



**HAL**  
open science

# Hyperspectral Image Super-Resolution via Deep Prior Regularization with Parameter Estimation

Xiuheng Wang, Jie Chen, Qi Wei, Cédric Richard

► **To cite this version:**

Xiuheng Wang, Jie Chen, Qi Wei, Cédric Richard. Hyperspectral Image Super-Resolution via Deep Prior Regularization with Parameter Estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, pp.1-18. 10.1109/TCSVT.2021.3078559 . hal-03352836v1

**HAL Id: hal-03352836**

**<https://hal.science/hal-03352836v1>**

Submitted on 23 Sep 2021 (v1), last revised 7 Apr 2022 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Hyperspectral Image Super-Resolution via Deep Prior Regularization with Parameter Estimation

Xiuheng Wang, *Student Member, IEEE*, Jie Chen, *Senior Member, IEEE*, Qi Wei, *Member, IEEE*, and Cédric Richard, *Senior Member, IEEE*,

## Abstract

Hyperspectral image (HSI) super-resolution is commonly used to overcome the hardware limitations of existing hyperspectral imaging systems on spatial resolution. It fuses a low-resolution (LR) HSI and a high-resolution (HR) conventional image of the same scene to obtain an HR HSI. In this work, we propose a method that integrates a physical model and deep prior information. Specifically, a novel, yet effective two-stream fusion network is designed to serve as a regularizer for the fusion problem. This fusion problem is formulated as an optimization problem whose solution can be obtained by solving a Sylvester equation. Furthermore, the regularization parameter is simultaneously estimated to automatically adjust contribution of the physical model and the learned prior to reconstruct the final HR HSI. Experimental results on both simulated and real data demonstrate the superiority of the proposed method over other state-of-the-art methods on both quantitative and qualitative comparisons.

## Index Terms

HSI super-resolution, deep learning, fusion, Sylvester equation, regularization parameter estimation.

## I. INTRODUCTION

**H**YPERSPECTRAL imaging simultaneously captures images of the same scene at different wavelengths. Rich spectral characteristics provided by hyperspectral images are important in remote sensing [1]. HSIs have been demonstrated to improve the performance of tasks including tracking [2], classification [3], [4], segmentation [5] and clustering [6]. HSIs contain richer spectral information of real scenes compared with conventional images such as color or gray-scale images. However, the high spectral resolution of hyperspectral images needs to make a compromise with spatial resolution to ensure an acceptable signal-to-noise ratio (SNR) [7]. This in turn restricts applications of HSIs. To cope with this issue, a critical task is thus to improve the spatial resolution of HSIs. Besides, conventional approaches such as push-broom spectral imaging sensors are often time-consuming and limit applications of HSIs. Recently, researchers designed dual-camera compressive hyperspectral imagers that can capture full HSIs with a single signal exposure, and acquire hyperspectral videos with successive exposures. These techniques reconstruct a full hyperspectral cube relying on the compressive sensing theory and computational imaging algorithms; detailed discussions can be found in [8]–[10].

Due to various physical limitations at hardware level, it is often challenging to develop hyperspectral cameras that simultaneously achieve high spatial and high spectral resolutions. On the other hand, conventional cameras capture RGB (red, green and blue) or panchromatic images with higher spatial resolution but lower spectral resolution by integrating the scene response over broad spectral bands [7]. Consequently, researchers have proposed computational methods to restore a high-resolution (HR) HSI by integrating a low-resolution (LR) HSI and an HR conventional image [11], [12]. This procedure is often referred to as *HSI super-resolution* or *HSI fusion*. A class of fusion approaches is based on component substitution such as the intensity, hue and saturation (IHS) technique [13], principal component analysis (PCA) [14] and wavelet transform [15], [16]. These methods are fast but tend to introduce spectral distortion when estimating the scene response of the latent HR HSIs from panchromatic images.

It has been popular to formulate the fusion problem as an image restoration problem [7], [17]–[27]. In such approaches, following a physical degradation model, the input LR HSI and HR conventional images are seen as the spatially and spectrally degraded observations (i.e. linear down-sampled versions) of the latent HR HSI respectively. Note that HSI super-resolution is a highly ill-posed problem owing to the large scaling factors in both spatial and spectral domain. Thus, it is important to incorporate prior information to constrain the solution space. Depending on the structures of prior information used in the optimization problem, existing techniques can be roughly divided into three categories: spectral unmixing based approaches [7], [18], [19], sparse representation based approaches [20]–[23] and tensor factorization based approaches [24]–[27]. A brief review on those methods is given in Section II. However, these predefined priors-based methods with handcrafted regularizers have some inherent disadvantages e.g., a complex regularizer may introduce extra difficulties in solving the related optimization problem.

X. Wang and J. Chen are with with Centre of Intelligent Acoustics and Immersive Communications at School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, and also with China Key Laboratory of Ocean Acoustics and Sensing, Ministry of Industry and Information Technology, Xi'an 710072, China (e-mail: xiuheng.wang@mail.nwpu.edu.cn; dr.jie.chen@ieec.org).

Q. Wei is with University of Toulouse, IRIT/INP-ENSEEIH, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. (e-mail: bjweiqi@gmail.com).

C. Richard is with Université Côte d'Azur, OCA, CNRS, 75016 Paris, France (cedric.richard@unice.fr).

Recently, inspired by the successful application of deep learning to many computer vision tasks, especially single image super-resolution (SISR) [28], convolutional neural networks (CNNs) have been introduced to address the HSI super-resolution problem [29]–[34]. Compared with optimization methods based on predefined priors, these deep learning methods require fewer assumptions on the prior knowledge of the latent HR HSI, and can directly learn the relevant information from training data in an end-to-end way. However, these learning based methods ignore the blurring and down-sampling operators as well as the spectral response function in the degradation model, though this model has a clear physical interpretation that relates the LR HSI and the HR conventional image to the HR HSI. To tackle this issue, recent deep learning methods [35], [36] successfully improved the quality of generated HR HSIs by learning the blurring matrix and the spectral response function. The work in [35] formulated a fusion model considering both spatial and spectral degradation. A deep network called MHF-net was accordingly proposed to solve this problem in an iterative manner. In [36], a two-stage network based on unsupervised adaption learning (UAL) was proposed to learn the deep priors and estimate the unknown spatial degradation.

In order to leverage the advantages of both optimization and deep learning methods, several deep priors-based approaches (i.e., learning the priors via deep CNNs) have been recently proposed [37], [38]. In these methods, deep neural networks were designed to exploit both spatial and spectral characteristics of the latent HR HSI, and their outputs then served as regularizers. This allowed these approaches to achieve enhanced performance. However, they learned deep priors from the images produced by solving a Sylvester equation [39] rather than two observed images. Thus, the efficacy of deep priors relied on the accuracy of the produced images in their HSI super-resolution schemes.

Although numerous works have explored the prior structures of the latent HR HSI, few of them have investigated the way of balancing the contribution of the physical model and prior information in their frameworks. That is however a key point for the effective implementation of such methods. Some classic regularization parameter estimation methods are reviewed in Section II.

Here, we aim at performing hyperspectral super-resolution by combining the physical model-based optimization and deep learning for constructing data-driven priors. Before proceeding with the algorithm, we raise the following three critical points to guide the problem formulation and solving:

- How to design a proper deep neural network capable of learning prior information from multi-source data?
- How to formulate the fusion problem in a way that can leverage both the physical model and end-to-end learned prior?
- How to balance the contribution of the physical model and prior learned from data?

In this work, a new deep priors-based method is proposed to practically address the three issues pointed above. We summarize our contributions as follows:

- To avoid handcrafting regularizers as in existing predefined priors-based methods, we train a new effective two-stream fusion network (TSFN) to directly learn deep priors through the fusion of two input observed images rather than produced images.
- In order to integrate deep priors into the degradation model, the network output, which corresponds to the extracted deep priors, is plugged into the objective function of the fusion problem as a regularization term. The squared Euclidean distance is finally selected as the regularizer after many attempts with various regularizers.
- To adjust the contribution of the deep priors in the hyperspectral super-resolution problem, we adopt the minimum distance criterion (MDC) [40] to the response curve of the bi-objective optimization problem for automatic selection of the regularization parameter with a golden-section search strategy. To our best knowledge, this is the first work designed to balance deep learning approach and optimization approach in this mathematical way.
- Our experimental results demonstrate the effectiveness of the proposed HSI super-resolution strategy, offering an improvement over the results of the state-of-the-art methods.

The rest of this paper is organized as follows. Related literatures are reviewed in Section II. We formulate the HSI super-resolution problem in Section III. In Section IV, the proposed method for HSI super-resolution is explained in detail. Section V presents and discusses experimental results on four public datasets and ablation studies. Finally, conclusions and future works are given in Section VI.

## II. RELATED WORK

In this section, we will briefly review three categories of existing methods related to this work.

### A. Spectral Unmixing Based Approaches

Leveraging priors for spectral unmixing with some constraints (e.g., non-negativity, sum-to-one) has been experimentally demonstrated to be beneficial for HSI super-resolution. In [18], the coupled non-negative matrix factorization (CNMF) method was proposed to alternatively unmix an LR HSI and an HR conventional image to estimate an HR HSI. In [7], with a similar framework, by jointly unmixing two input images, the initial optimization problem was first decoupled into two constrained least-square problems and then solved. The method proposed in [19] reconstructed the latent HR HSI with respect to its endmembers and their abundances by using the alternating direction method of multipliers (ADMM) technique [41].

### B. Sparse Representation Based Approaches

Sparse representation is another promising technique for fusing hyperspectral and conventional images, aiming to sparsely encode the latent HR HSI with an appropriate spectral dictionary learned from the input HSI. To this end, in [20] the spectral dictionary was learned with K-SVD and sparse matrix factorization was used to fuse two input images. Considering the similarity of neighboring pixels in the latent HSI, the method proposed in [21] enforced group sparsity as well as non-negativity among pixels within small patches. The work [22] fused the hyperspectral and conventional images in a variational approach and designed a sparse regularizer by decomposing the scene on a spectral dictionary. The method of [23] developed a non-negative sparse coding algorithm that exploited not only the sparsity of each pixel but also the non-local spatial similarity of the latent HR HSI, leading to enhanced fusion performance.

### C. Tensor Factorization Based Approaches

Another important approach in HSI super-resolution is based on tensor factorization, which has been commonly used to simultaneously capture correlations among different modes. The method in [24] employed non-local sparse tensor factorization to decompose each cube of the HSI as a sparse core tensor and dictionaries of three modes. To exploit the non-local spatial self-similarities, similar cubes were grouped together and assumed to share the same dictionaries. The work [25] formulated the fusion problem as the coupled sparse Tucker decomposition, which alternatively updated the dictionaries of three modes and sparse core tensor. In [26], a low tensor-train rank (LTTR) based method was proposed to learn correlations among the spatial, spectral and non-local modes of the non-local similar HR HSI cubes to regularize the super-resolution problem. More recently, a graph Laplacian-guided coupled tensor decomposition model was proposed in [27]. It adopted the coupled Tucker decomposition to exploit jointly and seamlessly the intrinsic spatial-spectral information of hyperspectral and conventional images.

### D. Regularization Parameter Estimation approaches

A classical method for estimating regularization parameters is the generalized cross-validation (GCV) [42], which has been applied for image restoration in [43]. In [44], the regularization parameter was chosen by finding a point near the ‘‘corner’’ of the L-curve, which was formed by plotting the data fidelity term and the regularizer in log-log scale. However, the use of the log-log scale led to a convexity loss of the L-curve [45]. Recently, basis pursuit has been formulated as a constrained least-square problem, and the corresponding *Pareto front* [46] has been proved convex and continuously differentiable over all points of interest in [47]. Inspired by this method, the work [40] proposed the maximum curvature criterion (MCC) and the MDC on the response surface (i.e., the linear plot of the data fidelity and regularization costs) and confirmed their good performance in estimating the regularization parameter for non-negative HSI deconvolution.

## III. PROBLEM FORMULATION

Consider to reconstruct a high-resolution HSI  $X_{3D} \in \mathbb{R}^{B \times L \times W}$  based on a low-resolution HSI  $Y_{3D} \in \mathbb{R}^{B \times l \times w}$  and a high-resolution conventional image  $Z_{3D} \in \mathbb{R}^{b \times L \times W}$  over the same scenario, where  $B$  and  $b$  are numbers of spectral bands of the hyperspectral and conventional images ( $B > b$ );  $(L, W)$  and  $(l, w)$  are the height and width of the HR and LR images respectively ( $L > l, W > w$ ). For ease of mathematical formulation,  $X_{3D}, Y_{3D}$  and  $Z_{3D}$  are transformed in matrix forms  $\mathbf{X} \in \mathbb{R}^{B \times N}$ ,  $\mathbf{Y} \in \mathbb{R}^{B \times n}$  and  $\mathbf{Z} \in \mathbb{R}^{b \times N}$ , respectively, where  $N = L \times W, n = l \times w$  are numbers of pixels in each band of the HR and LR images. According to the linear degradation model,  $\mathbf{Y}$  can be viewed as a spatially down-sampled vision of  $\mathbf{X}$ , while  $\mathbf{Z}$  is a down-sampled observation of  $\mathbf{X}$  in spectral domain, so that:

$$\mathbf{Y} = \mathbf{XBS}, \quad \mathbf{Z} = \mathbf{RX}, \quad (1)$$

where  $\mathbf{B} \in \mathbb{R}^{N \times N}$  represents the blurring matrix,  $\mathbf{S} \in \mathbb{R}^{N \times n}$  is a uniform down-sampling operator with scaling factor  $s = N/n$  and  $\mathbf{R} \in \mathbb{R}^{b \times B}$  denotes the spectral response function (SRF) of the conventional camera sensor, which is often known or can be estimated *a priori*. Moreover, the blurring matrix  $\mathbf{B}$  is assumed to be a known block circulant matrix with circulant blocks [22]. In this widely admitted assumption,  $\mathbf{B}$  can be decomposed as  $\mathbf{B} = \mathbf{FDF}^H$ . Here,  $\mathbf{F} \in \mathbb{R}^{N \times N}$  is the discrete Fourier transform (DFT) matrix ( $\mathbf{FF}^H = \mathbf{I}_N$ , and  $\mathbf{I}_N$  is an  $N \times N$  identity matrix.) and  $\mathbf{D} \in \mathbb{R}^{N \times N}$  is a diagonal matrix containing the eigenvalues of  $\mathbf{B}$ .

Based on the degradation model in (1), we can estimate  $\mathbf{X}$  by introducing proper priors and seeking the minimum of the following bi-objective function:

$$\begin{aligned} \hat{\mathbf{X}} &= \arg \min_{\mathbf{X}} \mathcal{J}(\mathbf{X}) \\ &= \arg \min_{\mathbf{X}} \mathcal{J}_1(\mathbf{X}) + \mu \mathcal{J}_2(\mathbf{X}) \end{aligned} \quad (2)$$

where

$$\mathcal{J}_1(\mathbf{X}) = \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 \quad (3)$$

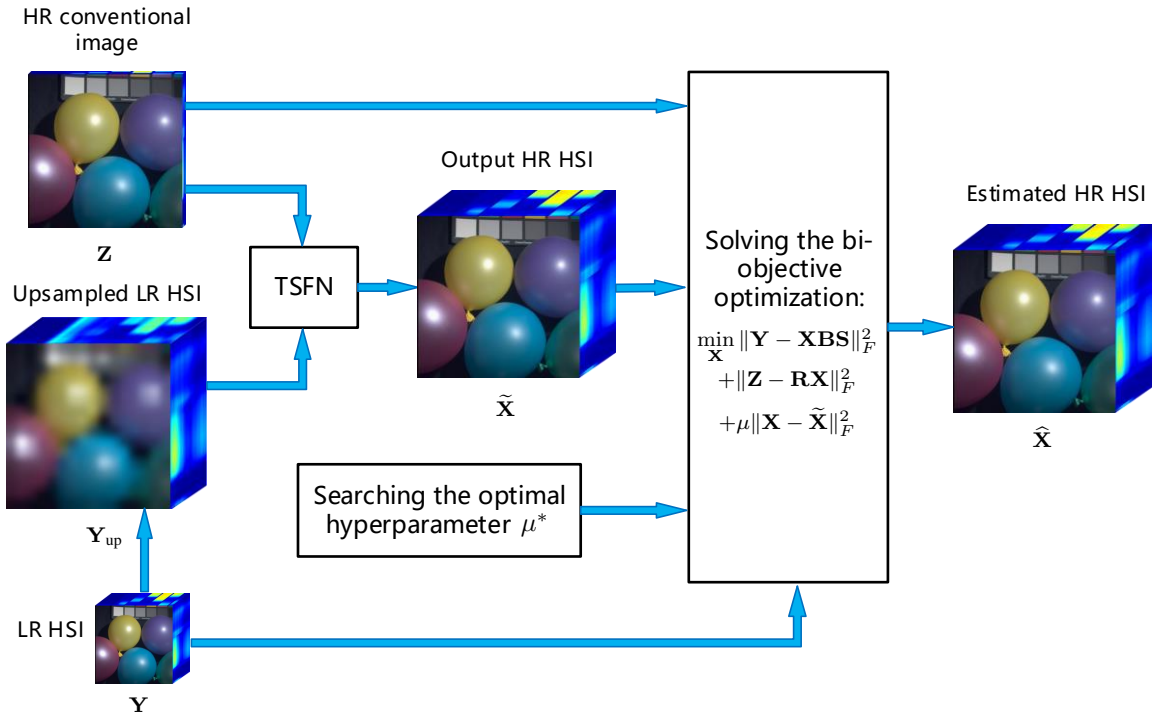


Fig. 1. Overall scheme of our proposed HSI super-resolution method. HR and LR represent high-resolution and low-resolution in spatial domains, respectively.

with  $\|\cdot\|_F$  denoting the matrix Frobenius norm.  $\mathcal{J}_1(\mathbf{X})$  is the data fidelity terms and  $\mathcal{J}_2(\mathbf{X})$  is a regularizer that enforces the desired property of the solution while  $\mu \geq 0$  is the hyperparameter to balance the data fidelity term  $\mathcal{J}_1$  and the regularization term  $\mathcal{J}_2$ . As shown in (2), prior information on the latent HR HSI is encoded in  $\mathcal{J}_2(\mathbf{X})$ . Handcrafting a powerful regularizer  $\mathcal{J}_2(\mathbf{X})$  is non-trivial. Recently, benefitting from the variable splitting techniques, plug-and-play methods have been proposed to solve various hyperspectral image inverse problems [48]–[50]. On the contrary, in this work, we construct a regularizer leveraging the network output  $\tilde{\mathbf{X}} \in \mathbb{R}^{B \times N}$  to enforce the solution of (3) to be close to the learned deep priors. More specifically, let  $\mathcal{J}_2(\mathbf{X})$  be the squared Euclidean distance between  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$ :

$$\mathcal{J}_2(\mathbf{X}) = \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 \quad (4)$$

Then (2) is given by

$$\arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 + \mu \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 \quad (5)$$

Compared to other possible formulations, (5) allows a straight-forward solver due to the differentiability of the Frobenius norm. Using a powerful priors structure for  $\tilde{\mathbf{X}}$ , appropriately integrating the priors into the degradation model as well as balancing contribution of the priors are three key points to obtain a good estimation of the latent HR HSI.

#### IV. PROPOSED METHOD

In this section, we present the proposed deep priors-based HSI super-resolution approach in detail. Firstly, we introduce the architecture design of the two-stream fusion network. Then, we elaborate on an efficient approach to the bi-objective optimization problem in (5) based on a Sylvester analytical solver. Finally, we balance the contribution of the learned deep priors by seeking the optimal value of hyperparameter  $\mu^*$  under the MDC. The overall scheme of our method is illustrated in Fig. 1.

##### A. Deep Priors Learning and Network Design

Instead of using handcrafted regularizers, we propose to learn the priors of the latent HR HSI from hyperspectral datasets by designing a deep CNN with the aforementioned two-stream architecture, i.e., TSFN.

In contrast to the single-stream architecture used in [37], [38], the two-stream deep neural networks extract and combine features of bimodal data at the feature level, which has been proved to be a powerful tool in applications including videos action recognition [51], [52], hyperspectral classification [53] and hyperspectral superresolution [32]–[34]. Specifically, we employ residual blocks [54] and a skip-connection [55] in our network. These techniques have been shown useful in boosting the performance in SISR approaches [56], [57].

As illustrated in Fig. 2, the proposed TSFN architecture takes an up-sampled image  $\mathbf{Y}_{\text{up}} \in \mathbb{R}^{B \times N}$  (produced from an LR HSI  $\mathbf{Y}$  via the bicubic interpolation) and an HR conventional image  $\mathbf{Z}$  as two inputs. The feature maps of  $\mathbf{Z}$  and  $\mathbf{Y}$  extracted by two corresponding streams, containing  $P$  and  $Q$  residual blocks respectively, are concatenated. Then, the concatenated features are fed into the next convolutional layer. The core of our network is the residual block, containing two convolutional (Conv) layers followed by batch-normalization (BN) layers [58] and ParametricReLU (PReLU) [59] as the activation function (see Fig. 3). The batch-normalization layer is used to speed up the training process as well as boost the image restoration performance. The skip connection passes input feature maps to the output of the residual block via element-wise sum. A skip connection operator is also employed to add the shallowest feature maps extracted from  $\mathbf{Y}_{\text{up}}$  to the deepest feature maps of the network to reconstruct the final  $\tilde{\mathbf{X}}$ . The last convolutional layer contains  $B$  filters and the others are composed of 64 filters, and the kernel size of each filter is  $3 \times 3$  with a stride of 1. The matrix  $\ell_1$ -norm based loss function is used:

$$\ell(\Theta) = \sum_{m=1}^M \|\mathcal{F}(\mathbf{Y}_{\text{up},m}, \mathbf{Z}_m; \Theta) - \mathbf{X}_m\|_1 \quad (6)$$

where  $\{(\mathbf{Y}_{\text{up},m}, \mathbf{Z}_m; \mathbf{X}_m)\}_{m=1}^M$  represent  $M$  training image (patch) pairs used to train the network function  $\mathcal{F}$  parameterized by  $\Theta$ . It has been demonstrated that  $\ell_1$ -norm is more appropriate than  $\ell_2$ -norm in low-level image processing tasks as it experimentally provides better performance [60]. In Subsection V-G, we conduct related ablation experiments on the loss function in Eq. (6) and give both quantitative and qualitative comparisons to further demonstrate this point. We also show additional ablation studies on the effect of identical residual block numbers  $P, Q$  and of the skip connection operator to validate the design choice in our network architecture.

It is worth noting that the proposed TSFN can be applied to effectively learn priors of the latent HR HSI in degraded scenarios with various scaling factor settings (e.g., 8, 16 and 32). Our network architecture bears some resemblance to the one in [34]. The main difference is that [34] used the attention residual block while our model extracts features through the residual block without the attention mechanism. Further, we avoid adopting the deconvolution layer and train our model with a mixture of image pairs with different scaling factors. In this way, our trained model can flexibly handle different scenarios.

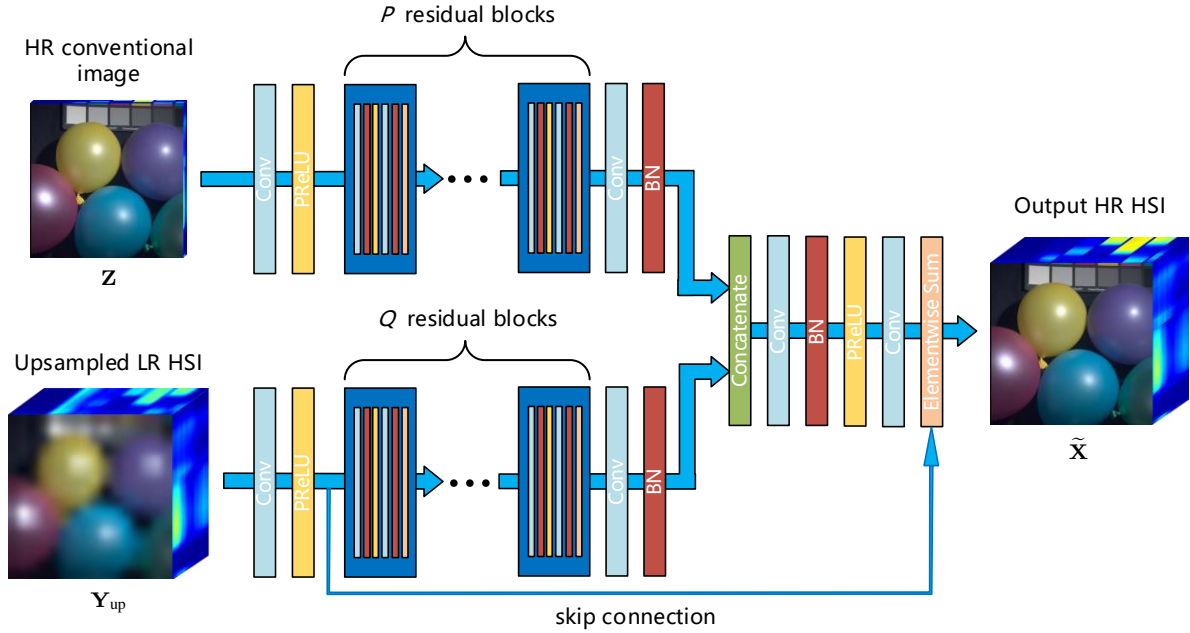


Fig. 2. Two-stream architecture for hyperspectral image super-resolution.

### B. Integrating Deep Priors into the Degradation Model

The fusion result inferred from the proposed TSFN, referred to  $\tilde{\mathbf{X}}$ , is used in the regularizer  $\mathcal{J}_2$ . More specifically,  $\tilde{\mathbf{X}}$ , a representation of the prior spatial-spectral information of the latent  $\mathbf{X}$ , is used to regularize the final estimation  $\hat{\mathbf{X}}$ , as illustrated in the bi-objective optimization problem (5). To solve (5), we force the derivative of the loss function w.r.t.  $\mathbf{X}$  to be zero. Thus, the optimum of (5) is defined by the solution of following Sylvester equation:

$$\mathbf{C}_1 \hat{\mathbf{X}} + \hat{\mathbf{X}} \mathbf{C}_2 = \mathbf{C}_3 \quad (7)$$

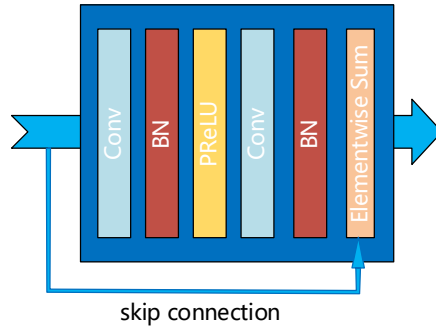


Fig. 3. Detail components of the residual block.

---

**Algorithm 1** Solution by solving the Sylvester Equation w.r.t.  $\tilde{\mathbf{X}}$ .

---

**Input:**  $\mathbf{Y}, \mathbf{Z}, \mathbf{B}, \mathbf{S}, \mathbf{R}, \tilde{\mathbf{X}}, \mu$ .

**Output:**  $\hat{\mathbf{X}}$ .

Initialize  $\mathbf{C}_1 = \mathbf{R}^T \mathbf{R} + \mu \mathbf{I}_B$ ,

$\mathbf{C}_2 = (\mathbf{BS})(\mathbf{BS})^T$ ,

$\mathbf{C}_3 = \mathbf{R}^T \mathbf{Z} + \mathbf{Y}(\mathbf{BS})^T + \mu \tilde{\mathbf{X}}$ .

(a) Eigen-decomposition of  $\mathbf{B}$ :

$$\mathbf{B} = \mathbf{FDF}^H$$

(b)  $\bar{\mathbf{D}} = \mathbf{D}(\mathbf{1}_s \otimes \mathbf{I}_n)$

(c) Eigen-decomposition of  $\mathbf{C}_1$ :

$$\mathbf{C}_1 = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

(d)  $\bar{\mathbf{C}}_3 = \mathbf{Q}^{-1} \mathbf{C}_3 \mathbf{F}$

(e) Compute auxiliary matrix  $\bar{\mathbf{X}}$  band by band

**for**  $k = 1$  to  $B$  **do**

$$\bar{\mathbf{X}}_k = \lambda_k^{-1} (\bar{\mathbf{C}}_3)_k - \lambda_k^{-1} (\bar{\mathbf{C}}_3)_k \bar{\mathbf{D}} (\lambda_k s \mathbf{I}_n + \sum_{t=1}^s \mathbf{D}_t^2) \bar{\mathbf{D}}^H$$

**end for**

(f)  $\hat{\mathbf{X}} = \mathbf{Q} \bar{\mathbf{X}} \mathbf{F}^H$

---

where

$$\mathbf{C}_1 = \mathbf{R}^T \mathbf{R} + \mu \mathbf{I}_B$$

$$\mathbf{C}_2 = (\mathbf{BS})(\mathbf{BS})^T$$

$$\mathbf{C}_3 = \mathbf{R}^T \mathbf{Z} + \mathbf{Y}(\mathbf{BS})^T + \mu \tilde{\mathbf{X}}$$

(8)

where  $\mathbf{I}_B$  represents the identity matrix of size  $B \times B$ . According to the well-known conclusion in [39], the Sylvester equation in (7) has a unique solution if and only if an arbitrary sum of the eigenvalues of  $\mathbf{C}_1$  and  $\mathbf{C}_2$  is not equal to zero. Note that  $\mathbf{R}^T \mathbf{R}$  and  $\mathbf{I}_B$  are both positive-definite matrices, and thus,  $\mathbf{C}_1$  is positive-definite. Further considering that  $\mathbf{C}_2$  is positive semi-definite, an arbitrary sum of the eigenvalues of  $\mathbf{C}_1$  and  $\mathbf{C}_2$  is greater than zero, which ensures the uniqueness of the solution in (7). The fast algorithm for solving (7) can be achieved by referring to our previous work [61], with detailed steps summarized in Algorithm 1.

### C. Adjusting the Contribution from Deep Priors

Different from previous deep priors-based approaches [37], [38], achieving a good balance between the data fidelity and regularization terms is also considered in our work. Intuitively, if  $\mu$  tends toward infinity, the regularization term  $\|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2$  will be minimized, i.e., the solution will be  $\tilde{\mathbf{X}}$ . On the contrary, the solution tends to minimize the data fidelity terms  $\|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2$  when  $\mu$  tends to zero. In this subsection, we propose to estimate the regularization parameter with the MDC that is initially used for hyperspectral image deconvolution [40]. We estimate the response curve of (5) that describes the characteristic of the solution set obtained by varying hyperparameter value  $\mu$ . Then, MDC is adopted to the estimated response curve to seek the optimal value of  $\mu$  in the hyperspectral super-resolution problem by using the golden-section search method.

Problem (5) can also be considered as a bi-objective optimization in form of

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} (\mathcal{J}_1(\mathbf{X}), \mathcal{J}_2(\mathbf{X})) \quad (9)$$



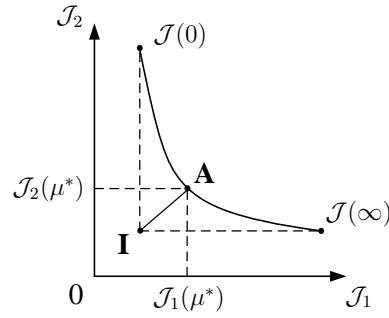


Fig. 4. Representation of the response curve. The optimal point is denoted as  $\mathbf{A} (\mathcal{J}_1(\mu^*), \mathcal{J}_2(\mu^*))$ .

1) *Pareto Front*: The definition of *Pareto front* depends on the notion of *domination* defined in [46]. This notion is important in the bi-objective optimization since it gives a criterion of judging a better solution. Let  $\widehat{\mathbf{X}}^{(1)}$  and  $\widehat{\mathbf{X}}^{(2)}$  be two different solutions of the bi-objective optimization in (9). When we say  $\widehat{\mathbf{X}}^{(1)}$  dominates  $\widehat{\mathbf{X}}^{(2)}$ , it means that  $\mathcal{J}_i(\mathbf{X}^{(1)})$  is not larger than  $\mathcal{J}_i(\mathbf{X}^{(2)})$  for all  $i \in \{1, 2\}$  and  $\mathcal{J}_j(\mathbf{X}^{(1)})$  is smaller than  $\mathcal{J}_j(\mathbf{X}^{(2)})$  for at least one  $j \in \{1, 2\}$ :

$$\widehat{\mathbf{X}}^{(1)} \preceq \widehat{\mathbf{X}}^{(2)} \text{ iff } \begin{cases} \mathcal{J}_i(\mathbf{X}^{(1)}) \leq \mathcal{J}_i(\mathbf{X}^{(2)}), & \forall i \in \{1, 2\} \\ \exists i \in \{1, 2\} & \text{s.t. } \mathcal{J}_i(\mathbf{X}^{(1)}) < \mathcal{J}_i(\mathbf{X}^{(2)}) \end{cases} \quad (10)$$

Apart from that, the solution  $\widehat{\mathbf{X}}^{(1)}$  does not dominate  $\widehat{\mathbf{X}}^{(2)}$ . The solution is said to be non-dominated or *Pareto optimal* for a bi-objective problem if all other solutions in the feasible region have higher values for at least one objective. The set containing all the non-dominated solutions is defined as the *Pareto front*. It means that any point in the *Pareto front* cannot be said to dominate any other. The shape of *Pareto front* denotes the set of all the achievable trade-offs between two objectives  $\mathcal{J}_1(\mathbf{X})$  and  $\mathcal{J}_2(\mathbf{X})$ .

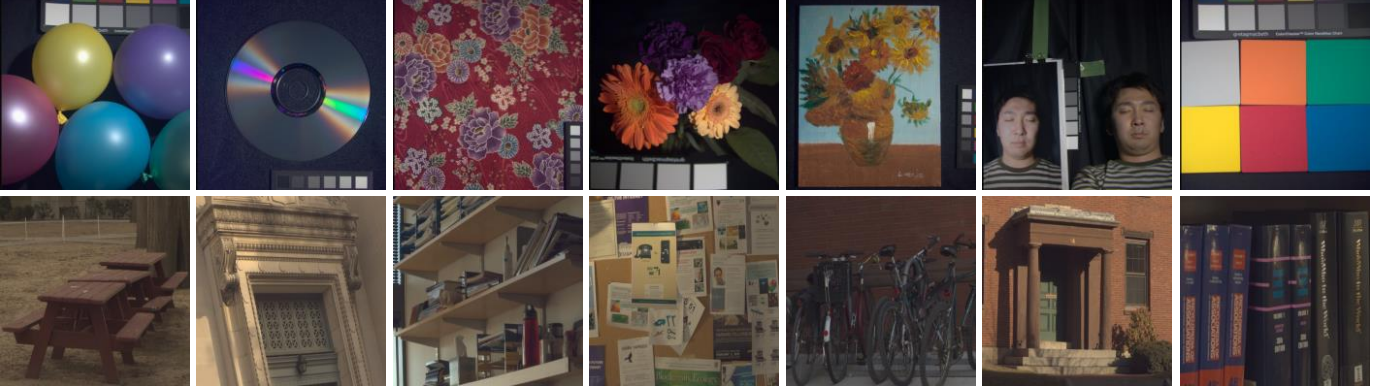


Fig. 5. Some color images from the CAVE dataset [62] (the first row) and the Harvard dataset [63] (the second row).

In our case, there is a trade-off between  $\mathcal{J}_1(\mathbf{X})$  and  $\mathcal{J}_2(\mathbf{X})$  controlled by the regularization parameter  $\mu$ . By solving (5) via Algorithm 1, each value of  $\mu$  yields a solution:

$$\widehat{\mathbf{X}}_\mu = \arg \min_{\mathbf{X}} \mathcal{J}(\mathbf{X}) = \arg \min_{\mathbf{X}} \mathcal{J}_1(\mathbf{X}) + \mu \mathcal{J}_2(\mathbf{X}) \quad (11)$$

and a point  $(\mathcal{J}_1(\widehat{\mathbf{X}}_\mu), \mathcal{J}_2(\widehat{\mathbf{X}}_\mu))$  of the response curve. According to the theorem in [40], this response curve is convex and exactly coincides with the *Pareto front* because  $\mathcal{J}(\mathbf{X})$  is convex. For notation simplicity, we will write the point as  $(\mathcal{J}_1(\mu), \mathcal{J}_2(\mu))$  and  $\mathcal{J}(\widehat{\mathbf{X}}_\mu)$  as  $\mathcal{J}(\mu)$ .

2) *Minimum Distance Criterion*: The minimum distance criterion [40] allows us to search the optimal point  $\mathbf{A} (\mathcal{J}_1(\mu^*), \mathcal{J}_2(\mu^*))$  on the response curve with  $\mu^*$  being the optimal parameter value. As represented in Fig. 4, this optimal point  $\mathbf{A}$  is at minimum distance to the *ideal point*  $\mathbf{I}$ , which is defined as  $(I_1, I_2)$  and corresponds to the point whose coordinates are minimum in the two objectives:

$$\begin{aligned} I_1 &= \mathcal{J}(0) = \arg \min_{\mathbf{X}} \mathcal{J}_1(\mathbf{X}) \\ I_2 &= \mathcal{J}(\infty) = \arg \min_{\mathbf{X}} \mathcal{J}_2(\mathbf{X}) \end{aligned} \quad (12)$$



---

**Algorithm 2** Deep Priors-Based HR Hyperspectral Image Super-Resolution.
 

---

**Input:**  $\mathbf{Y}, \mathbf{Z}, \mathbf{B}, \mathbf{S}, \mathbf{R}, \Theta, \alpha, a, b, \delta, \epsilon$ .

**Output:**  $\tilde{\mathbf{X}}$ .

 Initialize  $\mathbf{Y}_{\text{up}}$  from  $\mathbf{Y}$  with the bicubic interpolation.

(a) Obtain representation of learned deep-priors via TSFN:

$$\tilde{\mathbf{X}} = \mathcal{F}(\mathbf{Y}_{\text{up}}, \mathbf{Z}; \Theta)$$

(b) Search the optimal hyperparameter:

**repeat**

$$\mu_1 = a + \delta(b - a);$$

$$\mu_2 = b - \delta(b - a);$$

 Compute  $\tilde{\mathbf{X}}_{\mu_1}$  and  $\tilde{\mathbf{X}}_{\mu_2}$  via (5) using Algorithm 1;

 Obtain  $\mathcal{D}(\mu_1)$  and  $\mathcal{D}(\mu_2)$  via (13);

**if**  $\mathcal{D}(\mu_1) < \mathcal{D}(\mu_2)$  **then**

$$b = \mu_2$$

**else**

$$a = \mu_1$$

**end if**
**until**  $b - a < \epsilon$ 

 (c)  $\mu^* = \alpha(a + b)/2$ ;

 (d) Output  $\tilde{\mathbf{X}} = \tilde{\mathbf{X}}_{\mu^*}$  via (5) using Algorithm 1;
 

---

Unlike the HSI deconvolution problem in [40], the value of  $\mathcal{J}_1$  is typically much smaller than  $\mathcal{J}_2$  in the HSI super-resolution problem since  $\mathcal{J}_1$  involves less pixels ( $B \times n \times n$  and  $b \times N \times N$ ) compared to  $\mathcal{J}_2$  with  $B \times N \times N$  pixels. Considering the large scale difference between  $\mathcal{J}_1$  and  $\mathcal{J}_2$ , we propose to use a scaled distance as

$$\mathcal{D}(\mu) = (\mathcal{J}_1(\mu) - I_1)^2 + \alpha(\mathcal{J}_2(\mu) - I_2)^2 \quad (13)$$

where  $\alpha = (b/B)^2 + (1/s^2)^2$  is the scaling factor. Then the optimal hyperparameter is given by

$$\mu^* = \alpha \arg \min_{\mu} \mathcal{D}(\mu). \quad (14)$$

The varying range of  $\mu$  is  $(0, \infty)$ . In practice, the value of  $\mu$  cannot be set to  $\infty$  but can be fixed to a large value  $b$ . Meanwhile, the lower bound of  $\mu$  is set to a small value  $a$  rather than zero to avoid leading to highly ill-posed problems.

3) *Optimal Point Search:* The MDC is proved unimodal and always admits a unique minimum since the response curve is convex [40]. It is possible to design a fast approach aiming at finding the optimal point  $\mathbf{A}$  on the response curve. We propose to use the golden-section search method that is able to efficiently find the optimum point for the unimodal function [64]. This method operates in the interval  $[a, b]$  and generates two intermediate points :

$$\begin{aligned} \mu_1 &= a + \delta(b - a) \\ \mu_2 &= b - \delta(b - a) \end{aligned} \quad (15)$$

where  $\delta = 0.618$  is the golden ratio. The evaluated values of  $\mathcal{D}(\mu_1)$  and  $\mathcal{D}(\mu_2)$  are then compared and, if  $\mathcal{D}(\mu_1) < \mathcal{D}(\mu_2)$ , then  $\mu_2$  replaces  $b$  (else,  $\mu_1$  replaces  $a$ ). This procedure is repeated in the new smaller interval  $[a, b]$  until  $b - a < \epsilon$  where  $\epsilon > 0$  is an allowable final length of uncertainty. Finally, the estimated optimal hyperparameter is given by

$$\mu^* = \alpha(a + b)/2 \quad (16)$$

The proposed overall HSI super-resolution framework is summarized in Algorithm 2.

## V. EXPERIMENTS

In this section, several experiments are conducted to illustrate the effectiveness of our framework. The results provided by the proposed method are compared with those of state-of-the-art HSI super-resolution methods from both quantitative and qualitative perspectives.

### A. Dataset and Experimental Setup

In this study, three simulated hyperspectral datasets, i.e., the CAVE dataset [62], the Harvard dataset [63] and the Chikusei dataset [12], were used to evaluate the performance of our proposed method. The CAVE dataset consists of 32 indoor HSI recorded under controlled illuminations, each of which is of size  $512 \times 512$  in spatial domain and contains 31 spectral bands ranging from 400 nm to 700 nm at a wavelength interval of 10 nm. In the Harvard dataset, there are 50 indoor and outdoor HSI

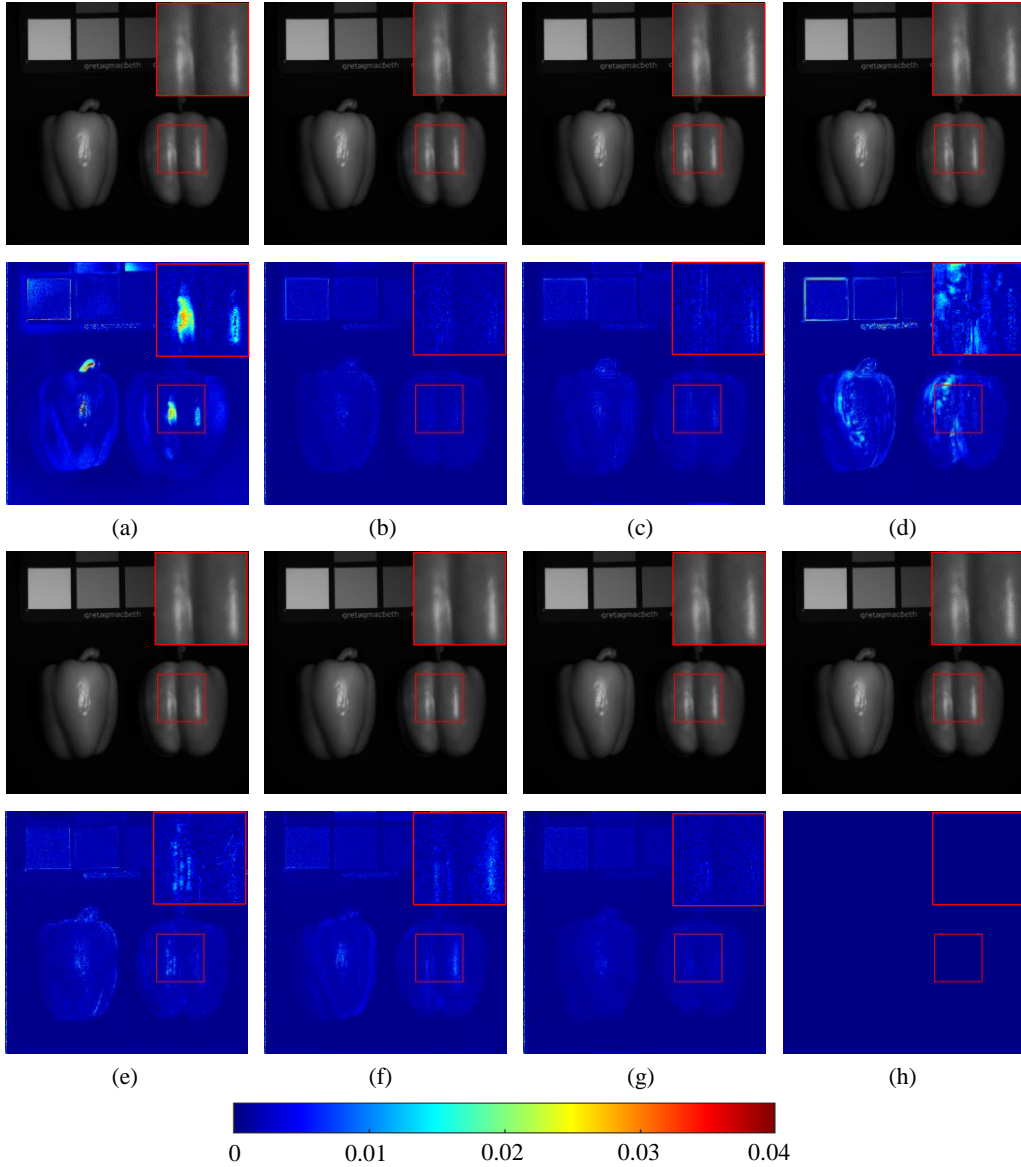


Fig. 6. Reconstructed images and corresponding error maps of the image *Real and fake peppers* from the CAVE dataset ( $s = 8$ ) at 550 nm band. Methods (RMSE/PSNR): (a) CSU (0.91/ 48.93). (b) NSSR (0.54/53.41). (c) DHSIS (0.58/52.91). (d) CMS (0.86/49.42). (e) LTR (0.61/52.49). (f) UAL (0.58/52.87). (g) Proposed (**0.51/54.01**). (h) Ground Truth.

TABLE I  
AVERAGE RMSE, PSNR, ERGAS, SAM OF DIFFERENT METHODS WITH THREE SCALING FACTORS ( $s = 8, 16, 32$ ) ON THE CAVE DATASET.

	Metric	CSU	NSSR	DHSIS	CMS	LTR	UAL	Proposed
$s = 8$	RMSE	2.57	1.47	1.36	1.65	1.49	1.42	<b>1.12</b>
	PSNR	41.76	46.66	47.01	45.29	46.55	46.47	<b>48.63</b>
	ERGAS	1.196	0.665	0.621	0.734	0.672	0.640	<b>0.511</b>
	SAM	6.27	3.72	3.74	3.89	3.84	3.60	<b>3.23</b>
$s = 16$	RMSE	2.82	1.77	1.79	1.99	1.84	1.57	<b>1.39</b>
	PSNR	41.01	45.31	44.74	43.94	44.87	45.80	<b>47.02</b>
	ERGAS	0.643	0.398	0.391	0.445	0.415	0.345	<b>0.310</b>
	SAM	6.47	4.32	4.57	4.37	4.63	3.84	<b>3.76</b>
$s = 32$	RMSE	3.02	2.24	2.45	2.35	2.28	1.85	<b>1.80</b>
	PSNR	40.44	43.49	42.34	42.66	43.33	44.66	<b>45.07</b>
	ERGAS	0.336	0.244	0.257	0.257	0.248	0.196	<b>0.190</b>
	SAM	6.83	5.22	5.87	5.04	5.46	<b>4.33</b>	4.59

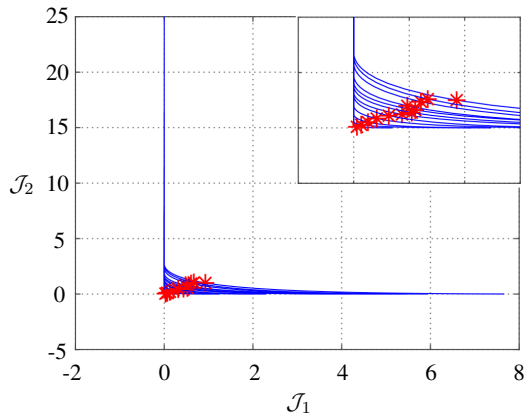


Fig. 7. Response curves (blue) and optimal points (red) with the scaling factor  $s = 8$  on the CAVE dataset. Each curve corresponds to a test image.

captured under daylight illumination. These images consist of  $1392 \times 1040$  pixels, with 31 spectral bands of 10 nm, covering the visible spectrum 420 to 720 nm. The top left  $1024 \times 1024$  pixels were cropped and extracted for our experiments. Examples of color images from these two datasets are shown in Fig. 5. The Chikusei dataset contains an airborne HSI taken by a Visible and Near-Infrared (NIR) imaging sensor over agricultural and urban areas in Chikusei, Ibaraki, Japan. The hyperspectral dataset has 128 bands in the spectral range from 363 nm to 1018 nm and the scene consists of  $2517 \times 2335$  pixels. After removing the black boundaries in the spatial domain, centering  $2048 \times 2048$  pixels were cropped and extracted for our experiments.

In our experiments, we followed the standard HSI super-resolution setups as in [23] and [17]. More specifically, the HSIs from three datasets were scaled into the range of  $[0, 1]$ , then served as the ground truth of  $\mathbf{X}$ . The LR HSI  $\mathbf{Y}$  was generated by down-sampling the ground truth over disjointing  $s \times s$  blocks, where scaling factor  $s$  was set to 8, 16 and 32. The HR conventional (RGB) image  $\mathbf{Z}$  was simulated by down-sampling the ground truth in the spectral domain using the spectral response function  $\mathbf{R}$ , which was derived from the response of a Nikon D700 camera<sup>1</sup>. For the Chikusei dataset, considering the diversity of hyperspectral sensors, the spectral response function  $\mathbf{R}$  of the IKONOS satellite<sup>2</sup> was employed to generate the HR conventional (RGB-NIR) image  $\mathbf{Z}$ . As in [37], we selected the first 20 images from the CAVE dataset as the training set and the rest 12 images as the test set. In the Harvard dataset, the first 30 images were used for training while the other 20 ones were used for testing. In the Chikusei dataset, we selected a  $1024 \times 2048$ -pixel-size image from the top area of the image for training while the remaining part of the image were cropped into 8 non-overlap  $512 \times 512 \times 128$  as testing data.

Besides, real sample images *Roman Colosseum* acquired by World View-2 (WV-2) were used in these experiments. This dataset contains an LR HSI and an HR RGB image, which are scaled into the range of  $[0, 1]$ . The HR RGB image is of size  $2632 \times 1676$  in spatial domain and the LR HSI is of size  $658 \times 419$  with 8 bands in spectral range from 400 to 1040 nm. Following [27], the blurring matrix and the spectral response function were estimated using the method in [65]. The ground-truth HR HSI in real data is not available so that training image (patch) pairs  $\{(\mathbf{Y}_{\text{up}, m}, \mathbf{Z}_m; \mathbf{X}_m)\}_{m=1}^M$  were used to be generated for training the network function. To tackle this issue, we considered the LR HSI as the ground truth  $\mathbf{X}$  and down-sampled it to generate network inputs  $\mathbf{Y}_{\text{up}}$  and  $\mathbf{Z}$  using the estimated blurring matrix and the spectral response function.

## B. Implementation Details

We implemented the proposed two-stream fusion network using PyTorch framework and initialized the model using the method *HeUniform* [59]. The Adam optimizer [66] was utilized to minimize the loss function in (6) with an initial learning rate 0.0002 and a mini-batch of 16 in 500 epochs. In the training phase, each original HR HSI was cropped into 100, 400, 800 patches of size  $128 \times 128$  for the CAVE, Harvard, Chikusei dataset, respectively. Each patch was randomly flipped and rotated for data augmentation before randomly generating an LR HSI and an HR conventional image in different degraded scenarios. When testing, we removed batch-normalization [58]. Each stream contains 6 identical residual blocks ( $P, Q=6$ ) in our model. Note that the proposed two-stream fusion network was trained separately on four datasets. For the golden-section search method, we set  $a = 10^{-8}$ ,  $b = 1$  and  $\epsilon = 0.01$ .

## C. Compared Methods and Quantitative Metrics

We compared the results of our method with six state-of-the-art HSI super-resolution methods: Coupled Spectral Unmixing (CSU) method [7], Non-negative Structured Sparse Representation (NSSR) method [23], Deep Hyperspectral Image Sharpening (DHSIS) method [37], Clustering Manifold Structure (CMS) method [17], Low Tensor-Train Rank (LTTR) method [26] and

<sup>1</sup><https://www.maxmax.com/faq/camera-tech/spectral-response/nikon-d700-study>.

<sup>2</sup><https://www.satimagingcorp.com/satellite-sensors/ikonos/>

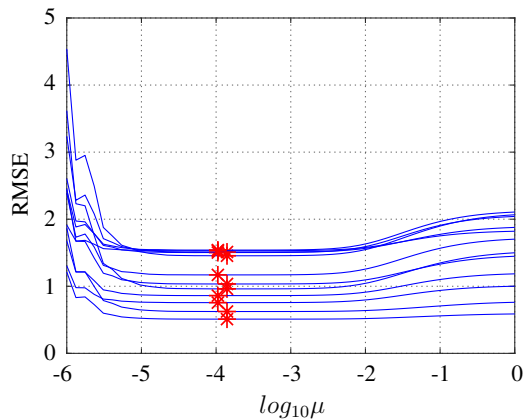


Fig. 8. RMSE curves (blue) and estimated points (red) with the scaling factor  $s = 8$  on the CAVE dataset. Each curve corresponds to a test image.

TABLE II  
AVERAGE RMSE, PSNR, ERGAS, SAM OF DIFFERENT METHODS WITH THREE SCALING FACTORS ( $s = 8, 16$  AND  $32$ ) ON THE HARVARD DATASET.

	Metric	CSU	NSSR	DHSIS	CMS	LTTR	UAL	Proposed
$s = 8$	RMSE	1.91	1.67	1.80	1.65	1.69	1.73	<b>1.60</b>
	PSNR	45.10	46.29	45.63	46.49	46.10	46.16	<b>46.73</b>
	ERGAS	1.423	1.219	1.415	1.355	1.319	1.270	<b>1.127</b>
	SAM	3.65	3.46	3.69	3.57	3.52	3.46	<b>3.36</b>
$s = 16$	RMSE	2.01	1.78	1.80	1.76	1.81	1.77	<b>1.72</b>
	PSNR	44.59	45.95	45.88	46.06	45.66	46.03	<b>46.32</b>
	ERGAS	0.739	0.653	0.654	0.721	0.709	0.633	<b>0.601</b>
	SAM	3.73	3.58	3.61	3.68	3.65	3.50	<b>3.48</b>
$s = 32$	RMSE	2.14	1.87	1.92	1.83	1.91	1.83	<b>1.80</b>
	PSNR	43.91	45.54	45.41	45.79	45.26	45.81	<b>45.99</b>
	ERGAS	0.394	0.363	0.339	0.381	0.375	0.323	<b>0.313</b>
	SAM	3.80	3.73	3.81	3.76	3.81	3.58	<b>3.59</b>

Unsupervised Adaption Learning (UAL) method [36]. CSU aims to regularize the HSI super-resolution problem by considering the spectral unmixing constraints. NSSR focuses on the sparse representation of the latent HR HSI as the prior structure for super-resolution. In DHSIS, the priors of latent HR HSI are learned by a deep neural network using massive data pairs without the use of pre-trained models. CMS exploits the manifold structure to capture the spatial correlation of the latent HR HSI to constrain the super-resolution scheme. LTTR learns correlations among the spatial, spectral and nonlocal modes of the nonlocal similar HR HSI cubes. UAL learns the deep priors and estimate the unknown spatial degradation in a unsupervised manner. Among these compared methods, DHSIS is the closest one to ours. The main difference is that it learned deep priors from a single pre-processed image while we directly learned the prior from two observed images. All these methods were implemented with their published codes online. Note that we separately trained the deep neural network for different scaling factors, blurring kernels and datasets in DHSIS. To evaluate the quality of reconstructed hyperspectral images, four quantitative metrics including the root-mean-square error (RMSE), the peak-signal-to-noise-ratio (PSNR), the erreur relative globale adimensionnelle de synthèse (ERGAS) [67] and the spectral angle mapper (SAM) [68] were used.

#### D. Performance Evaluation on Simulated Data

We validated the proposed method to show its effectiveness in terms of super-resolution performance over other compared methods on three public datasets.

1) *Evaluation on CAVE Dataset:* Firstly, we conducted and compared all methods to restore each latent HR HSI  $\mathbf{X}$  from the corresponding LR HSI  $\mathbf{Y}$  and HR conventional image  $\mathbf{Z}$  with scale factor  $s$  in the CAVE dataset. The average values of RMSE, PSNR, ERGAS and SAM of the compared methods with different scaling factors on the CAVE dataset are reported in Table I. It is clear that the proposed method outperforms all competing methods for all different settings. Especially when the scaling factor is small (e.g.,  $s = 8$ ), the superiority and robustness of our method are more significant. The improvement mainly stems from the effectiveness of the learned priors considering both spatial context and spectral attributes of the latent HR HSIs. Furthermore, the regularization parameter estimation also contributes to the satisfactory performance by giving a good trade-off

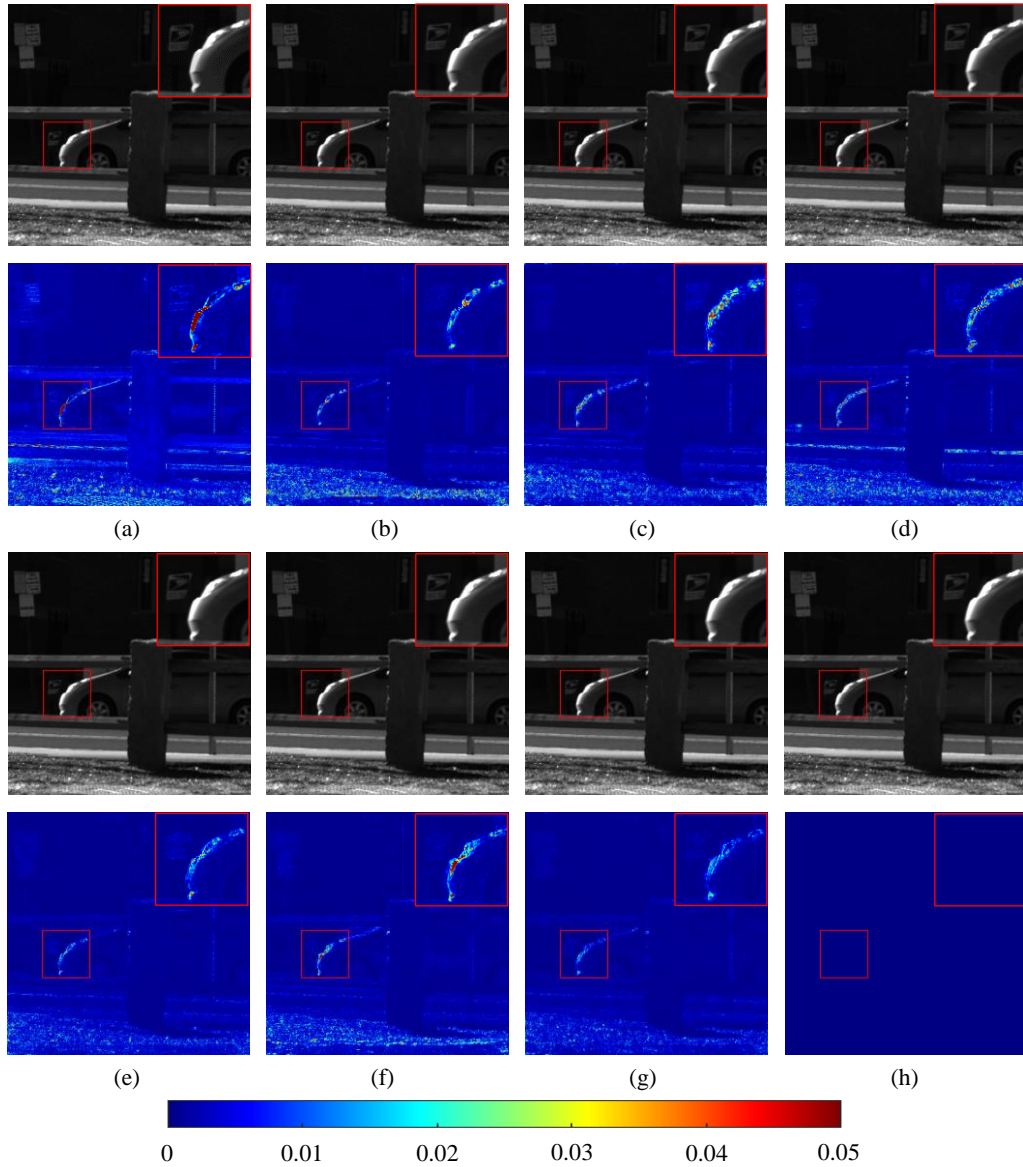


Fig. 9. Reconstructed images and corresponding error maps of the image *imgf1* from the Harvard dataset ( $s = 8$ ) at 650 nm band. Methods (RMSE/PSNR): (a) CSU (1.49/ 44.69). (b) NSSR (1.12/47.13). (c) DHSIS (0.94/48.70). (d) CMS (1.15/46.94). (e) LTR (0.82/49.91). (f) UAL (0.91/48.90). (g) Proposed (**0.73/50.89**). (h) Ground Truth.

TABLE III  
AVERAGE RMSE, PSNR, ERGAS, SAM OF DIFFERENT METHODS WITH THREE SCALING FACTORS ( $s = 8, 16$  AND  $32$ ) ON THE CHIKUSEI DATASET.

	Metric	CSU	NSSR	DHSIS	CMS	LTR	UAL	Proposed
$s = 8$	RMSE	0.89	1.36	0.72	2.42	2.09	0.85	<b>0.66</b>
	PSNR	51.16	49.90	53.16	44.64	45.67	52.47	<b>54.02</b>
	ERGAS	1.409	1.457	1.181	1.904	2.072	1.474	<b>1.128</b>
	SAM	1.79	2.36	1.58	2.97	3.33	1.85	<b>1.43</b>
$s = 16$	RMSE	0.98	1.47	0.78	2.72	2.58	0.73	<b>0.72</b>
	PSNR	50.32	49.27	52.64	43.80	44.18	53.10	<b>53.41</b>
	ERGAS	0.757	0.793	0.620	1.040	1.195	0.610	<b>0.607</b>
	SAM	1.91	2.55	1.71	3.40	4.15	1.58	<b>1.55</b>
$s = 32$	RMSE	1.11	1.64	0.84	3.13	3.41	<b>0.78</b>	<b>0.78</b>
	PSNR	48.85	48.26	51.96	42.77	42.30	52.74	<b>52.82</b>
	ERGAS	0.411	0.434	0.332	0.583	0.710	<b>0.314</b>	0.322
	SAM	2.07	2.89	1.83	4.17	5.40	<b>1.66</b>	1.67



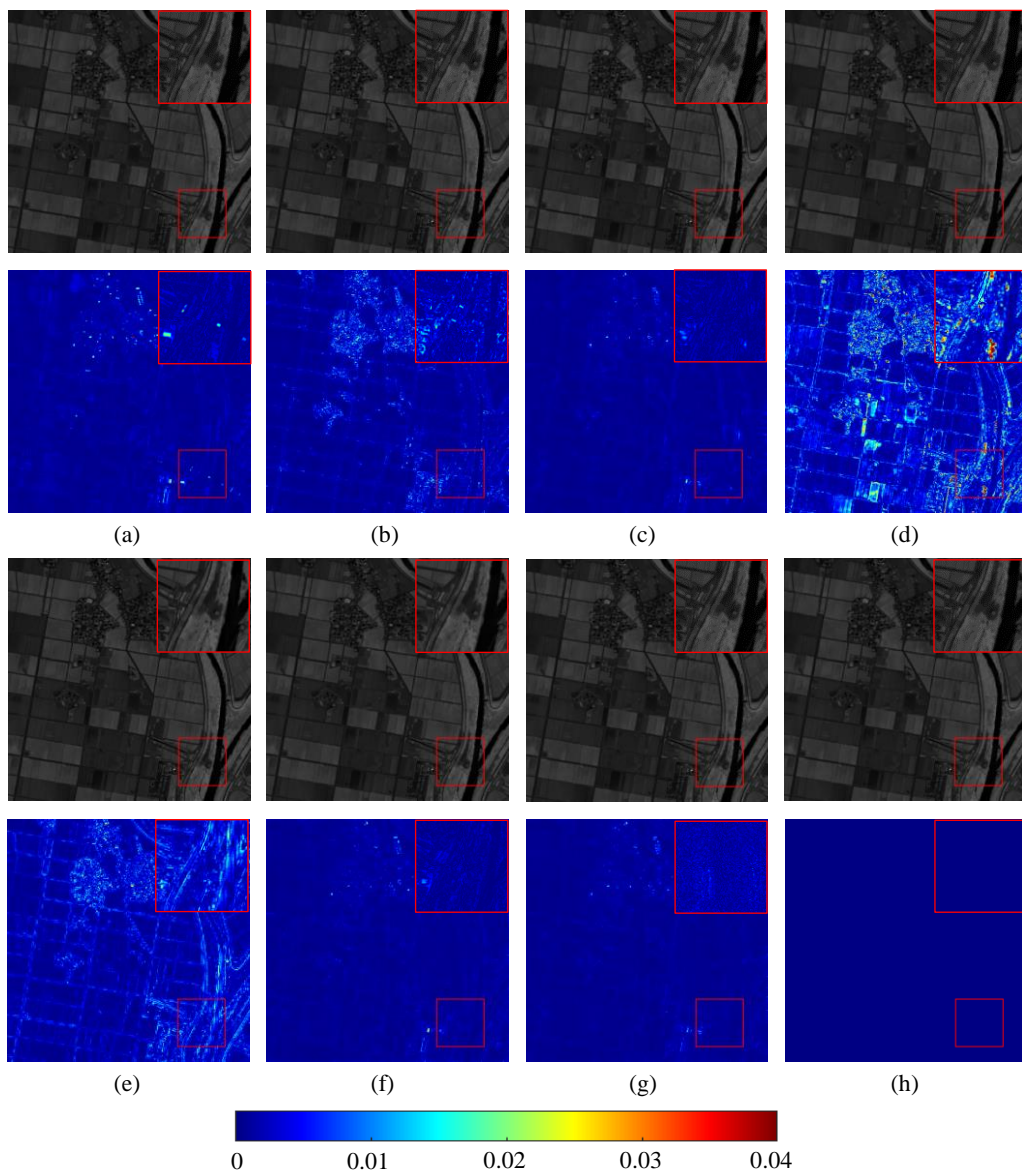


Fig. 10. Reconstructed images and corresponding error maps of a test image from the Chikusei dataset ( $s = 8$ ) at 800 nm band. Methods (RMSE/PSNR): (a) CSU (0.75/ 50.66). (b) NSSR (0.75/50.59). (c) DHSIS (0.46/54.89). (d) CMS (1.69/43.59). (e) LTR (0.97/48.38). (f) UAL (0.47/54.73). (g) Proposed (0.44/55.35). (h) Ground Truth.

between the data fidelity and regularization terms. Note that the regularization parameter of all competing methods have been fixed for all tested images.

For quality comparison, we took the scaling factor  $s = 8$  for example. Fig. 6 illustrates the reconstructed images and corresponding error maps of the test image *Real and fake pepper* at 550 nm. Visually, our method provides the best results in recovering the details of the latent HR HSIs and the corresponding error maps are closer to zero. This demonstrates the efficacy of our method in recovering the spatial information of the latent HR HSIs.

We evaluated the response curves with the scaling factor  $s = 8$  on the CAVE dataset and search their optimal points under MDC. For each response curve, the hyperparameter  $\mu$  was sampled on a 50 common logarithmic scale varying from  $10^{-6}$  to 1. Fig. 7 shows the response curves and the points with the optimal parameter values. Also, we evaluated the RMSE curves by varying the hyperparameter  $\mu$  values and the estimated optimal values under MDC. Fig. 8 shows the RMSE curves varying with  $\mu$  and the estimated optimal points. For better illustration, the horizontal axis is plotted in a common logarithmic scale. As can be observed, when  $\mu$  equals the optimal value estimated under MDC, each RMSE curve reaches corresponding minimum point, respectively. The optimal parameter values corresponding to different images with different scaling factors are slightly different but unique in our case.

2) *Evaluation on Harvard Dataset*: We further evaluated the proposed method on the Harvard dataset. Different from the CAVE dataset, the spatial size of images from the Harvard dataset is larger and slightly moving objects are not correctly

TABLE IV  
AVERAGE RMSE, PSNR, ERGAS, SAM OF DIFFERENT METHODS WITH THE SCALING FACTOR  $s = 8$  (GAUSSIAN BLUR KERNEL) ON THE CAVE,  
HARVARD AND CHIKUSEI DATASETS.

	Metric	CSU	NSSR	DHSIS	CMS	LTTR	UAL	Proposed
CAVE	RMSE	2.55	1.88	1.36	1.57	1.51	1.41	<b>1.13</b>
	PSNR	41.85	43.82	46.95	45.88	46.46	46.53	<b>48.56</b>
	ERGAS	1.138	0.838	0.619	0.686	0.679	0.636	<b>0.512</b>
	SAM	6.43	4.07	3.75	3.75	3.85	3.58	<b>3.22</b>
Harvard	RMSE	1.91	1.85	1.70	1.63	1.69	1.87	<b>1.58</b>
	PSNR	45.07	45.29	46.29	46.64	46.12	45.52	<b>46.77</b>
	ERGAS	1.420	1.275	1.228	1.306	1.306	1.722	<b>1.120</b>
	SAM	3.72	3.50	3.45	3.54	3.51	3.73	<b>3.34</b>
Chikusei	RMSE	0.88	1.52	0.73	2.19	2.10	1.26	<b>0.66</b>
	PSNR	51.04	47.86	52.93	46.00	46.65	50.53	<b>53.97</b>
	ERGAS	1.422	1.562	1.216	1.839	2.086	2.573	<b>1.124</b>
	SAM	1.82	2.59	1.58	3.23	3.33	2.73	<b>1.43</b>

TABLE V  
AVERAGE RMSE, PSNR, ERGAS, SAM OF DIFFERENT METHODS WITH THE SCALING FACTOR  $s = 8$  (GAUSSIAN BLUR KERNEL) ON THE CAVE  
DATASET UNDER NOISE CONDITIONS.

Method	RMSE	PSNR	ERGAS	SAM
CSU	2.68	41.01	1.231	6.96
NSSR	2.09	42.61	0.922	6.53
DHSIS	1.66	44.65	0.741	6.33
CMS	1.78	44.30	0.767	6.01
LTTR	1.70	44.90	0.749	5.70
UAL	1.51	45.72	0.673	<b>4.17</b>
Proposed	<b>1.34</b>	<b>46.42</b>	<b>0.604</b>	5.31

aligned in the neighboring bands. As illustrated in Table II, with the same experimental setups as discussed above, the numerical results of the proposed method surpass other competing methods in most cases. Fig. 9 displays the reconstructed images and corresponding error maps of image *imgfl* at 650 nm band with  $s = 8$ . It can be seen that our proposed method reconstructs more details and consequently produces smaller reconstruction error.

3) *Evaluation on Chikusei Dataset*: A public hyperspectral dataset with more bands, i.e., the Chikusei dataset, was used to evaluate the performance including the preservation of spectral information of the proposed method. The quantitative metrics of comparing methods with different scaling factors on the Chikusei dataset are reported in Table III. It can be observed that the numerical results of the proposed method surpass other competing methods in most cases. Figure 10 displays the reconstructed images and corresponding error maps of a test image at 800 nm band with  $s = 8$ . It can be seen that our method reconstructs more details and the corresponding error maps are closer to zero.

4) *Evaluation with Optics Blur*: In the above experiments, we applied the uniform blurring matrix  $\mathbf{B}$  of size  $s \times s$  to each band of the ground truth  $\mathbf{X}$  before the down-sampling operator. In the real-world hyperspectral imaging system, however, the optics blur may exist and it can be modeled by a Gaussian blurring kernel. To demonstrate the advantage of the proposed method when optics kernel occurs, we also simulate the LR HSI  $\mathbf{Y}$  by first applying  $8 \times 8$  Gaussian kernel with standard deviation 3 to the latent HR HSI  $\mathbf{X}$  before down-sampling in both the vertical and horizontal directions with scaling factor 8. The numerical results on the CAVE, Harvard and Chikusei datasets are reported in Table IV. We observe that the proposed method outperforms other considered methods.

#### E. Performance Evaluation under Noisy Conditions

We conducted some experiments under noise conditions to evaluate the robustness of the proposed method. These experiments were conducted on the CAVE dataset with the scaling factor  $s = 8$  and Gaussian blur kernel. We set the SNR to 40 dB to generate noise interruptions, which were added to both two input images. In Table V, it can be seen that the performance of all the methods declined compared with that in the noise-free condition shown in Table IV. However, the proposed method still provided the best quantitative results for the noisy LR HSI and HR conventional image.



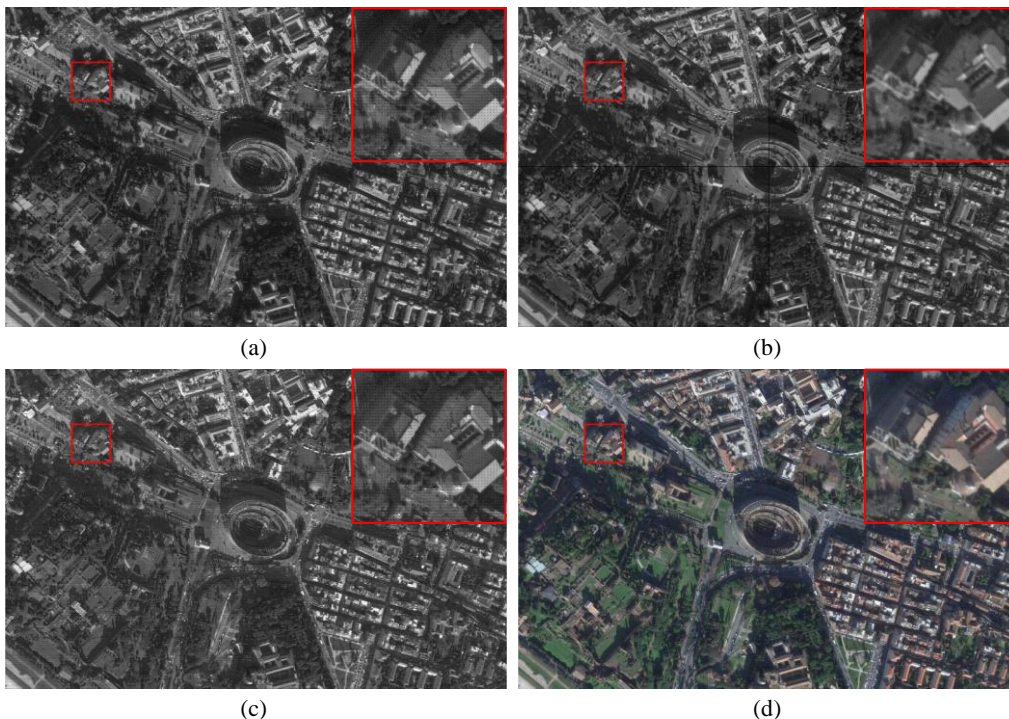


Fig. 11. Reconstructed images of the real dataset at 605 nm band and the HR RGB image. Methods: (a) LTTR. (b) UAL. (c) Proposed. (d) HR RGB.

TABLE VI  
AVERAGE RMSE, PSNR, ERGAS, SAM OF SOME COMPARED METHODS WITH THE SCALING FACTOR  $s = 8$  (GAUSSIAN BLUR KERNEL) ON THE CAVE DATASET.

Method	RMSE	PSNR	ERGAS	SAM
Bicubic	9.63	29.42	4.220	6.67
TSFN   $l_2$ -norm loss	1.80	44.54	0.807	4.30
TSFN w/o skip-con	2.58	42.08	1.092	4.87
TSFN	<b>1.52</b>	<b>46.01</b>	<b>0.674</b>	<b>3.81</b>
Proposed w/o TSFN	1.69	45.35	0.776	4.10
Proposed	<b>1.13</b>	<b>48.56</b>	<b>0.512</b>	<b>3.32</b>

### F. Performance Evaluation on Real Data

We conducted some experiments on real data to evaluate the efficiency of the proposed method. In this dataset, as the ground-truth HR HSI is not available, we generated the training data using the strategy described in Subsection V-A, and compared the proposed method with two most competitive methods, namely, LTTR [26] and UAL [36]. Due to the limit of GPU memory, we divided input images of UAL into four parts and concatenated the output images back to obtain the final estimated HR HSI.

Figure 11 shows the reconstructed images of the real dataset at 605nm and the HR RGB image. We can observe that the proposed method recovers more details and gives better visual effect. This demonstrates the applicability of our method in real-world scenarios.

### G. Ablation Study

We conducted ablation studies on the CAVE dataset. Without loss of generality, we conducted all these experiments with scaling factor 8 and Gaussian blur kernel mentioned above.

1) *Effect of Different Loss Functions*: To compare the performances of  $l_1$ -norm and  $l_2$ -norm for Eq. (6), we train the proposed TSFN loss functions with  $l_1$ -norm and  $l_2$ -norm, respectively. They are termed as “TSFN” and “TSFN |  $l_2$ -norm loss”. Table VI presents the average values of RMSE, PSNR, ERGAS and SAM of these two methods. We can see that  $l_1$ -norm shows significant superiority when compared to  $l_2$ -norm. The qualitative comparisons are shown in the Figure 12. It visualizes the reconstructed images and corresponding error maps of the test image *thread spools* at 550 nm. Visually, the  $l_1$ -norm based loss function provides better result in recovering the texture details of the latent HR HSI, and produces a smaller reconstruction error.

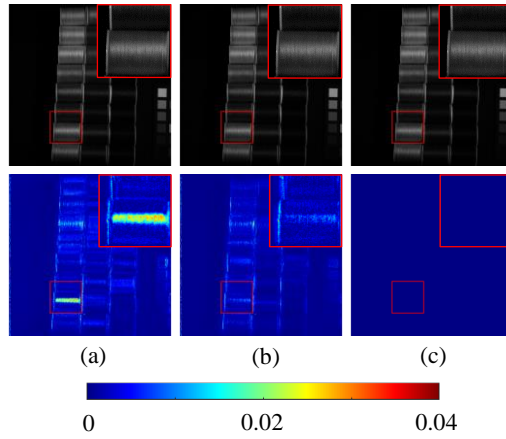


Fig. 12. Reconstructed images and corresponding error maps of the image *thread spools* from the CAVE dataset with  $s = 8$  (Gaussian blur kernel) at 550 nm band. Columns: (a) TSFN( $\ell_2$ -norm loss) (RMSE: 0.65, PSNR: 51.89). (b) TSFN (RMSE: 0.86, PSNR: 59.41). (c) Ground Truth.

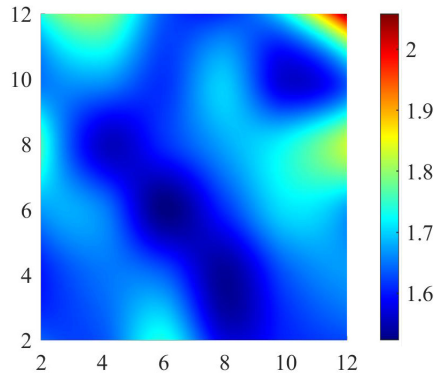


Fig. 13. 2D view of RMSE values with different selections of residual block numbers  $P$  and  $Q$ .

2) *Effect of Values of  $P$  and  $Q$  in TSFN*: The selection of numbers of residual blocks in two streams ( $P$  and  $Q$ ) are crucial in our network architecture design. The average RMSE values on test images were used to evaluate the performance. Figure 13 shows the two dimensional (2D) view of RMSE values with different settings of  $P$  and  $Q$ . Obviously, small values of  $P$  and  $Q$  will cause a poor inference ability of the network, yielding worse performance. On the other hand, the increase of numbers of residual blocks for feature extraction does not promise better results. It can be observed that the best performance occurred when the values of  $P, Q$  are both set around 6.

3) *Effect of the Skip Connection in TSFN*: The skip connection operator is the key component of our proposed TSFN. To show the effect of this operator, we utilized a variant of TSFN that removes the skip connection operator, which is termed as “TSFN w/o skip-con”. As reported in Table VI, removing the skip connection operator deteriorates the performance.

4) *Effect of TSFN in the Proposed Framework*: As stated in Section IV, we propose to learn the priors of the latent HR HSI via TSFN rather than using handcrafted regularizers. Here, we illustrate the effectiveness of TSFN, as well as its important role in our HSI super-resolution framework. First, we compared the up-sampled images  $\mathbf{Y}_{\text{up}}$  with the bicubic interpretation and  $\tilde{\mathbf{X}}$ , the outputs of our proposed TSFN. In addition, we evaluated the results of the proposed HSI super-resolution framework without TSFN (termed as “Proposed w/o TSFN”). More specifically, we used the up-sampled images  $\mathbf{Y}_{\text{up}}$  as the approximation of the latent HR HSI, and compared with the result of our complete framework with TSFN.

As shown in Table VI, the results of TSFN significantly outperformed those of the bicubic interpretation. This illustrates the effectiveness of the priors representation of our method learned via TSFN. Moreover, the performance of the proposed framework with TSFN exceeds that of the framework without TSFN, thus demonstrating TSFN is an indispensable part of our proposed HSI super-resolution method.

## VI. CONCLUSION

In this paper, we presented a deep priors-based HSI super-resolution method. Instead of using handcrafted image prior structures, we utilized the spatial-spectral priors learned by the proposed TSFN. The output of TSFN was leveraged to regularize the ill-posed problem. In addition, the regularization parameter was automatically estimated by adopting MDC on the response curve of the associated problem with a golden-section search approach. Experimental results on both simulated and real data demonstrated that our method can effectively handle scenarios with various scaling factor and blurring kernel setups.

In future works, the proposed HSI super-resolution method will be further extended in two directions. On the one hand, modeling the data fidelity term by estimating the blurring matrix and the spectral response function using deep learning techniques will be of interest. On the other hand, how to accelerate hyperparameter tuning will be investigated. These attempts will further enhance the applicability of our proposed method in real-world scenarios.

## REFERENCES

- [1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, 2013.
- [2] H. Van Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. -Workshops. (CVPRW)*. IEEE, 2010, pp. 44–51.
- [3] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, 2012.
- [4] J. Xie, N. He, L. Fang, and P. Ghamisi, "Multiscale densely-connected fusion networks for hyperspectral images classification," *IEEE Trans. Circuits Syst, Video Technol.*, 2020.
- [5] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, 2010.
- [6] J. Lei, X. Li, B. Peng, L. Fang, N. Ling, and Q. Huang, "Deep spatial-spectral subspace clustering for hyperspectral image," *IEEE Trans. Circuits Syst, Video Technol.*, 2020.
- [7] C. Lanaras, E. Baltasavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 3586–3594.
- [8] L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu, "Simultaneous depth and spectral imaging with a cross-modal stereo system," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 28, no. 3, pp. 812–817, 2016.
- [9] S. Zhang, H. Huang, and Y. Fu, "Fast parallel implementation of dual-camera compressive hyperspectral imaging system," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 29, no. 11, pp. 3404–3414, 2018.
- [10] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral computational imaging via collaborative tucker3 tensor decomposition," *IEEE Trans. Circuits Syst, Video Technol.*, 2020.
- [11] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, 2007.
- [12] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, 2017.
- [13] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.
- [14] P. Kwarteng and A. Chavez, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 339-348, p. 1, 1989.
- [15] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.
- [16] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, 1999.
- [17] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5969–5982, 2018.
- [18] 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, 2011.
- [19] 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, 2016.
- [20] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1693–1704, 2013.
- [21] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2014, pp. 63–78.
- [22] 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, 2015.
- [23] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, 2016.
- [24] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5344–5353.
- [25] 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, 2018.
- [26] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, 2019.
- [27] Y. Bu, Y. Zhao, J. Xue, J. C.-W. Chan, S. G. Kong, C. Yi, J. Wen, and B. Wang, "Hyperspectral and multispectral image fusion via graph laplacian-guided coupled tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, 2020.
- [28] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia.*, vol. 21, no. 12, pp. 3106–3121, 2019.
- [29] 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, 2017.
- [30] 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. (ICCV)*, 2017, pp. 5449–5457.
- [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, 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, 2019.
- [33] X. Deng and P. L. Dragotti, "Deep convolutional neural network for multi-modal image restoration and fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.
- [34] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-J. Zha, and F. Wu, "Deep residual attention network for spectral image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

- [35] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1585–1594.
- [36] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 3073–3082.
- [37] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, no. 99, pp. 1–11, 2018.
- [38] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, 2019.
- [39] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation  $ax + xb = c$  [4]," *Commun. ACM.*, vol. 15, no. 9, pp. 820–826, 1972.
- [40] Y. Song, D. Brie, E.-H. Djermoune, and S. Henrot, "Regularization parameter estimation for non-negative hyperspectral image deconvolution," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5316–5330, 2016.
- [41] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [42] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [43] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Process.*, vol. 1, no. 3, pp. 322–336, 1992.
- [44] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.
- [45] L. Kaufman and A. Neumaier, "Regularization of ill-posed problems by envelope guided conjugate gradients," *J. Comput. Graph. Statist.*, vol. 6, no. 4, pp. 451–463, 1997.
- [46] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, 2001, vol. 16.
- [47] E. Van Den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2009.
- [48] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imaging*, vol. 2, no. 4, pp. 408–423, 2016.
- [49] A. Teodoro, J. Bioucas-Dias, and M. Figueiredo, "A convergent image fusion algorithm using scene-adapted gaussian-mixture-based denoising," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 451–463, 2018.
- [50] X. Wang, J. Chen, C. Richard, and D. Brie, "Learning spectral-spatial prior via 3DDnCNN for hyperspectral image deconvolution," in *IEEE Int. Conf. on Acoust. Speech, Signal Process (ICASSP)*. IEEE, 2020, pp. 2403–2407.
- [51] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2014, pp. 568–576.
- [52] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 1933–1941.
- [53] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, 2017.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [55] —, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2016, pp. 630–645.
- [56] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 1637–1645.
- [57] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4681–4690.
- [58] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [59] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1026–1034.
- [60] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imaging*, vol. 3, no. 1, pp. 47–57, 2016.
- [61] 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, 2015.
- [62] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, 2010.
- [63] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 193–200.
- [64] J. Kim, "Iterated grid search algorithm on unimodal criteria," Ph.D. dissertation, Virginia Tech, 1997.
- [65] 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, 2014.
- [66] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [67] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" 2000.
- [68] R. H. Yuhua, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," 1992.