The optimization (14) is strongly convex and thus has a unique solution. Therefore, we take the derivative w.r.t. $\boldsymbol{A}$ as zero and obtain the following equation:

$$\boldsymbol{H}_1\boldsymbol{A} + \boldsymbol{A}\boldsymbol{H}_2 = \boldsymbol{H}_3, \tag{15}$$

where $\boldsymbol{H}_1$, $\boldsymbol{H}_2$, and $\boldsymbol{H}_3$ satisfie

$$\begin{gathered}
\boldsymbol{H}_1 = (\boldsymbol{R}\boldsymbol{S})^{\mathrm{T}}\boldsymbol{R}\boldsymbol{S} + 2\mu\boldsymbol{I},\\
\boldsymbol{H}_2 = (\boldsymbol{B}\boldsymbol{D})(\boldsymbol{B}\boldsymbol{D})^{\mathrm{T}},\\
\boldsymbol{H}_3 = (\boldsymbol{R}\boldsymbol{S})^{\mathrm{T}}\boldsymbol{Y} + \boldsymbol{S}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{B}\boldsymbol{D})^{\mathrm{T}} + \mu\left(\widehat{\boldsymbol{A}} - \frac{\widehat{\boldsymbol{G}}}{2\mu} + \overline{\boldsymbol{A}} - \frac{\overline{\boldsymbol{G}}}{2\mu}\right).
\end{gathered} \tag{16}$$

Eq. (15) is a Sylvester equation, which can be addressed by the conjugate gradient descent method. Following previous studies [55, 59], we obtain a closed-form solution by exploiting the properties of the convolution blur and downsampling matrix. From (16), we know that $\boldsymbol{H}_1$ is positive and symmetric, and therefore it can be diagonalized by eigenvalue decomposition, i.e.,

$$\boldsymbol{H}_1 = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}. \tag{17}$$

where diagonal matrix $\boldsymbol{\Lambda}$ is written as

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_L \end{bmatrix}. \tag{18}$$

By left multiplying by $\boldsymbol{Q}^{-1}$ and right multiplying (15) by the DFT matrix $\boldsymbol{F}$ for (15), we obtain the following equation:

$$\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{F} + \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{F}\boldsymbol{\Sigma}^{\mathrm{H}} = \boldsymbol{Q}^{-1}\boldsymbol{H}_3\boldsymbol{F}, \tag{19}$$

By substituting (24) into (19), we obtain the following equation:

$$\boldsymbol{\Lambda}\tilde{\boldsymbol{A}} + \tilde{\boldsymbol{A}}\boldsymbol{K}\boldsymbol{K}^{\mathrm{H}} = \boldsymbol{C}, \tag{20}$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{F}$, $\boldsymbol{K} = \boldsymbol{\Sigma}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{D}$, and $\boldsymbol{C} = \boldsymbol{Q}^{-1}\boldsymbol{H}_3\boldsymbol{F}$. Eq. (20) can be considered a new Sylvester equation for $\tilde{\boldsymbol{A}}$. $\tilde{\boldsymbol{A}}$ and $\boldsymbol{C}$ are written as $\tilde{\boldsymbol{A}} = [\tilde{\boldsymbol{a}}_1, \ldots, \tilde{\boldsymbol{a}}_L]^{\mathrm{T}}$ and $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_L]^{\mathrm{T}}$, respectively, where $\boldsymbol{c}_i$ and $\tilde{\boldsymbol{a}}_i$ denote the i-th row of $\boldsymbol{C}$ and $\tilde{\boldsymbol{A}}$, respectively. In this way, we estimate $\tilde{\boldsymbol{A}}$ in a row-by-row manner, i.e.,

$$\lambda_i\tilde{\boldsymbol{a}}_i + \tilde{\boldsymbol{a}}_i\boldsymbol{K}\boldsymbol{K}^{\mathrm{H}} = \boldsymbol{c}_i,\ \text{for } i = 1, \ldots, L. \tag{21}$$

We can obtain

$$\tilde{\boldsymbol{a}}_i = \boldsymbol{c}_i(\lambda_i\boldsymbol{I}_n + \boldsymbol{K}\boldsymbol{K}^{\mathrm{H}})^{-1},\ \text{for } i = 1, \ldots, L. \tag{22}$$

By using the matrix inverse formula $(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A})^{-1}\boldsymbol{B}$, $\tilde{\boldsymbol{a}}_i$ is equivalent to

$$\tilde{\boldsymbol{a}}_i = \lambda_i^{-1}\boldsymbol{c}_i - \lambda_i^{-1}\boldsymbol{c}_i\boldsymbol{K}(\lambda_i\boldsymbol{I}_n + \boldsymbol{K}^{\mathrm{H}}\boldsymbol{K})^{-1}\boldsymbol{K}^{\mathrm{H}}. \tag{23}$$

Lemma 1 (Wei et al. [60])**.** According to the properties of $\boldsymbol{D}$, the equation is obtained

$$\boldsymbol{F}^{\mathrm{H}}\boldsymbol{D} = \frac{1}{d}(\mathbf{1}_{d^2} \otimes \boldsymbol{I}_n), \tag{24}$$

where $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix, and $\mathbf{1}_{d^2} \in \mathbb{R}^{d^2}$ is a vector of ones.

By using Lemma 1, we can obtain $\boldsymbol{K}\boldsymbol{K}^{\mathrm{H}}$, which is also a diagonal matrix and can be computed as $\boldsymbol{K}^{\mathrm{H}}\boldsymbol{K} = \sum_{i=1}^{d^2}\boldsymbol{\Sigma}_i^2$, where $\boldsymbol{\Sigma}_i$ is the submatrix of $\boldsymbol{\Sigma}$. The diagonal matrix $\boldsymbol{\Sigma}$ can be represented as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 & \cdots & 0\\ 0 & \boldsymbol{\Sigma}_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \boldsymbol{\Sigma}_{d^2} \end{bmatrix}, \tag{25}$$

where $\Sigma_i \in \mathbb{C}^{n\times n}$. In this way, the computational complexity of $(\lambda_i \boldsymbol{I}_n + \boldsymbol{K}^{\mathrm{H}}\boldsymbol{K})^{-1}$ is reduced from $O(n^3)$ to $O(n)$. Therefore, we have

$$\tilde{\boldsymbol{a}}_i = \lambda_i^{-1}\boldsymbol{c}_i - \lambda_i^{-1}\boldsymbol{c}_i\boldsymbol{K}\left(\lambda_i\boldsymbol{I}_n + \sum_{i=1}^{d^2}\boldsymbol{\Sigma}_i^2\right)^{-1}\boldsymbol{K}^{\mathrm{H}}. \tag{26}$$

After obtaining $\tilde{\boldsymbol{A}}$, $\boldsymbol{A}$ is estimated as

$$\boldsymbol{A} = \boldsymbol{Q}\tilde{\boldsymbol{A}}\boldsymbol{F}^{\mathrm{H}}. \tag{27}$$

3.3.2 *Optimization of $\widehat{\boldsymbol{A}}$ (inter prior learning)*

The optimization of $\widehat{\boldsymbol{A}}$ can be written as

$$\begin{aligned}\widehat{\boldsymbol{A}} &= \underset{\widehat{\boldsymbol{A}}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{A}, \widehat{\boldsymbol{A}}, \overline{\boldsymbol{A}}, \widehat{\boldsymbol{G}}, \overline{\boldsymbol{G}}) \\ &= \underset{\widehat{\boldsymbol{A}}}{\operatorname{argmin}}\, \mu \left\|\boldsymbol{A} - \widehat{\boldsymbol{A}} + \frac{\widehat{\boldsymbol{G}}}{2\mu}\right\|_{\mathrm{F}}^2 + \lambda_1\phi_1(\widehat{\boldsymbol{A}}).\end{aligned} \tag{28}$$

Eq. (28) models the internal prior of $\boldsymbol{A}$. The internal prior represents the characteristics of the HSI and MSI to be fused, which differ from those of other types of images. We propose a local low-rank representation to learn the local low-rank prior. A scene often contains similar materials in a local area, which results in local similarities in the HSI. Because an MSI preserves the main spatial information of a high-resolution HSI, we learn the local spatial similarities from the MSI. Specifically, we first use the superpixel method [61] to divide the MSI into several superpixels $\boldsymbol{Z}^i, i = 1,\ldots,T$ based on the similarities of the adjacent pixels, where $\boldsymbol{Z}^i$ denotes the i-th superpixel, and T is the number of superpixels. On the basis of the learned spatial structures in the MSI, the abundances are also divided into several superpixels $\boldsymbol{A}^i, i = 1,\ldots,T$. The elements in each superpixel are very similar to each other, which leads to the low-rank prior. Therefore, we use the local low-rank regularization to learn the internal prior. On the basis of the proposed internal prior, Eq. (28) can be written as

$$\widehat{\boldsymbol{A}} = \underset{\widehat{\boldsymbol{A}}}{\operatorname{argmin}}\, \mu \left\|\boldsymbol{A} - \widehat{\boldsymbol{A}} + \frac{\widehat{\boldsymbol{G}}}{2\mu}\right\|_{\mathrm{F}}^2 + \lambda_1 \sum_{i=1}^{T} ||\boldsymbol{A}^i||_*, \tag{29}$$

where $||\cdot||_*$ represents the nuclear norm, which is defined as the sum of singular values and is a convex relation for the matrix rank. Since different superpixels do not have overlaps, the optimization (29) can be solved for each superpixel, that is,

$$\widehat{\boldsymbol{A}}^i = \underset{\widehat{\boldsymbol{A}}^i}{\operatorname{argmin}}\, \mu \left\|\boldsymbol{A}^i - \widehat{\boldsymbol{A}}^i + \frac{\widehat{\boldsymbol{G}}^i}{2\mu}\right\|_{\mathrm{F}}^2 + \lambda_1||\boldsymbol{A}^i||_*, \tag{30}$$

which has the following closed-form solution:

$$\widehat{\boldsymbol{A}}^i = \boldsymbol{U}\left(\boldsymbol{\Sigma} - \frac{\lambda_1}{2\mu}\right)_+ \boldsymbol{V}^{\mathrm{T}}, \tag{31}$$

where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ is the SVD of $\boldsymbol{A}^i + \frac{\widehat{\boldsymbol{G}}^i}{2\mu}$ and $(\cdot)^+$ is the positive part. After acquiring each superpixel, $\widehat{\boldsymbol{A}}$ is obtained by merging all superpixels.

3.3.3 *Optimization of $\widehat{\boldsymbol{A}}$ (external prior learning)*

The optimization of $\widehat{\boldsymbol{A}}$ can be written as

$$\begin{aligned}\overline{\boldsymbol{A}} &= \underset{\overline{\boldsymbol{A}}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{A}, \widehat{\boldsymbol{A}}, \overline{\boldsymbol{A}}, \widehat{\boldsymbol{G}}, \overline{\boldsymbol{G}}) \\ &= \underset{\overline{\boldsymbol{A}}}{\operatorname{argmin}}\, \mu \left\|\boldsymbol{A} - \overline{\boldsymbol{A}} + \frac{\overline{\boldsymbol{G}}}{2\mu}\right\|_{\mathrm{F}}^2 + \lambda_2\phi_2(\overline{\boldsymbol{A}}).\end{aligned} \tag{32}$$

Eq. (33) models the external prior of $\boldsymbol{A}$. The external prior denotes the generalized prior of the HSI, which is learned from the additional training data. The optimization (33) can be considered a denoising problem, which aims to remove the white additive Gaussian noise of variation $\sigma^2 = \frac{\lambda_2}{2\mu}$ from $\boldsymbol{A} + \frac{\overline{\boldsymbol{G}}}{2\mu}$. Therefore, we adopt FFDNet [62], a flexible and fast CNN for image denoising, to solve (33). FFDNet has three types of modules: a 3×3 convolution layer (Conv), batch normalization [63], and rectified linear units [64]. Furthermore, a tunable noise level and noisy images are simultaneously input into the network, and therefore FFDNet can adaptively handle noisy images of different noise levels without retraining. Each row of $\boldsymbol{A}$ preserves the spatial structures of the high-resolution HSI, and each row of abundances is independent because of the subspace representation. In this way, we estimate each row of $\overline{\boldsymbol{A}}$ by applying FFDNet to each row of $\boldsymbol{H} = \boldsymbol{A} + \frac{\overline{\boldsymbol{G}}}{2\mu}$, i.e.,

$$
\begin{aligned}
\overline{\boldsymbol{A}}_i &= \underset{\overline{\boldsymbol{A}}_i}{\operatorname{argmin}}\, \mu \left\| \boldsymbol{A}_i - \overline{\boldsymbol{A}}_i + \frac{\overline{\boldsymbol{G}}_i}{2\mu} \right\|_{\mathrm{F}}^2 + \lambda_2 \phi_2(\overline{\boldsymbol{A}}_i) \\
&= \mathcal{F}\left(\boldsymbol{H}_i, \frac{\lambda_2}{2\mu}; \Theta \right), \ i =, 1, 2, \ldots, L,
\end{aligned} \tag{33}
$$

where Θ denotes the parameters of FFDNet, and $\boldsymbol{H}_i$, $\overline{\boldsymbol{A}}_i$, $\overline{\boldsymbol{G}}_i$, and $\boldsymbol{A}_i$ denote the i-th row of $\boldsymbol{H}$, $\overline{\boldsymbol{A}}$, $\overline{\boldsymbol{G}}$, and $\boldsymbol{A}$, respectively. The FFDNet method is trained for gray images, and the images are scaled to $[0, 1]$. Therefore, we should also first scale each row of $\boldsymbol{H}$ to $[0, 1]$ by linear transform. Correspondingly, the noise variance should be changed as $c_i\sigma^2$. Finally, the denoising results are transformed back. In this way, we estimate $\boldsymbol{V}$ in (33) as

$$
\begin{aligned}
&\hat{\boldsymbol{H}}_i = c_i \boldsymbol{H}_i + b_i, \\
&\hat{\boldsymbol{V}}_i = \mathcal{F}\left(\boldsymbol{H}_i, \frac{c_i \lambda_2}{2\mu}; \Theta \right), \\
&\boldsymbol{V}_i = \frac{\hat{\boldsymbol{V}}(i,:) - b_i}{c_i},
\end{aligned} \tag{34}
$$

where $c_i = \frac{1}{\mathtt{max}(\boldsymbol{H}_i) - \mathtt{min}(\boldsymbol{H}_i)}$ and $b_i = -c_i \times \mathtt{min}(\boldsymbol{H}_i)$.

3.3.4 *Optimization of* $\hat{\boldsymbol{G}}$ *and* $\overline{\boldsymbol{G}}$

The Lagrangian multipliers $\hat{\boldsymbol{G}}$ and $\overline{\boldsymbol{G}}$ are updated through

$$
\begin{aligned}
&\hat{\boldsymbol{G}} = \hat{\boldsymbol{G}}_1 + 2\mu(\boldsymbol{A} - \hat{\boldsymbol{A}}), \\
&\overline{\boldsymbol{G}} = \overline{\boldsymbol{G}}_2 + 2\mu(\boldsymbol{A} - \overline{\boldsymbol{A}}).
\end{aligned} \tag{35}
$$

3.3.5 *Update for* μ

Because the noise level in each iteration is $\sigma^2 = \frac{\lambda}{2\mu}$, the parameter μ plays an important role in FFDNet denoising. The image tends to be closer to the clean image with increasing iterations. Hence, the noise level σ^2 should decrease with increasing iterations. In this way, we increase the value of μ in the iteration by

$$
\mu = \gamma\mu, \tag{36}
$$

where $\gamma > 1$. In this way, we can decrease the noise level.

4 Experiments

To verify the effectiveness of the proposed network, we test the compared approaches on two simulated datasets and one real dataset.

4.1 Dataset and data simulation

4.1.1 *Pavia university dataset*

The Pavia University dataset [65] was obtained by the ROSIS sensor during a flight campaign over Pavia University, Italy. The HSI has 115 spectral bands and 610×340 pixels, and only 93 bands are used by removing the bands of low signal-to-noise ratio (SNR). The HSI has a spatial resolution of 1.5 m. The low-resolution HSI is simulated by using a Gaussian filter (size of 7×7 and stand variation 3) and then downsampling the image with a factor of 4. The MSI is obtained by the IKONOS-like spectral response. We add the i.i.d. noise of 30 and 25 dB to the MSI and HSI, respectively. For HSI imaging, an increase in the number of spectral bands often results in a decrease in the SNR. Therefore, we add the serve noise to the HSI in the simulation.

4.1.2 *Indian pines dataset*

The Indian Pines dataset was acquired by NASA's airborne visible and infrared imaging spectrometer [66]. The HSI has a size of $128 \times 128 \times 224$ and 224 spectral bands and covers a wavelength range of 400–2500 nm. The number of spectral bands is reduced to 200 by removing bands 104–108, 150–163, and 220 of the image because of the extremely low SNR in these bands. A low-resolution HSI of size $32 \times 32 \times 200$ is produced by applying a 7×7 Gaussian blur (standard deviation of 3) and then by downsampling the blurred image with a ratio of 4. The MSI with six bands is simulated by a Landsat7-like spectral response. We add the i.i.d. noise of 30 and 25 dB to the MSI and HSI, respectively.

4.1.3 *Gaofen 5*

We also conduct experiments on a real dataset. The HSI is carried by the Chinese satellite Gaofen (GF) 5, which has a spatial resolution of 30 m. The HSI has 330 bands and a spectral range of 390–2513 nm. To match the MSI, only 150 spectral bands are used, corresponding to a spectral range of 390–1000 nm. The MSI is carried by the Chinese satellite GF 1, which has a spatial resolution of 16 m. The MSI has four bands and a spectral range of 400–1000 nm. The HSI and MSI are of size $600 \times 600 \times 150$ and $300 \times 300 \times 4$, respectively.

4.2 Compared approaches and settings

We compare the proposed method with seven recent state-of-the-art methods for HSI super-resolution, comprising the Gram-Schmidt adaptive (GSA) [67][1)], HySure [26][2)], CNMF [68][3)], coupled sparse tensor factorization (CSTF) [33][4)], global-local low-rank promoting algorithm (GLORIA) [30][5)], factor smoothed tensor ring decomposition (FSTRD) [38], and CNN fusion method (CNN-Fus) [51][6)]. Among the compared methods, the CSTF, GLORIA, and FSTRD exploit the internal priors, and CNN-Fus exploits the external priors. For the CSTF, we rewrite the initialization procedure of two spatial dictionaries by using the SVD, which makes the algorithm converge faster without deteriorating the fusion performance. We set $L = 10$, $T = 200$, $\lambda_1 = 1.5 \times 10^{-3}$, and $\lambda_2 = 1 \times 10^{-3}$ for the proposed method. The parameters of the other approaches are carefully tuned for optimal performance.

4.3 Quality metrics

To evaluate the quality of high-resolution HSIs obtained by HSI approaches, five quality metrics are used in our study, comprising the peak SNR (PSNR), structural similarity index (SSIM) [69], relative dimensionless global error in synthesis (ERGAS) [70], universal image quality index (UIQI) [71], spectral angle mapper (SAM), and running time in seconds. Larger values of PSNR and UIQI mean higher quality metrics, and smaller values of ERGAS and SAM mean higher quality metrics. The running time is used to evaluate the computational efficiency of the compared methods. All elements of the image are scaled to [0, 255] when the quality metrics are calculated.

---

1) https://openremotesensing.net/knowledgebase/hyperspectral-and-multispectral-data-fusion/.
2) https://github.com/alfaiate/HySure.
3) http://naotoyokoya.com/Download.html.
4) https://github.com/renweidian/CSTF.
5) https://github.com/REIYANG/GLORIA.
6) https://github.com/renweidian/CNN-FUS.

Table 1 Quantitative indices of the test approaches on Pavia University dataset [65]

Method	Pavia University dataset					
	PSNR	SAM	SSIM	UIQI	ERGAS	Running time (s)
Reference	$+\infty$	0	1	1	0	0
GSA [67]	35.950	5.479	0.935	0.944	2.941	**1.746**
CNMF [68]	39.633	3.716	0.957	0.967	2.111	164.848
Hysure [26]	40.935	3.302	0.965	0.973	1.876	129.930
CSTF [33]	38.357	4.590	0.941	0.957	2.451	117.345
GLORIA [30]	40.898	3.354	0.965	0.973	1.912	183.414
FSTRD [38]	40.206	3.527	0.961	0.971	2.016	490.207
CNN-Fus [51]	41.434	3.158	0.969	0.975	1.857	8.466
ExInL (our method)	**41.810**	**2.952**	**0.971**	**0.977**	**1.753**	29.532

Table 2 Quantitative indices of the test approaches on the Indian Pines dataset [72]

Method	Indian Pines dataset					
	PSNR	SAM	SSIM	UIQI	ERGAS	Running time (s)
Reference	$+\infty$	0	1	1	0	0
GSA [67]	33.646	3.898	0.815	0.539	3.104	2.378
CNMF [68]	36.458	2.882	0.861	0.654	1.818	19.327
Hysure [26]	38.380	2.479	0.894	0.720	1.426	13.59
CSTF [33]	35.070	3.428	0.814	0.602	2.260	9.686
GLORIA [30]	39.895	2.245	0.932	0.777	1.192	33.596
FSTRD [38]	40.087	2.322	0.928	0.777	1.185	34.565
CNN-Fus [51]	38.620	2.645	0.896	0.732	1.412	**2.288**
ExInL (our method)	**42.091**	**2.013**	**0.946**	**0.829**	**0.967**	3.081

4.4 Experiments on simulated data fusion

Table 1 [26, 30, 33, 38, 51, 65, 67, 68] lists the quality metrics of the compared approaches on the Pavia University dataset. The optimal results are highlighted in bold. Table 1 shows that the proposed ExInL consistently achieves the optimal results on all quality metrics, verifying the superiority of the proposed method. In addition, CNN-Fus, which exploits the external priors, obtains suboptimal results. For computational efficiency, the GSA method has high computational efficiency because it does not need to solve a complex optimization problem. Figure 1 displays the false-color images constituted by the 60th, 25th, and 4th bands and the corresponding error images of fused Pavia University images through test approaches. Compared with the low-resolution HSI, all compared approaches can provide sharp spatial details, which demonstrate the effectiveness of the HSI and MSI fusion. The error images show that the GSA method has a relatively large error, and the proposed ExInL method and CNN-Fus have fewer errors than other approaches.

Table 2 [26, 30, 33, 38, 51, 67, 68, 72] reports the average quality metrics of the compared approaches on the Indian Pines dataset. Table 2 shows that the proposed method obtains significantly better results than the other compared approaches on all quality metrics. The main reason is that the Indian Pines dataset has strong local similarities, which are effectively exploited by the proposed method. GLORIA and FSTRD provide suboptimal results. For computational efficiency, the GSA method and CNN-Fus need less running time than the other compared approaches. The false-color images constituted by the 33rd, 19th, and 5th bands and the corresponding error images of fused Pavia University images through the test approaches are shown in Figure 2. Figure 2 shows that the proposed ExInL has considerably fewer errors than the other compared approaches. GLORIA and FSTRD also outperform the other approaches.

4.5 Experiments on real data fusion

To further verify the performance of the proposed method, we test the compared methods on real data fusion. We estimate the spatial and spectral responses through the proposed method and use them in the compared fusion methods. For real data validation, there is no ground truth for calculating quality

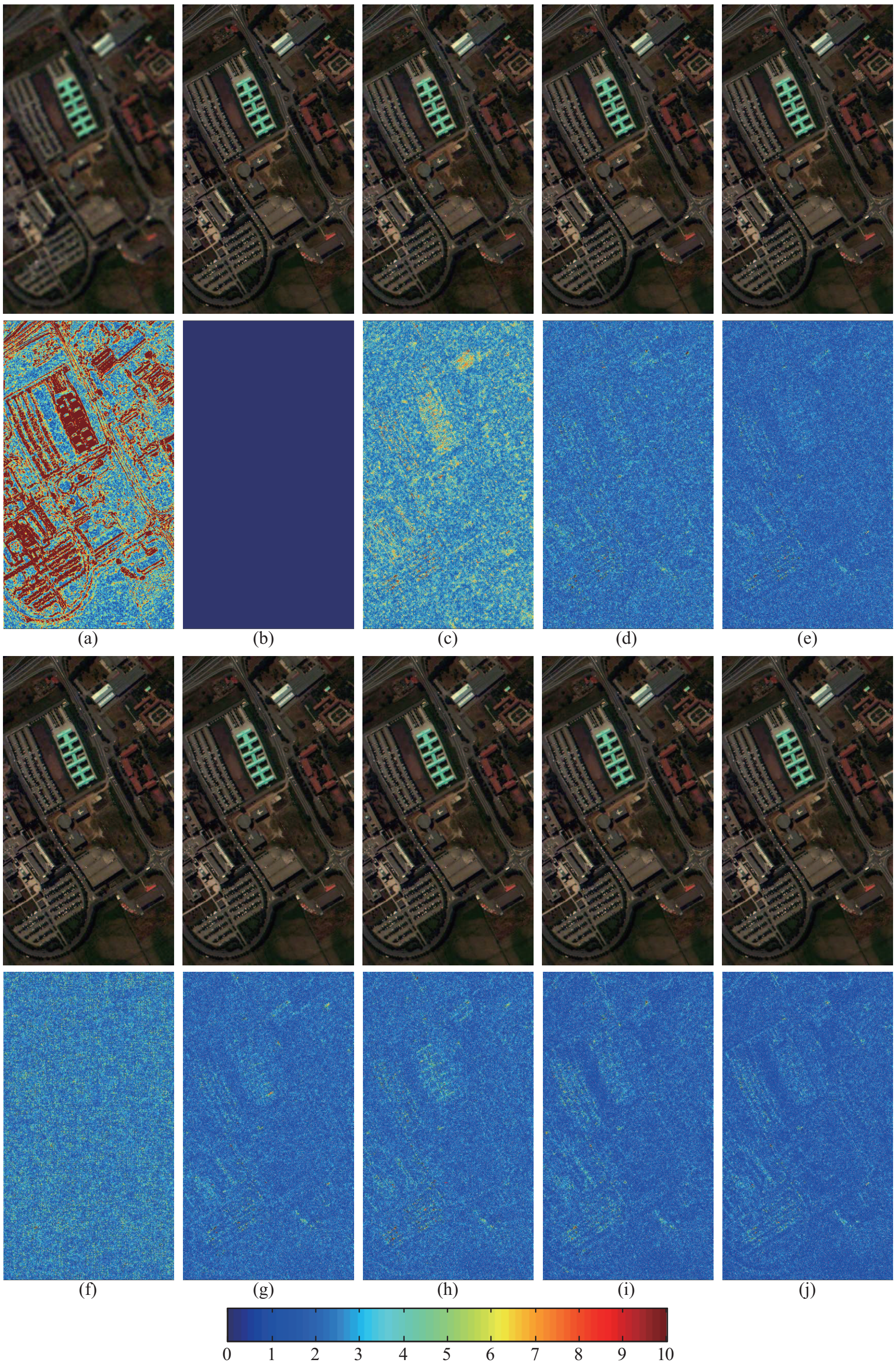

Figure 1 (Color online) False-color images constituted by the 60th, 25th, and 4th bands and the corresponding error images of fused Pavia University images using the test approaches. (a) Low-resolution HSI; (b) ground truth; (c) GSA [67]; (d) CNMF [68]; (e) Hysure [26]; (f) CSTF [33]; (g) GLORIA [30]; (h) FSTRD [38]; (i) CNN-Fus [51]; (j) ExInL.

metrics, and therefore the visual comparison is essential to measure the compared approaches. Figure 3 shows the false-color images constituted by the 61st, 38th, and 14th bands of fused GF 5 and GF 1 by test approaches, and a meaningful region is marked and magnified for visual comparison. The magnified region shows that the results of GSA and CNMF lose some important details. The proposed method and GLORIA perform the best in reconstructing the high-resolution details, which provides evidence of the superiority of our method.

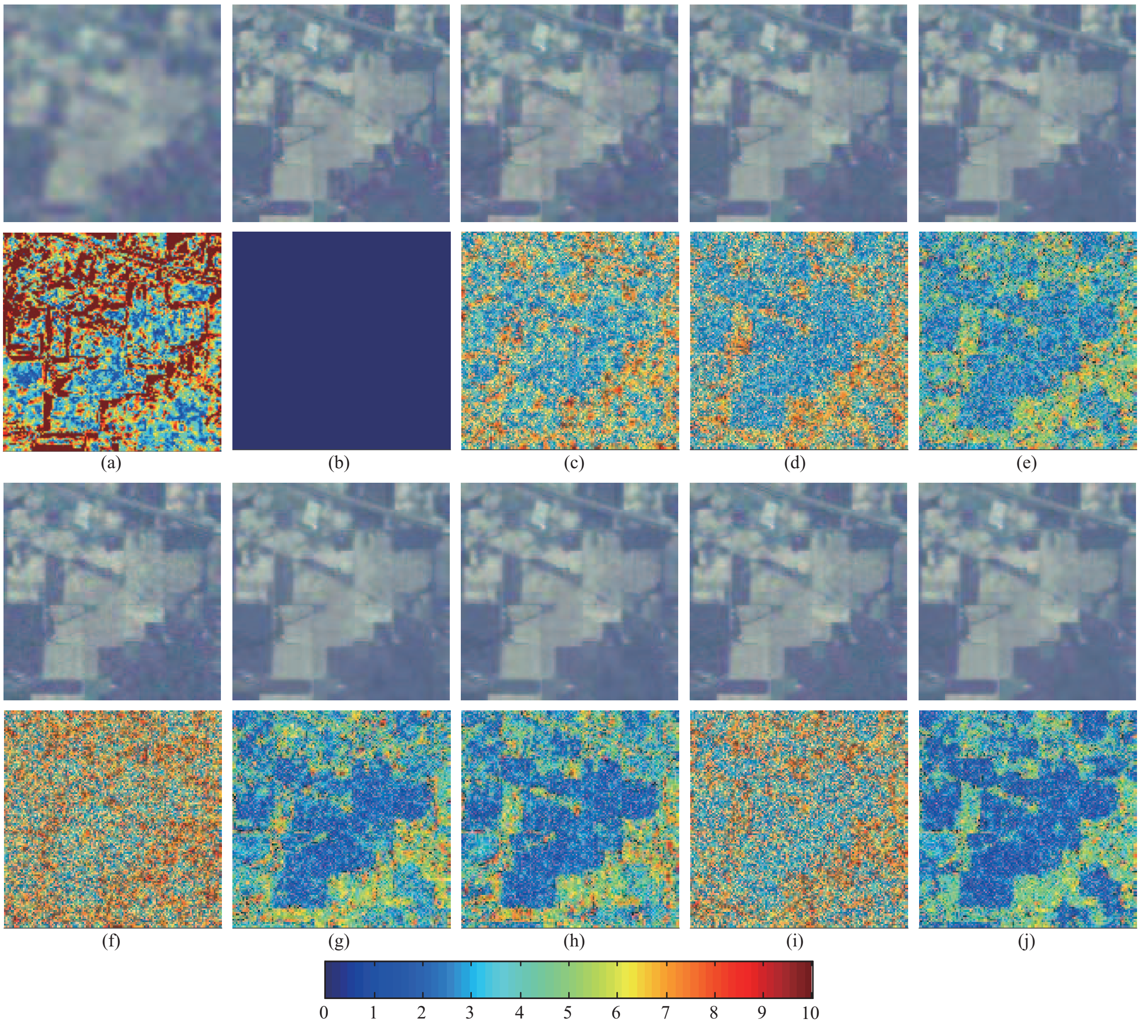

Figure 2 (Color online) False-color images constituted by the 33rd, 19th, and 5th bands and the corresponding error images of the fused Indian Pines dataset using the test approaches. (a) Low-resolution HSI; (b) ground truth; (c) GSA [67]; (d) CNMF [68]; (e) Hysure [26]; (f) CSTF [33]; (g) GLORIA [30]; (h) FSTRD [38]; (i) CNN-Fus [51]; (j) ExInL.

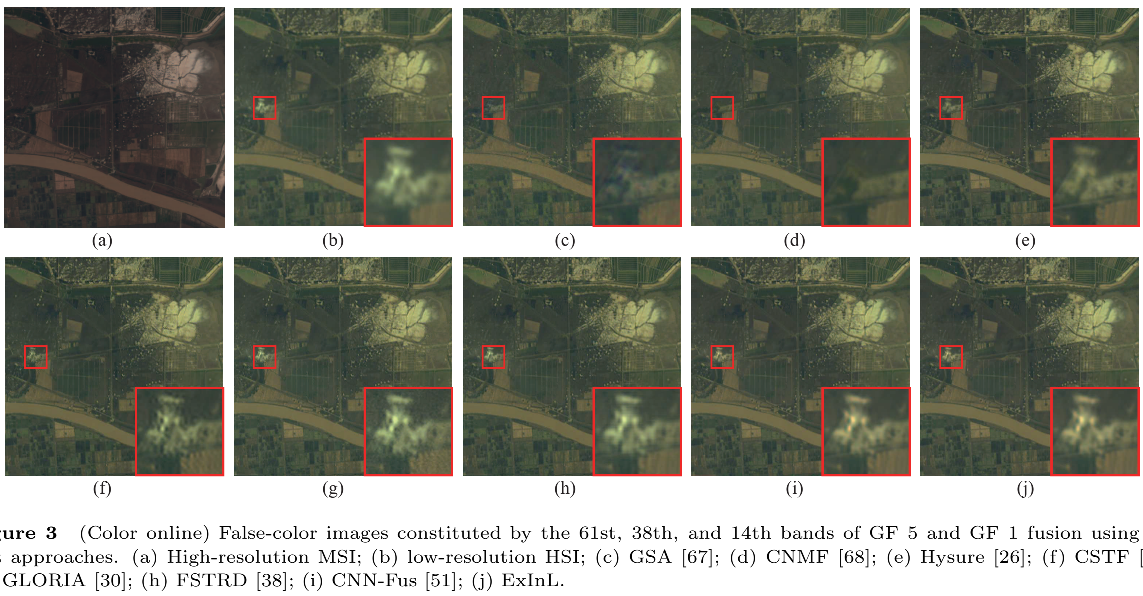

Figure 3 (Color online) False-color images constituted by the 61st, 38th, and 14th bands of GF 5 and GF 1 fusion using the test approaches. (a) High-resolution MSI; (b) low-resolution HSI; (c) GSA [67]; (d) CNMF [68]; (e) Hysure [26]; (f) CSTF [33]; (g) GLORIA [30]; (h) FSTRD [38]; (i) CNN-Fus [51]; (j) ExInL.

Table 3 Ablation analysis for the proposed method

Method	Pavia University					Indian Pines				
	PSNR	SAM	SSIM	UIQI	ERGAS	PSNR	SAM	SSIM	UIQI	ERGAS
Reference	$+\infty$	0	1	1	0	$+\infty$	SAM(0)	SSIM(1)	UIQI(1)	ERGAS(0)
None	36.482	6.402	0.894	0.929	3.287	31.413	6.436	0.639	0.452	3.413
External	41.393	3.152	0.969	0.975	1.869	38.491	2.709	0.891	0.728	1.436
Internal	40.218	3.638	0.956	0.968	2.001	40.679	2.217	0.930	0.793	1.110
Internal+External	**41.810**	**2.952**	**0.971**	**0.977**	**1.753**	**42.091**	**2.013**	**0.946**	**0.829**	**0.967**

4.6 Ablation analysis

As mentioned above, the proposed method exploits the external and internal priors to solve the fusion problem. In this subsection, we illustrate the effectiveness of the external and internal priors. To verify their effectiveness, we conduct the experiments only with external priors, only with internal priors, and without both priors, while keeping other settings the same. Table 3 reports the quality metrics of the proposed method only with external priors, only with internal priors, without both priors (denoted by none), and with both priors. Table 3 shows that the proposed method obtains poor results without both priors and achieves superior results with both priors, which demonstrates the effectiveness of the external and internal priors.

5 Conclusion

In this paper, we present a novel multispectral and hyperspectral image fusion method based on external and internal prior learning. Specifically, we learn the external learning from the additional training data by using a convolutional neural network. Moreover, the external prior is learned from the high-resolution MSI by exploiting the local low-rank property. Finally, we incorporate the learned external and internal priors into a unified optimization framework. Experiments on simulated and real datasets demonstrate the effectiveness of the proposed method.

Acknowledgements This work was supported by National Key Research and Development Program of China (Grant No. 2021YFA0715203), National Natural Science Foundation of China (Grant Nos. 62201205, 62221002, 61890962), Key Laboratory of Visual Perception and Artificial Intelligence Fund of Hunan Province (Grant No. 2018TP1013), and Changsha Natural Science Foundation of China (Grant No. kq2202170).

References

1 Gu Y, Liu T, Gao G, et al. Multimodal hyperspectral remote sensing: an overview and perspective. Sci China Inf Sci, 2021, 64: 121301
2 Xu F, Hu C, Li J, et al. Special focus on deep learning in remote sensing image processing. Sci China Inf Sci, 2020, 63: 140300
3 Mei S, Geng Y, Hou J, et al. Learning hyperspectral images from RGB images via a coarse-to-fine CNN. Sci China Inf Sci, 2022, 65: 152102
4 Hong D, Gao L, Yokoya N, et al. More diverse means better: multimodal deep learning meets remote-sensing imagery classification. IEEE Trans Geosci Remote Sens, 2020, 59: 4340–4354
5 Hong D, Yokoya N, Chanussot J, et al. CoSpace: common subspace learning from hyperspectral-multispectral correspondences. IEEE Trans Geosci Remote Sens, 2019, 57: 4349–4359
6 Hou Z, Li W, Tao R, et al. Collaborative representation with background purification and saliency weight for hyperspectral anomaly detection. Sci China Inf Sci, 2022, 65: 112305
7 Akbari H, Kosugi Y, Kojima K, et al. Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging. IEEE Trans Biomed Eng, 2010, 57: 2011–2017
8 Pan Z H, Healey G, Prasad M, et al. Face recognition in hyperspectral images. IEEE Trans Pattern Anal Machine Intell, 2003, 25: 1552–1560
9 Wang L, Xiong Z, Huang H, et al. High-speed hyperspectral video acquisition by combining nyquist and compressive sampling. IEEE Trans Pattern Anal Mach Intell, 2018, 41: 857–870
10 Wang L, Xiong Z, Shi G, et al. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. IEEE Trans Pattern Anal Mach Intell, 2016, 39: 2104–2111
11 Dian R, Fang L, and Li S. Hyperspectral image super-resolution via non-local sparse tensor factorization. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017. 5344–5353
12 Qu Y, Qi H, Ayhan B, et al. Does multispectral/hyperspectral pansharpening improve the performance of anomaly detection? In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2017. 6130–6133
13 Gomez-Chova L, Tuia D, Moser G, et al. Multimodal classification of remote sensing images: a review and future directions. Proc IEEE, 2015, 103: 1560–1584
14 Ferraris V, Dobigeon N, Wei Q, et al. Robust fusion of multiband images with different spatial and spectral resolutions for change detection. IEEE Trans Comput Imag, 2017, 3: 175–186
15 Dian R, Li S, Sun B, et al. Recent advances and new guidelines on hyperspectral and multispectral image fusion. Inf Fusion, 2021, 69: 40–51

16 Deng L J, Feng M, Tai X C. The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior. Inf Fusion, 2019, 52: 76–89
17 Ma J, Yu W, Chen C, et al. Pan-GAN: an unsupervised pan-sharpening method for remote sensing image fusion. Inf Fusion, 2020, 62: 110–120
18 Meng X, Xiong Y, Shao F, et al. A large-scale benchmark data set for evaluating pansharpening performance: overview and implementation. IEEE Geosci Remote Sens Mag, 2021, 9: 18–52
19 Meng X, Shen H, Yuan Q, et al. Pansharpening for cloud-contaminated very high-resolution remote sensing images. IEEE Trans Geosci Remote Sens, 2018, 57: 2840–2854
20 Zhou Z M, Ma N, Li Y X, et al. Variational PCA fusion for Pan-sharpening very high resolution imagery. Sci China Inf Sci, 2014, 57: 112107
21 Chen Z, Pu H, Wang B, et al. Fusion of hyperspectral and multispectral images: a novel framework based on generalization of pan-sharpening methods. IEEE Geosci Remote Sens Lett, 2014, 11: 1418–1422
22 Selva M, Aiazzi B, Butera F, et al. Hyper-sharpening: a first approach on SIM-GA data. IEEE J Sel Top Appl Earth Obs Remote Sens, 2015, 8: 3008–3024
23 Dian R, Li S, Fang L, et al. Multispectral and hyperspectral image fusion with spatial-spectral sparse representation. Inf Fusion, 2019, 49: 262–270
24 Dong W, Fu F, Shi G, et al. Hyperspectral image super-resolution via non-negative structured sparse representation. IEEE Trans Image Process, 2016, 25: 2337–2352
25 Han X H, Shi B, Zheng Y. Self-similarity constrained sparse representation for hyperspectral image super-resolution. IEEE Trans Image Process, 2018, 27: 5625–5637
26 Simoẽs M, Bioucas-Dias J, Almeida L B, et al. A convex formulation for hyperspectral image superresolution via subspace-based regularization. IEEE Trans Geosci Remote Sens, 2015, 53: 3373–3388
27 Nascimento J M P, Bioucasdias J M B. Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE Trans Geosci Remote Sens, 2005, 43: 898–910
28 Zhou Y, Feng L, Hou C, et al. Hyperspectral and multispectral image fusion based on local low rank and coupled spectral unmixing. IEEE Trans Geosci Remote Sens, 2017, 55: 5997–6009
29 Liu J, Wu Z, Xiao L, et al. A truncated matrix decomposition for hyperspectral image super-resolution. IEEE Trans Image Process, 2020, 29: 8028–8042
30 Wu R, Ma W K, Fu X, et al. Hyperspectral super-resolution via global-local low-rank matrix estimation. IEEE Trans Geosci Remote Sens, 2020, 58: 7125–7140
31 Ren K, Sun W, Meng X, et al. A locally optimized model for hyperspectral and multispectral images fusion. IEEE Trans Geosci Remote Sens, 2022, 60: 1–15
32 Dian R, Li S, Fang L, et al. Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion. IEEE Trans Cybern, 2019, 50: 4469–4480
33 Li S, Dian R, Fang L, et al. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. IEEE Trans Image Process, 2018, 27: 4118–4130
34 Chang Y, Yan L, Zhao X L, et al. Weighted low-rank tensor recovery for hyperspectral image restoration. IEEE Trans Cybern, 2020, 50: 4558–4572
35 Xu Y, Wu Z, Chanussot J, et al. Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion. IEEE Trans Geosci Remote Sens, 2019, 58: 348–362
36 Liu N, Li L, Li W, et al. Hyperspectral restoration and fusion with multispectral imagery via low-rank tensor-approximation. IEEE Trans Geosci Remote Sens, 2021, 59: 7817–7830
37 He W, Chen Y, Yokoya N, et al. Hyperspectral super-resolution via coupled tensor ring factorization. 2021. ArXiv:2001.01547
38 Chen Y, Zeng J, He W, et al. Hyperspectral and multispectral image fusion using factor smoothed tensor ring decomposition. IEEE Trans Geosci Remote Sens, 2022, 60: 1–17
39 Fu X, Wang W, Huang Y, et al. Deep multiscale detail networks for multiband spectral image sharpening. IEEE Trans Neural Netw Learn Syst, 2021, 32: 2090–2104
40 Dong W, Zhang T, Qu J, et al. Laplacian pyramid dense network for hyperspectral pansharpening. IEEE Trans Geosci Remote Sens, 2022, 60: 1–13
41 Hu J F, Huang T Z, Deng L J, et al. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. IEEE Trans Neural Netw Learn Syst, 2022, 33: 7251–7265
42 Wang Z, Chen B, Lu R, et al. FusionNet: an unsupervised convolutional variational network for hyperspectral and multispectral image fusion. IEEE Trans Image Process, 2020, 29: 7565–7577
43 Wang W, Fu X, Zeng W, et al. Enhanced deep blind hyperspectral image fusion. IEEE Trans Neural Netw Learn Syst, 2021. doi: 10.1109/TNNLS.2021.3105543
44 Wang W, Zeng W, Huang Y, et al. Deep blind hyperspectral image fusion. In: Proceedings of IEEE/CVF International Conference on Computer Vision, 2019. 4149–4158
45 Dian R, Li S, Guo A, et al. Deep hyperspectral image sharpening. IEEE Trans Neural Netw Learn Syst, 2018, 29: 5345–5355
46 Zheng Y, Li J, Li Y, et al. Edge-conditioned feature transform network for hyperspectral and multispectral image fusion. IEEE Trans Geosci Remote Sens, 2022, 60: 1–15
47 Dong W, Zhou C, Wu F, et al. Model-guided deep hyperspectral image super-resolution. IEEE Trans Image Process, 2021, 30: 5754–5768
48 Xie Q, Zhou M, Zhao Q, et al. MHF-Net: an interpretable deep network for multispectral and hyperspectral image fusion. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 1457–1473
49 Xie Q, Zhou M, Zhao Q, et al. Multispectral and hyperspectral image fusion by MS/HS fusion net. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2019. 1585–1594
50 Yang J, Zhao Y Q, Chan J. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. Remote Sens, 2018, 10: 800
51 Dian R, Li S, Kang X. Regularizing hyperspectral and multispectral image fusion by CNN denoiser. IEEE Trans Neural Netw Learn Syst, 2021, 32: 1124–1135
52 Yang J, Xiao L, Zhao Y Q, et al. Variational regularization network with attentive deep prior for hyperspectral-multispectral image fusion. IEEE Trans Geosci Remote Sens, 2022, 60: 1–17
53 Zheng K, Gao L, Hong D, et al. NonRegSRNet: a nonrigid registration hyperspectral super-resolution network. IEEE Trans Geosci Remote Sens, 2022, 60: 1–16

54 Huang T, Dong W, Wu J, et al. Deep hyperspectral image fusion network with iterative spatio-spectral regularization. IEEE Trans Comput Imag, 2022, 8: 201–214

55 Dian R, Li S. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. IEEE Trans Image Process, 2019, 28: 5135–5146

56 Zhuang L, Bioucas-Dias J M. Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations. IEEE J Sel Top Appl Earth Obs Remote Sens, 2018, 11: 730–742

57 Zhuang L, Ng M K, Fu X, et al. Hy-demosaicing: hyperspectral blind reconstruction from spectral subsampling. IEEE Trans Geosci Remote Sens, 2022, 60: 1–15

58 Boyd S. Distributed optimization and statistical learning via the alternating direction method of multipliers. FNT Machine Learn, 2011, 3: 1–122

59 Dian R, Li S, Fang L. Learning a low tensor-train rank representation for hyperspectral image super-resolution. IEEE Trans Neural Netw Learn Syst, 2019, 30: 2672–2683

60 Wei Q, Dobigeon N, Tourneret J Y. Fast fusion of multi-band images based on solving a sylvester equation. IEEE Trans Image Process, 2015, 24: 4109–4121

61 Liu M, Tuzel O, Ramalingam S, et al. Entropy rate superpixel segmentation. In: Proceedings of Conference on Computer Vision and Pattern Recognition, 2011. 2097–2104

62 Zhang K, Zuo W, Zhang L. FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans Image Process, 2018, 27: 4608–4622

63 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015. 448–456

64 Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning, 2010. 807–814

65 Dell'Acqua F, Gamba P, Ferrari A, et al. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. IEEE Geosci Remote Sens Lett, 2004, 1: 322–326

66 Green R O, Eastwood M L, Sarture C M, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). Remote Sens Environ, 1998, 65: 227–248

67 Aiazzi B, Baronti S, Selva M. Improving component substitution pansharpening through multivariate regression of MS +Pan data. IEEE Trans Geosci Remote Sens, 2007, 45: 3230–3239

68 Yokoya N, Yairi T, Iwasaki A. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans Geosci Remote Sens, 2012, 50: 528–537

69 Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 2004, 13: 600–612

70 Wald L. Quality of high resolution synthesised images: is there a simple criterion? In: Proceedings of the 3rd Conference Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images, 2000. 99–103

71 Wang Z, Bovik A C. A universal image quality index. IEEE Signal Process Lett, 2002, 9: 81–84

72 Yokoya N, Grohnfeldt C, Chanussot J. Hyperspectral and multispectral data fusion: a comparative review of the recent literature. IEEE Geosci Remote Sens Mag, 2017, 5: 29–56