Hyperspectral Image Superresolution via Subspace-Based Deep Prior Regularization

Jianwei Zheng [ID], *Member, IEEE*, Pengfei Li [ID], Honghui Xu [ID], Jiawei Jiang [ID], Yuchao Feng, and Zhi Liu [ID]

Abstract—Hyperspectral imaging is able to provide a finer delivery of various material properties than conventional imaging systems. Yet in reality, an optical system can only generate data with high spatial resolution but low spectral one, or vice versa, at video rates. As a result, an issue that fuses low-resolution hyperspectral and high-resolution multispectral images has gained great attention. However, most fusion approaches depend purely on hand-crafted regularizers or data-driven priors, leading to the issues of tricky parameter selection or poor interpretability. In this work, a subspace-based deep prior regularization is proposed to tackle these problems, which takes both hand-crafted regularizer and data-driven prior into account. Specifically, we leverage the spectral correlation of the images and transfer them from the original space to the subspace domain, within which a modified U-net-based deep prior learning network (SDPL-net) is designed for the fusion issue. Moreover, instead of taking the output of SDPL-net directly as the result, we further feed the output back to the model-based optimization. Under such prior regularization, the recovered high-resolution hyperspectral image holds a high consistency to its inherent structure and hence tends to present enhanced reliability and accuracy. Experimental results on simulated and real data reveal that the proposed method excels other state-of-the-art methods in both quantitative and qualitative metrics.

Index Terms—Deep learning (DL), fusion, hyperspectral image (HSI) superresolution, low-rank subspace, Sylvester equation.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) comprise various bands acquired by the sensor at different wavelengths, which allow more reliable dissemination of information in real scenes than conventional images with only a few bands. The rich spectrum of HSIs can facilitate the characterization of imaging scenes and greatly improve the performance of numerous computer vision applications, e.g., object detection, classification, tracking, and segmentation [1], [2], [3], [4], [5], [6], [7], [8]. However, owing to the finite incident energy, there exists a crucial tradeoff between spatial and spectral resolution for real-world applications. Specifically, a typical optical system may only generate images with high spatial resolution but a

limited amount of spectra [i.e., high-resolution multispectral (HRMS) images], or images with plenty of spectral bands but a low spatial resolution [i.e., low-resolution hyperspectral (LRHS) images] [9]. Therefore, a research topic of HSI superresolution, which fuses these two kinds of images to produce high-resolution hyperspectral (HRHS) images has received much attention [10], [11]. In practice, following a physically delicate degradation model, the input LRHS image and HRMS image are considered as spectrally and spatially degraded observations (i.e., linear down-sampled versions) of latent HRHS images, respectively.

Commonly, HSI fusion is a highly ill-posed issue due to large scaling factors in both spectra and space. Most traditional approaches consider the incorporation of certain hand-crafted prior to bound the solution space, which holds the potential of achieving acceptable performance without the support form extra training samples [12], [13]. According to the structure of the prior knowledge employed in optimization, available methodologies can be broadly categorized into three types: spectral unmixing-based methods [12], [14], [15], [16], sparse representation-based methods [17], [18], [19], [20], [21], and tensor decomposition-based methods [13], [22], [23], [24], [25], [26], [27]. However, these prior constrained approaches are still inherently flawed in three ways. First, the involved optimization often requires excessive iterations, which further demands prolonged time to recover an HRHS image. Second, a single prior tends to express only one aspect of the visual properties, hence is insufficient to fully reveal the complex structure of HSIs. Third, the blend of multiple hand-crafted priors may pose further hardships in dealing with the concerned optimization issue.

In recent years, inspired by the strong fitting capacity of deep learning (DL) in many computer vision applications, most convolutional neural networks (CNNs) have been introduced to the superresolution field [28], [29], [30], [31], [32], [33]. In comparison to optimization methodologies based on predefined regularizer, the DL approaches demand little assumption on the prior knowledge of general HSIs, putting in their best effort to learn latent information from the training data in an end-to-end fashion. However, these approaches often ignore the blurring and down-sampling operators as well as the spectral response function during degradation process, even these components have a clear physical interpretation that relates LRHS and HRMS to the HRHS image. To relieve this problem, more current end-to-end methods [34], [35], [36] succeed in improving the quality of the recovered images by learning spectral response function and down-sampling operator. Moreover, several deep

prior-based methods are lately presented [37], [38], [39], which marry the merits of both degradation and DL approaches.

Albeit attaining encouraging performance, the DL-based superresolution methods still suffer from some drawbacks. The most critical one is the ignorance of the explicit priors, which are known as the inherent properties of most HSIs, e.g., spectral low-rank property [40]. This may lead to a deviation of the network output from the general prior configuration and impact negatively on the recovery accuracy. Another significant challenge facing the current deep prior-based method lies in the overfitting issue. Mostly, due to labor and hardware costs, only a limited amount of paired LRHS and HRMS can be collected. As a result, the mismatch issue of training-testing bias often exists in practical scenarios. Accordingly, the overfitting issue tends to occur in this circumstance. Other drawbacks lie in the specific design of the network framework for two inputs differing in sizes and features, which inevitably results in complex architectures, tremendous parameters, and information distortions.

Against the abovementioned issues, we attempt to perform hyperspectral fusion by using jointly the handcrafted prior as well as the data-driven knowledge, which naturally holds the opportunity of providing better performance and meanwhile generalizing well [10], [41]. Note that the joint consideration also taps the potential of using simpler network architecture for the fusion problem. The specific and primary contributions of this work can be highlighted as follows.

1) Since the bands of HRHS image are highly correlated, the spectral vectors often live in a low-dimensional subspace. On that basis, we propose a subspace-based deep prior regularization to recover the HRHS images using two factor matrices. The first is produced via a simple singular value decomposition (SVD) of the LRHS image, with the guarantees of leveraging the spectral correlation and retaining the low-rank property. The second matrix can be readily leaned by a relatively concise network due to the narrowed band size.

2) To avoid the hand-crafted regularizer as used in most traditional approaches, we develop a U-net-based model, namely SDPL-net, to learn deep priors in the subspace domain instead of the original one. Note that to highlight the individual features of each band, the RCAB [42] block is further introduced, permitting the low-frequency structures to be bypassed and enabling more concentration on high-frequency information. As a result, the recovered HRHS image finely complies with the intrinsically visual structure and guarantees better reliability, benefitting both from the explicitly low-rank property as well as the implicitly learned prior.

3) Instead of using SDPL-net directly as the outcome producer, we further feed the output, behaving as a prior-learned regularizer, back to the optimization-based model. Such a prior-inspired method enjoys promising generalization capability. When the spatial and spectral information in the training and test images is constrained, our proposed method can automatically extract low-rank and deep priors from latent HRHS images. For more tough circumstances in which data are captured from diverse optical scenarios, even different sensors with variant spatial and

spectral information, SDPR can be readily generalized to new test samples and perform well.

The rest of this article is organized as follows. The relevant literature is reviewed in Section II. In Section III, the proposed HSI fusion method is described in details. Section IV displays and analyzes experiments on several publicly available datasets. Finally, Section V concludes this article and conceives some future investigations.

II. RELATED WORK

This section provides a brief review of several existing approaches relevant to our work.

A. Spectral Unmixing Based Methods

Leveraging prior to spectral unmixing has been experimentally instantiated to be conducive for HSI fusion, mostly under certain constraints, such as nonnegativity and sum-to-one. The work [14] develops a coupled nonnegative matrix factorization for alternately unmixing LRHS and HRMS images in order to recover the HRHS one. With a similar framework, [15] decouples the optimization issue into two restricted least square problems, followed by a solver that jointly unmixes the two input images. Concerning the endmembers and their abundances, in [16], the latent HRHS image is reconstructed employing the alternating direction method of multipliers [43]. By fully exploiting the specific properties of matrix decomposition, a four-stage fusion framework dubbed as MDF is proposed in [12].

B. Sparse Representation Based Methods

Sparse representation is another promising scheme for the fusion of LRHS and HRMS images, which sparsely encodes the latent HRHS image using a proper spectral dictionary acquired from the inputs. For instance, [17] merges hyperspectral and multispectral images following a variational framework, in which a sparse regularizer is designed by factorizing the scenes over a given dictionary. More recently, a nonnegative sparse coding scheme is proposed [18], which employs not only the pixelwise sparsity but also the nonlocal spatial similarity, yielding an improved performance. Similarly, in [20], a fusion method based on nonlocal low-rank tensor approximation (LRTA) and sparse representation is given, in which the sparsity of abundances is highlighted by a delicately devised constraint. Depending on a newly structured low-rank representation, the approach in [21] represents images as linear combinations of bases in an appropriately learned dictionary, inducing sparse property inferred from certain subspace decomposition of the affinity matrix.

C. Tensor Decomposition-Based Methods

Another crucial technique for HSI fusion relies on tensor decomposition, which has been widely employed to reveal the correlations among different modes. In [22], the fusion issue is formulated as a coupled sparse Tucker decomposition, alternately updating dictionaries of three modes as well as the sparse core tensor. The work in [23] develops a low tensor train rank for the learning of correlations among spatial, spectral, and

nonlocal modes of all analogous cubes. A novel subspace-based low-tensor multirank regularization (LTMR) is proposed in [24], which leverages both the spectral correlation and the nonlocal similarity in HRHS images. Through a coupled tensor canonical polyadic decomposition, [25] probes into the intrinsic relationship between HRHS and LRMS images. The approach in [26] develops a novel low-rank tensor ring decomposition for HSI superresolution. By exploiting the idea from the field of tensor completion, [13] develops an LRTA method to directly enforce spatial and spectral low rankness while avoiding computationally complicated steps, such as patch clustering and dictionary learning.

D. End-to-End Learning Methods

In [34], an unsupervised adaptation learning (UAL) method is proposed, which develops a two-stage CNN to leverage characteristics of latent HRHS targets and estimate the unknown down-sampling operator simultaneously. With the aid of attention mechanism and pixel shuffling, in [35], a simple yet effective CNN architecture is presented holding a high reliability of outputting spatially high-quality details. In [36], a new fusion model is formulated that considers both spectral and spatial degradation. Accordingly, a deep CNN dubbed MHF-net is raised to tackle the fusion issue iteratively. Unfortunately, these end-to-end learning methods still have evident drawbacks. For instance, given limited samples, the data-driven approaches are sensitive to variations in spatial and spectral information, suffering severely from the mismatch between training and test sets [35]. Such an overfitting problem happens frequently, since real data are generally acquired under changing environmental settings or even with different sensors.

E. Deep Prior-Based Methods

The work in [37] shows that neural networks can naturally capture the visual prior implicitly. On that basis, acceptable results can be guaranteed in standard inverse problems, such as denoising, superresolution, and inpainting. Following this idea, in [38], a deep hyperspectral image sharpening (DHSIS) method is presented to directly learn deep prior via CNN-based residual learning. In [39], a novel network architecture is developed to leverage both the spectral and spatial characteristics of latent HRHS images, whose outputs serve to regularize the superresolution issue. Albeit with acceptable performance, the explicit priors, such as spectral correlation, possessed by most HSIs, are often ignored by such "black-box" deep models, leaving considerable room for further performance improvements. Our approach is inspired by this line of research. As a whole, we aim to construct a framework that incorporates the explicitly hand-crafted regularizer and the implicitly learned prior in our investigation.

III. PROPOSED METHOD

A. Preliminaries

The desired HRHS image is denoted as $\mathcal{X} \in \mathbb{R}^{W \times H \times S}$, where W, H, and S are the width, height, and band number

of the target image, respectively. Accordingly, $\mathcal{Y} \in \mathbb{R}^{w \times h \times S}$ denotes the observed LRHS image with $w \times h$ pixels ($w < W$, $h < H$) and S spectral bands. Following the physically delicate degeneracy model, $\mathcal{Y}$ is a spatially degraded observation of $\mathcal{X}$:

$$\mathbf{Y}_{(3)} = \mathbf{X}_{(3)}\mathbf{BS} \tag{1}$$

where, the matrices $\mathbf{X}_{(3)} \in \mathbb{R}^{S \times WH}$ and $\mathbf{Y}_{(3)} \in \mathbb{R}^{S \times wh}$ are acquired by unfolding $\mathcal{X}$ and $\mathcal{Y}$ along the third mode, respectively. $\mathbf{S} \in \mathbb{R}^{WH \times wh}$ denotes the down-sampling matrix. The blurring matrix is represented by $\mathbf{B} \in \mathbb{R}^{WH \times WH}$. Practically, $\mathbf{B}$ can be diagonalized as

$$\mathbf{B} = \mathbf{F}\mathbf{K}\mathbf{F}^{-1} \tag{2}$$

where, $\mathbf{F}$ denotes the fast Fourier transform (FFT) and $\mathbf{F}^{-1}$ represents inverse FFT ($\mathbf{F}\mathbf{F}^{-1} = \mathbf{I}_{WH}$, and $\mathbf{I}_{WH}$ is a $WH \times WH$ identity matrix). Eigenvalues of $\mathbf{B}$ are held by diagonal matrix $\mathbf{K}$.

Within the same scene, the HRMS image with $W \times H$ pixels and s spectral bands ($s < S$) is represented by $\mathcal{Z} \in \mathbb{R}^{W \times H \times s}$. $\mathcal{Z}$ can be modeled as a spectrally degenerate observation of $\mathcal{X}$, that is

$$\mathbf{Z}_{(3)} = \mathbf{R}\mathbf{X}_{(3)} \tag{3}$$

where, $\mathbf{Z}_{(3)} \in \mathbb{R}^{s \times WH}$ is the matrix acquired by unfolding $\mathcal{Z}$ along the third mode. $\mathbf{R} \in \mathbb{R}^{s \times S}$ is the spectral response matrix.

Due to the fact that HSI owns a close correlation among different bands, the globally spectral low-rank property is a dominant prior for most subsequent applications. Particularly, the pixels located at the same spatial position but with different spectral bands often live in a low-dimensional subspace, whose ambient size is much less than that of the original one. Therefore, it is a natural operation to reveal the spectral low-rank property by using the subspace decomposition technology, i.e.,

$$\mathbf{X}_{(3)} = \mathbf{D}\mathbf{C}_{(3)}. \tag{4}$$

Here, $\mathbf{D} \in \mathbb{R}^{S \times L}$ denotes the domain switching matrix that records the mapping from the low-rank subspace domain to the original image domain. L is the number of atoms resulted from the subspace representation. The matrix $\mathbf{C}_{(3)} \in \mathbb{R}^{L \times WH}$ is acquired by unfolding $\mathcal{C} \in \mathbb{R}^{W \times H \times L}$ with the third mode, where $\mathcal{C}(i, j, :)$ is the coefficient of spectral pixel $\mathcal{X}(i, j, :)$. Regarding the subspace, two remarks are made: a) value of L is small, that is $L < S$, which means that the spectral vectors lie in a low-dimensional subspace. Generally, the reduction of dimension favors a more efficient computation; b) columns of subspace are orthogonal, implying that each spectral vector $\mathcal{X}(i, j, :)$ holds the same Frobenius norm as the corresponding coefficient $\mathcal{C}(i, j, :)$. In this manner, self-similarity among spectra can be translated into the subspace domain.

Thus, the superresolution issue can be rephrased into estimating the switching matrix $\mathbf{D}$ and coefficients $\mathcal{C}$ from two input images. Since the LRHS image retains most of the spectral information from the HRHS image, the two images are supposed to lie in the same spectral subspace. Therefore, by borrowing the simple SVD decomposition, we can estimate the domain
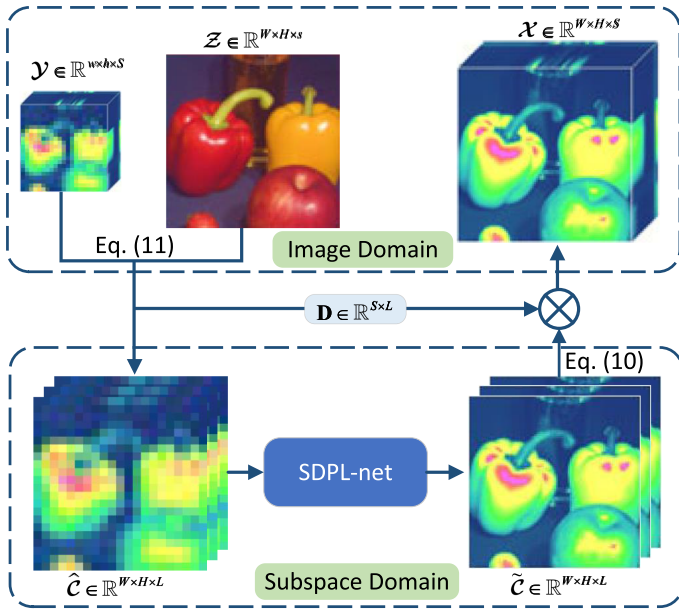
Fig. 1. Overall scheme of our proposed SDPR method.

switching matrix from the LRHS image

$$\mathbf{Y}_{(3)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \qquad (5)$$

where, $\mathbf{U}$ and $\mathbf{V}$ are both orthogonal matrices and the diagonal matrix $\boldsymbol{\Sigma}$ contains singular values aligned in descending order. By holding the L largest singular values and eliminating the remaining ones, the learned subspace matrix $\mathbf{D}$ is given as

$$\mathbf{D} = \mathbf{U}(:, 1 : L). \qquad (6)$$

With all the preliminaries prepared, the overall flowchart of our method is given in Fig. 1. Generally, there are three main steps, including a preacquisition of the initial estimation, a learning of subspace-based deep prior, and an integration of the prior into the degradation model. In the sequel, we first provide the integration idea, followed by a sequential presentations of different steps.

B. Integrating Prior Into the Degradation Model

Using the degenerate model in (1), (3), and (4), $\mathcal{C}$ can be estimated by introducing appropriate priors and penalizing the following biobjective function:

$$\widehat{\mathbf{C}}_{(3)} = \underset{\mathbf{C}_{(3)}}{\arg\min} \, \mathcal{J}(\mathbf{C}_{(3)})$$

$$= \underset{\mathbf{C}_{(3)}}{\arg\min} \, \mathcal{J}_1(\mathbf{C}_{(3)}) + \mu \mathcal{J}_2(\mathbf{C}_{(3)}) \qquad (7)$$

where

$$\mathcal{J}_1(\mathbf{C}_{(3)}) = \left\| \mathbf{Y}_{(3)} - \mathbf{D}\mathbf{C}_{(3)}\mathbf{B}\mathbf{S} \right\|_F^2 + \left\| \mathbf{Z}_{(3)} - \mathbf{R}\mathbf{D}\mathbf{C}_{(3)} \right\|_F^2 \qquad (8)$$

with $\| \cdot \|_F^2$ representing the matrix Frobenius norm. $\mathcal{J}_1(\mathbf{C}_{(3)})$ denotes a data fidelity term. $\mathcal{J}_2(\mathbf{C}_{(3)})$ represents a regularizer that shrinks the solution space and enforces certain prior properties. Moreover, $\mu > 0$ is a hyperparameter that balances $\mathcal{J}_1$

and $\mathcal{J}_2$. As demonstrated in (7), the subspace-based deep prior structure of latent HRHS images is encoded in $\mathcal{J}_2(\mathbf{C}_{(3)})$. As known, it is nontrivial to handcraft a sophisticated regularizer. Recently, benefiting from the rocketing development of the variable splitting technique, plug-and-play approaches have been introduced to address diverse HSI inversion issues [44], [45], [46]. In contrast, in this article, we design a regularizer that leverages network output $\widetilde{\mathcal{C}} \in \mathbb{R}^{W \times H \times L}$ to ensure a close solution of (8) to the learned subspace-based deep prior. Concretely, we let $\mathcal{J}_2(\mathbf{C}_{(3)})$ be the squared Euclidean distance between $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ in the subspace coefficient domain

$$\mathcal{J}_2(\mathbf{C}_{(3)}) = \left\| \mathbf{C}_{(3)} - \widetilde{\mathbf{C}}_{(3)} \right\|_F^2 \qquad (9)$$

where, $\widetilde{\mathbf{C}}_{(3)} \in \mathbb{R}^{L \times WH}$ is the matrix resulted from the unfolding of $\widetilde{\mathcal{C}}$ along with the third mode. Then (7) can be rewritten as

$$\underset{\mathbf{C}_{(3)}}{\arg\min} \left\| \mathbf{Y}_{(3)} - \mathbf{D}\mathbf{C}_{(3)}\mathbf{B}\mathbf{S} \right\|_F^2 + \left\| \mathbf{Z}_{(3)} - \mathbf{R}\mathbf{D}\mathbf{C}_{(3)} \right\|_F^2$$

$$+ \mu \left\| \mathbf{C}_{(3)} - \widetilde{\mathbf{C}}_{(3)} \right\|_F^2. \qquad (10)$$

Compared to other possible formulations, cost function (10) allows a straightforward solution owing to the differentiability of the Frobenius norm. Employing a promising deep prior structure for $\widetilde{\mathcal{C}}$ is the key point to achieve a satisfactory recovery of the latent HRHS image.

C. Preoptimization

Although deep CNN can efficiently learn the prior from existing HSIs, two inputs $\mathcal{Y}$ and $\mathcal{Z}$ that contain redundant information and differ in sizes are difficult to map to $\mathcal{C}$ directly. Not that only the spatial and spectral features are tough to extract individually, but the information distortion in feature merging of different dimensions is also hard to implement. We tackle this issue by a preacquisition of estimated $\widehat{\mathcal{C}} \in \mathbb{R}^{W \times H \times L}$ by adequately exploiting $\mathcal{Y}$ and $\mathcal{Z}$, followed by an immediate feed into SDPL-net. Similar to (10), the practical implementation simply replaces the third term of $\mathcal{J}_2$ with the Euclidean distance between $\mathbf{D}\widehat{\mathbf{C}}_{(3)}$ and $\mathbf{Y}_{up(3)}$ in the original image domain

$$\underset{\widehat{\mathbf{C}}_{(3)}}{\arg\min} \left\| \mathbf{Y}_{(3)} - \mathbf{D}\widehat{\mathbf{C}}_{(3)}\mathbf{B}\mathbf{S} \right\|_F^2 + \left\| \mathbf{Z}_{(3)} - \mathbf{R}\mathbf{D}\widehat{\mathbf{C}}_{(3)} \right\|_F^2$$

$$+ \lambda \left\| \mathbf{D}\widehat{\mathbf{C}}_{(3)} - \mathbf{Y}_{up(3)} \right\|_F^2 \qquad (11)$$

where, $\widehat{\mathbf{C}}_{(3)} \in \mathbb{R}^{L \times WH}$ and $\mathbf{Y}_{up(3)} \in \mathbb{R}^{S \times WH}$ are, respectively, the unfolded versions of the expected estimation $\widehat{\mathcal{C}}$ and the upsampled LRHR image $\mathcal{Y}_{up} \in \mathbb{R}^{W \times H \times S}$. $\lambda > 0$ is the tradeoff parameter. In our implementation, the bicubic interpolation scheme is selected as the upsampling operation. Note that $\widetilde{\mathbf{C}}_{(3)}$ and $\widehat{\mathbf{C}}_{(3)}$ denote different stages of the subspace coefficient $\mathbf{C}_{(3)}$ when the algorithm is running, so that different superscript hats are used for ease of identification. That is to say, $\widetilde{\mathbf{C}}_{(3)}$ and $\widehat{\mathbf{C}}_{(3)}$ also satisfy the properties of $\mathbf{C}_{(3)}$, e.g., $\mathbf{X}_{(3)} = \mathbf{D}\widehat{\mathbf{C}}_{(3)}$, during the optimization stage.

Algorithm 1: Solution by Solving the Sylvester Equation w.r.t. $\mathcal{Y}_{\text{up}}$.

Input: $\mathbf{Y}_{(3)}, \mathbf{Z}_{(3)}, \mathbf{B}, \mathbf{S}, \mathbf{R}, \mathbf{D}, \mathbf{Y}_{up(3)}, \lambda$.

1 **Initialization**: $\mathbf{H}_1 = (\mathbf{RD})^T\mathbf{RD} + \lambda\mathbf{I}_L$,
$\mathbf{H}_2 = \mathbf{BS}(\mathbf{BS})^T$,
$\mathbf{H}_3 = \mathbf{D}^T\mathbf{Y}_{(3)}(\mathbf{BS})^T + (\mathbf{RD})^T\mathbf{Z}_{(3)} + \lambda\mathbf{D}^T\mathbf{Y}_{up(3)}$.

2 Eigen-decomposition of $\mathbf{B}$: $\mathbf{B} = \mathbf{FKF}^H$;

3 $\widetilde{\mathbf{K}} = \mathbf{K}(\mathbf{1}_d \otimes \mathbf{1}_{wh})$;

4 Eigen-decomposition of $\mathbf{H}_1$: $\mathbf{H}_1 = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}$;

5 $\widetilde{\mathbf{H}}_3 = \mathbf{Q}^{-1}\mathbf{H}_3\mathbf{F}$;

6 **for** $l = 1$ to L **do**

7 $\quad \mathbf{o}_l = \lambda_l^{-1}(\widetilde{\mathbf{H}}_3)_l - \lambda_l^{-1}(\widetilde{\mathbf{H}}_3)_l\widetilde{\mathbf{K}}(\lambda_l d\mathbf{I}_n + \sum_{t=1}^{d}\mathbf{K}_t^2)\widetilde{\mathbf{K}}^H$;

8 **end**

9 Set $\widehat{\mathbf{C}}_{(3)} = \mathbf{QOF}^H$;

Output: $\widehat{\mathbf{C}}_{(3)}$.

To solve (11), we enforce the derivative of (11) w.r.t. $\widehat{\mathcal{C}}$ to be zero. As a result, the solution of (11) can be further achieved by the following Sylvester equation:

$$\mathbf{H}_1\widehat{\mathbf{C}}_{(3)} + \widehat{\mathbf{C}}_{(3)}\mathbf{H}_2 = \mathbf{H}_3. \tag{12}$$

Since the domain switching matrix is acquired by SVD, it fulfills $\mathbf{D}^T\mathbf{D} = \mathbf{I}_L$. Thus, we are able to derive the following equation:

$$\mathbf{H}_1 = (\mathbf{RD})^T\mathbf{RD} + \lambda\mathbf{I}_L$$
$$\mathbf{H}_2 = \mathbf{BS}(\mathbf{BS})^T$$
$$\mathbf{H}_3 = \mathbf{D}^T\mathbf{Y}_{(3)}(\mathbf{BS})^T + (\mathbf{RD})^T\mathbf{Z}_{(3)} + \lambda\mathbf{D}^T\mathbf{Y}_{\text{up}(3)} \tag{13}$$

where, $\mathbf{I}_L$ is the identity matrix with size $L \times L$. According to the theoretical analysis in [47], the Sylvester equation (12) has a unique solution if and only if the summation of eigenvalues of $\mathbf{H}_1$ and $\mathbf{H}_2$ is nonzero. It is known that both $(\mathbf{RD})^T\mathbf{RD}$ and $\mathbf{I}_L$ are positive definite, hence $\mathbf{H}_1$ is positive definite. Moreover, considering that $\mathbf{H}_2$ is a semipositive definite matrix, then the arbitrary summation of eigenvalues of $\mathbf{H}_1$ and $\mathbf{H}_2$ is surely larger than zero, guaranteeing a unique solution of (12). On that basis, we borrow the idea from [48] to analytically and efficiently get the solution of $\widehat{\mathbf{C}}_{(3)}$. The concrete solver is given in Algorithm 1.

In Algorithm 1, convolution blurring $\mathbf{B}$ can be factorized as illustrated in (2), in which the diagonal matrix $\mathbf{K} \in \mathbb{C}^{WH \times WH}$ is given as

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{K}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{K}_d \end{bmatrix} \tag{14}$$

where, $\mathbf{K}_i \in \mathbb{C}^{wh \times wh}$ and d denotes the spatial down-sampling factor. The eigen decomposition of $\mathbf{H}_1$ yields the unitary matrix $\mathbf{Q}$ and the diagonal matrix $\boldsymbol{\Lambda}$, i.e., $\mathbf{H}_1 = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}$, where $\boldsymbol{\Lambda}$ is

expressed as

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_L \end{bmatrix} \tag{15}$$

with λ_l representing the eigenvalue of $\mathbf{H}_1$ and $\mathbf{Q}$ holding the eigenvectors of $\mathbf{H}_1$ in its columns. $\mathbf{o}_l$ denotes the lth row of $\mathbf{O} = \mathbf{Q}^{-1}\mathbf{C}_{(3)}\mathbf{F}$, that is, $\mathbf{O} = [\mathbf{o}_1^T, \mathbf{o}_2^T, \ldots, \mathbf{o}_L^T]^T$. $\mathbf{1}_d \in \mathbb{R}^d$ represents a vector whose elements are all 1.

D. Subspace-Based Deep Prior Learning

Instead of using the hand-crafted regularizer, we propose to learn the prior in the subspace coefficient domain of latent HRHS images from hyperspectral datasets. Note that, after preoptimization, we obtain the initial estimation $\widehat{\mathcal{C}}$ with the same size as the target $\mathcal{C}$ in the subspace coefficient domain. Taking $\widehat{\mathcal{C}}$ and $\mathcal{C}$ as input and output, respectively, in the sequel, we construct an end-to-end deep network dubbed SDPL-net following a U-Net architecture.

U-Net has been proven to be an effective tool for various tasks, including magnetic resonance spectroscopy imaging [49] and biomedical image segmentation [50]. In contrast to another well-known model Resnet [51], U-Net primarily operates by performing pixelwise transformation on input images, which is applicable to the superresolution problem. Using convolutional layers, U-Net first continuously down samples the input image to a small size, which aids in extracting globally valuable features. Next, the image is scaled up through a reverse operation. This process helps to identify locally vital features so as to refine the image at a finer resolution. Unfortunately, the features in U-Net architecture contain redundant low-frequency information equally across all channels, which inevitably hinders the representability. To solve this problem, we employ RCAB in our network. In particular, RCAB permits the low-frequency structure to be bypassed via the multiple skip connections, enabling CNN to concentrate more on high-frequency information. Meanwhile, RCAB adaptively rescales channelwise features taking interdependencies among channels into account.

As shown in Fig. 2, the overall network follows an encoder–decoder design. To begin, a convolution layer without activation function is initially employed to extract the shallow features. In addition, the modules of encoder and decoder are sequentially appended. Note that the encoder module intends to learn the blurring and down-sampling operators, which is much simpler than the decoder module for the learning of upscaling mapping. In practice, the encoder block is designed with only two convolutional layers and a LeakyReLU activation layer, which is considerably efficient, yet it functions well in our experiment. As for the decoder block, unlike the baseline U-Net, RCAB is utilized to boost the model capacity. Besides, shortcut connections, which consider the hierarchical nature of the features, are employed in the feature extraction process. We try this connection since typical convolutional networks tend to finalize their output with high-level features, neglecting the importance of low-level features [8]. In general, high-level features refer to
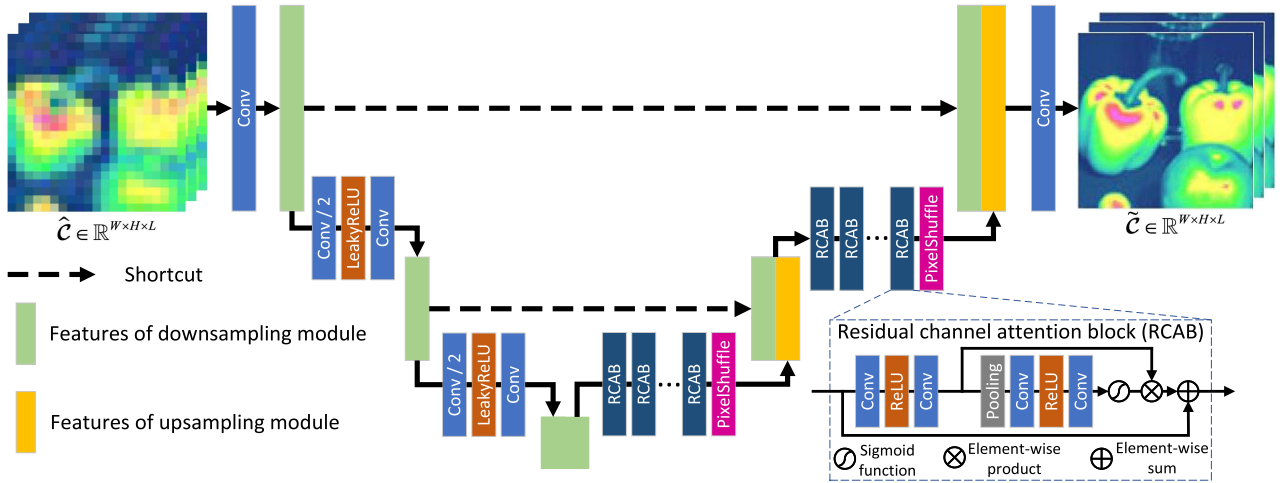
Fig. 2. Architecture of the SDPL-net. The green boxes represent the feature maps of the encoding module and the yellow boxes denote the feature maps of the decoding module. We utilize shortcut connections to concatenate the corresponding shallow and deep feature maps, following the U-Net framework.

semantic knowledge, whereas low-level features mostly facilitate information about contours, boundaries, etc. We argue that low-level features are indispensable for HSI superresolution. Therefore, the features from the encoder block are merged into the decoder block in the feature extraction process.

It has been proven that ℓ_1-norm is more applicable than ℓ_2-norm in low-level visual tasks since it experimentally delivers improved performance [52]. Therefore, the loss function based on ℓ_1-norm is employed

$$\ell(\Theta) = \sum_{m=1}^{M} \left\| \mathcal{F}\left(\widehat{\mathcal{C}}_m; \Theta\right) - \mathcal{C}_m \right\|_1 \qquad (16)$$

where, $\{(\widehat{\mathcal{C}}_m; \mathcal{C}_m)\}_{m=1}^{M}$ represents M training pairs for the learning of parameter Θ in SDPL-net $\mathcal{F}$.

The superresolution result derived from our SDPL-net, referred to $\widetilde{\mathcal{C}}$ in subspace, is again employed in the regularizer $\mathcal{J}_2$, as shown in the biobjective optimization formulation (10). More specifically, $\widetilde{\mathcal{C}}$ behaves as a spatial-spectral prior and contributes positively to the finer estimation of $\mathcal{C}$. The solver of (10) is similar to that of (11) and will not be repeated here. Finally, the desired result of the HRHS image $\mathcal{X}$ can be obtained by calculating (4).

Overall, the joint usage of low-rank decomposition, deep priors, and degradation knowledge is considered in our model. Recall that the hand-crafted prior-based optimization methods could guarantee an acceptable performance given no training samples [12], [13]. Besides, using a network to recover a subspace variable instead of the original one would compress the model parameters. Both of these two merits favor the relief of the overfitting issue. Moreover, the fast inference of Sylvester equation given in Algorithm 1 would guarantee an efficient solution.

IV. EXPERIMENTS

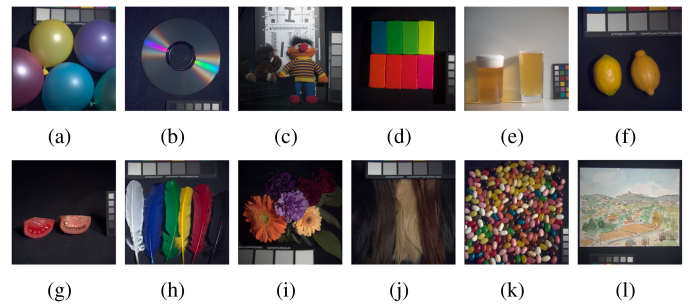In this section, the presented SDPR approach is compared with several SOTA HSI fusion methods: LRTA fusion [13],

Fig. 3. 12 test HSIs from CAVE dataset: (a) *balloons*, (b) *cd*, (c) *chart and stuffed toy*, (d) *clay*, (e) *fake and real beers*, (f) *fake and real lemons*, (g) *fake and real tomatoes*, (h) *feathers*, (i) *flowers*, (j) *hairs*, (k) *jelly beans*, (l) *watercolors*.

UAL framework [34], matrix decomposition fusion (MDF) [12], LTMR [24], DHSIS [38], and MHF-net [36].

A. Data and Experimental Settings

In this study, experiments are implemented on two simulated datasets to assess the efficacy of our presented approach: Columbia computer vision laboratory (CAVE) [53] and Harvard [54]. Moreover, we also utilize the University of Houston (UH) [55] as a real dataset to evaluate the model performance.

CAVE dataset holds 32 high-quality indoor images recorded by a generalized and assorted pixel camera. These HSIs have a spatial resolution of 512 × 512 and 31 spectral bands with wavelengths ranging from 400 to 700 nanometers. Since the first two spectra of CAVE suffer ambiguity, we remove them to avoid bias. 20 HSIs are selected at random to train the network, and the rest 12 HSIs are considered for the test. The example HSIs of CAVE are given in Fig. 3.

Harvard dataset consists of 77 HSIs covering indoor and outdoor scenarios of various targets, materials, and scales under daylight illumination. Each image has 31 bands with wavelengths ranging from 420 to 720 nanometers and spatial size of 1392 × 1040. For the test, ten images are randomly selected, whose representatives are shown in Fig. 4.
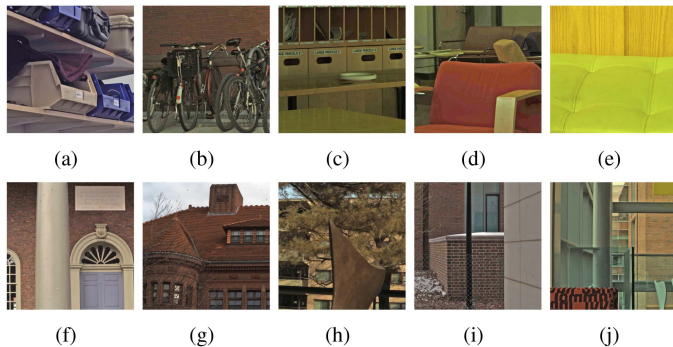
Fig. 4. 10 test HSIs from Harvard dataset: (a) *baskets*, (b) *bikes*, (c) *cabinet*, (d) *chairs*, (e) *cushion*, (f) *door*, (g) *roof*, (h) *tree*, (i) *wall*, (j) *window*.

UH dataset is provided by the IEEE Geoscience and Remote Sensing Society in 2018, including an HRMS image and an LRHS image with size $83440 \times 24040 \times 3$ and $4172 \times 1202 \times 48$, respectively. For uniformity, the HRMS image is resized to $33376 \times 9616 \times 3$. As a result, the spatial ratio of HRMS image to LRHS image is varied to 8. Moreover, we crop a 1024×1024 subimage and a corresponding 128×128 subimage from the HRMS and LRHS images, respectively, for testing.

The images from two simulated datasets are employed as the ground truths (GTs). LRHS images are obtained from each simulated dataset by employing an 8×8 Gaussian filter with a standard deviation of 2 and then down sampling with a factor of 8. The spectral down-sampling matrix $\mathbf{R}$ for the CAVE and Harvard datasets is obtained from the response of the Nikon D700 camera, which produces HRMS image as RGB image. In the nonblind case, we treat the convolutional blur $\mathbf{B}$ and the spectral response matrix $\mathbf{R}$ as known prior knowledge to guide the generation of superresolution images. Yet in the blind case, we follow most well-known works, such as [12], [24] that borrow HySure [56] to estimate $\mathbf{B}$ and $\mathbf{R}$.

B. Quality Measures

Four prevalent metrics are selected in the experiment to assess the quality of superresolution results.

1) Peak signal-to-noise ratio (PSNR): The average spatial similarity between produced and reference images across all bands can be described by the PSNR. The greater PSNR, the better spatial quality.
2) Spectral angle mapper (SAM): SAM measures the angle averaged across the whole spatial domain to determine the spectral quality of the fusion result. The lower degree, the less spectral distortion.
3) Structural similarity index measure (SSIM): SSIM computes the average structure similarity in the spatial domain between created and reference images. The higher SSIM, the better preservation of spatial structure.
4) Root-mean-squared error (RMSE): The discrepancy between produced and reference images can be described by the RMSE. Obviously, the smaller value denotes the better result.

TABLE I
QUANTITATIVE RESULTS OF ALL COMPETING APPROACHES ON CAVE DATASET

Methods	CAVE dataset [53]			
	PSNR	SAM	SSIM	RMSE
LRTA [13]	42.61	9.35	0.9720	2.55
UAL [34]	42.91	5.32	0.9915	2.70
MDF [12]	43.16	11.27	0.9728	2.45
LTMR [24]	46.88	5.22	0.9918	1.44
DHSIS [38]	47.18	<u>4.87</u>	0.9925	1.34
MHF-net [36]	<u>47.83</u>	5.36	<u>0.9933</u>	<u>1.27</u>
SDPR	**48.85**	**4.53**	**0.9944**	**1.01**

C. Parameter Selection

In our method, three key parameters require to be manually adjusted, including subspace dimension L as well as two tradeoff hyperparameters λ and μ. To discuss the effect of different atoms L on CAVE and Harvard, as shown in Fig. 5(a), we exhibit the mean PSNR curves of the fusion results against varying L. As can be observed, with L varied from 4 to 8, the PSNR values climb rapidly for both CAVE and Harvard. The results remain relatively stable as L increases further. It is well-known that Harvard is less challenging than CAVE. Accordingly, the PSNR curve for the former is substantially smoother than that for CAVE. As a result, the setting of 8 on subspace dimension is sufficient to retain the most spectral information, validating the claim that spectral vectors indeed lie in a low-dimensional subspace.

Regarding to the quality of $\widehat{\boldsymbol{\mathcal{X}}}$ ($\widehat{\mathbf{X}}_{(3)} = \mathbf{D}\widehat{\mathbf{C}}_{(3)}$), parameter λ is employed for preoptimization. Similarly, parameter μ is utilized to influence the quality of the final superresolution result $\boldsymbol{\mathcal{X}}$. Therefore, these two parameters need to be properly tuned for a pleasing performance. Unquestionably, both λ and μ shall be greater than 0. In consequence, we traverse these parameters from {1e-5, 1e-4,...,1} to report a better result. As illustrated in Fig. 5(b), when λ consistently increases, the PSNR values on two simulated datasets first remain flat and then decrease. Note that a larger PSNR for $\widehat{\boldsymbol{\mathcal{X}}}$ does not assure a better ultimate result, which is more dependent on the prior learning of the subsequent operations. That is, λ can be adjusted within a coarse range. In our implementation, $\lambda = 10^{-5}$ is chosen to ensure an acceptable overall quality. Likewise, in Fig. 5(c), the PSNR values go stably as μ rises. We attribute this to the fact that $\widetilde{\boldsymbol{\mathcal{C}}}$ is quite close to the true value $\boldsymbol{\mathcal{C}}$ in subspace, which leads to a reduced fluctuation of PSNR values. In all our experiments, we simply select the intermediate value, i.e., $\mu = 0.001$, to produce the fusion results. Note that the traversed span is notably loose for practical applications, which demonstrates the property of parameter insensitivity of our method. A finer tuned parameter would surely promote the performance, yet we leave it a relaxed setting for easy model selection.

D. Performance Comparisons

1) Performance Comparison on CAVE: Table I displays the average PSNR, SAM, SSIM, and RMSE for the 12 test HSIs, with the best values highlighted by bold fonts and the next best ones underlined for clarity. It is demonstrated in the table
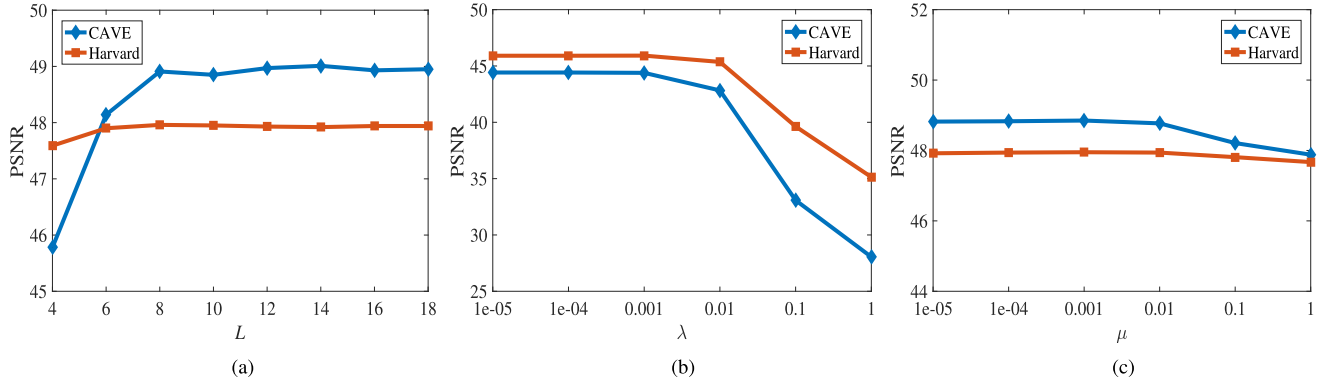
Fig. 5. Average PSNR values against different setting of parameters L, λ, and μ, respectively: (a) is for L, (b) is for λ, (c) is for μ.
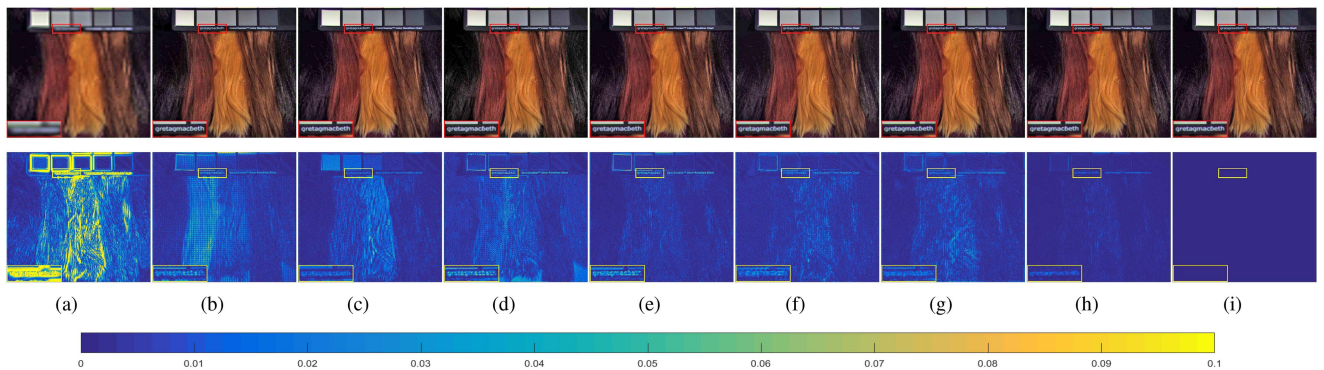

Fig. 6. Visual comparison of CAVE dataset. (a) LRHS. (b) LRTA. (c) UAL. (d) MDF. (e) LTMR. (f) DHSIS. (g) MHF-net. (h) SDPR. (i) GT. The top and bottom rows display the fused pseudocolor images (R-29, G-19, B-1) and the reconstruction errors, respectively.

that the suggested SDPR approach exceeds other competing methodologies across all evaluation measures. Presumably, our approach can better maintain both spectral and spatial structures. Furthermore, taking *hairs* (an image in CAVE) as an example for visual comparison of the competing methods, Fig. 6 displays the fused pseudocolor results and their corresponding error maps. The pseudocolor images are formed by bands 29, 19, and 1 of the recovered images, while the error maps are the differences between the recovery and the GT. For ease of observation, a representative portion of each fused HSI is marked and magnified by a factor of 2. This together with the error maps shows that SDPR recovers complex structures more efficiently and has less distortion than all other competitors. Fig. 7(a) further shows the PSNR curves of all competing approaches to compare the quality along different bands. As can be seen from this subfigure, SDPR outperforms the other competitors in almost all spectral bands.

2) Performance Comparison on Harvard: Table II shows the average performance of all competing approaches on the 10 test HSIs of Harvard. We would like to point out that the DL-based approaches in this experiment are all trained on CAVE and then tested on other datasets without any retraining or fine-tuning steps. Therefore, the results from these methods on Harvard stands as a representative of the generalization ability [35]. As can be seen from the table, SDPR remains the best in most

TABLE II
QUANTITATIVE RESULTS OF ALL COMPETING APPROACHES
ON HARVARD DATASET

Methods	Harvard dataset [54]			
	PSNR	SAM	SSIM	RMSE
LRTA [13]	45.02	3.76	0.9708	2.10
UAL [34]	37.44	4.7	0.9158	11.26
MDF [12]	44.24	4.09	0.9634	2.49
LTMR [24]	46.82	**2.92**	0.9799	1.43
DHSIS [38]	44.33	3.71	0.9719	1.82
MHF-net [36]	41.56	6.06	0.9587	3.39
SDPR	**47.95**	2.93	**0.9822**	**1.07**

evaluation metrics. Moreover, the margins between SDPR and other DL-based methods become critically larger. Taking the clipped portion of test image *bikes* (an HSI in Harvard data) as an example, Fig. 8 illustrates the fused images and the corresponding error maps from all competing approaches. A representative region is also marked and magnified. From this figure, the results achieved by LTMR and SDPR are similar and are very close to the GT. Evidently, they generate the minimum fusion error in both the edges and the smoothed areas. Fig. 7(b) further depicts the PSNR curves against the spectral bands of the whole Harvard dataset. As can be seen, albeit with a similar visual presentation, LTMR lags behind SDPR numerically in
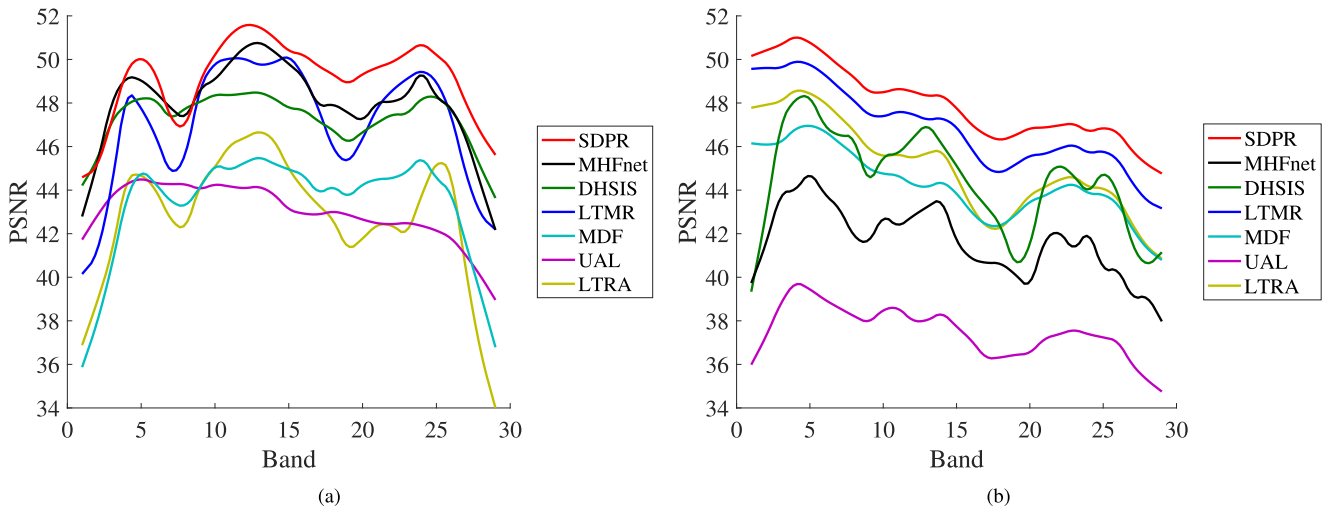
Fig. 7. PSNR of all competing methods against different spectral bands. (a) CAVE. (b) Harvard.
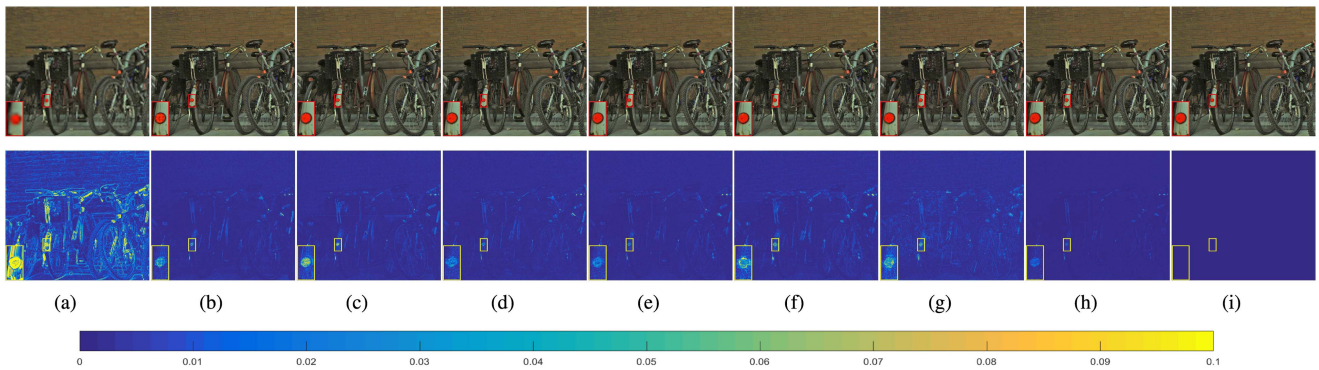

Fig. 8. Visual comparison of Harvard dataset. (a) LRHS. (b) LRTA. (c) UAL. (d) MDF. (e) LTMR. (f) DHSIS. (g) MHF-net. (h) SDPR. (i) GT. The top and bottom rows display the fused images (R-29, G-16, B-7) and the reconstruction errors, respectively.

all spectral bands. On the whole, both the quantitative metrics and qualitative visuals demonstrate that most DL-based methods (e.g., UAL and MHF-net) exhibit the weaker generalization than our model. The main reason owes to the overfitting issue since the model parameters fit in closely with the features of CAVE, yet the model perform suboptimally on Harvard dataset.

3) Performance Comparison on UH: In this case, both convolutional blurring $\mathbf{B}$ and spectral response matrix $\mathbf{R}$ are unknown. For those methods that need $\mathbf{B}$ and $\mathbf{R}$ given, we estimate them via HySure. Fig. 9 depicts a portion of the superresolution results, which are RGB images created by assessed spectral response function. The visual inspection clearly demonstrates that the output achieved by SDPR is more realistic and presents much sharper details.
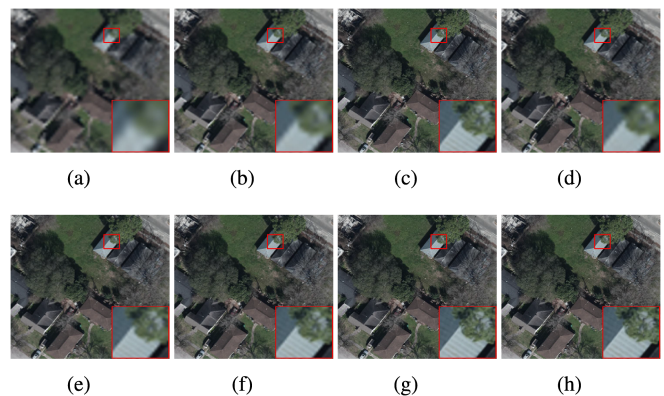
Fig. 9. Visual comparison of UH dataset. (a) LRHS. (b) LRTA. (c) UAL. (d) MDF. (e) LTMR. (f) DHSIS. (g) MHF-net. (h) SDPR.

E. Analysis and Discussion

1) Computational Efficiency: To elucidate the computational efficiency of SDPR, we examine the running times of each competitive methodology on three datasets. A GeForce GTX 3080 16 GB graphics card is used for all DL-based methods. All fusion programs are run on an Intel(R) Core(TM) i7-11800H CPU 2.30 GHz and 32 GB RAM. Table III shows the average running times of each competing approach. The results demonstrate that our SDPR method is faster than all traditional methods including

TABLE III
RUNNING TIME OF ALL COMPARED METHODS (IN SECONDS)

Methods	LRTA	UAL	MDF	LTMR	DHSIS	MHF-net	SDPR
CAVE	40.31	57.75	16.37	155.30	4.50	1.77	2.72
Harvard	253.85	315.13	102.15	767.35	25.98	6.28	14.39
PU	346.94	381.61	140.84	1048.68	32.39	7.53	14.50
Average	213.70	251.50	86.45	657.11	20.96	5.19	10.53

TABLE IV
QUANTITATIVE RESULTS ON $\mathcal{Y}_{\mathrm{up}}$, $\widehat{\mathcal{X}}$, $\widetilde{\mathcal{X}}$ AND $\mathcal{X}$

Datesets	Metrics	PSNR	SAM	SSIM	RMSE
CAVE	$\mathcal{Y}_{\mathrm{up}}$	27.21	8.57	0.8287	3.42
	$\widehat{\mathcal{X}}$	44.42	6.04	0.9857	1.33
	$\widetilde{\mathcal{X}}$	47.83	5.02	0.9936	1.08
	$\mathcal{X}$	**48.85**	**4.53**	**0.9944**	**1.01**
Harvard	$\mathcal{Y}_{\mathrm{up}}$	34.27	3.91	0.8649	2.46
	$\widehat{\mathcal{X}}$	45.91	3.42	0.9775	1.22
	$\widetilde{\mathcal{X}}$	47.63	2.98	0.9816	1.09
	$\mathcal{X}$	**47.95**	**2.93**	**0.9822**	**1.07**

TABLE V
IMPACT OF THE PROPOSED SUBSPACE LEARNING SCHEME ON THE FUSION PERFORMANCE

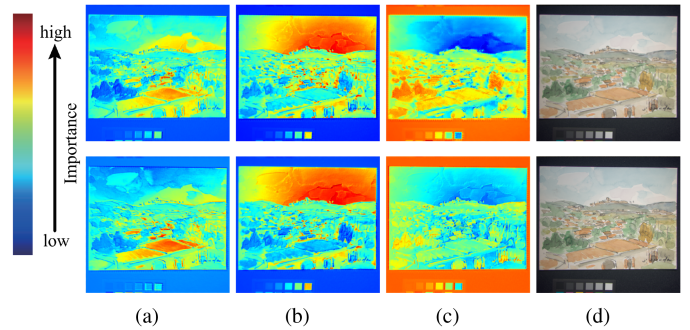| Datesets	Models	Subspace	PSNR	SAM	SSIM	RMSE
CAVE	SDPR	×	47.76	4.64	0.9922	1.09
		√	**48.85**	**4.53**	**0.9944**	**1.01**
	MHF-net	×	47.83	5.36	0.9933	1.27
		√	**48.58**	**4.99**	**0.9936**	**1.16**
Harvard	SDPR	×	41.18	5.94	0.9126	1.69
		√	**47.95**	**2.93**	**0.9822**	**1.07**
	MHF-net	×	41.56	6.06	0.9587	3.39
		√	**43.92**	**4.40**	**0.9711**	**2.00**

Fig. 10. Comparison of feature visualization on CAVE dataset. (a)–(c) Feature visualizations. (d) Fused images (R-29, G-19, B-1). The top and bottom rows are from the SDPR method without a subspace scheme (i.e., deep prior only) and with a subspace scheme (i.e., hybrid prior), respectively.

LRTA, MDF, LTMR, and most DL-based methods, such as UAL and DHSIS. In other words, SDPR is second only to MHF-net in terms of running efficiency. We attribute the advantage mainly to the enforcement of network on subspace factor, which results in a remarkably reduced dimension of target variable. Besides, the concise design of the network architecture also saves the computational cost. Inevitably, generating initial estimation with preoptimization, integrating prior into the degradation model, and estimating $\mathbf{B}$ and $\mathbf{R}$ would add some computational effort. Overall, we believe that the speed rank of our SDPR is within an acceptable range due to the improved fusion quality.

2) Effectiveness of Different Steps: As mentioned, our approach can be divided into three steps: preoptimization to generate the initial estimation, using SDPL-net to learn subspace-based deep prior, and integrating prior into the degradation model. Table IV provides the average quantitative results of $\mathcal{Y}_{\mathrm{up}}$, $\widehat{\mathcal{X}}$, $\widetilde{\mathcal{X}}$ ($\widetilde{\mathbf{X}}_{(3)} = \mathbf{D}\widetilde{\mathbf{C}}_{(3)}$), and $\mathcal{X}$ on two simulated datasets, evaluating the effectiveness of these three steps from an objective perspective. The table shows that $\widehat{\mathcal{X}}$ exceeds $\mathcal{Y}_{\mathrm{up}}$ with a huge margin on both datasets, demonstrating that the preoptimization step is qualified to retain both spatial and spectral structures when the two inputs are initially integrated as a whole. The result $\widetilde{\mathcal{X}}$ generated by SDPL-net yields improved quantification compared to $\widehat{\mathcal{X}}$. This indicates that prior learning plays an effective and helpful role in the superresolution task. Furthermore, the final superresolution result $\mathcal{X}$ based on $\widetilde{\mathcal{X}}$ achieves a further rise, showing that integrating prior into the degradation model has a positive impact on quality improvement. Overall, these three steps are all indispensable and layerwisely benefit the performance. Note that more iterates of the three steps would surely bring more benefit. However, the improvement is marginal yet with considerably more computational cost.

3) Ablation Study of Subspace Learning Scheme: We further carry out an ablation study of the subspace learning scheme, and the quality of the fusion results are shown in Table V. Taking MHF-net [36] that ranks second in most experiments as an example, we also equip it with subspace learning scheme to validate its general transferability. Specifically, with the same SVD operation, the output of MHF-net is mapped to the subspace domain, which can be further fed into (10) for the final result. From this table, the models equipped with the subspace learning scheme clearly produces improved results in comparison to the baselines. Especially, the gaps of our SDPR with and without the scheme are 6.77, 3.01, 0.07, and 0.62 in metrics PSNR, SAM, SSIM, and RMSE, respectively, on the Harvard dataset.

Furthermore, taking *watercolors* (an image in CAVE) as an example, Fig. 10 visualizes some feature maps as well as the fused images with and without the equipment of subspace scheme. As can be seen, the more focused and clearer features can be learned from the model equipped with the subspace scheme. For instance, in the first row of Fig. 10(b), more scattered hot areas can be observed around the central regions. In the first row of Fig. 10(c), the features of most trees are in a close tangle with the border areas. All these findings demonstrate the advantage of the subspace learning scheme in facilitating the capture of more semantic information and benefiting the generalization capability.

TABLE VI
COMPARISON OF THE PERFORMANCE ON STANDARD U-NET AND SDPL-NET

Datesets	Models	PSNR	SAM	SSIM	RMSE
CAVE	U-net	47.42	5.22	0.9928	1.12
	SDPL-net	**48.85**	**4.53**	**0.9944**	**1.01**
Harvard	U-net	47.14	3.06	0.9802	1.13
	SDPL-net	**47.95**	**2.93**	**0.9822**	**1.07**

4) Comparison Between U-Net and SDPL-Net: By replacing SDPL-net with a standard U-Net, we also conduct an ablation study and report the superresolution results in Table VI. Evidently, compared to U-Net, SDPL-net yields better performance on both datasets. In particular, on the CAVE dataset, the improvement of SDPR-net over U-Net is 1.43, 0.69, and 0.11 in terms of PSNR, SAM, and RMSE, respectively. These improvements demonstrate that the introduction of RCAB indeed benefits the feature learning and enriches the prior knowledge.

V. CONCLUSION

In this article, we present an explicit–implicit prior-based HSI superresolution approach called SDPR. Instead of learning a deep prior in the original image domain, we leverage the spectral correlation to transform the image into a low-rank subspace domain. In addition, rather than using a hand-crafted prior constraint, we utilize explicit–implicit joint knowledge learned by a newly proposed SDPL-net. Moreover, the output of SDPL-net is leveraged to regularize the ill-posed fusion mission. In contrast to most existing DL approaches that always confront the overfitting problem on the training data, our SDPR method can readily allow for the capture of invariant reconstruction rules under response variations. This enables it to be finely generalized to the test HSIs, even tremendous change happens in terms of band number, illumination intensity, and spectral information. Experiments conducted on a variety of synthetic and real-world datasets have confirmed the superiority of our SDPR method over both the traditional model-based and the more prevalent DL-based methodologies.

Full-resolution quality remains an open problem. In fact, there leaves considerable room for a further performance improvement. On the one hand, it makes sense to integrate more traditional explicit prior and deeper implicit knowledge to solve the ill-posed issue. On the other hand, ways to model data fidelity terms by estimating spectral response function and convolution blur matrix will be investigated. These efforts will further favor the adaptability of the presented approach to real-world circumstances.

REFERENCES

[1] Y. Wang, X. Chen, F. Wang, M. Song, and C. Yu, "Meta-learning based hyperspectral target detection using siamese network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 5527913.

[2] J. Zheng, Y. Feng, C. Bai, and J. Zhang, "Hyperspectral image classification using mixed convolutions and covariance pooling," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 522–534, Jan. 2021.

[3] Y. Feng, J. Zheng, M. Qin, C. Bai, and J. Zhang, "3D octave and 2D vanilla mixed convolutional neural network for hyperspectral image classification with limited samples," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4407.

[4] F. Xiong, J. Zhou, and Y. Qian, "Material based object tracking in hyperspectral videos," *IEEE Trans. Image Process.*, vol. 29, pp. 3719–3733, 2020.

[5] J. Zheng, J. Jiang, H. Xu, Z. Liu, and F. Gao, "Manifold-based nonlocal second-order regularization for hyperspectral image inpainting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 224–236, 2021.

[6] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[7] J. Xue, Y. Zhao, W. Liao, J. C.-W. Chan, and S. G. Kong, "Enhanced sparsity prior model for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4567–4581, Nov. 2020.

[8] Y. Feng, H. Xu, J. Jiang, H. Liu, and J. Zheng, "ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 4410213.

[9] S. Michel, P. Gamet, and M.-J. Lefevre-Fonollosa, "Hypxim–A hyperspectral satellite defined for science, security and defence users," in *Proc. 3rd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2011, pp. 1–4.

[10] T. Huang, W. Dong, J. Wu, L. Li, X. Li, and G. Shi, "Deep hyperspectral image fusion network with iterative spatio-spectral regularization," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 201–214, 2022.

[11] J. Jiang, Y. Feng, H. Xu, G. Shen, and J. Zheng, "Hyperspectral and multispectral data fusion via joint local-nonlocal modeling and truncation operator," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5880–5893, 2022.

[12] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "A truncated matrix decomposition for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 8028–8042, 2020.

[13] N. Liu, L. Li, W. Li, R. Tao, J. E. Fowler, and J. Chanussot, "Hyperspectral restoration and fusion with multispectral imagery via low-rank tensor-approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7817–7830, Sep. 2021.

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

[15] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3586–3594.

[16] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.

[17] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[18] W. Dong et al., "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[19] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4469–4480, Oct. 2020.

[20] X. Li, Y. Yuan, and Q. Wang, "Hyperspectral and multispectral image fusion via nonlocal low-rank tensor approximation and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 550–562, Jan. 2021.

[21] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.

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

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

[24] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

[25] Y. Xu, Z. Wu, J. Chanussot, P. Comon, and Z. Wei, "Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 348–362, Jan. 2020.

[26] H. Xu, M. Qin, S. Chen, Y. Zheng, and J. Zheng, "Hyperspectral-multispectral image fusion via tensor ring and subspace decompositions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8823–8837, 2021.

[27] H. Xu, J. Jiang, Y. Feng, Y. Jin, and J. Zheng, "Tensor completion via hybrid shallow-and-deep priors," *Appl. Intell.*, to be published, doi: 10.1007/s10489-022-04331-4.

[28] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017.

[29] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1753–1761.

[30] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-J. Zha, and F. Wu, "Deep residual attention network for spectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 1–14.

[31] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[32] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Joint image filtering with deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1909–1923, Aug. 2019.

[33] X. Deng and P. L. Dragotti, "Deep convolutional neural network for multimodal image restoration and fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3333–3348, Oct. 2021.

[34] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3070–3079.

[35] J. Hu, T. Huang, L. Deng, T. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7251–7265, Dec. 2022.

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

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

[38] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.

[39] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020.

[40] H. Xu, J. Zheng, X. Yao, Y. Feng, and S. Chen, "Fast tensor nuclear norm for structured low-rank visual inpainting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 2, pp. 538–552, Feb. 2022.

[41] W. Wang, W. Zeng, Y. Huang, X. Ding, and J. Paisley, "Deep blind hyperspectral image fusion," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4149–4158.

[42] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.

[43] V. Hryhorenko, D. Klyushin, and S. Lyashko, "Multiblock ADMM in machine learning," in *Proc. IEEE Int. Conf. Adv. Trends Inf. Theory*, 2019, pp. 461–464.

[44] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022, doi: 10.1109/TPAMI.2021.3088914.

[45] M. B. Alver, A. Saleem, and M. Çetin, "Plug-and-play synthetic aperture radar image formation using deep priors," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 43–57, 2021.

[46] M. Zhao, X. Wang, J. Chen, and W. Chen, "A plug-and-play priors framework for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 5501213.

[47] N. Li, "Iterative algorithm for the reflexive solution of a Sylvester quaternion k-conjugate matrix equation," in *Proc. IEEE 3rd Int. Conf. Safe Prod. Informatization*, 2020, pp. 526–529.

[48] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

[49] Z. Iqbal, D. Nguyen, G. Hangel, S. Motyka, W. Bogner, and S. Jiang, "Super-resolution 1 h magnetic resonance spectroscopic imaging utilizing deep learning," *Front. Oncol.*, vol. 9, 2019, Art. no. 1010.

[50] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[51] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 105–114.

[52] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[53] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

[54] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 193–200.

[55] Y. Xu et al., "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1709–1724, Jun. 2019.

[56] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

Jianwei Zheng (Member, IEEE) received the B.S. degree in electronic and computer engineering and the Ph.D. degree in control theory and control engineering from the Zhejiang University of Technology, Hangzhou, China, in 2005 and 2010, respectively.

He is currently an Associate Professor with the College of Computer Science and Technology, Zhejiang University of Technology. He has authored or coauthored more than 60 journal and conference papers in these areas. His research interests include machine learning and compressive sensing.

Pengfei Li received the B.E. degree in security engineering in 2019 from the Zhejiang University of Technology, Hangzhou, China, where he is currently working toward the B.S. degree with the School of Computer Science and Engineering.

His research interests include image processing and optimization algorithm.

Honghui Xu received the B.E. degree in mechanical engineering in 2018 from the Zhejiang University of Technology, Hangzhou, China, where he is currently working toward the Ph.D. degree in computer science with the School of Computer Science and Engineering.

His research interests include image processing and optimization algorithm.

Jiawei Jiang received the B.E. degree in electricity engineering from China Jiliang University, Hangzhou, China, in 2018. He is currently working toward the Ph.D. degree with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China.

His research interests include multidimensional visual data process and the its application.

Yuchao Feng received the B.E. degree in electronics information from Wenzhou University, Wenzhou, China, in 2019. He is currently working toward the Ph.D. degree with the School of Computer Science and Engineering, Zhejiang University of Technology, Hangzhou, China.

His research interests include remote sensing image process and pattern recognition.

Zhi Liu received the B.S. degree in automatic control and the M.S. degree in system engineering from Xi'an Jiaotong University, Xi'an, China, in 1991 and 1994, respectively. She received the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2001.

She is currently a Professor with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China. Her research interests include 3-D model retrieval, image processing, and intelligent transportation system.

Dr. Liu is a Member of China Computer Federation.