

Spectral Superresolution of Multispectral Imagery with Joint Sparse and Low-Rank Learning

Lianru Gao, *Senior Member, IEEE*, Danfeng Hong, *Member, IEEE*, Jing Yao, Bing Zhang, *Fellow, IEEE*, Paolo Gamba, *Fellow, IEEE*, and Jocelyn Chanussot, *Fellow, IEEE*

Abstract—Extensive attention has been widely paid to enhance the spatial resolution of hyperspectral (HS) images with the aid of multispectral (MS) images in remote sensing. However, the ability in the fusion of HS and MS images remains to be improved, particularly in large-scale scenes, due to the limited acquisition of HS images. Alternatively, we super-resolve MS images in the spectral domain by the means of partially overlapped HS images, yielding a novel and promising topic: spectral superresolution (SSR) of MS imagery. This is challenging and less investigated task due to its high ill-posedness in inverse imaging. To this end, we develop a simple but effective method, called joint sparse and low-rank learning (J-SLoL), to spectrally enhance MS images by jointly learning low-rank HS-MS dictionary pairs from overlapped regions. J-SLoL infers and recovers the unknown hyperspectral signals over a larger coverage by sparse coding on the learned dictionary pair. Furthermore, we validate the SSR performance on three HS-MS datasets (two for classification and one for unmixing) in terms of reconstruction, classification, and unmixing by comparing with several existing state-of-the-art baselines, showing the effectiveness and superiority of the proposed J-SLoL algorithm.

Index Terms—Dictionary learning, hyperspectral, joint learning, low-rank, multispectral, remote sensing, sparse representation, superresolution.

I. INTRODUCTION

WITH the rapid development and enormous breakthroughs in imaging technology, a variety of image

This work was supported by the National Natural Science Foundation of China under Grant 41722108 and Grant 91638201 as well as with the support of the AXA Research Fund. (*Corresponding author: Danfeng Hong*).

L. Gao is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China. (e-mail: gaolr@aircas.ac.cn)

D. Hong and J. Yao are with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling 82234, Germany, and Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich 80333, Germany. (e-mail: danfeng.hong@dlr.de, jing.yao@dlr.de)

D. Hong is also with the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, Grenoble 38000, France.

B. Zhang is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China. (e-mail: zb@radi.ac.cn)

P. Gamba is with Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia 27100, Italy. (e-mail: paolo.gamba@unipv.it)

J. Chanussot is with the Univ. Grenoble Alpes, INRIA, CNRS, Grenoble INP, LJK, Grenoble 38000, France, with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik 101, Iceland, as well as with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China. (e-mail: jocelyn@hi.is)

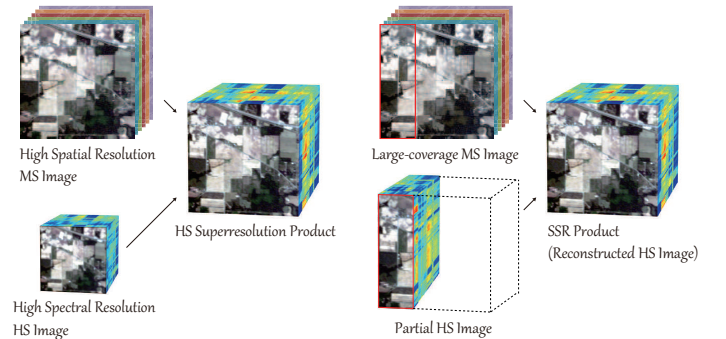


Fig. 1. Illustrative comparison for the tasks of HS-SR and SSR of MS images.

products, e.g., hyperspectral (HS) data, multispectral (MS) data, synthetic aperture radar (SAR), light detection and ranging (LiDAR), have been widely and successfully applied to many practical applications in remote sensing (RS) and geoscience [1]–[3], such as mineral exploration, urban planning and management, previous framing, water quality assessment, disaster pre-warning and prevention, to name a few.

Characterized by very rich and diverse spectral information, HS images, which enable to identify the materials on the surface of the Earth more accurately, have been paid an increasing attention in various HS RS tasks: feature extraction and embedding [4]–[11], spectral unmixing [12]–[14], data fusion [15]–[19], target detection [20]–[22], and multimodal data analysis [23], [24]. However, although currently operational, e.g., Earth Observing-1 (EO-1), DLR Earth Sensing Imaging Spectrometer (DESIS), or upcoming imaging spectroscopy, e.g., Environmental Mapping and Analysis Program (EnMAP), satellite missions and airborne imaging sensors can provide HS data of high spectral resolution, yet its low spatial resolution limits the performance to be further improved.

For this reason, enormous effects have been recently made to enhance the spatial resolution of HS images by fusing corresponding MS images [25], yielding high-quality HS products (high spatial and spectral resolutions). Yokoya *et al.* [26] developed a classic but very effective work for the fusion of HS and MS images by coupled spectral unmixing. Such a strategy has been proven to be effective by many follow-up works. For example, authors in [27] re-formulated the coupled spectral unmixing model as a constrained optimization problem for HS superresolution (HS-SR). Beyond the pixel-based fusion, Xu *et al.* [16] proposed a nonlocal coupled spectral unmixing in a tensorized manner by considering a larger perspective field of each pixel to reconstruct the fused HS image. The same investigators [28] further extended their

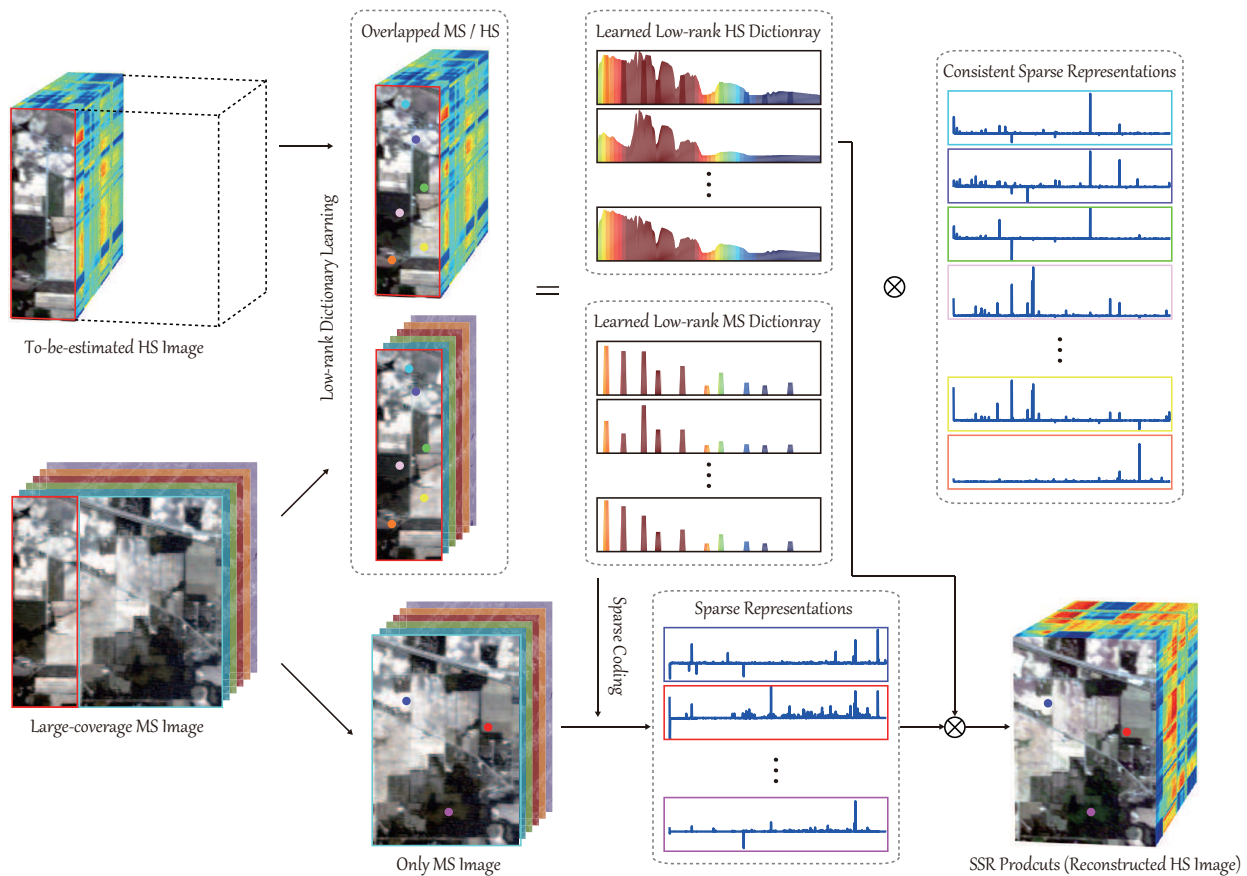


Fig. 2. An illustration of the proposed J-SLoL model for the SSR task. It consists of two parts: low-rank dictionary learning (learned HS-MS dictionary pair) and sparse coding for the final SSR product (reconstructed HS image).

work by adaptively learning response functions instead of pre-given ones. Besides, there are many other types of HS-MS fusion methods, e.g., dictionary learning [29], sparse representation [30], Bayesian fusion [31], **subspace learning** [19], [32], deep learning-based approaches [33], etc.

It is well known, however, that the HS data fail to be acquired in a large-covered area due to the limitations of imaging devices. This would bring a big challenge in collecting the same-size HS-MS image pairs for the fusion task. But fortunately, we may expect to have the MS data [34] at a large and even global scale, owing to its easy-availability. We have to admit that the relatively poor spectral information makes difficulty for the MS data to detect and recognize the materials at a more accurate level, particularly for those classes that hold very similar visual cues. As a trade-off, spectral superresolution (SSR) of MS images might be an alternative solution, as illustrated in Fig. 1. This topic is novel and promising. **Although some famous works [35]–[39] have been studied by the attempts to challenge the similar topic from the perspective of methodology in the computer vision field, yet it is less investigated by researchers in RS.** There are only several tentative works. For example, authors of [40] simply assumed the existence of a regression matrix between overlapped HS and MS regions and utilized the estimated transformation to reconstruct the unknown HS signals. Further, Sun *et al.* [41] learned multiple transformation matrices from

the grouped HS image and screened out an optimal one by a weighted spectral angle distance. Another recent work related to this task was presented in [42], in which sparse representation is used to recover the large-scale HS image from partially overlapped HS and MS images. It should be noted that without any priors (or regularizations), estimating simple regression matrix from the limited HS-MS pairs (e.g., [40] and [41]) is a highly ill-conditioned problem, thereby leading to poor reconstruction for HS images. On the other hand, sparse representation in [42] is capable of recovering the unknown HS signals well. Nevertheless, reconstruction coefficients (or sparse representations) are only estimated on the MS data taken as the dictionary (or basis) and directly transferred into the HS reconstruction. Consequently, the two kinds of approaches could be effective for the SSR task to some extent, but the potential in fully making use of overlapped HS-MS images remains limited.

To overcome these difficulties, we propose a simple but effective learning algorithm, called **joint sparse and low-rank learning (J-SLoL)**, for the task of SSR. As the name suggests, J-SLoL jointly learns the low-rank HS-MS dictionary pair and its corresponding sparse representations. The learned HS and MS dictionaries are consistent in the activated locations of sparse coefficients for each pixel. Such consistency enables the reconstruction of HS images at a more accurate level. More specifically, main contributions of this paper can be

highlighted as follows

- We equivalently convert the problem of HS-SR to that of the SSR of MS images. Compared to the former, the latter has great potential in recovering high-quality HS images, particularly in large-coverage case. To our best knowledge, this is the first time to investigate the SSR problem in RS.
- We propose a J-SLoL model for addressing the SSR problem effectively. J-SLoL can learn low-rank overcomplete dictionaries with respect to HS and MS data, respectively, and consistent sparse representations from the overlapped HS-MS images and further reconstruct the unknown HS images by using shared sparse coefficients obtained from the corresponding MS parts.
- Reconstruction, classification, and unmixing are used to evaluate the product quality. Extensive experiments conducted on three HS-MS datasets demonstrate the superiority of the proposed J-SLoL model in the SSR case.

The remaining part of this paper is organized as follows. We briefly state the SSR problem formulation and present the proposed J-SLoL model and its optimizer in Section II. Section III provides extensive experiments as well as corresponding results and analysis. Finally, we make a conclusion to this paper with a future outlook in Section IV.

II. JOINT SPARSE AND LOW-RANK LEARNING

A. Problem Statement

It is clear that our goal is to obtain the HS product of high spatial and high spectral resolutions. Therefore, HS-SR and SSR of MS images are two feasible solutions, as shown in Fig. 1. It should be noted, however, that the main advantage of SSR using MS images over HS-SR lies in easy availability of MS images on a larger geospatial coverage. On the other hand, the HS imagery is generally acquired in the form of a very narrow strip, as the spectrum within one pixel consists of hundreds of wavelength bands, limiting the imaging range of HS images.

Unlike the HS-SR task that the same geospatial region for HS and MS images is needed, SSR aims to enhance spectral resolution of MS images only using partially overlapped HS and MS images (see Fig. 1). This can save the cost in data preparation well and enables the generation of high-quality HS images in an easier way. Please note that a simplified case of SSR in this paper is investigated by using partially HS-MS images at a same ground sampling distance (GSD).

B. Problem Formulation

Fig. 2 illustrates the proposed J-SLoL model, where we reasonably assume each spectrum out of the overlapped region, either in the HS or the MS images, can be well reconstructed by an identical sparse combination of atoms on the low-rank HS (or MS) dictionary learned from the overlapped HS and MS images.

Let the spectral signatures of the HS and MS images in the overlapped part be $\mathbf{H}_{in} = [\mathbf{h}_{in}^1, \dots, \mathbf{h}_{in}^i, \dots, \mathbf{h}_{in}^N] \in$

$\mathbb{R}^{P \times N}$ with P spectral bands by N pixels and $\mathbf{M}_{in} = [\mathbf{m}_{in}^1, \dots, \mathbf{m}_{in}^i, \dots, \mathbf{m}_{in}^N] \in \mathbb{R}^{Q \times N}$ with Q channels by N pixels, respectively. Moreover, we define those pixels out of the overlapped region as $\mathbf{H}_{out} = [\mathbf{h}_{out}^1, \dots, \mathbf{h}_{out}^j, \dots, \mathbf{h}_{out}^{N_1}] \in \mathbb{R}^{P \times N_1}$ and $\mathbf{M}_{out} = [\mathbf{m}_{out}^1, \dots, \mathbf{m}_{out}^j, \dots, \mathbf{m}_{out}^{N_1}] \in \mathbb{R}^{Q \times N_1}$ for the HS and MS images, respectively. Our J-SLoL model consists of two steps: low-rank dictionary learning and sparse recovery, for the SSR task.

1) *Low-Rank Dictionary Learning (D-Step)*: This process can be formulated by solving the following constrained optimization problem

$$\begin{aligned} \min_{\mathbf{D}_h, \mathbf{D}_m, \mathbf{X}} \quad & \frac{1}{2} \|\mathbf{H}_{in} - \mathbf{D}_h \mathbf{X}\|_F^2 + \frac{\alpha}{2} \|\mathbf{M}_{in} - \mathbf{D}_m \mathbf{X}\|_F^2 + \beta \|\mathbf{X}\|_{1,1} \\ & + \gamma (\|\mathbf{D}_h\|_* + \|\mathbf{D}_m\|_*), \\ \text{s.t.} \quad & \mathbf{D}_h \succeq \mathbf{0}, \quad \mathbf{D}_m \succeq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{X} = \mathbf{1}, \end{aligned} \quad (1)$$

where $\mathbf{D}_h \in \mathbb{R}^{P \times L}$ and $\mathbf{D}_m \in \mathbb{R}^{Q \times L}$ are the to-be-learned low-rank dictionaries with respect to HS and MS data, and $\mathbf{X} \in \mathbb{R}^{L \times N}$ denotes the consistent sparse representations on the two dictionaries. $\|\bullet\|_*$ and $\|\bullet\|_{1,1}$ represent the nuclear norm to approximate the rank of matrix and the sparsity-promoting term approximately estimated by ℓ_1 -norm, respectively. Moreover, $\mathbf{1}$ and $\mathbf{0}$ are the unit vector and the zero matrix, respectively. α , β , and γ are penalty parameters to balance the importance of different terms in Eq. (1).

2) *Sparse Recovery (S-Step)*: Owing to the consistent sparse representations in dictionary learning, the learned HS and MS dictionaries are well applicable to the unknown HS reconstruction. Therefore, there will be two main parts in sparse recovery, i.e., sparse coding on \mathbf{D}_m and HS reconstruction using \mathbf{D}_h .

The sparse coefficients of \mathbf{M}_{out} can be encoded on the MS dictionary (\mathbf{D}_m), the resulting sparse coding problem can be then written as

$$\min_{\mathbf{Y}} \quad \frac{1}{2} \|\mathbf{M}_{out} - \mathbf{D}_m \mathbf{Y}\|_F^2 + \eta \|\mathbf{Y}\|_{1,1}, \quad \text{s.t.} \quad \mathbf{1}^\top \mathbf{Y} = \mathbf{1}, \quad (2)$$

where \mathbf{Y} denotes the sparse coefficients with respect to the variable \mathbf{M}_{out} .

Once the coding results (\mathbf{Y}) are given, it is straightforward to derive the HS reconstruction (or SSR of MS images), denoted as $\hat{\mathbf{H}}_{out} = \mathbf{D}_h \mathbf{Y}$.

C. Model Optimization

The proposed J-SLoL model consists of two optimization problems: low-rank dictionary learning and sparse coding, where the latter is a special case of the former.

1) *D-Step Solver*: Despite the non-convexity in Eq. (1), the sub-problems for each variable is solvable. For that, an alternating direction method of multipliers (ADMM) algorithm is designed to solve this model fast and effectively. To facilitate the use of ADMM optimizer, an equivalent form of Eq. (1) is converted by introducing several auxiliary variables, i.e., \mathbf{Z} ,

Algorithm 1: J-SLoL solver: *D-Step*

Input: $\mathbf{H}_{in}, \mathbf{M}_{in}$, parameters α, β , and \maxIter .
Output: $\mathbf{D}_h, \mathbf{D}_m$, and \mathbf{X} .

- 1 **Initialization:** $\mathbf{J}^0 = \mathbf{0}, \mathbf{K}^0 = \mathbf{0}, \mathbf{Z}^0 = \mathbf{0}, \{\Lambda_i^0\}_{i=1}^3 = \mathbf{0}, \mu^0 = 10^{-3}, \mu_{max} = 10^6, \xi = 1.5, \varepsilon = 10^{-6}, t = 1$.
- 2 **while not converged** or $t > \maxIter$ **do**
- 3 Update the variable \mathbf{X}^t by Eq. (6);
- 4 Update the variable \mathbf{D}_h^t by Eq. (8);
- 5 Update the variable \mathbf{D}_m^t by Eq. (10);
- 6 Update the variables \mathbf{J}^t and \mathbf{K}^{t+1} by Eq. (13);
- 7 Update the variable \mathbf{Z}^t by Eq. (15);
- 8 Update Lagrange multipliers $\{\Lambda_i^t\}_{i=1}^3$ by Eq. (16);
- 9 Update the parameter μ^t by $\min(\xi\mu^{t-1}, \mu_{max})$;
- 10 Check the convergence conditions:
- 11 **if** $\|\mathbf{Z}^t - \mathbf{X}^t\|_F < \varepsilon$ **and** $\|\mathbf{J}^t - \mathbf{D}_h^t\|_F < \varepsilon$ **and** $\|\mathbf{K}^t - \mathbf{D}_m^t\|_F < \varepsilon$ **then**
- 12 | Stop iteration;
- 13 **else**
- 14 | $t \leftarrow t + 1$;
- 15 **end**
- 16 **end**

\mathbf{J} , and \mathbf{K} , to replace the to-be-estimated variables \mathbf{X} , \mathbf{D}_h , and \mathbf{D}_m , respectively, we then have

$$\begin{aligned} \min_{\mathcal{S}} \frac{1}{2} \|\mathbf{H}_{in} - \mathbf{D}_h \mathbf{X}\|_F^2 + \frac{\alpha}{2} \|\mathbf{M}_{in} - \mathbf{D}_m \mathbf{X}\|_F^2 + \beta \|\mathbf{Z}\|_{1,1} \\ + \gamma (\|\mathbf{J}\|_* + \|\mathbf{K}\|_*) + l_R^+(\mathbf{J}) + l_R^+(\mathbf{K}), \\ \text{s.t. } \mathbf{D}_h = \mathbf{J}, \quad \mathbf{D}_m = \mathbf{K}, \quad \mathbf{X} = \mathbf{Z}, \quad \mathbf{1}^\top \mathbf{X} = \mathbf{1}, \end{aligned} \quad (3)$$

where the variable set $\mathcal{S} = \{\mathbf{X}, \mathbf{D}_h, \mathbf{D}_m, \mathbf{Z}, \mathbf{J}, \mathbf{K}\}$, and the symbol $()^+$ is an element-wise positive operator that truncates the non-negative part of the vector or matrix, i.e., $l_R^+(\bullet)$ means $\bullet \succeq \mathbf{0}$. Further, the augmented Lagrangian function of problem (3) \mathcal{L} can be equivalently written in the form of

$$\begin{aligned} \mathcal{L}(\mathcal{S}, \{\Lambda_i\}_{i=1}^3) = \frac{1}{2} \|\mathbf{H}_{in} - \mathbf{D}_h \mathbf{X}\|_F^2 + \frac{\alpha}{2} \|\mathbf{M}_{in} - \mathbf{D}_m \mathbf{X}\|_F^2 \\ + \beta \|\mathbf{Z}\|_{1,1} + \gamma (\|\mathbf{J}\|_* + \|\mathbf{K}\|_*) + l_R^+(\mathbf{J}) + l_R^+(\mathbf{K}) \\ + \Lambda_1^\top (\mathbf{Z} - \mathbf{X}) + \Lambda_2^\top (\mathbf{J} - \mathbf{D}_h) + \Lambda_3^\top (\mathbf{K} - \mathbf{D}_m) \\ + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{X}\|_F^2 + \frac{\mu}{2} \|\mathbf{J} - \mathbf{D}_h\|_F^2 + \frac{\mu}{2} \|\mathbf{K} - \mathbf{D}_m\|_F^2, \\ \text{s.t. } \mathbf{1}^\top \mathbf{X} = \mathbf{1}, \end{aligned} \quad (4)$$

where $\{\Lambda_i\}_{i=1}^3$ and μ are the Lagrange multipliers and the positive penalty parameter, respectively. These variables (\mathcal{S} and $\{\Lambda_i\}_{i=1}^3$) in \mathcal{L} can be successively minimized as follows:

Optimization with respect to \mathbf{X} : We solve the following constrained optimization problem for the variable \mathbf{X} :

$$\begin{aligned} \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{H}_{in} - \mathbf{D}_h \mathbf{X}\|_F^2 + \frac{\alpha}{2} \|\mathbf{M}_{in} - \mathbf{D}_m \mathbf{X}\|_F^2 \\ + \Lambda_1^\top (\mathbf{Z} - \mathbf{X}) + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{X}\|_F^2, \quad \text{s.t. } \mathbf{1}^\top \mathbf{X} = \mathbf{1}. \end{aligned} \quad (5)$$

The closed-form solution of problem (5) is given by

$$\mathbf{X} \leftarrow \mathbf{A}^{-1} \mathbf{B} - \mathbf{C} (\mathbf{I}_{1 \times L} \mathbf{A}^{-1} \mathbf{B} - \mathbf{1}), \quad (6)$$

Algorithm 2: J-SLoL solver: *S-Step*

Input: $\mathbf{M}_{out}, \mathbf{D}_h, \mathbf{D}_m$, parameters η , and \maxIter .
Output: \mathbf{Y} .

- 1 **Initialization:** $\mathbf{O}^0 = \mathbf{0}, \Delta^0 = \mathbf{0}, \rho^0 = 10^{-3}, \rho_{max} = 10^6, \xi = 1.5, \varepsilon = 10^{-6}, t = 1$.
- 2 **while not converged** or $t > \maxIter$ **do**
- 3 Update the variable \mathbf{Y}^t by Eq. (19);
- 4 Update the variable \mathbf{O}^t by Eq. (20);
- 5 Update Lagrange multiplier Δ^t by Eq. (21);
- 6 Update the parameter ρ^t by $\min(\xi\rho^{t-1}, \rho_{max})$;
- 7 Check the convergence conditions:
- 8 **if** $\|\mathbf{O}^t - \mathbf{Y}^t\|_F < \varepsilon$ **then**
- 9 | Stop iteration;
- 10 **else**
- 11 | $t \leftarrow t + 1$;
- 12 **end**
- 13 **end**

where

$$\begin{aligned} \mathbf{A} &= \mathbf{D}_h^\top \mathbf{D}_h + \alpha \mathbf{D}_m^\top \mathbf{D}_m + \mu \mathbf{I}_{L \times L}, \\ \mathbf{B} &= \mathbf{D}_h^\top \mathbf{H}_{in} + \alpha \mathbf{D}_m^\top \mathbf{M}_{in} + \mu \mathbf{Z} + \Lambda_1, \\ \mathbf{C} &= \mathbf{A}^{-1} \mathbf{I}_{L \times 1} (\mathbf{I}_{1 \times L} \mathbf{A}^{-1} \mathbf{I}_{L \times 1})^{-1}. \end{aligned}$$

Optimization with respect to \mathbf{D}_h : The optimization problem of \mathbf{D}_h can be expressed by

$$\min_{\mathbf{D}_h} \frac{1}{2} \|\mathbf{H}_{in} - \mathbf{D}_h \mathbf{X}\|_F^2 + \Lambda_2^\top (\mathbf{J} - \mathbf{D}_h) + \frac{\mu}{2} \|\mathbf{J} - \mathbf{D}_h\|_F^2, \quad (7)$$

which has the following analytical resolution

$$\mathbf{D}_h \leftarrow (\mathbf{H}_{in} \mathbf{X}^\top + \mu \mathbf{J} + \Lambda_2) \times (\mathbf{X} \mathbf{X}^\top + \mu \mathbf{I}_{L \times L})^{-1}. \quad (8)$$

Optimization with respect to \mathbf{D}_m : Similarly to problem (7), the objective function for the variable \mathbf{D}_m is

$$\min_{\mathbf{D}_m} \frac{\alpha}{2} \|\mathbf{M}_{in} - \mathbf{D}_m \mathbf{X}\|_F^2 + \Lambda_3^\top (\mathbf{K} - \mathbf{D}_m) + \frac{\mu}{2} \|\mathbf{K} - \mathbf{D}_m\|_F^2, \quad (9)$$

whose solution can be directly derived by

$$\mathbf{D}_m \leftarrow (\alpha \mathbf{M}_{in} \mathbf{X}^\top + \mu \mathbf{K} + \Lambda_3) \times (\alpha \mathbf{X} \mathbf{X}^\top + \mu \mathbf{I}_{L \times L})^{-1}. \quad (10)$$

Optimization with respect to \mathbf{J} and \mathbf{K} : The low-rank problem can be effectively solved by the *Singular Value Thresholding (SVT)* used in [43], [44]. The proximal operator can be generalized to the following steps.

- Step 1. Given a variable \mathbf{G} with a r -rank, we first perform singular value decomposition (SVD) on \mathbf{G} :

$$\text{SVD}(\mathbf{G}) := \mathbf{U} \mathbf{S} \mathbf{V}, \quad \mathbf{S} = \text{diag}(\{s_k\}_{1 \leq k \leq r}). \quad (11)$$

- Step 2. For any $\tau \geq 0$, a soft-thresholding operator, denoted as \mathcal{D}_τ , is further adopted as follows:

$$\mathcal{D}(\mathbf{G}) := \mathbf{U} \mathcal{D}_\tau(\mathbf{S}) \mathbf{V}, \quad \mathcal{D}_\tau(\mathbf{S}) = \text{diag}(\{s_k - \tau\}^+). \quad (12)$$

- Step 3. The nuclear norm of \mathbf{G} , namely $\|\mathbf{G}\|_*$, can be obtained by $\|\mathcal{D}_\tau(\mathbf{S})\|_{1,1}$.

Thereby, the update rule of \mathbf{J} and \mathbf{K} are given by

$$\begin{aligned} \mathbf{J} &\leftarrow \max\{\mathbf{0}, \mathcal{D}_\beta(\mathbf{D}_h - \Lambda_2/\mu)\}, \\ \mathbf{K} &\leftarrow \max\{\mathbf{0}, \mathcal{D}_\beta(\mathbf{D}_m - \Lambda_3/\mu)\}, \end{aligned} \quad (13)$$

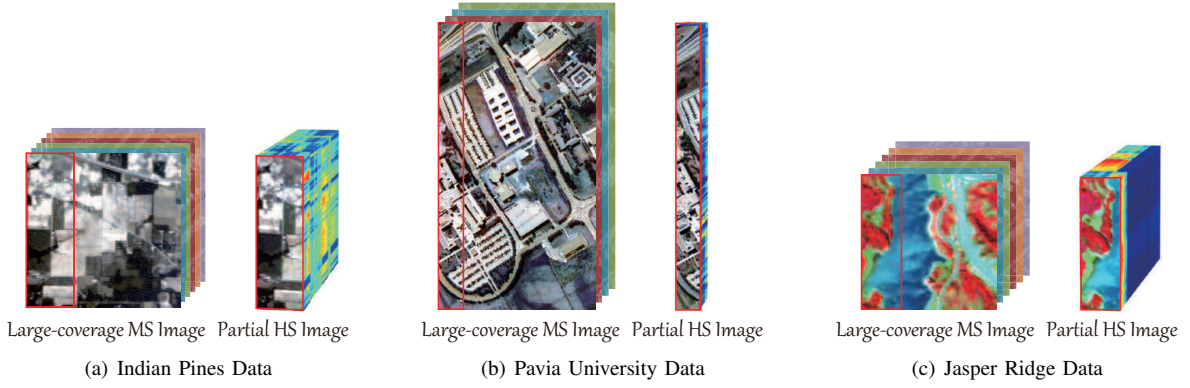


Fig. 3. Three used HS-MS datasets for performance evaluation of SSR products in terms of reconstruction, classification, and unmixing.

where $\mathcal{D}_\beta(\bullet)$ is defined as the general *SVT* operator, i.e., $\mathbf{U}\mathbf{S}_r\mathbf{V}$, where $\mathbf{S}_r = \text{diag}(\max\{0, s_k - \beta/\mu\})$.

Optimization with respect to \mathbf{Z} : The ℓ_1 -norm of \mathbf{Z} can be optimized by solving the following problem

$$\min_{\mathbf{Z}} \beta \|\mathbf{Z}\|_{1,1} + \Lambda_1 (\mathbf{Z} - \mathbf{X}) + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{X}\|_{\mathbb{F}}^2, \quad (14)$$

which can be well solved using a well-known *soft threshold* operator [45], [46], that is,

$$\mathbf{Z} \leftarrow \text{sign}(\mathbf{X} - \Lambda_1/\mu) \odot \max\{0, |\mathbf{X} - \Lambda_1/\mu| - \alpha/\mu\}, \quad (15)$$

where \odot denotes the element-wise Schur-Hadamard product.

Optimization with respect to $\{\Lambda_i\}_{i=1}^3$: The Lagrange multipliers can be updated by

$$\begin{aligned} \Lambda_1 &\leftarrow \Lambda_1 + \mu(\mathbf{Z} - \mathbf{X}), \\ \Lambda_2 &\leftarrow \Lambda_2 + \mu(\mathbf{J} - \mathbf{D}_h), \\ \Lambda_3 &\leftarrow \Lambda_3 + \mu(\mathbf{K} - \mathbf{D}_m). \end{aligned} \quad (16)$$

More specific optimization procedures for *D-Step* as shown in Eq. (3) are detailed in **Algorithm 1**.

2) *S-Step Solver*: The optimization problem is nothing but a classic sparse coding, which has been well solved by many excellent works [45], [47]–[49]. The ADMM-based optimizer has been proven to be effective for a fast and accurate solution. Similarly, we introduce an additional auxiliary variable \mathbf{O} into the problem (2) to replace the variable \mathbf{Y} in the term of ℓ_1 -norm, yielding the following augmented Lagrangian function

$$\begin{aligned} \mathcal{L}(\mathbf{Y}, \mathbf{O}, \Delta) &= \frac{1}{2} \|\mathbf{M}_{out} - \mathbf{D}_m \mathbf{Y}\|_{\mathbb{F}}^2 + \eta \|\mathbf{O}\|_{1,1} \\ &+ \Delta^\top (\mathbf{O} - \mathbf{Y}) + \frac{\rho}{2} \|\mathbf{O} - \mathbf{Y}\|_{\mathbb{F}}^2, \quad \text{s.t. } \mathbf{1}^\top \mathbf{Y} = 1, \end{aligned} \quad (17)$$

where Δ and ρ denote the Lagrange multiplier and the Lagrange penalty parameter, respectively.

To solve the problem (17), we then have

Optimization with respect to \mathbf{Y} : The sparse coefficients can be estimated by solving the constrained least squares regression as follows:

$$\begin{aligned} \min_{\mathbf{Y}} & \frac{1}{2} \|\mathbf{M}_{out} - \mathbf{D}_m \mathbf{Y}\|_{\mathbb{F}}^2 + \Delta^\top (\mathbf{O} - \mathbf{Y}) + \frac{\rho}{2} \|\mathbf{O} - \mathbf{Y}\|_{\mathbb{F}}^2, \\ \text{s.t. } & \mathbf{1}^\top \mathbf{Y} = 1. \end{aligned} \quad (18)$$

Similarly to the problem (5), we have the closed-form solution of the problem (18):

$$\mathbf{Y} \leftarrow \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{B}} - \tilde{\mathbf{C}} (\mathbf{I}_{1 \times L} \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{B}} - 1), \quad (19)$$

where

$$\begin{aligned} \tilde{\mathbf{A}} &= \mathbf{D}_m^\top \mathbf{D}_m + \rho \mathbf{I}_{L \times L}, \\ \tilde{\mathbf{B}} &= \mathbf{D}_m^\top \mathbf{M}_{out} + \rho \mathbf{O} + \Delta, \\ \tilde{\mathbf{C}} &= \tilde{\mathbf{A}}^{-1} \mathbf{I}_{L \times 1} (\mathbf{I}_{1 \times L} \tilde{\mathbf{A}}^{-1} \mathbf{I}_{L \times 1})^{-1}. \end{aligned}$$

Optimization with respect to \mathbf{O} : Likewise, the variable \mathbf{O} can be updated by following the same rule as Eq. (15)

$$\mathbf{O} \leftarrow \text{sign}(\mathbf{Y} - \Delta/\rho) \odot \max\{0, |\mathbf{Y} - \Delta/\rho| - \eta/\rho\}. \quad (20)$$

Optimization with respect to Δ : By employing the same rule in Eq. (16), the Lagrange multiplier in each step is written as

$$\Delta \leftarrow \Delta + \rho(\mathbf{O} - \mathbf{Y}). \quad (21)$$

Algorithm 2 gives more optimization details for the *S-Step*.

D. Convergence Analysis

The solution of the optimization problems both (3) and (17) can be well obtained by using a ADMM solver. Actually, the used ADMM in this paper can be generalized to *inexact Augmented Lagrange Multiplier (ALM)* [50], whose convergence has been theoretically guaranteed as long as the number of block is less than three. Although the multi-block ADMM optimization, e.g., our problem (3), is still lack of a *strictly mathematical proof*, yet its convergence has been well proven and maintained by tons of practical cases, such as [51]–[54].

III. EXPERIMENTS

We evaluate the quality of SSR product of MS images from three different perspectives.

- **Reconstruction.** We directly measure the differences between the real HS image and the SSR product.
- **Classification.** The goal of SSR is to generate high-quality HS products for the subsequent high-level applications, e.g., classification. Classification can thus be seen as an effective tool to verify spectrally physical properties of reconstructed HS images (or SSR products).
- **Unmixing.** Material mixture is also an unique to the HS image. As a result, the evaluation of SSR products

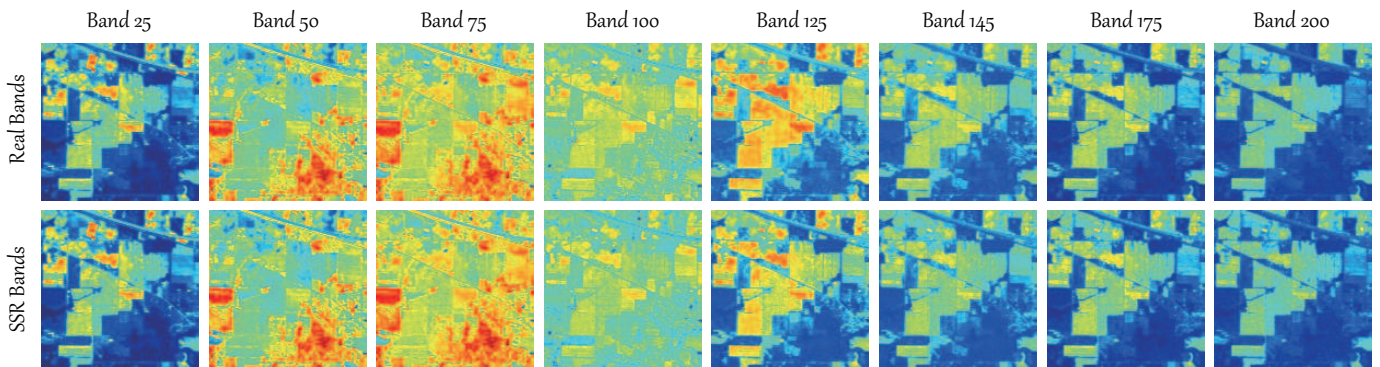


Fig. 4. Selective band visualization of the real HS image and the SSR product using the proposed J-SLoL model on the Indian Pines data.

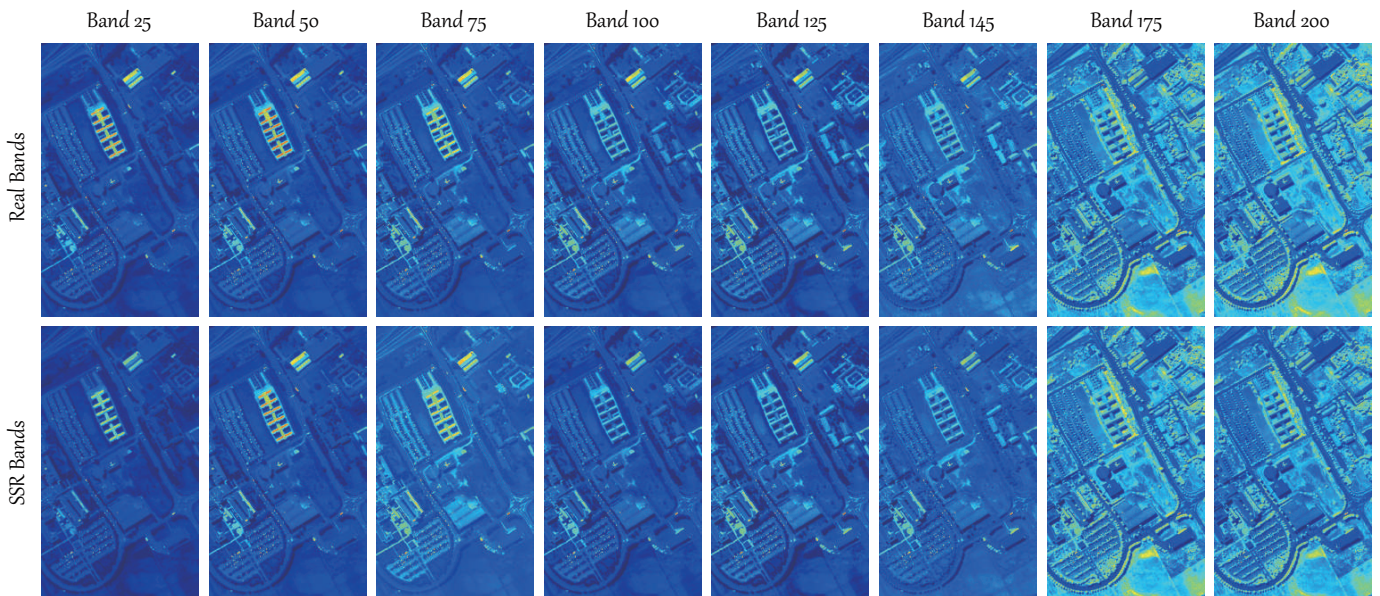


Fig. 5. Selective band visualization of the real HS image and the SSR product using the proposed J-SLoL model on the Pavia University data.

TABLE I

RECONSTRUCTION PERFORMANCE COMPARISON OF DIFFERENT METHODS IN TERMS OF FIVE INDICES ON THE HS-MS INDIAN PINES DATASET. THE BEST RESULTS ARE SHOWN IN BOLD.

Methods	RMSE	PSNR	SAD	SSIM	ERGAS
PwC	0.0057	38.7302	0.0125	0.7932	1.0092
CRISP [40]	0.0039	39.3170	0.0092	0.7782	0.8820
Sun's [41]	0.0039	41.4753	0.0093	0.7999	0.7110
Yokoya's [42]	0.0036	42.2450	0.0087	0.8201	0.6546
Arad's [35]	0.0035	42.4517	0.0084	0.8344	0.6405
J-SLoL	0.0032	43.1547	0.0080	0.8688	0.6044
Ideal Value	↓ 0	↑ ∞	↓ 0	↑ 1	↓ 0

can, to a great extent, be determined by the unmixing performance.

Moreover, three HS-MS datasets are used for the SSR task, where the first two are applied for the performance comparison in terms of reconstruction and classification, and the last one is used for unmixing evaluation.

A. Dataset Description

1) *Indian Pines Data*: This widely-used HS image was collected by the optical sensor – Airborne Visible / Infrared

TABLE II

RECONSTRUCTION PERFORMANCE COMPARISON OF DIFFERENT METHODS IN TERMS OF FIVE INDICES ON THE HS-MS PAVIA UNIVERSITY DATASET. THE BEST RESULTS ARE SHOWN IN BOLD.

Methods	RMSE	PSNR	SAD	SSIM	ERGAS
PwC	0.0133	38.1921	0.0487	0.8804	3.4181
CRISP [40]	0.0074	45.8438	0.0367	0.9608	1.9097
Sun's [41]	0.0070	46.2677	0.0355	0.9584	1.8140
Yokoya's [42]	0.0068	46.4856	0.0351	0.9619	1.7705
Arad's [35]	0.0069	46.4749	0.0350	0.9674	1.7782
J-SLoL	0.0060	47.4549	0.0328	0.9618	1.6403
Ideal Value	↓ 0	↑ ∞	↓ 0	↑ 1	↓ 0

Imaging Spectrometer (AVIRIS) – over the Indiana state, USA. It consists of 145×145 pixels with 220 spectral channels. To meet the requirement of our SSR's problem setting, the corresponding MS image is simulated by using the spectral response functions (SRFs) of Sentinel-2 and a sub-image is selected with the size of 145×45 as the partially overlapped region of HS and MS images, as shown in Fig. 3(a). Furthermore, there are 16 to-be-investigated categories in the studied scene, and they will be used for the part of classification evaluation. More details about training and test sets can be found in [55].

TABLE III

CLASSIFICATION PERFORMANCE COMPARISON OF DIFFERENT METHODS IN TERMS OF OA (%), AA (%), AND κ (%) AS WELL AS THE ACCURACY FOR EACH CLASS ON THE INDIAN PINES DATASET. R-HS MEANS THE REAL HS IMAGE. THE BEST RESULTS ARE SHOWN IN BOLD.

	PwC	CRISP	Sun's	Yokoya's	Arad's	J-SLoL	R-HS
OA	52.92	63.74	60.77	61.60	64.67	65.14	65.89
AA	65.34	73.04	71.15	71.90	76.27	73.37	75.71
κ	47.31	59.00	55.69	56.62	60.10	60.53	61.48
C1	35.55	44.94	43.28	46.10	43.64	48.55	51.66
C2	42.47	48.34	43.75	43.88	50.38	48.60	57.40
C3	65.76	63.59	61.96	61.96	69.57	67.39	70.65
C4	68.46	88.14	74.72	74.94	86.80	86.13	88.14
C5	80.63	81.78	82.93	83.21	83.93	83.36	81.78
C6	78.82	93.62	90.89	93.62	92.71	93.39	95.90
C7	48.69	71.24	66.45	67.10	78.65	73.42	66.56
C8	43.26	53.14	51.08	50.74	54.76	56.12	55.21
C9	44.5	48.94	42.73	44.15	50.53	46.10	53.01
C10	94.44	97.53	97.53	97.53	96.30	97.53	98.15
C11	68.17	83.28	82.23	83.68	78.94	83.92	82.88
C12	39.70	52.42	47.27	47.27	55.76	50.30	50.91
C13	95.56	97.78	100	100	97.78	97.78	97.78
C14	66.67	82.05	71.79	74.36	89.74	79.49	79.49
C15	72.73	81.82	81.82	81.82	90.91	81.82	81.82
C16	100	80.00	100	100	100	80.00	100

2) *Pavia University Data*: The second HS data was captured by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor covering the Pavia University, Pavia, Italy, which was used for IEEE GRSS data fusion contest 2008 [56]. The image comprises 103 spectral bands covering the wavelength range from 430nm to 860nm. Also, this scene consists of 610×340 pixels at a 1.3m GSD, including 9 classes used for the land cover classification. To make a relatively fair comparison, we adopt a set of fixed training and test samples widely used in many researches [57]. Similarly, the simulated MS image is generated by using the SRFs of QuickBird, yielding the size of 610 × 340 × 4, 610 × 50 of which are selected as the HS-MS overlapped region (see Fig. 3(b)).

3) *Jasper Ridge Data*: The HS scene was acquired using the AVIRIS instrument over a rural area at Jasper Ridge, California, USA. A widely-used region of interest (ROI) with 100 × 100 pixels at a GSD of 20m and 198 spectral bands in the range of 380nm to 2500nm is used in our experiments. A Sentinel-2 MS product is simulated using SRFs on the HS image, and there is the size of 100 × 30 pixels in the overlapped part between HS and MS images, as shown in Fig. 3(c). Four main materials, such as #1 Tree, #2 Water, #3 Soil, and #4 Road, are involved in this scene with ground truth of abundance maps given from the website¹.

B. Reconstruction-based Evaluation

Several important indices from the perspective of reconstruction, e.g., root mean square error (RMSE), peak signal to noise ratio (PSNR), spectral angle distance (SAD), structural similarity index (SSIM), and erreur relative globale adimensionnelle de synthèse (ERGAS), are employed to quantitatively evaluate the performance of SSR of MS images. In addition, we select four state-of-the-art baselines related to the SSR task, including pixel-wise copy (PwC)², color resolution improve-

¹<https://rslab.ut.ac.ir/data>

²We directly copy the HS pixel from the overlapped region into the unknown HS pixel, according to the similarity measurement in the MS image.

TABLE IV

CLASSIFICATION PERFORMANCE COMPARISON OF DIFFERENT METHODS IN TERMS OF OA (%), AA (%), AND κ (%) AS WELL AS THE ACCURACY FOR EACH CLASS ON THE PAVIA UNIVERSITY DATASET. R-HS MEANS THE REAL HS IMAGE. THE BEST RESULTS ARE SHOWN IN BOLD.

	PwC	CRISP	Sun's	Yokoya's	Arad's	J-SLoL	R-HS
OA	66.29	70.86	70.68	70.50	70.39	71.15	71.85
AA	76.09	80.41	80.31	80.35	80.25	80.48	81.15
κ	58.53	63.85	63.60	63.43	63.30	64.16	65.01
C1	71.56	73.88	73.56	73.77	73.91	73.64	73.17
C2	53.91	59.66	59.62	59.00	58.79	60.42	61.32
C3	55.45	56.88	56.88	56.69	56.41	56.74	60.17
C4	96.64	97.39	97.42	97.45	97.42	97.29	97.39
C5	98.88	99.18	99.03	99.11	99.11	99.11	99.26
C6	65.52	71.11	69.78	70.25	70.17	71.13	73.35
C7	71.05	83.53	84.51	84.59	84.36	83.98	84.96
C8	81.75	86.37	86.61	86.58	86.50	86.42	85.36
C9	90.07	95.67	95.35	95.67	95.56	95.56	95.35

ment software package (CRISP) [40], Sun's [41], Yokoya's [42], and Arad's [35], in comparison with our J-SLoL model.

Table I lists the quantitative results of the aforementioned compared algorithms in five indices on the Indian Pines data. Overall, the PwC approach yields the poor reconstruction performance in nearly all indices, except the SSIM value that is slightly higher than CRISP's. By making use of grouping strategy, the spectrally-enhanced performance of Sun's algorithm is further improved compared to the original CRISP, specifically in PSNR and ERGAS. Inspired by the current success and good theoretical support in sparse representation, Yokoya's and Arad's methods show great potential in the SSR task, yielding a moderate improvement in all measures. **Please note that the main difference between Yokoya's and Arad's methods lies in the MS dictionary construction. The former directly takes the overlapped MS data as the dictionary, while the latter generates the MS dictionary by performing the linear interpolation on HS data. As a result, the MS dictionary obtained by Arad's method might be more correlated with the HS's, yielding relatively higher reconstruction results compared to Yokoya's.** Remarkably, the proposed J-SLoL model performs better than others at a comprehensive increase of around 0.0004 RMSE, 1 PSNR, 0.0007 SAD, 0.04 SSIM, and 0.05 ERGAS, compared to the second best approach. Furthermore, Fig. 4 visualizes several selected bands, where there is a very small visual difference between the real HS image and reconstructed one obtained by J-SLoL. This, to a great extent, demonstrates the effectiveness of the proposed method.

Similarly, there is a basically identical trend between the Indian Pines data and the Pavia University data, in either quantitative (see Table II) or qualitative results (see Fig. 5). The main difference lies in that the second datasets are more challenging due to the larger image size and less spectral bands, leading to the limitations in reconstructing detailed information, e.g., texture. This can be well explained by different indices. Compared to those in the Indian Pines data, RMSE, SAD, and ERGAS are relatively low while PSNR and SSIM reflecting the image structural information are higher.

Parameter Sensitivity Analysis. As the quality of the SSR product is closely associated with the parameter setting in our J-SLoL model, i.e., α , β , γ in *D-Step* and η in *S-Step*, hence the performance gain in terms of RMSE is investigated by

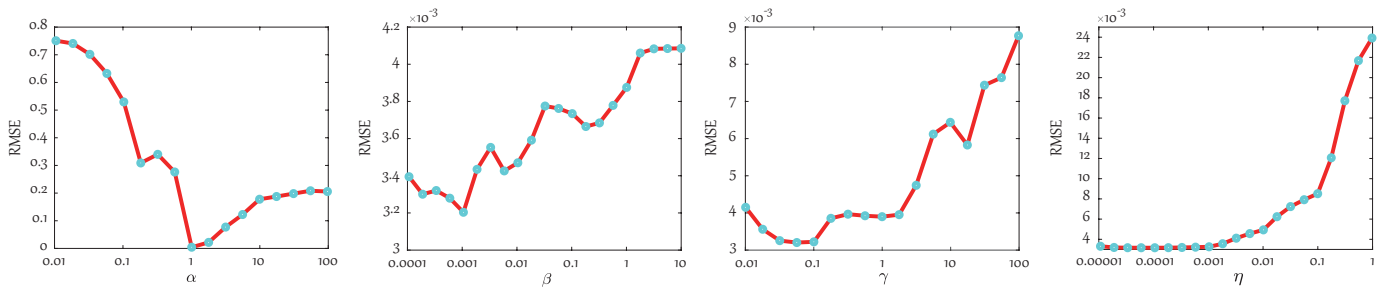


Fig. 6. Parameter sensitivity analysis of the proposed J-SLoL model in terms of four regularization parameters, e.g., α , β , γ in Eq. (1), and η in Eq. (2) on the Indian Pines data.

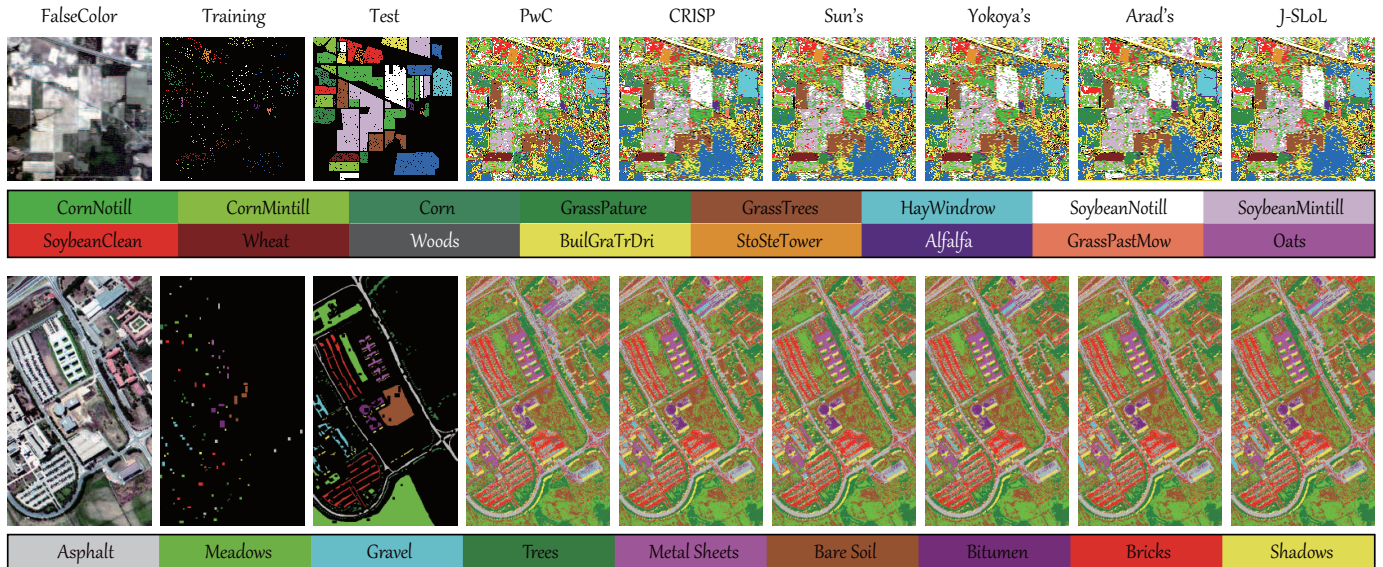


Fig. 7. Visualization of false-color images, training and test sample distribution, and classification maps by using different SSR algorithms on the two datasets: Indian Pines (top) and Pavia University (bottom).

changing these parameters in a proper range on the Indian Pines data. We can see from the Fig. 6 that the parameter α plays a dominant role in dictionary learning, while other parameters in D -Step and S -Step have also important effects on the whole SSR process. In detail, the optimal parameter combination ($\alpha, \beta, \gamma, \eta$) can be given as (1, 0.001, 0.1, 0.0001) as shown in Fig. 6. We also found that this set of parameter setting is relatively stable, we therefore apply them in the rest of datasets for simplicity.

C. Classification-based Evaluation

Classification, as a potential high-level application, has been proven to be effective for model performance assessment, where three common indices: overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) are used in our experiments. Note that a simple nearest neighbor (NN) classifier is applied for the classification task. This is because if more advanced classifiers are used, we might confuse the additional performance gain from these classifiers or our SSR products.

Tables III and IV list the quantitative comparison between the real HS image and different SSR algorithms in terms of OA, AA, and κ as well as the accuracy for each class on the two same datasets, and Fig. 7 shows the corresponding

classification maps. Due to only pixel-based copy operation, the PwC method fails to reconstruct high-quality HS data well, yielding poor classification performance on both datasets. Conversely, the J-SLoL model as expected outperforms other competitors, despite only a slight improvement in classification accuracies. It should be noted, however, that the results of our J-SLoL method are very close to those using the real HS image. This might directly demonstrate the superiority of the proposed strategy. Furthermore, there is also a similar trend in the visual comparison in terms of classification maps (*cf.* Figs. 7). Intuitively, the classification maps of the proposed J-SLoL model are more similar to those obtained by the real HS image from either structural information or detailed textures. Moreover, our approach is capable of making the classification maps relatively smooth in certain classes, such as *HayWindrow* and *SoybeanNotill* in the first data, and *Soil* in the second one.

D. Unmixing-based Evaluation

Due to the low spatial resolution, multiple materials are highly mixed within one pixel in the HS image. As a result, spectral unmixing can be regarded as a feasible solution for quality assessment of SSR of MS images. More specifically, the fully constrained least squares unmixing (FCLSU) [58]

TABLE V
RECONSTRUCTION AND UNMIXING PERFORMANCE COMPARISON OF DIFFERENT METHODS ON THE JASPER RIDGE DATASET. R-HS MEANS THE REAL HS IMAGE. THE BEST RESULTS ARE SHOWN IN BOLD.

Model	Reconstruction					Unmixing		
	RMSE	PSNR	SAD	SSIM	ERGAS	aRMSE	rRMSE	aSAM
PwC	0.0370	28.1639	0.0772	0.7508	8.2955	0.1802 ± 0.0147	0.0234 ± 0.0069	0.1224 ± 0.0233
CRISP	0.0320	35.6180	0.0764	0.9304	6.8960	0.1789 ± 0.0118	0.0222 ± 0.0070	0.1239 ± 0.0279
Sun's	0.0298	36.1892	0.0598	0.9204	6.7216	0.1767 ± 0.0139	0.0238 ± 0.0068	0.1178 ± 0.0243
Yokoya's	0.0292	36.3338	0.0590	0.9219	6.5876	0.1764 ± 0.0133	0.0235 ± 0.0067	0.1171 ± 0.0243
Arad's	0.0282	36.5349	0.0578	0.9320	6.3419	0.1763 ± 0.0132	0.0232 ± 0.0067	0.1162 ± 0.0243
J-SLoL	0.0271	36.7630	0.0565	0.9311	6.0605	0.1762 ± 0.0135	0.0229 ± 0.0068	0.1152 ± 0.0244
R-HS	0	∞	0	1	0	0.1760 ± 0.0126	0.0206 ± 0.0061	0.1175 ± 0.0215

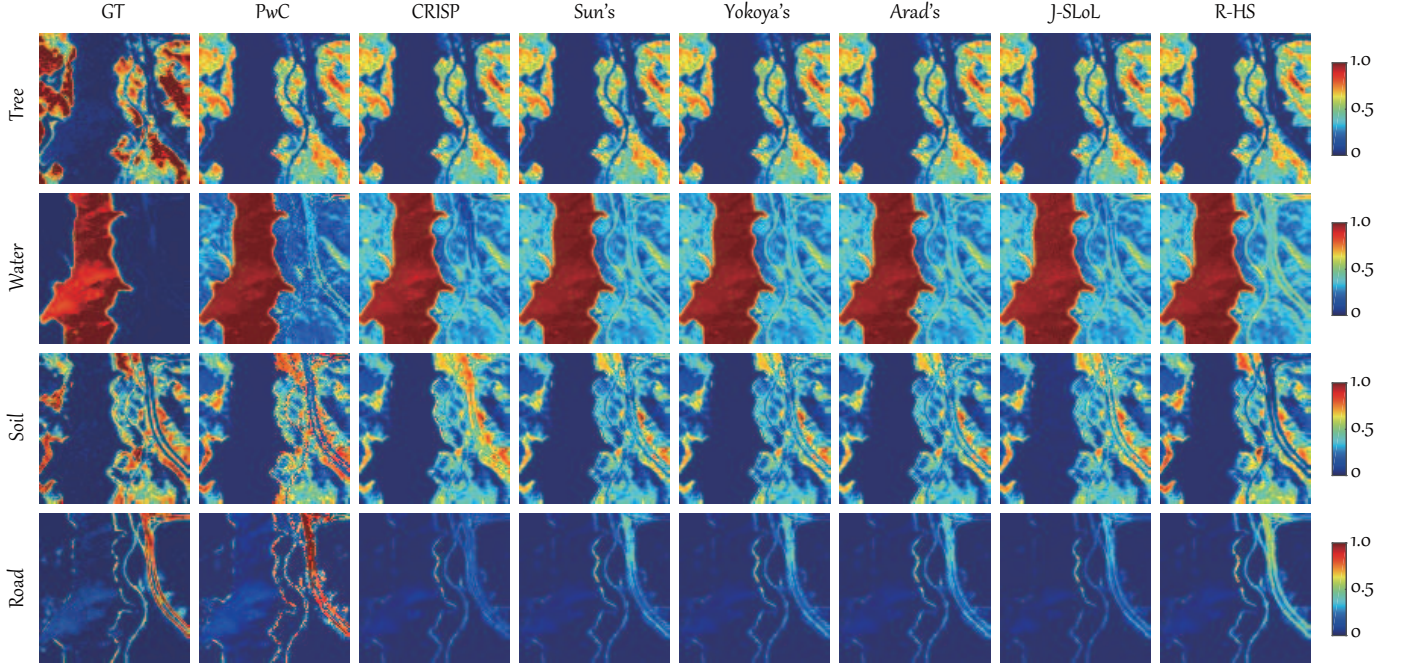


Fig. 8. Abundance maps of four materials using different SSR algorithms and the real HS image.

algorithm with three popular criteria [14], including abundance overall root mean square error (aRMSE), reconstruction overall root mean square error (rRMSE), and average spectral angle mapper (aSAM), is used to quantify the unmixing performance in the following experiments.

Reconstruction and unmixing are successively conducted to quantitatively evaluate the algorithm performance, as listed in Table V. Roughly, there is a consistent trend in performance gain from PwC to Arad's in the Jasper Ridge data. The PwC holds relatively bad performance in both reconstruction and unmixing. For the CRISP approach, it brings a dramatic improvement on the basis of the PwC, while its modified model, i.e., Sun's, obtains better results. Inspired by the sparsity-promoting assumption, Yokoya's algorithm achieves a competitive performance. Similarly, Arad's method performs slightly better than Yokoya's, possibly owing to the use of high-quality spectral dictionary. Not unexpectedly, our proposed J-SLoL observably exceeds other compared methods, particularly in several important indices, such as RMSE and SSIM in the reconstruction task, and aRMSE in the unmixing task. Additionally, a direct proof is given by the fact that

the unmixing results using the SSR product from J-SLoL are comparable to those using the real HS image under all three measures, showing the great potential of the proposed method.

We also make a visual comparison in terms of abundance maps, as illustrated in Fig. 8. From the figure, we can see that the abundance maps estimated by PwC and CRISP are more different from those of R-HS, particularly the materials of *Water*, *Soil*, and *Road*. By contrary, the later three methods have better visual effect, in which our J-SLoL performs more similar abundance maps, e.g., *Water* and *Soil*. Despite a big difference between J-SLoL and GT, this might result from the limitations of the FCLSU algorithm itself. In summary, both visual and numerical unmixing evaluation can also demonstrate that the J-SLoL is well applicable to the SSR task.

E. Comparison of Computational Cost

The computational cost of our J-SLoL model in Eqs. (1) and (2) is dominated by matrix products. More specifically, the update of \mathbf{X} , \mathbf{D}_h , and \mathbf{D}_m in *D-Step* consists of matrix multiplications and matrix inversions, yielding complexity with $\mathcal{O}(L^3 + L^2N + LNP)$, $\mathcal{O}(L^3 + L^2N + LNP)$, and

TABLE VI
COMPARISON OF COMPUTATIONAL COST FOR DIFFERENT ALGORITHMS.

Methods	Computational Cost
PwC	$\mathcal{O}(N_1NP)$
CRISP	$\mathcal{O}(NPQ)$
Sun's	$\mathcal{O}(NPQ + KPQ^2)$
Yokoya's	$\mathcal{O}(L^3 + L^2N_1 + LN_1P)$
Arad's	$\mathcal{O}(L^3 + L^2N_1 + LN_1P)$
J-SLoL	$\mathcal{O}(L^3 + L^2N + LNP + LN_1P)$

$\mathcal{O}(L^3 + L^2N + LNQ)$, respectively, while updating \mathbf{J} and \mathbf{K} both require computing SVDs with the orders of cost as $\mathcal{O}(\min(P^2L, PL^2))$ and $\mathcal{O}(\min(Q^2L, QL^2))$. Therefore, optimizing the variables \mathbf{X} and \mathbf{D}_h are the most expensive computational cost steps in problem (4), yielding a dominant complexity $\mathcal{O}(L^3 + L^2N + LNP)$ with respect to **Algorithm 1**. For *S-Step*, the main per-iteration cost of **Algorithm 2** lies in the update of \mathbf{Y} , being similar to the update of \mathbf{X} , leading to a $\mathcal{O}(L^3 + L^2N_1 + LN_1P)$ complexity.

To demonstrate the efficiency and effectiveness of the proposed J-SLoL model, we make an approximated comparison in computational cost. As listed in Table VI, PwC holds a very high cost of $\mathcal{O}(N_1NP)$ due to its pixel-to-pixel matching operation. The complexity in CRISP lies in the estimation of the regression matrix between overlapped HS and MS images, yielding a $\mathcal{O}(NPQ)$ computational cost. Sun's method involves an additional spectral matching cost behind CRISP, yielding a complexity of $\mathcal{O}(KPQ^2)$, where K is a predefined number of materials. In fact, the Yokoya's and Arad's methods can be approximately seen as our *S-Step*, hence its computational cost is $\mathcal{O}(L^3 + L^2N_1 + LN_1P)$. Although our J-SLoL performs relatively higher than Yokoya's and Arad's methods, yet the overall computational cost is acceptable, since the number of HS-MS samples in the overlapped region, i.e., N , are limited.

IV. CONCLUSION

In this paper, a novel and promising topic – SSR – is introduced to enhance the spectral resolution of MS images, which has a great potential as a better alternative of the classic HS and MS fusion task (or HS-SR) with only need of partially HS data. Inspired by the effectiveness and recent success of sparse representation, we propose an effective MS enhancement algorithm, called J-SLoL, by jointly learning a low-rank dictionary pair from the overlapped HS and MS region and further inferring the unknown HS image by sharing the sparse coefficients estimated by using MS data. Beyond previous models, the proposed J-SLoL is capable of fully making use of the correspondences between HS and MS images to learn more completed HS and MS dictionaries, further yielding a more accurate HS recovery. We have to admit, however, that the linearized sparse technique remains limited in data representation and fitting, especially in large-scale and complex cases. In the future work, we would like to develop more advanced reconstruction and recovery strategies by the means of nonlinear techniques, e.g., deep learning, or by introducing the new data source, e.g., LiDAR, SAR, to further improve the model's generalization ability.

REFERENCES

- [1] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, J. Gao, and L. Zhang, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, pp. 111716, 2020.
- [2] L. Gao, D. Yao, Q. Li, L. Zhuang, B. Zhang, and J. Bioucas-Dias, "A new low-rank representation based hyperspectral image denoising method for mineral mapping," *Remote Sens.*, vol. 9, no. 11, pp. 1145, 2017.
- [3] Q. Yuan, H. Xu, T. Li, H. Shen, and L. Zhang, "Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental us," *J. Hydrol.*, vol. 580, pp. 124351, 2020.
- [4] L. Gao, B. Zhao, X. Jia, W. Liao, and B. Zhang, "Optimized kernel minimum noise fraction transformation for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 6, pp. 548, 2017.
- [5] D. Hong, N. Yokoya, and X. Zhu, "Learning a robust local manifold representation for hyperspectral dimensionality reduction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2960–2975, Jul 2017.
- [6] J. Yao, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Robust subspace clustering via penalized mixture of gaussians," *Neurocomputing*, vol. 278, pp. 4–11, 2018.
- [7] D. Hong, N. Yokoya, J. Xu, and X. Zhu, "Joint & progressive learning from high-dimensional data for multi-label classification," in *Proc. ECCV*, 2018, pp. 469–484.
- [8] G. Licciardi and J. Chanussot, "Spectral transformation based on nonlinear principal component analysis for dimensionality reduction of hyperspectral images," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 375–390, 2018.
- [9] X. Cao, J. Yao, X. Fu, H. Bi, and D. Hong, "An enhanced 3-dimensional discrete wavelet transform for hyperspectral image classification," *IEEE Geosci. and Remote Sens. Lett.*, 2020, 10.1109/LGRS.2020.2990407.
- [10] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep (overview and toolbox)," *IEEE Geosci. Remote Sens. Mag.*, 2020, DOI: 10.1109/MGRS.2020.2979764.
- [11] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, 2020, DOI:10.1109/TGRS.2020.2964627.
- [12] M. Tang, L. Gao, A. Marinoni, P. Gamba, and B. Zhang, "Integrating spatial information in the normalized p-linear algorithm for nonlinear hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1179–1190, 2017.
- [13] J. Yao, D. Meng, Q. Zhao, W. Cao, and Z. Xu, "Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2991–3006, 2019.
- [14] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. on Image Process.*, vol. 28, no. 4, pp. 1923–1938, 2019.
- [15] X. Liu, C. Deng, J. Chanussot, D. Hong, and B. Zhao, "Stfnnet: A two-stream convolutional neural network for spatiotemporal image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6552–6564, 2019.
- [16] 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, 2019.
- [17] J. Hu, D. Hong, and X. Zhu, "MIMA: Mapper-induced manifold alignment for semi-supervised fusion of optical image and polarimetric sar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9025–9040, 2019.
- [18] Y. Wang, Q. Yuan, T. Li, H. Shen, L. Zheng, and L. Zhang, "Large-scale modis aod products recovery: Spatial-temporal hybrid fusion considering aerosol variation mitigation," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 1–12, 2019.
- [19] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral computational imaging via collaborative tucker3 tensor decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, 2020.
- [20] X. Wu, D. Hong, J. Chanussot, Y. Xu, R. Tao, and Y. Wang, "Fourier-based rotation-invariant feature boosting: An efficient framework for geospatial object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 302–306, 2020.
- [21] Y. Wu, S. López, B. Zhang, F. Qiao, and L. Gao, "Approximate computing for onboard anomaly detection from hyperspectral images," *J. Real-Time Image Process.*, vol. 16, no. 1, pp. 99–114, 2019.

- [22] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, "ORSIm Detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5146–5158, 2019.
- [23] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, "Classification of hyperspectral and lidar data using coupled cnns," *IEEE Trans. Geosci. Remote Sens.*, 2020, DOI: 10.1109/TGRS.2020.2969024.
- [24] D. Hong, J. Chanussot, N. Yokoya, J. Kang, and X. Zhu, "Learning shared cross-modality representation using multispectral-lidar and hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, 2020, DOI:10.1109/LGRS.2019.2944599.
- [25] 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.
- [26] 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.
- [27] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. ICCV*, 2015, pp. 3586–3594.
- [28] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral images super-resolution via learning high-order coupled tensor ring representation," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.
- [29] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. ECCV*, 2014, pp. 63–78.
- [30] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015.
- [31] Q. Wei, N. Dobigeon, and J. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, 2015.
- [32] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, 2019.
- [33] 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. CVPR*, 2019, pp. 1585–1594.
- [34] D. Hong, W. Liu, J. Su, Z. Pan, and G. Wang, "A novel hierarchical approach for multispectral palmprint recognition," *Neurocomputing*, vol. 151, pp. 511–521, 2015.
- [35] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *Proc. ECCV*. Springer, 2016, pp. 19–34.
- [36] J. Aeschbacher, J. Wu, and R. Timofte, "In defense of shallow learned spectral reconstruction from rgb images," in *Proc. ICCV Workshops*, 2017, pp. 471–479.
- [37] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, and K. Schindler, "Learned spectral super-resolution," *arXiv preprint arXiv:1703.09470*, 2017.
- [38] Y. Can and R. Timofte, "An efficient cnn for spectral reconstruction from rgb images," *arXiv preprint arXiv:1804.04647*, 2018.
- [39] Naveed Akhtar and Ajmal Mian, "Hyperspectral recovery from rgb images using gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 100–113, 2018.
- [40] M. Winter, E. Winter, S. Beaven, and A. Ratkowski, "Hyperspectral image sharpening using multispectral data," in *IEEE Aerospace Conference*. IEEE, 2007, pp. 1–9.
- [41] X. Sun, L. Zhang, H. Yang, T. Wu, Y. Cen, and Y. Guo, "Enhancement of spectral resolution for remotely sensed multispectral image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2198–2211, 2014.
- [42] N. Yokoya, U. Heiden, and M. Bachmann, "Spectral enhancement of multispectral imagery using partially overlapped hyperspectral data and sparse signal representation," in *Proc. WHISPERS*. IEEE, 2018.
- [43] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, 2012.
- [44] D. Hong and X. Zhu, "SULoRA: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1351–1363, 2018.
- [45] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. WHISPERS*. IEEE, 2010, pp. 1–4.
- [46] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, 2019.
- [47] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, no. Jan, pp. 19–60, 2010.
- [48] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Found. Trends® Comput. Graph. Vis.*, vol. 8, no. 2-3, pp. 85–283, 2014.
- [49] J. Kang, D. Hong, J. Liu, G. Baier, N. Yokoya, and B. Demir, "Learning convolutional sparse coding on complex domain for interferometric phase restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020, DOI: 10.1109/TNNLS.2020.2979546.
- [50] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.
- [51] Y. Xu, W. Yin, Z. Wen, and Y. Zhang, "An alternating direction algorithm for matrix completion with nonnegative factors," *Front. Math. China*, vol. 7, no. 2, pp. 365–384, 2012.
- [52] Y. Zhang, Z. Jiang, and L. Davi, "Learning structured low-rank representations for image classification," in *Proc. CVPR*, 2013, pp. 676–683.
- [53] P. Zhou, C. Zhang, and Z. Lin, "Bilevel model based discriminative dictionary learning for recognition," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1173–1187, 2017.
- [54] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. Zhu, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 35–49, 2019.
- [55] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5384–5394, 2019.
- [56] G. Licciardi, F. Pacifici, D. Tuija, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, and J. Chanussot, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRSS data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, 2009.
- [57] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, 2020.
- [58] D. Heinz and C. Wang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, 2001.