

Structure-Prior-Constrained Low-Rank and Sparse Representation With Discriminative Incremental Dictionary for Hyperspectral Image Classification

Xiangyu Nie, Zhaohui Xue^{ID}, Member, IEEE, Cong Lin^{ID}, Ling Zhang, and Hongjun Su^{ID}, Senior Member, IEEE

Abstract—Low-rank and sparse representation (LRSR) model has gained popularity in hyperspectral image (HSI) classification. However, most existing LRSR models are limited by the highly nonlinear correlation of hyperspectral data, which leads to poor subspace segmentation performance. Furthermore, current LRSR methods usually directly used labeled samples to build the dictionary, whereas low discriminative labeled samples may degrade the representation ability of the dictionary. To solve the above issues, we propose a novel structure-prior-constrained LRSR with discriminative incremental dictionary (SPCLSR-DID) method for HSI classification. First, global and local data structures are maintained by low-rank and sparsity constraints, while a structural prior constraint is introduced to explore the intrinsic spectral-spatial structural information of HSI, improving the subspace segmentation ability of the model. Second, a discriminative incremental dictionary (DID) method is presented to find reliable and discriminative augmented atoms to improve the completeness and representation power of the dictionary. In DID, the incremental dictionary size is controllable to suit different tasks. Finally, the class label of each target sample is determined by jointly considering contextual information within a certain local range, which ensures the accuracy and smoothness of the classification map. Experimental results based on four popular hyperspectral datasets demonstrate that the proposed SPCLSR-DID method significantly outperforms other related comparison methods in terms of classification accuracy and generalization performance.

Index Terms—Classification, discriminative incremental dictionary (DID), hyperspectral image (HSI), low-rank and sparse representation (LRSR), structural prior constraint.

NOMENCLATURE

D denotes a matrix, **y** denotes a vector, and **b** denotes a scalar.

Manuscript received 28 October 2023; revised 22 December 2023; accepted 9 January 2024. Date of publication 12 January 2024; date of current version 5 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42271324, Grant 42201406, and Grant 41971279; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20221506; and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 22KJD420001. (*Corresponding author: Zhaohui Xue*.)

Xiangyu Nie, Zhaohui Xue, and Hongjun Su are with the School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China, also with the College of Geography and Remote Sensing, Hohai University, Nanjing 211100, China, and also with the Jiangsu Province Engineering Research Center of Water Resources and Environment Assessment Using Remote Sensing, Hohai University, Nanjing 211100, China (e-mail: zhaohui.xue@hhu.edu.cn).

Cong Lin is with the Nanjing Research Institute of Surveying, Mapping and Geotechnical Investigation Company Ltd., Nanjing 210019, China.

Ling Zhang is with the School of Naval Architecture and Ocean Engineering, Jiangsu Maritime Institute, Nanjing 211100, China.

Digital Object Identifier 10.1109/TGRS.2024.3353370

H	Hyperspectral dataset.
D	Dictionary.
D_c	Subdictionary of the <i>c</i> th class.
Y	Test set.
Ŷ	Denoised test set.
N_i	<i>i</i> th neighborhood sample set.
X	Coefficient matrix.
E	Noise matrix.
R	Reconstruction error matrix.
W	Structural prior matrix.
Ŵ	Constraint matrix of R .
D̄	Incremental dictionary.
D̄_c	Incremental subdictionary of the <i>c</i> th class.
y_i	<i>i</i> th target sample.
ŷ_i	<i>i</i> th denoised target sample.
d_i	<i>i</i> th dictionary atom.
d_{ci}	<i>i</i> th atom of D_c .
e_i	<i>i</i> th column vector of E .
C	Number of classes.
k	Number of pixels.
b	Number of bands.
m	Number of dictionary atoms.
m_c	Number of atoms in D_c .
n	Number of target samples.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) consists of hundreds of spectral features in narrow spectral bands, which can provide rich spectral information to accurately identify and distinguish various ground objects [1]. The classification task for HSI is to determine the class label of each pixel based on spectral and spatial information. As fundamental research, HSI classification has gained tremendous popularity in environmental monitoring, agricultural management, urban planning, mineral survey, and other fields [2].

Over the past few decades, advanced machine learning and pattern recognition methods have greatly facilitated the evolution of HSI classification techniques [3], [4], [5], [6], [7], [8], [9]. Recently, with the emerging paradigm of compressed sensing, research on HSI classification under representation methods has attracted great attention [10]. Typical prototypes of representation-based models for HSI classification include sparse representation (SR) [11],

low-rank representation (LRR) [12], collaborative representation (CR) [13], and low-rank and SR (LRSR) [14]. The purpose of SR is to represent a target sample with as few atoms as possible from a given dictionary such that the coefficients are as sparse as possible. In view of its powerful reconstruction ability, Chen et al. [15] took the lead in applying SR to the HSI classification scenario; meanwhile, they also proposed a joint SR (JSR) method to combine spatial–spectral information. Later, inspired by JSR, a series of improved algorithms have been developed. In [16], a kernel version of the JSR method was proposed, integrating higher order spatial context and distinctive spectral information to enhance classification performance. In addressing the issue of the assumption of joint sparsity within a fixed spatial window not always holding true in JSR, Peng et al. [17] employed the self-paced learning strategy along with an iterative updating approach to learn the weights of neighboring pixels. In addition, a maximum likelihood estimation-based JSR was given in [18], effectively suppressing interference from heterogeneous pixels within the neighborhood. Subsequently, advanced SR-based studies for HSI classification mainly focused on graph regularization [19], [20], dictionary learning [21], superpixel segmentation [22], [23], and multi-feature learning/extraction [24], [25]. Nevertheless, a common problem of SR-based methods is that they only consider local samples with similar characteristics as the target sample, neglecting the global data structure of HSI.

Compared with SR, the basic LRR model seeks the lowest rank representation of target samples based on a given complete dictionary, which can capture the global data structure and perform robust subspace segmentation [26]. For HSI classification, LRR-related methods are widely utilized. To suppress the intraclass spectral variations, Mei et al. [27] designed a simultaneous spatial and spectral LRR (S3LRR) method to robustly model spectral variations based on the spatial and spectral low-rank properties of HSI. For accurate classification in complex scenes, Su et al. [28] introduced the elastic network idea into the LRR model by replacing the rank function with a combination of the nuclear-norm and Frobenius-norm. Furthermore, in [29], a superpixel-guided discriminative LRR (SP-DLRR) method was proposed to explore the local spatial information and low-rank characteristics of HSI. In addition to the classification scenario, the LRR model is also applicable for denoising [30], unmixing [31], [32], super-resolution [33], band selection [34], and anomaly detection [35], [36]. However, the LRR model lacks an effective description of the local data structure, which limits its further development in HSI processing.

In this context, LRSR can simultaneously capture the global and local structural information of HSI; thus, it has better applicability than SR and LRR. Relevant research results validate the effectiveness of LRSR in various HSI processing tasks, such as combining mixtures of Gaussian for anomaly detection [37], [38], denoising and inpainting disturbed HSI data considering inherently sparse and low-rank structure characteristics [39], [40], [41], and extracting discriminative features [42], [43]. Furthermore, some studies are devoted to extending LRSR for spectral–spatial classification

of HSI [44], [45]. However, the subspace segmentation of the standard LRSR does not work well when the subspace independence assumption cannot be satisfied. In practice, the data structure of HSI exhibits a highly nonlinear correlation and does not satisfy the assumption.

As observed in [46] and [47], imposing effective structural constraints in LRR or LRSR can improve the segmentation ability. Accordingly, a spectral consistency-constrained LRSR model for HSI classification was presented in [48], which can ensure that similar hyperspectral pixels have similar encoding coefficients. Nevertheless, since the method only considers spectral information when constructing the structure matrix, it is susceptible to noisy or mixed pixels. In [49], a locality-constrained LRSR model was proposed to exploit the low-dimensional manifold structure of HSI, and the nonnegative and sum-to-one constraints were imposed to learn discriminative representation coefficients. However, the inclusion of multiple constraints increases the difficulty of optimization. Subsequently, Huang et al. [50] integrated dictionary learning into locality constrained LRSR framework. However, it cannot guarantee the completeness of the dictionary, especially for classes with limited labeled samples that are difficult to reflect their specific characteristics. In addition, the aforementioned methods lack consideration of the imbalanced problem of labeled samples in the classification process.

To address the aforementioned issues, we propose a structure-prior-constrained LRSR with the discriminative incremental dictionary (SPCLSR-DID) for HSI classification. In SPCLSR-DID, low-rank and sparse constraints are utilized to maintain the global and local structures of HSI, and the prior constraints applied to the coefficient matrix and reconstruction error matrix can effectively exploit the intrinsic spectral–spatial characteristics of HSI. Then, a discriminative incremental dictionary (DID) learning method is designed to construct a complete dictionary with high reliability and low redundancy. In DID, a regularized neighborhood is established by calculating the spectral angle to find the augmented atoms, and the incremental dictionary (ID) is screened by the average correlation between the subdictionary and augmented atoms. Moreover, we combine contextual information to finally determine the class of target pixels, which effectively improves the smoothness and accuracy of the classification map.

In our previous research works, we explored the importance of enhancing the representation ability of the dictionary for improving classification accuracy in the representation learning framework. We proposed three dictionary optimization schemes, namely, multifeature dictionary learning (MDL) [51], ID learning (IDL) [52], and self-balancing dictionary learning (SBDL) [53], to enhance the representation ability of the dictionary. Different from our previous work, this article considers the discriminative property of the augmented atoms in dictionary learning, which ensures that the optimized dictionary has both strong representation ability and low redundancy. More importantly, we propose the structural prior constraint to improve the subspace segmentation ability of the LRSR model. In addition, our study [54] also proposed a simple and effective spatial regularization method to remove heterogeneous samples in the local neighborhood, which is

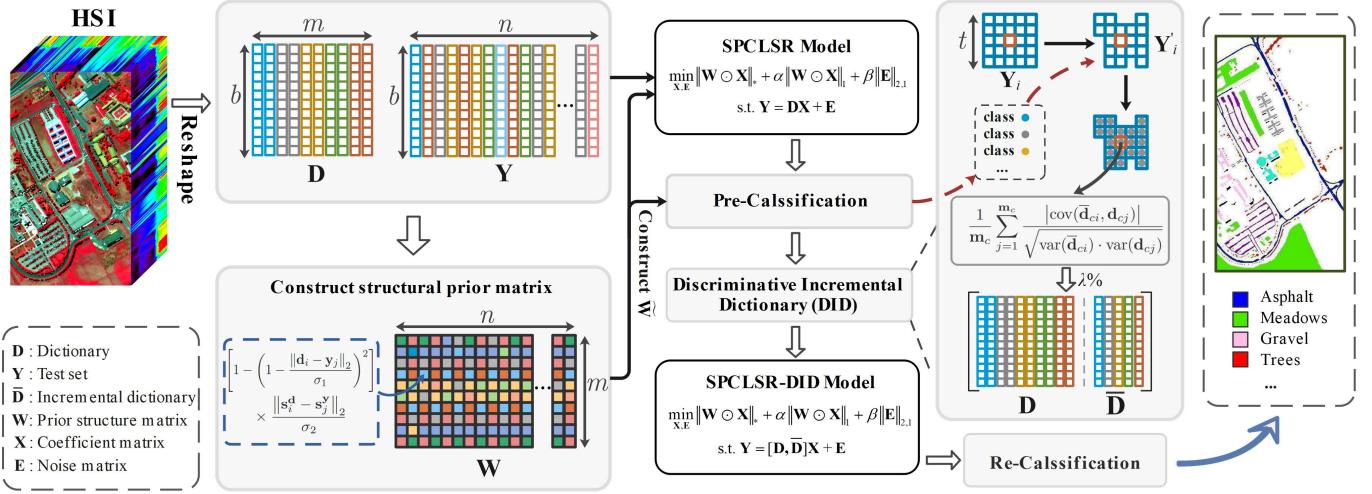


Fig. 1. Graphical illustration of the proposed SPCLSR-DID model for HSI classification.

introduced into the proposed DID learning process to construct a regularized neighborhood.

In this context, the main contributions of this article can be summarized as follows.

- 1) A novel SPCLSR-DID method is proposed for HSI classification, which can effectively maintain the global and local structures of the data while incorporating the inherent prior information. Compared with other related methods, the proposed SPCLSR-DID can achieve superior performance.
- 2) We propose a structural prior constraint to enhance the subspace segmentation ability of LRSR, which can increase intraclass correlations and interclass differences by combining spectral similarity and spatial structure information. Furthermore, classification based on constrained reconstruction errors effectively weakens the imbalance problem of labeled samples.
- 3) We design a DID scheme to improve the completeness of the dictionary. The DID scheme is able to introduce reliable and discriminative augmented atoms, while the size-controllable ID can avoid excessive computation consumption and is suitable for different tasks.

II. PROPOSED METHOD

Denote a hyperspectral dataset as $\mathbf{H} \in \mathbb{R}^{b \times k}$, where b is the number of bands and k is the number of pixels. Let $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_c, \dots, \mathbf{D}_C] \in \mathbb{R}^{b \times m}$ denote the dictionary constructed from the labeled samples, and m is the number of atoms in \mathbf{D} . In addition, \mathbf{D} can also be denoted as $[\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_i, \dots, \mathbf{d}_m]$, where $\mathbf{d}_i \in \mathbb{R}^b$ represents the i th atom in \mathbf{D} . Let $\mathbf{D}_c = [\mathbf{d}_{c1}, \mathbf{d}_{c2}, \dots, \mathbf{d}_{cm_c}] \in \mathbb{R}^{b \times m_c}$ denote the subdictionary containing only labeled samples of the c th class, and m_c is the number of atoms in \mathbf{D}_c . Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_i, \dots, \mathbf{y}_n] \in \mathbb{R}^{b \times n}$ denote the test set, where n is the number of samples in \mathbf{Y} and $\mathbf{y}_i \in \mathbb{R}^b$ represents the i th target sample. Let $\mathbf{N}_i = [\mathbf{y}_i, \mathbf{y}_{i1}, \dots, \mathbf{y}_{i(T-1)}] \in \mathbb{R}^{b \times T}$ denote the i th neighborhood sample set constructed with \mathbf{y}_i as the center sample and $T = t \times t$ as the window size. For clarity,

the notations utilized throughout this article are summarized in Nomenclature.

A graphical illustration of the proposed SPCLSR-DID model for HSI classification is shown in Fig. 1. First, the dictionary and test set are partitioned according to the reshaped HSI, and then, the structural prior matrix is constructed. Second, we solve the SPCLSR model and obtain a preclassification result according to the constrained reconstruction error. Third, we preliminarily determine the label of augmented atoms for each subdictionary based on the consistency of the samples in the regularized neighborhood. The incremental subdictionary for each class is then optimized by the average correlation. Finally, the SPCLSR-DID model is solved, and the final classification map can be obtained by jointly using the context information.

A. SPCLSR-DID Model Definition

The original objective function of LRSR is formulated as

$$\min_{\mathbf{X}} \text{rank}(\mathbf{X}) + \alpha \|\mathbf{X}\|_0, \text{ s.t. } \mathbf{Y} = \mathbf{DX} \quad (1)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ is the coefficient matrix, $\text{rank}(\cdot)$ stands for low-rank operation, α is the regularization parameter for the sparse term, and $\|\cdot\|_0$ means the number of nonzero elements.

For the convenience of solving, nuclear-norm ($\|\cdot\|_*$) and l_1 -norm ($\|\cdot\|_1$) are utilized to perform convex optimization on (1) [12], and the optimized equation can be reformulated as

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* + \alpha \|\mathbf{X}\|_1, \text{ s.t. } \mathbf{Y} = \mathbf{DX} \quad (2)$$

where $\|\cdot\|_*$ represents the sum of the singular values and $\|\cdot\|_1$ is the sum of the absolute values of all elements.

In LRSR, low-rankness and sparsity are combined to maintain global and local data structures simultaneously, but the underlying prior information in HSI is ignored. Therefore, we design a structural prior matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$ based on the spectral-spatial structural information between target samples and dictionary atoms, which can effectively improve the subspace segmentation ability of LRSR.

Represent the space coordinates of \mathbf{d}_i and \mathbf{y}_j as $\mathbf{s}_i^{\mathbf{d}}$ and $\mathbf{s}_j^{\mathbf{y}}$, respectively, and the element $\mathbf{W}_{i,j}$ in \mathbf{W} is

$$\mathbf{W}_{i,j} = \left[1 - \left(1 - \frac{\|\mathbf{d}_i - \mathbf{y}_j\|_2}{\sigma_1} \right)^2 \right] \times \frac{\|\mathbf{s}_i^{\mathbf{d}} - \mathbf{s}_j^{\mathbf{y}}\|_2}{\sigma_2} \quad (3)$$

where $\|\cdot\|_2$ is the l_2 -norm, $\sigma_1 = \max_{\forall i,j}(\|\mathbf{d}_i - \mathbf{y}_j\|_2)$, $\sigma_2 = \max_{\forall i,j}(\|\mathbf{s}_i^{\mathbf{d}} - \mathbf{s}_j^{\mathbf{y}}\|_2)$, and $\mathbf{W}_{i,j} \in (0, 1)$. The structural prior matrix reflects the similar relationship between dictionary atoms and target samples. A higher spectral similarity between \mathbf{d}_i and \mathbf{y}_j , as well as their spatial proximity, results in a smaller $\mathbf{W}_{i,j}$ value. In other words, the element $\mathbf{W}_{i,j}$ in \mathbf{W} can be regarded as a penalty parameter of the coefficient $\mathbf{X}_{i,j}$. During optimization, a smaller value of $\mathbf{W}_{i,j}$ means a lower punishment effect on the coefficient $\mathbf{X}_{i,j}$.

Then, the structural prior matrix is imposed on both the low-rank and sparse terms, and the objective function of the SPCLSR model is

$$\min_{\mathbf{X}} \|\mathbf{W} \odot \mathbf{X}\|_* + \alpha \|\mathbf{W} \odot \mathbf{X}\|_1, \text{ s.t. } \mathbf{Y} = \mathbf{D}\mathbf{X} \quad (4)$$

where \odot represents the Hadamard product.

However, HSI is usually noisy in real applications due to the interference of various factors [55], which will affect classification accuracy. Therefore, a robust SPCLSR objective function takes the form as

$$\begin{aligned} & \min_{\mathbf{X}, \mathbf{E}} \|\mathbf{W} \odot \mathbf{X}\|_* + \alpha \|\mathbf{W} \odot \mathbf{X}\|_1 + \beta \|\mathbf{E}\|_{2,1}, \\ & \text{s.t. } \mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{E} \end{aligned} \quad (5)$$

where $\mathbf{E} \in \mathbb{R}^{b \times n}$ is the noise matrix, β is its corresponding regularization parameter, $\|\mathbf{E}\|_{2,1} = \sum_i \|\mathbf{e}_i\|_2$ represents the $l_{2,1}$ -norm, and \mathbf{e}_i is the i th column vector of \mathbf{E} .

After the ID $\bar{\mathbf{D}}$ is obtained, we reconstruct \mathbf{W} according to \mathbf{D} and $\bar{\mathbf{D}}$ (denoted as $[\mathbf{D}, \bar{\mathbf{D}}]$), and then, the final SPCLSR-DID model is

$$\begin{aligned} & \min_{\mathbf{X}, \mathbf{E}} \|\mathbf{W} \odot \mathbf{X}\|_* + \alpha \|\mathbf{W} \odot \mathbf{X}\|_1 + \beta \|\mathbf{E}\|_{2,1}, \\ & \text{s.t. } \mathbf{Y} = [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} + \mathbf{E}. \end{aligned} \quad (6)$$

It is worth noting that the samples contained in $\bar{\mathbf{D}}$ have been removed from \mathbf{Y} .

B. SPCLSR-DID Model Solving

The alternating direction method of multipliers (ADMM) [56] is adopted to solve the SPCLSR-DID model in (6).

First, we introduce two auxiliary variables \mathbf{X}_1 and \mathbf{X}_2 to separate the low-rank and sparse terms in (6), and then, the new objective equation is

$$\begin{aligned} & \min_{\mathbf{X}, \mathbf{E}} \|\mathbf{W} \odot \mathbf{X}_1\|_* + \alpha \|\mathbf{W} \odot \mathbf{X}_2\|_1 + \beta \|\mathbf{E}\|_{2,1}, \\ & \text{s.t. } \mathbf{Y} = [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} + \mathbf{E}, \mathbf{X}_1 = \mathbf{X}, \mathbf{X}_2 = \mathbf{X}. \end{aligned} \quad (7)$$

Subsequently, the augmented Lagrange function corresponding to (7) is

$$\begin{aligned} L(\mathbf{X}, \mathbf{E}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3) \\ = \|\mathbf{W} \odot \mathbf{X}_1\|_* + \alpha \|\mathbf{W} \odot \mathbf{X}_2\|_1 + \beta \|\mathbf{E}\|_{2,1} \\ + \langle \mathbf{Q}_1, \mathbf{Y} - [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} - \mathbf{E} \rangle + \langle \mathbf{Q}_2, \mathbf{X} - \mathbf{X}_1 \rangle \end{aligned}$$

$$\begin{aligned} & + \langle \mathbf{Q}_3, \mathbf{X} - \mathbf{X}_2 \rangle + \frac{\mu}{2} (\|\mathbf{Y} - [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} - \mathbf{E}\|_F^2 \\ & + \|\mathbf{X} - \mathbf{X}_1\|_F^2 + \|\mathbf{X} - \mathbf{X}_2\|_F^2) \end{aligned} \quad (8)$$

where \mathbf{Q}_1 , \mathbf{Q}_2 , and \mathbf{Q}_3 are the Lagrange multipliers and μ is a penalty parameter. $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{A}^T \mathbf{B})$ represents the matrix inner product, and the superscript T means the matrix transpose.

Finally, the variables \mathbf{X}_1 , \mathbf{X}_2 , \mathbf{E} , and \mathbf{X} are updated alternately.

- 1) Fix \mathbf{X}_2 , \mathbf{E} , and \mathbf{X} , and update \mathbf{X}_1

$$\arg \min_{\mathbf{X}_1} \frac{1}{\mu} \|\mathbf{W} \odot \mathbf{X}_1\|_* + \frac{1}{2} \left\| \mathbf{X}_1 - \mathbf{X} - \frac{\mathbf{Q}_2}{\mu} \right\|_F^2. \quad (9)$$

According to the literature [48], the closed-form solution of (9) is

$$\mathbf{X}_1 = (\mu \mathbf{X} + \mathbf{Q}_2) \cdot / (\mu + 2(\mathbf{W} \odot \mathbf{W})) \quad (10)$$

where $\cdot /$ represents the elementwise division.

- 2) Fix \mathbf{X}_1 , \mathbf{E} , and \mathbf{X} , and update \mathbf{X}_2

$$\arg \min_{\mathbf{X}_2} \frac{\alpha}{\mu} \|\mathbf{W} \odot \mathbf{X}_2\|_1 + \frac{1}{2} \left\| \mathbf{X}_2 - \mathbf{X} - \frac{\mathbf{Q}_3}{\mu} \right\|_F^2. \quad (11)$$

Note that $\mathbf{X}_{i,j}$, $(\mathbf{Q}_3)_{i,j}$, and $\mathbf{W}_{i,j}$ are the i -row and j -column elements of \mathbf{X} , \mathbf{Q}_3 , and \mathbf{W} , respectively. Then, \mathbf{X}_2 can be updated by the shrinkage operator

$$\begin{aligned} (\mathbf{X}_2)_{i,j} = & \max \left(\mathbf{X}_{i,j} + \frac{1}{\mu} (\mathbf{Q}_3)_{i,j} - \frac{\alpha}{\mu} \mathbf{W}_{i,j}, 0 \right) \\ & + \min \left(\mathbf{X}_{i,j} + \frac{1}{\mu} (\mathbf{Q}_3)_{i,j} + \frac{\alpha}{\mu} \mathbf{W}_{i,j}, 0 \right). \end{aligned} \quad (12)$$

- 3) Fix \mathbf{X}_1 , \mathbf{X}_2 , and \mathbf{X} , and update \mathbf{E}

$$\arg \min_{\mathbf{E}} \frac{\beta}{\mu} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \left\| \mathbf{E} - \mathbf{Y} + [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} - \frac{\mathbf{Q}_1}{\mu} \right\|_F^2. \quad (13)$$

Let $\mathbf{G} = \mathbf{Y} - [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} + (\mathbf{Q}_1/\mu)$, and \mathbf{g}_i represents the i th column vector of \mathbf{G} , and then, the i th column vector of \mathbf{E} can be updated as

$$\mathbf{e}_i = \begin{cases} \frac{\|\mathbf{g}_i\|_2 - \frac{\beta}{\mu}}{\|\mathbf{g}_i\|_2} \mathbf{g}_i, & \frac{\beta}{\mu} < \|\mathbf{g}_i\|_2 \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

- 4) Fix \mathbf{X}_1 , \mathbf{X}_2 , and \mathbf{E} , and update \mathbf{X}

$$\begin{aligned} & \arg \min_{\mathbf{X}} \langle \mathbf{Q}_1, \mathbf{Y} - [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} - \mathbf{E} \rangle + \langle \mathbf{Q}_2, \mathbf{X} - \mathbf{X}_1 \rangle \\ & + \langle \mathbf{Q}_3, \mathbf{X} - \mathbf{X}_2 \rangle + \frac{\mu}{2} (\|\mathbf{Y} - [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} - \mathbf{E}\|_F^2 \\ & + \|\mathbf{X} - \mathbf{X}_1\|_F^2 + \|\mathbf{X} - \mathbf{X}_2\|_F^2). \end{aligned} \quad (15)$$

We introduce three intermediate variables (\mathbf{F}_1 , \mathbf{F}_2 , and \mathbf{F}_3) to simplify expressing the result

$$\begin{aligned} \mathbf{F}_1 &= [\mathbf{D}, \bar{\mathbf{D}}]^T \left(\frac{\mathbf{Q}_1}{\mu} - \mathbf{E} \right) \\ \mathbf{F}_2 &= [\mathbf{D}, \bar{\mathbf{D}}]^T \mathbf{Y} - \frac{1}{\mu} (\mathbf{Q}_2 + \mathbf{Q}_3) \\ \mathbf{F}_3 &= \mathbf{X}_1 + \mathbf{X}_2. \end{aligned} \quad (16)$$

Then, \mathbf{X} can be updated as

$$\mathbf{X} = \left([\mathbf{D}, \bar{\mathbf{D}}]^T [\mathbf{D}, \bar{\mathbf{D}}] + 2\mathbf{I} \right)^{-1} (\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3) \quad (17)$$

where \mathbf{I} represents the identity matrix.

The pseudocode corresponding to ADMM for SPCLSR-DID model solving is summarized in Algorithm 1.

Algorithm 1 ADMM for SPCLSR-DID Model Solving

Input: $[\mathbf{D}, \bar{\mathbf{D}}]$, $\mathbf{W}, \mathbf{Y}, \alpha, \beta$, max-iter

Initialize: $\mathbf{X} = \mathbf{X}_1 = \mathbf{X}_2 = 0$, $\mathbf{E} = 0$, $\mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{Q}_3 = 0$, $\mu = 10^{-4}$, $\mu_{max} = 10^2$, max-iter=200

While not max-iter **do**

 Update \mathbf{X}_1 according to Eq. (10)

 Update \mathbf{X}_2 according to Eq. (12)

 Update \mathbf{E} according to Eq. (14)

 Update \mathbf{X} according to Eq. (17)

 Update $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3$ and μ

$$\mathbf{Q}_1 = \mathbf{Q}_1 + \mu(\mathbf{Y} - [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} - \mathbf{E})$$

$$\mathbf{Q}_2 = \mathbf{Q}_2 + \mu(\mathbf{X}_1 - \mathbf{X})$$

$$\mathbf{Q}_3 = \mathbf{Q}_3 + \mu(\mathbf{X}_2 - \mathbf{X})$$

$$\mu = \min(1.2\mu, \mu_{max})$$

Output: \mathbf{X}, \mathbf{E}

C. SPCLSR-DID for HSI Classification

1) *Preclassification Based on Constrained Reconstruction Error*: The HSI classification models based on LRSR can directly determine the class label of a target sample by comparing reconstruction errors. However, in practical applications, the number of labeled samples of each class is usually unbalanced, and the classification result tends to classes with a large number of labeled samples. To alleviate this issue, we construct the constraint matrix of reconstruction error based on the structural prior matrix and utilize the constrained reconstruction error as the basis for classification.

According to $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_c, \dots, \mathbf{D}_C]$, the equation $\mathbf{Y} = \mathbf{DX} + \mathbf{E}$ can be expanded into

$$\mathbf{Y} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_c, \dots, \mathbf{D}_C]\mathbf{X} + \mathbf{E}. \quad (18)$$

Therefore, the test set $\mathbf{Y}_c = [\mathbf{y}_{c1}, \mathbf{y}_{c2}, \dots, \mathbf{y}_{ci}, \dots, \mathbf{y}_{cn}] \in \mathbb{R}^{b \times n}$ reconstructed based on the c th subdictionary can be expressed as

$$\mathbf{Y}_c = [0, \dots, \mathbf{D}_c, \dots, 0]\mathbf{X}. \quad (19)$$

Subsequently, solve (5) by ADMM to get the coefficient matrix \mathbf{X} and noise matrix \mathbf{E} . Then, the denoised test set $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_i, \dots, \hat{\mathbf{y}}_n] \in \mathbb{R}^{b \times n}$ can be obtained by

$$\hat{\mathbf{Y}} = \mathbf{DX} \text{ or } \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{E}. \quad (20)$$

Denote the reconstruction error matrix as $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_c, \dots, \mathbf{r}_C]^T \in \mathbb{R}^{C \times n}$, where the specific mathematical expression of \mathbf{r}_c is

$$\mathbf{r}_c = [\|\mathbf{y}_{c1} - \hat{\mathbf{y}}_1\|_1, \|\mathbf{y}_{c2} - \hat{\mathbf{y}}_2\|_1, \dots, \|\mathbf{y}_{cn} - \hat{\mathbf{y}}_n\|_1]^T. \quad (21)$$

It is worth noting that \mathbf{y}_{ci} and $\hat{\mathbf{y}}_i$ ($i \in [1, n]$) have been converted to unit vectors before calculating \mathbf{r}_c .

Next, corresponding to different class labels of \mathbf{D} , the structural prior matrix \mathbf{W} can be further divided into

$$\mathbf{W} = (\mathbf{W}^T)^T = [\mathbf{W}_1^T, \mathbf{W}_2^T, \dots, \mathbf{W}_c^T, \dots, \mathbf{W}_C^T]^T \quad (22)$$

where \mathbf{W}_c is the part of \mathbf{W} corresponding to the subdictionary \mathbf{D}_c . In this case, the constraint matrix $\tilde{\mathbf{W}} = [\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \dots, \tilde{\mathbf{w}}_c, \dots, \tilde{\mathbf{w}}_C]^T \in \mathbb{R}^{C \times n}$ corresponding to \mathbf{R} can be calculated, and the specific expression of $\tilde{\mathbf{w}}_c$ is

$$\tilde{\mathbf{w}}_c = [(\mathbf{w}_{c,1})_{min}, (\mathbf{w}_{c,2})_{min}, \dots, (\mathbf{w}_{c,n})_{min}]^T \quad (23)$$

where $\mathbf{w}_{c,i}$ is the i th column vector in \mathbf{W}_c and $(\cdot)_{min}$ means to return the minimum value of the input vector.

After obtaining the matrix \mathbf{R} and its constraint matrix $\tilde{\mathbf{W}}$, the corresponding elements in the two matrices are multiplied

$$\tilde{\mathbf{R}} = \mathbf{R} \odot \tilde{\mathbf{W}} \quad (24)$$

where $\tilde{\mathbf{R}} \in \mathbb{R}^{C \times n}$ is the constrained reconstruction error matrix. Denote $\tilde{\mathbf{r}}_i \in \mathbb{R}^C$ as the i th column vector in $\tilde{\mathbf{R}}$, and then, the class label of the i th target sample is

$$\text{class}(\mathbf{y}_i) = \min(\tilde{\mathbf{r}}_i)_{index} \quad (25)$$

where $\min(\tilde{\mathbf{r}}_i)_{index} \in [1, C]$ returns the index of the minimum value in $\tilde{\mathbf{r}}_i$. In this way, a preliminary classification result can be obtained for the entire test set.

2) *DID Learning*: In the preclassification process, the dictionary \mathbf{D} is fixed and directly composed of labeled samples. However, some labeled samples may lack discriminativeness or be seriously disturbed by noise, resulting in a suboptimal representation of the dictionary and difficulty in achieving satisfactory classification performance. To cope with this deficiency, we construct a DID based on the preclassification result combined with the regularized neighborhood and average correlation coefficients to enhance the completeness of the dictionary.

First, we center each target sample and construct its corresponding neighborhood sample set through a spatial window of size $T = t \times t$ and the neighborhood sample set corresponding to \mathbf{y}_i is $\mathbf{N}_i = [\mathbf{y}_i, \mathbf{y}_{i1}, \dots, \mathbf{y}_{iT-1}]$. Then, to strengthen the consistency of the samples in \mathbf{N}_i , we measure the spectral similarity of each neighborhood sample to the center sample by calculating the spectral angle

$$S_{ij} = \frac{(\mathbf{y}_i)^T \mathbf{y}_{ij}}{\|\mathbf{y}_i\|_2 \times \|\mathbf{y}_{ij}\|_2} \quad (26)$$

where a larger S_{ij} value represents a higher spectral similarity between \mathbf{y}_i and \mathbf{y}_{ij} . Set a similarity threshold S , when $S_{ij} < S$, \mathbf{y}_{ij} will be removed from \mathbf{N}_i . After regularization, the new neighborhood sample set is denoted as \mathbf{N}'_i . At this point, we assign the preclassification result to \mathbf{N}'_i ; if the class labels of all samples in \mathbf{N}'_i are consistent, assuming c , then add the center sample \mathbf{y}_i to the incremental subdictionary $\bar{\mathbf{D}}_c$. In this way, the ID $\bar{\mathbf{D}} = [\bar{\mathbf{D}}_1, \bar{\mathbf{D}}_2, \dots, \bar{\mathbf{D}}_c, \dots, \bar{\mathbf{D}}_C]$ can be obtained by traversing all target samples.

However, there are still two problems for the current ID $\bar{\mathbf{D}}$. On the one hand, there may be some wrong augmented atoms in $\bar{\mathbf{D}}$, which will seriously affect the classification accuracy. On the other hand, there is a high correlation between some

augmented atoms and labeled samples. These augmented atoms cannot bring additional discriminative information and are useless in enhancing the representation ability of the dictionary. Therefore, we optimize the ID according to the correlation between augmented atoms and labeled samples.

Assuming that $\bar{\mathbf{d}}_{ci}$ and \mathbf{d}_{cj} are the i th and j th samples in $\bar{\mathbf{D}}_c$ and \mathbf{D}_c , respectively, then the average correlation coefficient $\rho_{\bar{\mathbf{d}}_{ci}}$ between $\bar{\mathbf{d}}_{ci}$ and \mathbf{D}_c is

$$\begin{aligned} \rho_{\bar{\mathbf{d}}_{ci}} &= \frac{1}{m_c} \sum_{j=1}^{m_c} \frac{|\text{cov}(\bar{\mathbf{d}}_{ci}, \mathbf{d}_{cj})|}{\sqrt{\text{var}(\bar{\mathbf{d}}_{ci})} \cdot \text{var}(\mathbf{d}_{cj})} \\ &= \frac{1}{m_c} \sum_{j=1}^{m_c} \frac{\left| \left(\bar{\mathbf{d}}_{ci} - (\bar{\mathbf{d}}_{ci})_{\text{avg}} \right)^T \left(\mathbf{d}_{cj} - (\mathbf{d}_{cj})_{\text{avg}} \right) \right|}{\left\| \bar{\mathbf{d}}_{ci} - (\bar{\mathbf{d}}_{ci})_{\text{avg}} \right\|_2 \cdot \left\| \mathbf{d}_{cj} - (\mathbf{d}_{cj})_{\text{avg}} \right\|_2} \quad (27) \end{aligned}$$

where $\text{cov}(\cdot)$, $\text{var}(\cdot)$, and $(\cdot)_{\text{avg}}$ represent covariance, variance, and average operations, respectively.

After calculating the average correlation coefficients between $\bar{\mathbf{D}}_c$ and \mathbf{D}_c , sort the atoms in $\bar{\mathbf{D}}_c$ in ascending order based on their correlation values and remove a certain percentage of atoms from the top and bottom, respectively. The reason is that if the average correlation is low, the corresponding augmented atoms are likely to be wrong. Conversely, if the correlation is high, such augmented atoms are redundant and cannot improve the representation ability of the dictionary. Furthermore, after the dictionary is augmented, the computational complexity of the model also increases. In order to balance the classification accuracy and computational complexity to meet the needs of different tasks, we can randomly select a part from the remaining atoms of $\bar{\mathbf{D}}_c$ to augment \mathbf{D}_c . Assuming that the sampling rate is λ , thence the size of \mathbf{D} can be adjusted with λ , and thus, control the computational complexity of the model.

Throughout the entire DID learning process, we prioritize the high reliability of the acquired augmented atoms first through the spatial structural information constraint. Subsequently, optimization is performed on each class of augmented atoms through average correlation, enhancing the spectral distinctiveness among augmented atoms. As a result, the representation ability of the dictionary is effectively enhanced.

3) Reclassification by Integrating Contextual Information:

For HSI, the samples within a certain neighborhood usually belong to the same class and have similar spectral characteristics. Related studies [57], [58], [59] have shown that integrating context information can effectively improve the accuracy and smoothness of the classification map.

To this end, we first reconstruct the neighborhood sample set \mathbf{N}_i according to a new window size $T' = t' \times t'$ and $t' = \max(t - \Delta t, t_{\min})$, where t_{\min} represents the minimum window size and Δt is a window size variable used to control the scaling of the window. In the proposed method, the minimum window size is 3×3 (i.e., $t_{\min} = 3$). Then, inspired by our previous work [52], we empirically set the window size variable Δt to 4. In this case, the size of t' can be directly determined by t , without the need for separate adjustment. Moreover, it is worth noting that all the neighborhood sample sets here do not contain the atoms in $[\mathbf{D}, \bar{\mathbf{D}}]$. Next, solve the

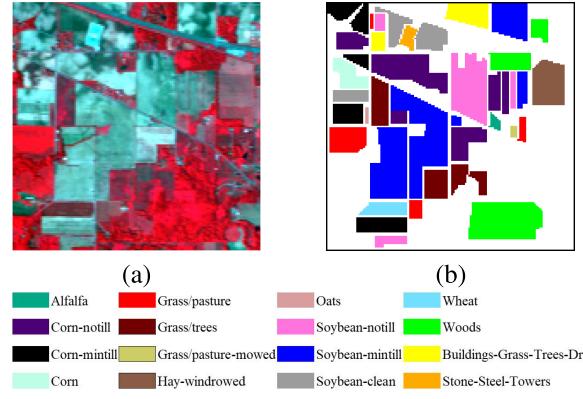


Fig. 2. Indian Pines. (a) False-color composite image (R: 50, G: 27, and B: 17). (b) Ground-truth map.

SPCLSR-DID model in (5) and recalculate the constrained reconstruction error matrix $\tilde{\mathbf{R}}$ according to (18)–(24). Finally, denote the part of $\tilde{\mathbf{R}}$ corresponding to \mathbf{N}_i as $\tilde{\mathbf{R}}_i$, and then, the class label of the central sample \mathbf{y}_i can be determined as

$$\text{class}(\mathbf{y}_i) = \min \left(\tilde{\mathbf{R}}_i \right)_{\text{row}} \quad (28)$$

where $\min(\tilde{\mathbf{R}}_i)_{\text{row}} \in [1, C]$ represents the row index of the minimum value in $\tilde{\mathbf{R}}_i$. Furthermore, it should be noted that there is a distinction between (25) and (28). Specifically, (28) incorporates contextual information to assist in the determination of the class label for the center sample, while (25) does not.

The pseudocode of the proposed SPCLSR-DID method for HSI classification is outlined in Algorithm 2.

Algorithm 2 SPCLSR-DID for HSI Classification

Input: Dictionary \mathbf{D} , test set \mathbf{Y} , parameters α, β, T, S .

1. Construct and solve the SPCLSR model in Eq. (5).
2. Calculate the reconstruction error matrix \mathbf{R} and its constraint matrix $\tilde{\mathbf{W}}$ by Eqs. (18)–(23).
3. Obtain the pre-classification result by Eq. (25).
4. Construct neighborhood sets, and then find augmented atoms according to the pre-classification result.
5. Calculate the average correlation coefficient by Eq. (27), and then optimize the augmented atoms.
6. Construct and solve the SPCLSR-DID model in Eq. (6).
7. Reclassify the target samples by Eq. (28).

Output: Class labels for \mathbf{Y} .

III. EXPERIMENTAL RESULTS

A. Hyperspectral Datasets

1) *Indian Pines*: The first dataset was acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over Northwestern Indiana's Indian Pines test site. The corresponding false-color composite image and ground-truth map are shown in Fig. 2. The size of the dataset is 145×145 pixels, the spatial resolution is 20 m and contains 220 bands in the wavelength range from 0.4 to 2.5 μm . After removing part of

TABLE I
NUMBER OF DICTIONARY AND TEST SAMPLES FOR INDIAN PINES

Class	Name	Dictionary	Test	Total
1	Alfalfa	3	43	46
2	Corn-notill	72	1,356	1,428
3	Corn-mintill	42	788	830
4	Corn	12	225	237
5	Grass/pasture	25	458	483
6	Grass/trees	37	693	730
7	Grass/pasture-mowed	2	26	28
8	Hay-windowed	24	454	478
9	Oats	1	19	20
10	Soybean-notill	49	923	972
11	Soybean-mintill	123	2,332	2,455
12	Soybean-clean	30	563	593
13	Wheat	11	194	205
14	Woods	64	1,201	1,265
15	Buildings-Grass-Trees-Drives	20	366	386
16	Stone-Steel-Towers	5	88	93
	Total	520	9,729	10,249

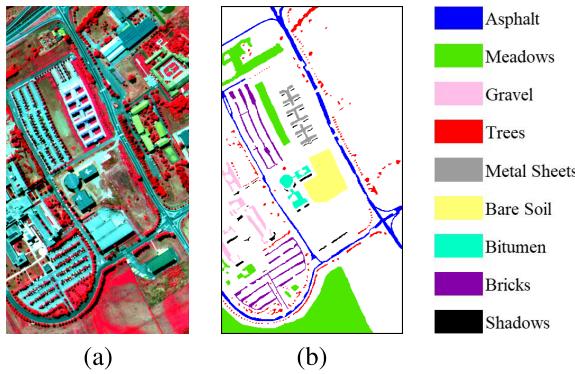


Fig. 3. University of Pavia. (a) False-color composite image (R: 102, G: 56, and B: 31). (b) Ground-truth map.

TABLE II
NUMBER OF DICTIONARY AND TEST SAMPLES
FOR THE UNIVERSITY OF PAVIA

Class	Name	Dictionary	Test	Total
1	Asphalt	67	6,564	6,631
2	Meadows	187	18,462	18,649
3	Gravel	21	2,078	2,099
4	Trees	31	3,033	3,064
5	Metal Sheets	14	1,331	1,345
6	Bare Soil	51	4,978	5,029
7	Bitumen	14	1,316	1,330
8	Bricks	37	3,645	3,682
9	Shadows	10	937	947
	Total	432	42,344	42,776

the noisy and water-absorbed bands, the remaining 200 bands are utilized for experiments. The ground-truth map includes 16 classes, and the specific information is listed in Table I.

2) *University of Pavia*: The second dataset was gathered by the Reflective Optics Spectrographic Imagine System (ROSIS) sensor over the urban area of the University of Pavia, Northern Italy. The dataset contains 610×340 pixels and has a high spatial resolution of 1.3 m. There are 103 radiance channels utilized in the experiments after removing 12 noisy bands. The corresponding false-color composite image and ground-truth map are shown in Fig. 3. The ground-truth map includes nine classes, and the specific quantity information for each class can be found in Table II.

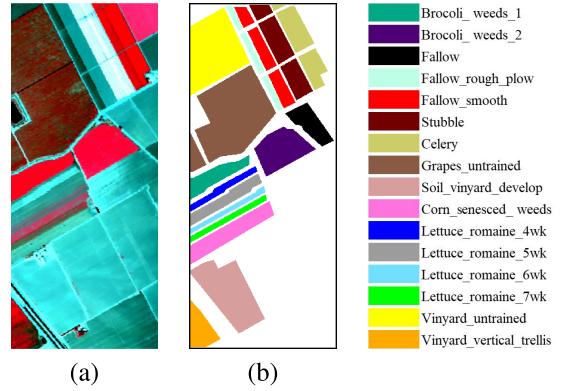


Fig. 4. Salinas. (a) False-color composite image (R: 50, G: 27, and B: 17). (b) Ground-truth map.

TABLE III
NUMBER OF DICTIONARY AND TEST SAMPLES FOR SALINAS

Class	Name	Dictionary	Test	Total
1	Brocoli-weeds-1	21	1,988	2,009
2	Brocoli-weeds-2	38	3,688	3,726
3	Fallow	20	1,956	1,976
4	Fallow-rough-plow	14	1,380	1,394
5	Fallow-smooth	27	2,651	2,678
6	Stubble	40	3,919	3,959
7	Celery	36	3,543	3,579
8	Grapes-untrained	113	11,158	11,271
9	Soil-vinyard-develop	63	6,140	6,203
10	Corn-senesced-weeds	33	3,245	3,278
11	Lettuce-romaine-4wk	11	1,057	1,068
12	Lettuce-romaine-5wk	20	1,907	1,927
13	Lettuce-romaine-6wk	10	906	916
14	Lettuce-romaine-7wk	11	1,059	1,070
15	Vinyard-untrained	73	7,195	7,268
16	Vinyard-vertical-trellis	19	1,788	1,807
	Total	549	53,580	54,129

3) *Salinas*: The third dataset was obtained by the AVIRIS sensor over Salinas Valley, CA, USA. Salinas has a spatial resolution of 3.7 m, including 512×217 pixels and 224 bands in the wavelength range from 0.4 to $2.5 \mu\text{m}$. There are 204 spectral bands retained after discarding the 20 water absorption bands. The corresponding false-color composite image and ground-truth map can be found in Fig. 4. The ground-truth map includes 16 classes, and the specific information is listed in Table III.

4) *WHU-LongKou*: The fourth dataset was acquired with an 8-mm focal length Headwall Nano-Hyperspec imaging sensor equipped on a DJI Matrice 600 Pro (DJI M600 Pro) UAV platform over Longkou town, Hubei province, China. It contains 270 bands in the wavelength range from 0.4 to $1 \mu\text{m}$. The spatial resolution of the dataset is about 0.463 m, and the size is 550×400 pixels. The corresponding false-color composite image and ground-truth map can be found in Fig. 5. The ground-truth map includes nine classes, and the specific information can be found in Table IV.

B. Experimental Settings

1) *Comparison Methods*: The classification methods that we compared include support vector machines with the composite kernel (SVMCK) [60], kernel JSR (KJSR) [61], fusing correlation coefficient and JSR (CCJSR) [62], SP-DLRR [29], global-local balanced low-rank approximation

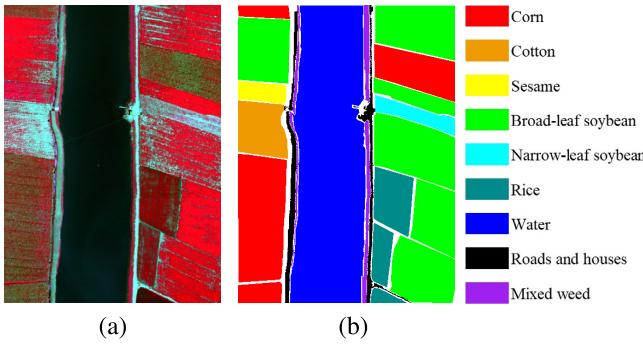


Fig. 5. WHU-LongKou. (a) False-color composite image (R: 174, G: 118, and B: 57). (b) Ground-truth map.

TABLE IV

NUMBER OF DICTIONARY AND TEST SAMPLES FOR WHU-LONGKOU

Class	Name	Dictionary	Test	Total
1	Corn	173	34,338	34,511
2	Cotton	42	8,322	8,374
3	Sesame	16	3,015	3,031
4	Broad-leaf soybean	317	62,895	63,212
5	Narrow-leaf soybean	21	4,130	4,151
6	Rice	60	11,794	11,854
7	Water	336	66,720	67,056
8	Roads and houses	36	7,088	7,124
9	Mixed weed	27	5,202	5,229
	Total	1,028	203,514	204,542

(GLB-LRA) [63], LRSR with spectral consistency constraint (LRSR-SCCE) [48], and sparse and LRR with key connectivity (SLRC) [44]. Furthermore, in subsequent experiments, we also include the preclassification based on constrained reconstruction as a comparison method, denoted as SPCLSR.

2) *Dictionary Settings*: For Indian Pines, 5% of the labeled samples are randomly selected to form the dictionary. For the University of Pavia and Salinas, their dictionaries consist of 1% labeled samples. For WHU-LongKou, the dictionary is composed of 0.5% labeled samples. The number of dictionary atoms for four datasets is listed in Tables I–IV. Furthermore, during the DID learning progress, we experientially set the removal ratio to 10%.

3) *Quantitative Metrics*: Three quantitative accuracy evaluation metrics are utilized to evaluate the classification performance of each method, including overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ). Moreover, all experimental data presented are the average of ten independent runs, and the dictionary for each run is generated by random sampling.

4) *Running Platform*: All implementations are carried out using MATLAB R2021b on a desktop PC equipped with Intel Core i7-11700F CPU, NVIDIA GeForce RTX 3060ti GPU, and 32-GB RAM. The source code will be available at <https://github.com/ZhaohuiXue/SPCLSR-DID>.

C. Parameter Sensitiveness

In this section, we analyze the parameters that affect the classification performance of SPCLSR-DID in detail, including regularization parameters α and β , neighborhood window $T = t \times t$, similarity threshold S , and incremental rate λ .

- 1) First, we tune parameters α and β simultaneously to verify their impact on classification accuracy. The values of α and β are taken from $[1e-3, 1e-2, 1e-1, 1, 1e, 1e2, 1e3]$ and $[2e-3, 2e-2, 2e-1, 2, 2e, 2e2, 2e3]$, respectively, and the changes in OA of preclassification are shown in Fig. 6. For Indian Pines, as shown in Fig. 6(a), the variation range of the OA is 91.18%–92.69% and the maximum value is obtained when $\alpha = 1$ and $\beta = 2e-2$. Then, as shown in Fig. 6(b) and (c), the OA for the University of Pavia and Salinas fluctuated between 96.62% and 97.57% and 95.57% and 96.87%, respectively. Meanwhile, the optimal α and β corresponding to the University of Pavia and Salinas are $\{\alpha = 1e3, \beta = 2e\}$ and $\{\alpha = 1e3, \beta = 2\}$. For WHU-LongKou, it can be inferred from Fig. 6(d) that the performance is optimal when $\{\alpha = 1e2, \beta = 2e2\}$. In the subsequent experiments, α and β are set by the above-mentioned optimal values for the four hyperspectral datasets.
- 2) Second, we sequentially optimize the window size T and threshold S . During the optimization of T , S is uniformly set to 0. λ is set to 20% for Indian Pines, 2% for WHU-LongKou, and 5% for the other two datasets. In Fig. 7(a), as T increases from 3×3 to 15×15 , the OA of the University of Pavia and Salinas shows a trend of first increasing and then decreasing, and both reach the highest at $T = 9 \times 9$. For Indian Pines and WHU-LongKou, the highest accuracy is achieved when $T = 3 \times 3$ and $T = 5 \times 5$, respectively, and then continues to decrease as the window size increases. Subsequently, during the optimization of S , its value ranges from 0.85 to 0.99 and the evolution of OA with S is shown in Fig. 7(b). From the changes in OA and error bar, it can be seen that an appropriate threshold S can effectively improve the classification accuracy and stability of the model. According to Fig. 7(b), the similarity thresholds for the Indian Pines, University of Pavia, Salinas, and WHU-LongKou are set to 0.95, 0.88, 0.96, and 0.97, respectively.
- 3) Third, we analyze the impact of λ on classification accuracy and computational complexity. In the analysis process, the neighborhood window in the reclassification step is fixed at 1×1 , and other parameters are set according to the optimal values. As illustrated in Fig. 8, the accuracy of four HSI datasets increases rapidly and then tends to be stable with the increase of λ . Furthermore, the computational complexity is positively correlated with λ , and the accuracy improvement brought by a larger λ is limited, but the running time is greatly increased. For instance, the running time and OA on the University of Pavia with $\lambda = 5\%$ are 132 s and 98.10%, respectively, and 220 s and 98.21% with $\lambda = 10\%$. In this case, an additional time consumption of about 70% only brings a 0.11% accuracy improvement. Thus, considering the accuracy and efficiency comprehensively, λ is set to 20% for Indian Pines, 5% for the University of Pavia and Salinas, and 2% for WHU-LongKou.

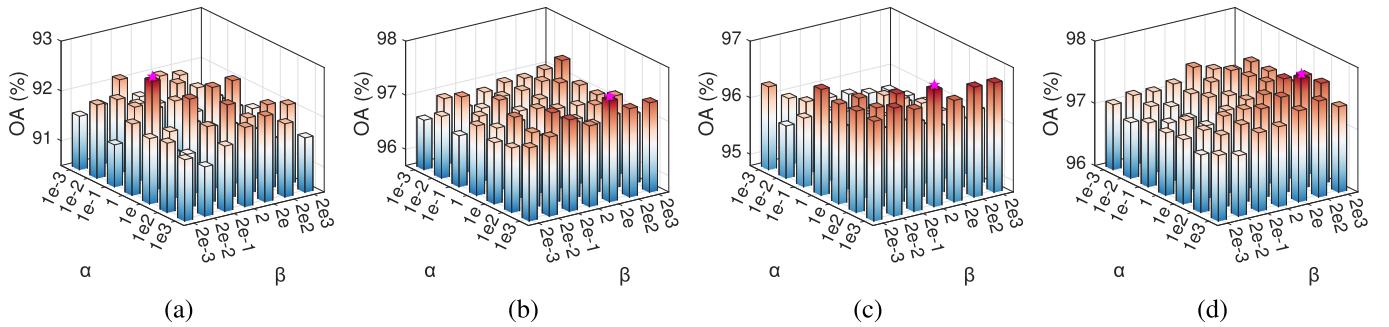


Fig. 6. Impact of parameters α and β on classification accuracy. (a) Indian Pines. (b) University of Pavia. (c) Salinas. (d) WHU-LongKou. Pentagram markers denote the optimal parameters' combination.

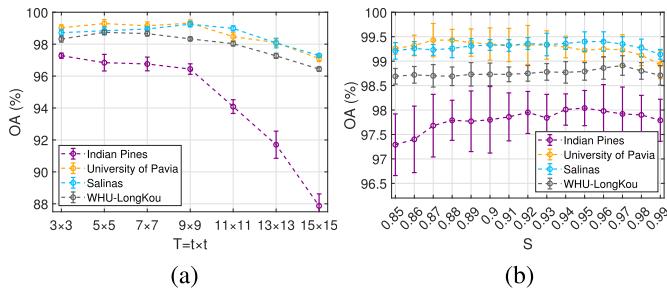


Fig. 7. Impact of (a) neighborhood window T and (b) similarity threshold S on classification accuracy.

D. Classification Results

1) *Indian Pines*: The classification results of each method on this dataset are listed in Table V. The proposed SPCLSR-DID achieves the highest OA of 98.04%, which is 0.61%–7.94% higher than other comparison methods. In addition, AA and κ are increased by 1.52%–24.29% and 0.67%–9.05%, respectively. For class-specific accuracies, SPCLSR-DID achieves the highest accuracy in 11 classes. SP-DLRR and GLB-LRA methods also show excellent classification performance and achieve the best classification results on multiple classes. For *Alfalfa* and *Grass/pasture-mowed* with very few labeled samples, both SPCLSR-DID and GLB-LRA methods have 100% accuracy, whereas SVMCK is only 25% and 0%. The classification maps corresponding to each method are exhibited in Fig. 9. Compared with SPCLSR-DID, there are obviously noise-like misclassified pixels in the map of SPCLSR. The main reason is that SPCLSR lacks context information and cannot effectively improve the smoothness and accuracy of the classification map.

2) *University of Pavia*: Table VI presents the classification results of each method at the University of Pavia, and corresponding classification maps are shown in Fig. 10. Based on 1% of labeled samples, our SPCLSR-DID has the highest OA, AA, and κ of 99.43%, 99.15%, and 0.9925, which yields 1.70%–20.79%, 2.37%–36.48%, and 2.27%–28.57% accuracy improvements. Among the nine classes of the dataset, SPCLSR-DID performs best on seven classes, while SVMCK and SP-DLRR have the best accuracy on the remaining *Trees* and *Shadows*, respectively. As shown in Fig. 10, each method has obvious misclassified pixels on *Bare Soil*, especially for JSR and CCJSR, which assign a large number of pixels

belonging to *Bare Soil* with the label of *Meadows*. In contrast, SPCLSR-DID significantly outperforms other methods in this class with an accuracy exceeding 99%. On the whole, the proposed SPCLSR-DID has the highest classification accuracy, and the classification map is closest to the ground-truth, proving its effectiveness and advancement.

3) *Salinas*: For this dataset, the specific classification results obtained by each method with 1% labeled samples are reported in Table VII, and the classification maps are placed in Fig. 11. Similarly, our SPCLSR-DID still performs best and brings improvements of 0.71%–6.70%, 1.24%–5.01%, and 0.78%–7.46% in three quantitative metrics. For class-specific accuracies, SPCLSR-DID has the highest accuracy on ten out of 16 classes and achieves 100% for *Brocoli-weeds-1*, *Brocoli-weeds-2*, and *Soil-vinyard-develop*. Moreover, different from the University of Pavia, it is interesting to note that KJSR performs well on Salinas and has 100% accuracy for classes *Brocoli-weeds-1*, *Celery*, and *Soil-vinyard-develop*. In Fig. 11, for the confusing classes *Grapes-untrained* and *Vinyard-untrained*, the classification performance of SPCLSR-DID and SP-DLRR is significantly better than the rest. However, the overall correctness and smoothness of the map corresponding to SPCLSR-DID are better than SP-DLRR, especially for *Brocoli-weeds-2* and *Fallow*.

4) *WHU-LongKou*: For this dataset, the classification results of each method are listed in Table VIII, and the corresponding classification maps are shown in Fig. 12. From the classification results, the OA, AA, and κ of the proposed SPCLSR-DID are the highest, which are 0.58%–7.97%, 1.49%–17.2%, and 0.76%–10.54% higher than the other methods, respectively. Among the nine types of ground objects included in this dataset, SPCLSR-DID achieved the highest accuracy for seven classes and 100% for *Bitumen*. Moreover, since the proposed SPCLSR-DID takes into account the imbalance problem between labeled samples, it also achieves excellent classification performance for some classes with a low proportion of labeled samples. For example, for the *Mixed weed*, SPCLSR-DID achieved the best accuracy of 86.04%. In comparison, the classification accuracies of SVMCK, KSR, and CCJSR are all lower than 70%, and the most competitive SP-DLRR method has an accuracy of only 82.66%. Similar to the other three datasets, the accuracy and smoothness of the classification map obtained by SPCLSR-DID are significantly better than the other comparison methods.

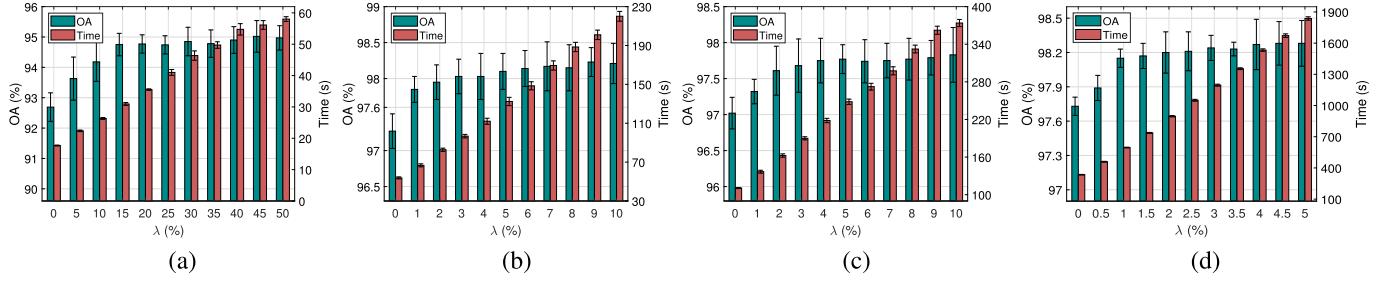


Fig. 8. Impact of incremental rate λ on classification accuracy and running time. (a) Indian Pines. (b) University of Pavia. (c) Salinas. (d) WHU-LongKou.

TABLE V
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS FOR INDIAN PINES (5% LABELED SAMPLES)

Class	SVMCK [60]	KJSR [61]	CCJSR [62]	SP-DLRR [29]	GLB-LRA [63]	LRSR-SCCE [48]	SLRC [44]	SPCLSR	SPCLSR-DID
1	25.00±35.70	82.27±8.43	83.49±14.09	100.00±0.00	91.86±8.59	78.37±20.57	93.95±9.20	93.49±4.26	100.00±0.00
2	82.25±3.11	91.43±3.00	89.08±3.18	93.64±2.16	96.98±1.79	92.01±2.51	93.78±2.02	90.64±1.08	95.98±1.23
3	90.10±4.23	91.65±2.93	87.03±1.67	97.01±1.24	98.91±0.95	90.34±3.25	93.01±3.04	82.83±1.91	95.57±2.00
4	61.07±10.35	90.55±3.20	81.29±7.01	90.18±8.64	90.22±9.18	95.02±3.69	93.82±3.30	81.64±4.98	96.18±2.33
5	90.33±3.28	90.30±5.53	93.25±2.96	98.21±1.90	92.88±4.78	93.49±3.24	90.28±4.36	94.34±2.35	97.21±1.78
6	98.44±0.98	94.75±2.74	93.85±2.09	98.40±1.39	98.34±1.63	96.77±1.67	98.67±0.85	98.83±0.63	100.00±0.00
7	0.00±0.00	66.30±20.78	85.38±11.38	100.00±0.00	93.08±10.15	95.77±6.76	99.23±2.31	98.46±1.88	100.00±0.00
8	99.54±0.53	97.54±1.91	98.85±0.89	100.00±0.00	100.00±0.00	99.76±0.33	100.00±0.00	99.69±0.78	100.00±0.00
9	0.00±0.00	6.32±14.47	43.16±17.89	100.00±0.00	90.53±9.65	70.53±18.71	70.00±20.12	77.37±15.08	93.68±16.61
10	85.99±2.66	91.14±1.75	91.24±2.10	95.40±3.47	96.93±3.15	91.14±1.43	91.93±3.47	89.45±1.88	95.33±1.78
11	94.94±2.56	95.59±1.13	92.88±1.31	95.35±3.12	98.83±0.71	94.19±0.84	96.32±1.30	92.81±0.81	99.13±0.55
12	80.07±4.30	87.99±3.98	81.10±5.01	92.50±4.83	96.34±2.14	88.17±4.73	89.47±5.12	89.98±1.59	98.19±1.35
13	95.64±2.85	90.06±5.89	87.68±4.02	99.33±0.24	90.88±10.28	96.80±3.32	99.28±1.20	99.54±0.43	100.00±0.00
14	98.90±0.78	98.40±1.29	96.71±2.07	99.63±0.36	99.93±0.03	98.86±0.97	99.73±0.48	99.65±0.16	100.00±0.00
15	92.89±4.95	90.71±5.19	83.06±5.79	97.70±2.20	96.94±4.28	95.46±2.48	92.54±4.65	91.12±3.44	98.25±2.34
16	84.77±9.05	88.30±6.89	83.75±7.69	87.05±13.10	75.34±18.49	93.98±5.33	95.68±4.31	97.16±2.40	99.09±0.52
OA	90.10±1.57	93.16±0.43	90.89±0.58	96.20±0.87	97.43±0.36	93.93±0.27	95.16±0.53	92.72±0.30	98.04±0.36
AA	73.75±3.38	84.58±1.73	85.74±2.08	96.52±1.00	94.25±1.60	91.92±1.13	93.61±1.00	92.31±0.93	98.04±0.92
$\kappa \times 100$	88.69±1.80	92.20±0.49	89.60±0.65	95.66±1.14	97.07±0.44	93.09±0.31	94.48±0.61	91.70±0.34	97.74±0.64
Time (s)	63	27	79	32	20	32	29	17	35

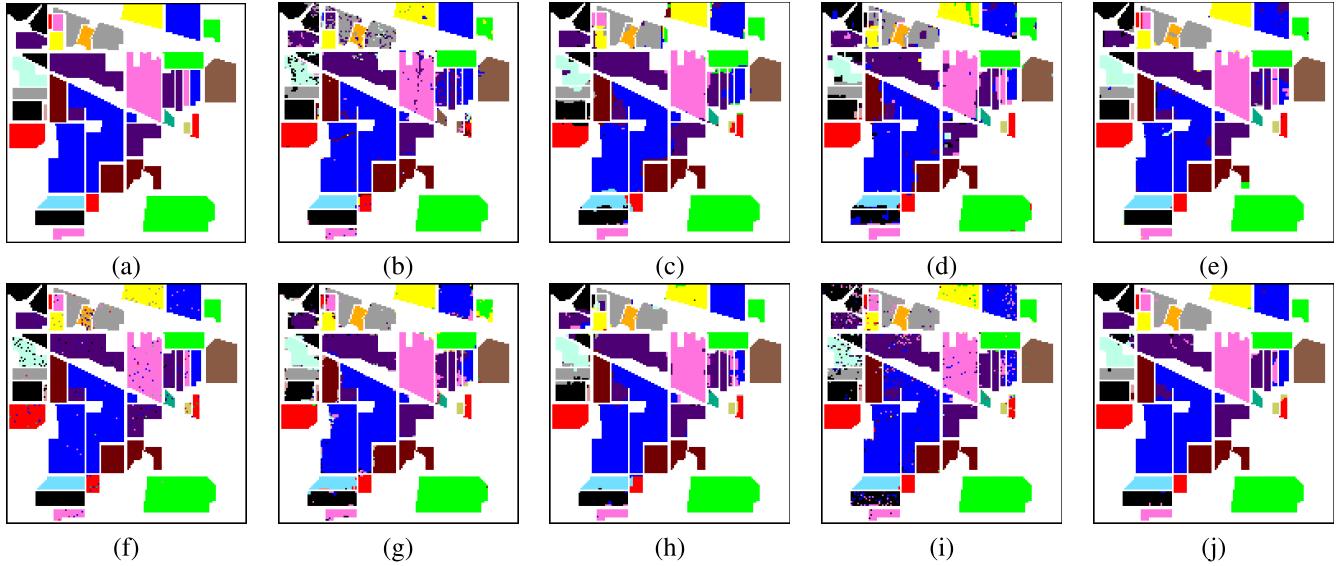


Fig. 9. Classification maps obtained by different methods for Indian Pines. (a) Ground truth. (b) SVMCK (90.10%). (c) KJSR (93.16%). (d) CCJSR (90.89%). (e) SP-DLRR (96.20%). (f) GLB-LRA (97.43%). (g) LRSR-SCCE (93.09%). (h) SLRC (95.16%). (i) SPCLSR (92.72%). (j) SPCLSR-DID (98.04%). The OAs are given in parentheses.

IV. DISCUSSION

A. Generalization Performance

To verify the generalization performance of the proposed SPCLSR-DID, we compare the classification accuracy of each method under different percentages of labeled samples. Specifically, we randomly select 1%–10% of the labeled samples

for the Indian Pines to construct the dictionary or training set, 0.2%–2% for the University of Pavia and Salinas, and 0.1%–1% for WHU-LongKou. As depicted in Fig. 13, the classification accuracy of each method increases smoothly with the increase in the number of labeled samples. Throughout the process, the OAs of SPCLSR-DID on the four datasets maintain the leading position, followed by GLB-LRA and

TABLE VI
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS FOR THE UNIVERSITY OF PAVIA (1% LABELED SAMPLES)

Class	SVMCK [60]	KJSR [61]	CCJSR [62]	SP-DLRR [29]	GLB-LRA [63]	LRSR-SCCE [48]	SLRC [44]	SPCLSR	SPCLSR-DID
1	93.79±1.71	83.77±1.27	61.15±3.26	97.05±1.06	94.84±1.65	95.41±1.16	97.19±1.50	93.93±1.53	98.92±0.80
2	98.63±0.60	97.25±0.73	98.24±0.59	99.66±0.20	99.11±0.54	99.26±0.20	99.94±0.07	99.65±0.10	99.98±0.07
3	94.35±1.58	44.36±3.49	76.91±4.57	94.98±9.12	92.13±4.21	93.42±1.79	91.95±3.52	96.40±1.05	98.53±2.15
4	99.47±0.24	73.45±6.30	77.22±3.58	88.59±3.35	91.08±4.22	91.84±1.02	94.47±0.87	94.84±1.03	97.93±0.40
5	65.53±14.81	82.33±4.91	99.03±0.68	99.11±0.24	99.68±0.28	99.95±0.03	99.17±1.23	99.83±0.07	99.96±0.05
6	93.96±3.03	68.58±3.13	77.21±3.43	98.29±0.59	98.82±0.86	92.59±1.23	94.82±1.48	94.75±1.07	99.55±0.47
7	95.31±2.53	55.74±4.74	93.20±4.77	96.95±1.98	97.01±2.59	94.16±1.36	99.94±0.07	95.84±0.92	99.98±0.05
8	95.14±2.32	33.43±1.93	86.80±4.70	96.57±0.93	90.29±2.53	87.66±2.14	96.76±1.81	95.43±1.18	99.07±1.32
9	61.04±19.21	25.09±2.99	22.40±2.08	99.81±0.11	97.67±1.94	95.90±1.02	95.46±0.42	98.07±1.33	98.47±0.41
OA	94.91±1.63	78.64±0.74	84.67±0.73	97.71±0.77	96.66±0.45	95.85±0.35	97.73±0.34	97.17±0.33	99.43±0.22
AA	88.58±4.03	62.67±1.28	76.91±1.03	96.78±1.44	95.63±0.78	94.46±0.40	96.63±0.50	96.53±0.29	99.15±0.29
$\kappa \times 100$	93.27±2.15	70.68±1.11	79.49±0.99	96.95±1.03	95.56±0.60	94.48±0.47	96.98±0.45	96.24±0.44	99.25±0.29
Time (s)	96	150	277	93	75	97	105	54	132

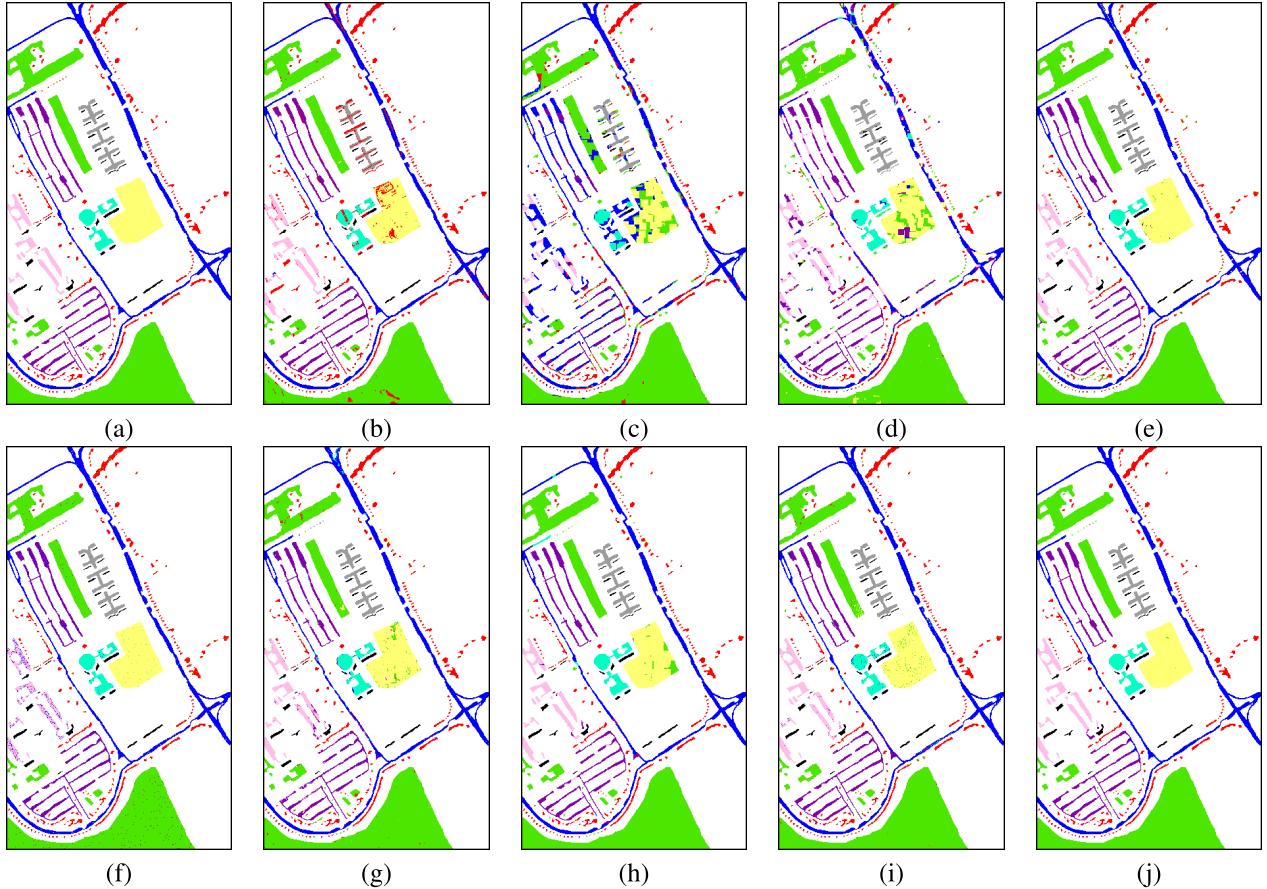


Fig. 10. Classification maps obtained by different methods for the University of Pavia. (a) Ground truth. (b) SVMCK (94.91%). (c) KJSR (78.64%). (d) CCJSR (84.67%). (e) SP-DLRR (97.71%). (f) GLB-LRA (96.66%). (g) LRSR-SCCE (95.85%). (h) SLRC (97.73%). (i) SPCLSR (97.17%). (j) SPCLSR-DID (99.43%). The OAs are given in parentheses.

SP-DLRR. This result demonstrates the better classification performance of SPCLSR-DID. Moreover, the advantages of the proposed method are more pronounced when the number of labeled samples used is relatively small. For example, when using only 0.2% of the labeled samples, SPCLSR-DID achieved OAs of 95.82% and 97.56% on the University of Pavia and Salinas datasets, respectively, which represents improvements of 4.68%–32.50% and 2.20%–12.34% over the comparison methods. For Indian Pines, the lead of SPCLSR-DID is significant when the percentage of labeled samples utilized ranges from 2% to 5%. Overall, the experiment demonstrates the remarkable generalization performance of the proposed method.

B. Ablation Analysis

To further investigate the factors that affect the classification performance of SPCLSR-DID, we present a series of variant methods including: 1) both the structural prior constraint and reconstruction error constraint are eliminated (Variant-1, V1); 2) preserve the structural prior constraint while removing the reconstruction error constraint (Variant-2, V2); 3) both DID and the contextual information are eliminated (Variant-3, V3); 4) preserve DID without combining the contextual information (Variant-4, V4); and 5) preserve the contextual information while removing DID (Variant-5, V5).

The classification accuracies obtained by each variant method are shown in Fig. 14. Among them, V1 performs

TABLE VII
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS FOR SALINAS (1% LABELED SAMPLES)

Class	SVMCK [60]	KJSR [61]	CCJSR [62]	SP-DLRR [29]	GLB-LRA [63]	LRSR-SCCE [48]	SLRC [44]	SPCLSR	SPCLSR-DID
1	98.98±0.70	100.00±0.00	99.99±0.03	99.95±0.08	99.42±0.50	99.41±0.32	100.00±0.00	99.49±0.79	100.00±0.00
2	96.38±3.01	99.97±0.04	99.67±0.22	99.37±0.72	98.79±1.84	99.84±0.09	99.83±0.09	99.98±0.02	100.00±0.00
3	98.13±1.89	98.29±2.87	97.86±2.50	99.44±1.00	99.47±0.61	98.68±1.80	99.31±0.93	97.60±1.80	97.59±2.54
4	97.00±2.14	73.45±3.55	84.47±1.88	97.62±4.27	96.52±2.27	98.37±0.79	89.36±2.42	98.90±1.03	98.52±0.48
5	97.74±0.75	93.06±1.03	95.90±0.67	98.69±0.36	92.31±2.39	97.40±0.99	95.46±1.23	98.84±0.56	99.54±0.26
6	98.11±1.78	95.88±0.33	99.21±0.39	99.63±0.63	99.49±0.30	99.92±0.15	99.86±0.13	99.86±0.06	99.79±0.06
7	98.44±0.95	100.00±0.00	99.89±0.15	99.19±0.54	98.57±0.98	99.34±0.39	99.69±0.13	99.86±0.06	99.99±0.02
8	93.42±1.42	97.38±1.42	85.31±2.83	99.15±0.92	99.03±0.92	92.36±1.52	95.23±0.66	94.60±1.07	99.62±0.23
9	98.77±0.94	100.00±0.00	99.99±0.02	99.92±0.05	99.91±0.20	99.90±0.08	99.93±0.06	99.98±0.02	100.00±0.00
10	94.47±4.11	93.78±0.95	96.02±1.24	94.80±2.83	96.52±1.76	95.46±1.26	97.65±1.21	95.40±1.24	98.01±0.98
11	93.58±3.86	93.15±1.42	99.65±0.26	95.78±4.13	83.25±11.07	96.77±1.79	97.49±2.05	97.85±0.93	99.80±0.44
12	99.08±1.65	92.56±2.41	95.89±1.20	99.71±0.46	98.59±0.96	99.77±0.38	96.09±0.86	99.22±0.81	99.98±0.05
13	94.56±3.96	90.96±5.73	94.01±2.55	98.01±1.00	89.04±5.59	98.32±1.17	98.95±0.32	98.26±0.46	99.97±0.10
14	87.01±6.23	99.54±0.67	98.56±0.40	89.99±7.15	91.53±6.54	96.08±2.81	94.96±2.21	93.33±1.16	97.43±1.73
15	89.16±2.86	81.13±3.78	78.27±3.26	98.74±1.24	98.75±0.82	89.87±3.58	92.73±2.24	90.95±1.48	98.69±0.78
16	93.96±5.03	99.54±0.47	99.04±0.79	98.89±0.30	99.16±1.62	98.75±0.66	98.82±0.53	99.36±0.17	99.80±0.28
OA	95.20±1.22	94.57±0.36	92.70±0.69	98.69±0.32	97.92±0.44	96.24±0.28	96.97±0.25	96.91±0.28	99.40±0.20
AA	95.55±1.64	94.29±0.28	95.23±0.33	98.06±0.59	96.27±1.02	97.52±0.20	97.21±0.25	97.72±0.21	99.30±0.23
$\kappa \times 100$	94.64±1.37	93.95±0.40	91.87±0.77	98.55±0.36	97.68±0.49	95.82±0.31	96.63±0.28	96.56±0.31	99.33±0.22
Time (s)	121	191	476	114	133	215	286	110	248

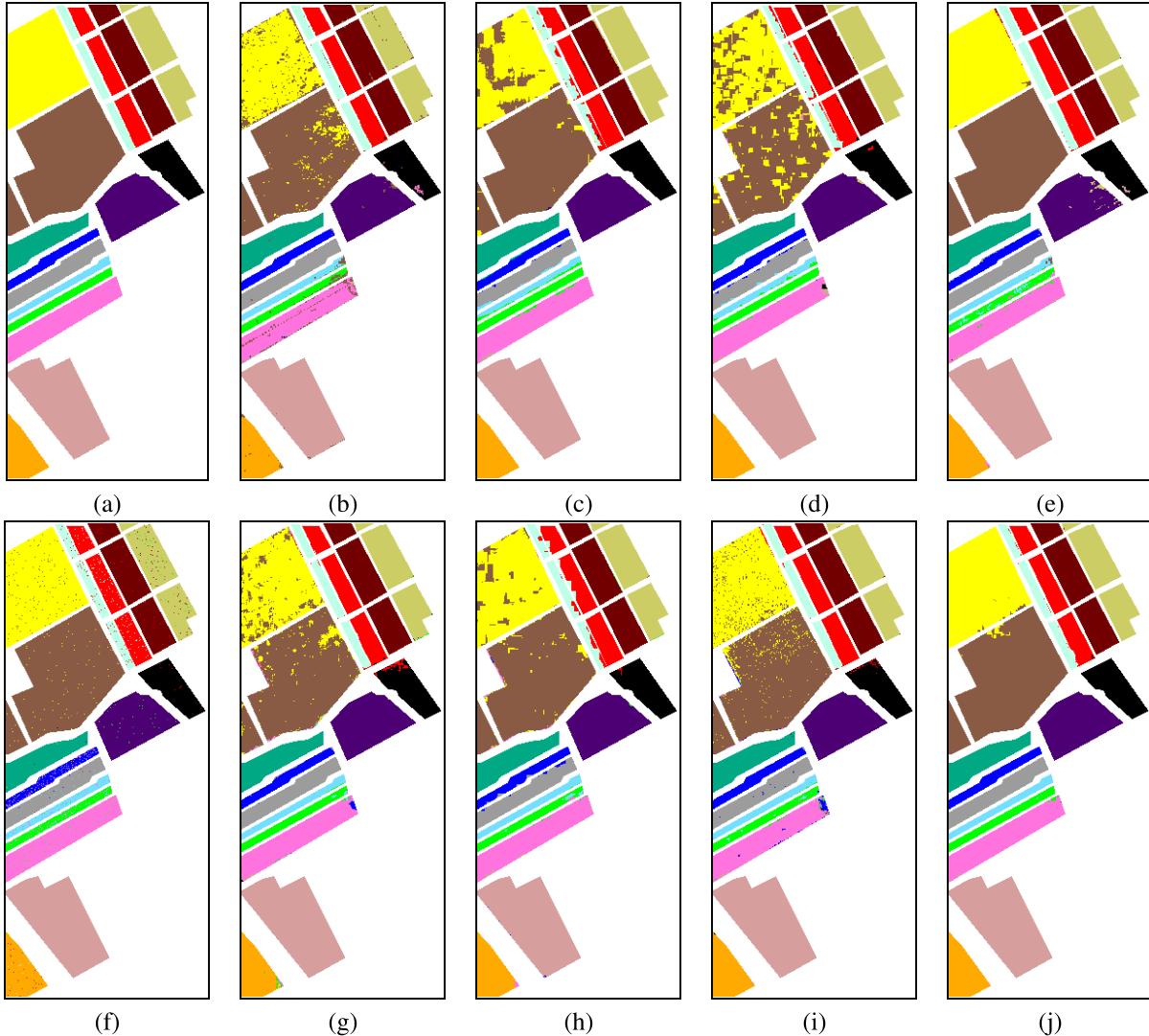


Fig. 11. Classification maps obtained by different methods for Salinas. (a) Ground truth. (b) SVMCK (95.20%). (c) KJSR (94.57%). (d) CCJSR (92.70%). (e) SP-DLRR (98.69%). (f) GLB-LRA (97.92%). (g) LRSR-SCCE (96.24%). (h) SLRC (96.97%). (i) SPCLSR (96.91%). (j) SPCLSR-DID (99.40%). The OAs are given in parentheses.

the worst, followed by V2. It proves that the structural prior constraint and the reconstruction error constraint in SPCLSR-DID are extremely important for improving

classification performance. As expected, the accuracies of V4 and V5 are higher than that of V3. Moreover, V4 is better than V5 in general, indicating that the improvement brought

TABLE VIII
CLASSIFICATION ACCURACIES OF DIFFERENT METHODS FOR WHU-LONGKOU (0.5% LABELED SAMPLES)

Class	SVMCK [60]	KJSR [61]	CCJSR [62]	SP-DLRR [29]	GLB-LRA [63]	LRSR-SCCE [48]	SLRC [44]	SPCLSR	SPCLSR-DID
1	98.25±0.94	90.80±1.47	93.82±0.61	99.45±0.53	99.10±0.30	99.55±0.10	98.39±0.44	99.54±0.13	99.91±0.06
2	84.56±5.67	84.31±2.82	87.51±1.80	95.93±2.13	98.49±1.20	93.15±2.76	69.42±3.98	95.38±1.95	98.60±1.38
3	71.07±11.33	93.81±2.24	88.15±1.21	94.36±6.24	96.03±1.82	92.63±2.34	71.06±7.23	91.66±2.30	94.93±2.75
4	97.43±0.75	99.27±0.17	98.06±0.36	98.13±0.47	99.11±0.33	96.88±0.37	93.62±0.57	96.79±0.22	98.44±0.54
5	80.81±6.12	70.62±7.43	82.15±1.92	96.42±3.28	82.68±5.20	92.62±1.39	77.43±4.15	92.16±0.76	98.56±0.37
6	96.30±2.70	90.28±2.31	95.87±0.45	99.62±0.42	98.01±1.26	99.19±0.23	99.28±0.61	99.76±0.08	99.95±0.05
7	99.87±0.07	93.57±0.55	98.68±0.10	99.95±0.01	99.80±0.15	99.87±0.04	99.94±0.02	99.97±0.01	100.00±0.00
8	88.68±6.18	29.92±4.01	91.41±2.14	94.55±2.74	90.92±1.88	90.74±1.84	91.26±1.19	94.37±0.83	98.05±0.79
9	69.67±11.22	67.03±4.88	65.14±2.64	82.66±5.39	78.93±5.43	80.32±3.04	74.09±2.41	80.20±1.53	86.04±2.49
OA	96.03±0.71	90.94±0.31	95.45±0.18	98.33±0.32	98.06±0.16	97.51±0.16	94.59±0.35	97.73±0.08	98.91±0.20
AA	87.40±2.10	79.96±1.22	88.98±0.45	95.67±1.22	93.67±0.65	93.88±0.37	86.05±1.16	94.43±0.43	97.16±0.40
$\kappa \times 100$	94.76±0.95	88.03±0.42	94±0.23	97.81±0.42	97.45±0.22	96.73±0.21	92.9±0.46	97.02±0.11	98.57±0.27
Time (s)	357	834	2758	337	473	1282	1475	355	897

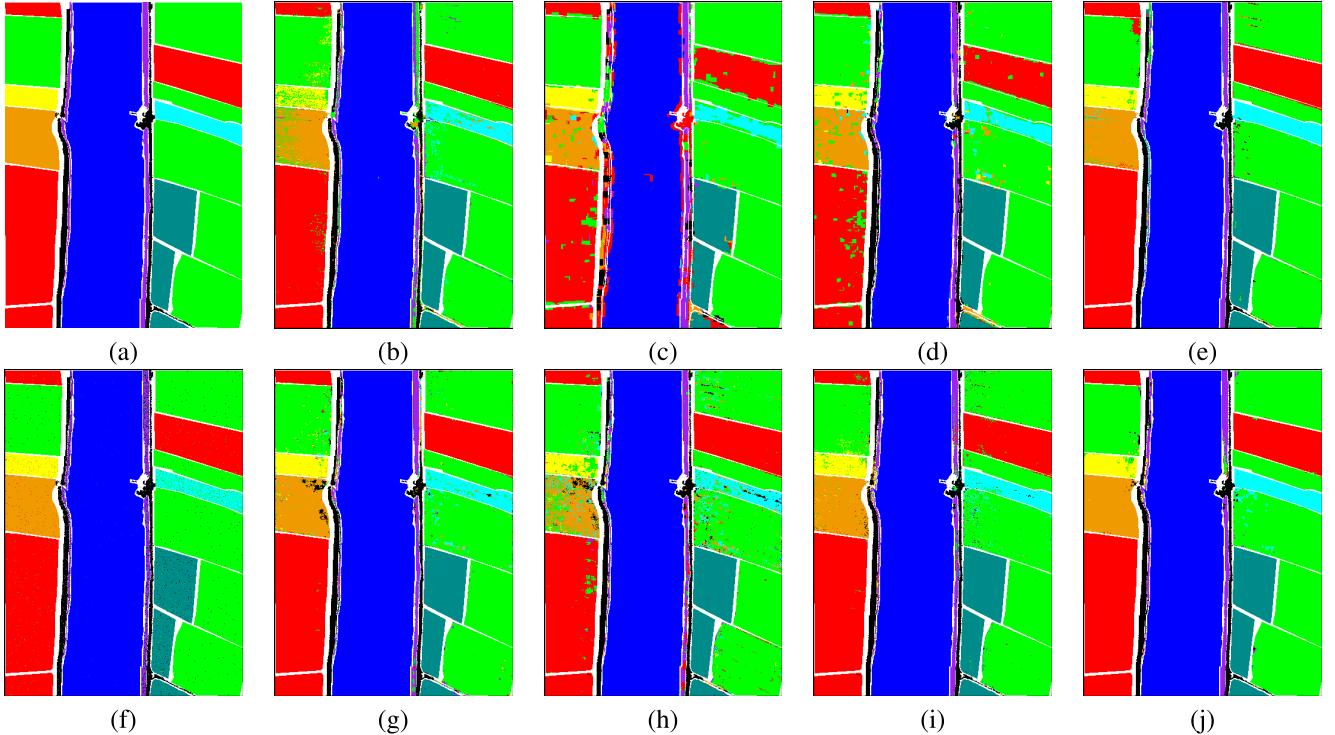


Fig. 12. Classification maps obtained by different methods for WHU-LongKou. (a) Ground truth. (b) SVMCK (96.03%). (c) KJSR (90.94%). (d) CCJSR (95.45%). (e) SP-DLRR (98.33%). (f) GLB-LRA (98.06%). (g) LRSR-SCCE (97.51%). (h) SLRC (94.59%). (i) SPCLSR (97.73%). (j) SPCLSR-DID (98.91%). The OAs are given in parentheses.

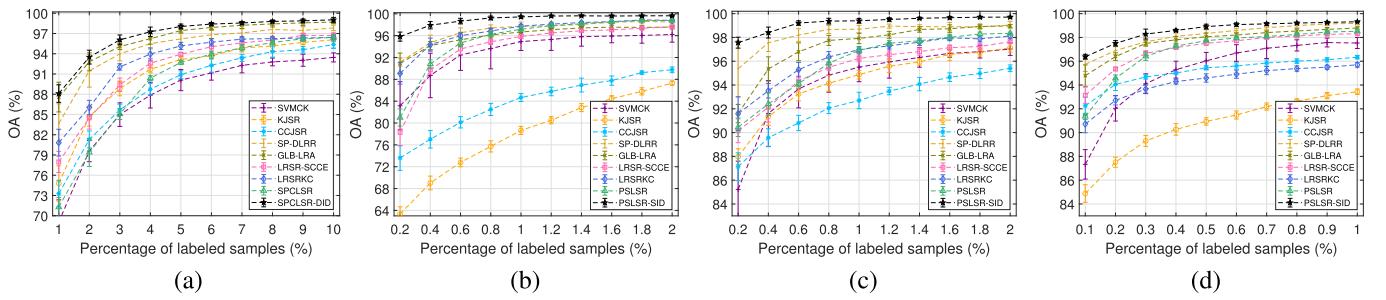


Fig. 13. Evolution of OAs obtained by different methods as a function of different percentages of labeled samples. (a) Indian Pines. (b) University of Pavia. (c) Salinas. (d) WHU-LongKou.

by DID is higher than that of the contextual information. The reason is that DID learning effectively enhances the completeness of the dictionary, and a dictionary with better completeness leads to superior classification performance for the model. For the proposed SPCLSR-DID method, OA, AA,

and κ are the highest on the four datasets. In addition, SPCLSR-DID exhibited the smallest standard deviation across ten random sampling experiments, demonstrating its excellent stability in classification. In summary, each component of SPCLSR-DID jointly affects the final classification accuracy,

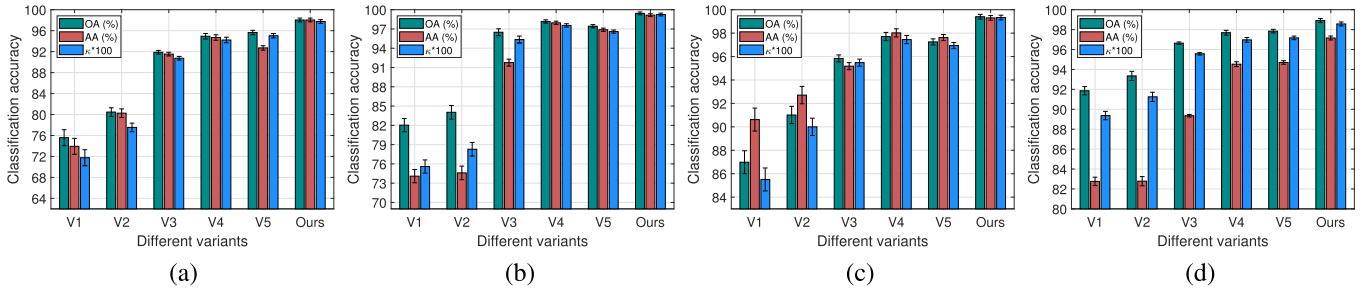


Fig. 14. Comparison of different variants of SPCLSR-DID in terms of OA, AA, and κ . (a) Indian Pines. (b) University of Pavia. (c) Salinas. (d) WHU-LongKou. V1 represents that both the structural prior constraint and reconstruction error constraint are eliminated; V2 represents that the reconstruction error constraint is eliminated; V3 represents that both the contextual information and DID are eliminated; V4 represents that the contextual information is eliminated; and V5 represents that DID is eliminated.

TABLE IX

NUMBER AND CORRECTNESS OF AUGMENTED ATOMS
INTRODUCED BY DID LEARNING FOR INDIAN PINES

Class	ID-B		ID-A ($\lambda=20\%$)	
	Introduced	Correct	Introduced	Correct
Alfalfa	35	35	6	6
Corn-notill	854	854	137	137
Corn-mintill	328	324	53	53
Corn	84	84	14	14
Grass/pasture	365	365	59	59
Grass/trees	643	643	103	103
Grass/pasture-mowed	19	19	4	4
Hay-windrowed	454	454	73	73
Oats	6	6	1	1
Soybean-notill	544	544	87	87
Soybean-mintill	1,617	1,608	259	257
Soybean-clean	344	344	55	55
Wheat	194	194	31	31
Woods	1,175	1,175	188	188
Buildings-Grass-Trees-Drives	193	193	31	31
Stone-Steel-Towers	72	72	12	12
Total	6,927	6,914	1,113	1,111
Accuracy	99.81%		99.82%	

TABLE XI

NUMBER AND CORRECTNESS OF AUGMENTED ATOMS
INTRODUCED BY DID LEARNING FOR SALINAS

Class	ID-B		ID-A ($\lambda=5\%$)	
	Introduced	Correct	Introduced	Correct
Brocoli-weeds-1	1,988	1,988	80	80
Brocoli-weeds-2	3,688	3,688	148	148
Fallow	1,624	1,624	65	65
Fallow-rough-plow	708	707	29	29
Fallow-smooth	1,751	1,751	70	70
Stubble	3,536	3,536	142	142
Celery	3,460	3,460	139	139
Grapes-untrained	5,790	5,790	232	232
Soil-vinyard-develop	6,100	6,100	245	245
Corn-senesced-weeds	1,997	1,997	80	80
Lettuce-romaine-4wk	528	528	22	22
Lettuce-romaine-5wk	1,273	1,273	51	51
Lettuce-romaine-6wk	356	356	15	15
Lettuce-romaine-7wk	229	229	10	10
Vinyard-untrained	848	848	34	34
Vinyard-vertical-trellis	1,704	1,704	69	69
Total	35,580	35,579	1,431	1,431
Accuracy	99.99%		100%	

TABLE X

NUMBER AND CORRECTNESS OF AUGMENTED ATOMS INTRODUCED BY DID LEARNING FOR THE UNIVERSITY OF PAVIA

Class	ID-B		ID-A ($\lambda=5\%$)	
	Introduced	Correct	Introduced	Correct
Asphalt	3,672	3,671	147	147
Meadows	17,077	17,075	684	684
Gravel	1,297	1,296	52	52
Trees	2,021	2,021	81	81
Metal Sheets	879	879	36	36
Bare Soil	1,561	1,561	63	63
Bitumen	398	398	16	16
Bricks	2,111	2,111	85	85
Shadows	649	649	26	26
Total	29,665	29,661	1,190	1,190
Accuracy	99.99%		100%	

which further validates the effectiveness of the proposed SPCLSR-DID.

C. Effectiveness of DID

In this discussion, to evaluate the reliability of the augmented atoms introduced by the DID learning, we analyze the number and correctness of augmented atoms for each class, and the specific results are reported in Tables IX–XII. In these tables, ID-B (ID-before) and ID-A (ID-after), respectively, represent the ID before and after optimization. The important point to note is that, as stated in the Experimental Settings

section, the removal rate for DID is set to 10%. This implies that during the construction of ID-A from ID-B, the top and bottom 10% of augmented atoms with the highest and lowest correlation coefficients will be optimized out. In this context, the specific relationship between the number of ID-A (N_c^A) and ID-B (N_c^B) in the c th class is $N_c^A = 0.8 \times N_c^B \times \lambda$.

Then, as shown in Fig. 15, we plot the average correlation coefficients for each augmented atom for each class in different datasets. This provides a visual presentation of the relationship between the correctness of augmented atoms and their average correlation coefficients. Moreover, in order to

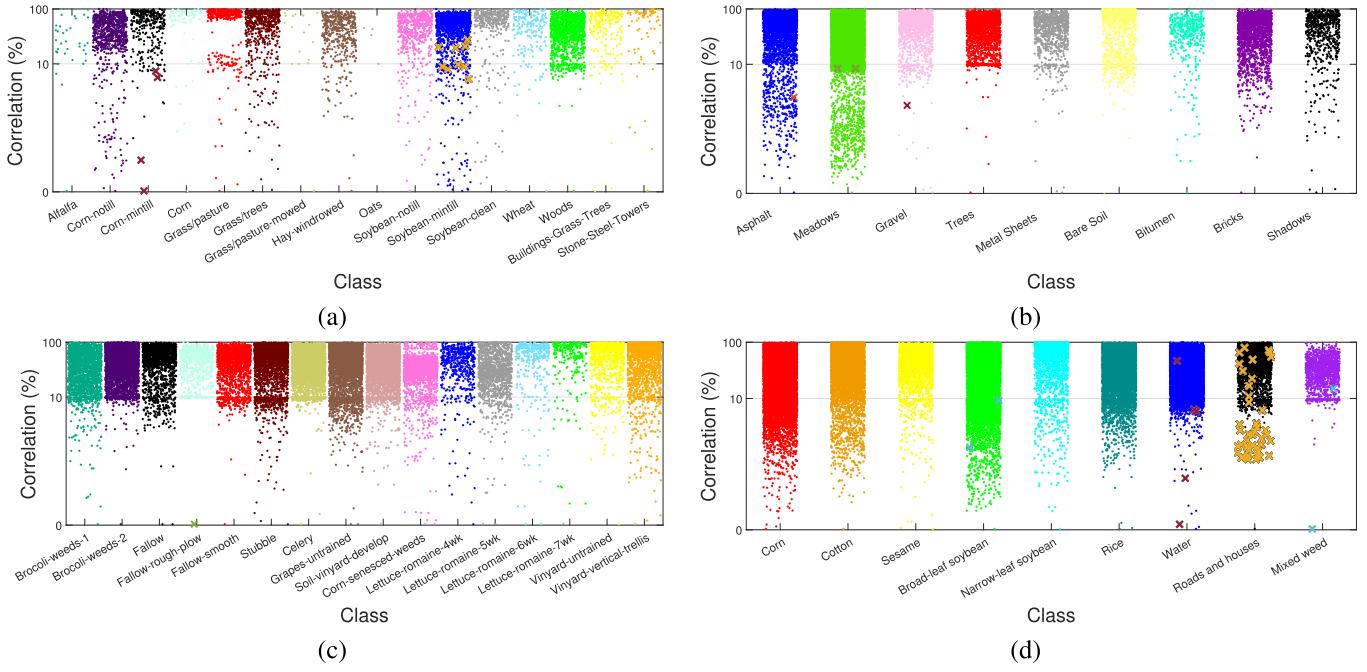


Fig. 15. Statistical analysis between each class of augmented samples and the average correlation coefficients in SPCLSR-DID. (a) Indian Pines. (b) University of Pavia. (c) Salinas. (d) WUH-LongKou. Cross marks indicate the incorrectly augmented atoms, and the color of the cross mark is a representation of the correct class to which it belongs. The position of the 10% scale on the y-axis is adjusted, and a horizontal line is added to better illustrate the distribution of incorrect augmented atoms under low correlation.

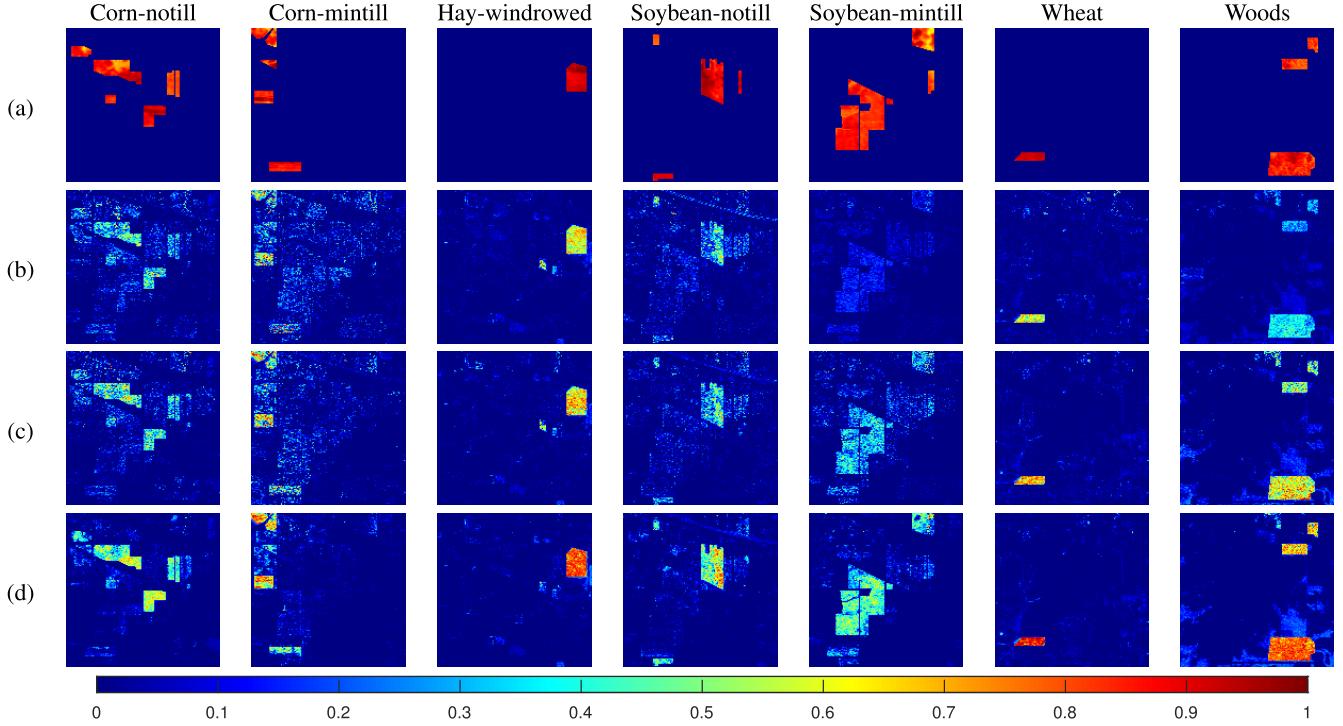


Fig. 16. Subspace feature maps of different classes on the Indian Pines dataset. (a) Ground truth. (b) Subspace feature maps corresponding to the original LRSR model. (c) Subspace feature maps corresponding to the SPCLSR model. (d) Subspace feature maps corresponding to the SPCLSR-DID model.

better illustrate the distribution of wrong augmented atoms under low correlation, we adjust the position of the 10% scale on the y-axis and add a horizontal line. During the subsequent DID learning process, all augmented atoms located below this horizontal line will be excluded from consideration.

- 1) For Indian Pines, there are 6927 atoms in ID-B, 13 of which are incorrectly introduced and mainly distributed

in classes *Corn-mintill* and *Soybean-mintill*. As shown in Fig. 15(a), the incorrectly augmented atoms in *Corn-mintill* are in the bottom 10% interval with low average correlation, and all of them can be removed in the process of DID learning. Then, the same is true for most incorrectly introduced atoms in *Soybean-mintill*. In the case of $\lambda = 20\%$, a total of 1113 augmented atoms

are introduced into the ID-A, and only two atoms are incorrect.

- 2) For the University of Pavia, four out of the 29 665 augmented atoms introduced in the ID-B are incorrect. There is one incorrect atom in both *Asphalt* and *Gravel*, and there are two incorrect atoms in class *Meadows*. As described in Fig. 15(b), sorted by average correlation, the correlations of all misintroduced atoms are in the bottom 10% interval and do not affect the correctness of atoms in ID-A. Therefore, in the case of $\lambda=5\%$, the 1190 augmented atoms introduced in ID-A achieve 100% correctness.
- 3) For Salinas, the augmented atoms of all classes have extremely high reliability, and there is only one incorrect atom. As clearly shown in Fig. 15(c), the incorrect augmented atom is in class *Fallow-rough-plow*, and the corresponding average correlation of the incorrect atom with the subdictionary atoms of *Fallow-rough-plow* is quite low. Similar to the University of Pavia, in the case of $\lambda=5\%$, the ID-A of Salinas contains 1431 augmented atoms, and the correctness is also 100%.
- 4) For WHU-LongKou, 182 994 augmented atoms are introduced in ID-B, with an accuracy of 99.97%. The wrong augmented atoms are mainly concentrated in the class of *Roads and houses*, with 44 of the 2878 augmented atoms being wrong. It can be found from Fig. 15(d) that most of the wrong atoms have a relatively low correlation, which can be easily removed. Then, in the case of $\lambda = 1\%$, 1470 discriminative augmented atoms are retained, and the accuracy is also 100%, which is the same as the University of Pavia and Salinas.

In addition, it can be observed from Fig. 15(a)–(d) that the augmented atoms within the top 90% interval of the correlation in the four datasets all exhibit remarkably high accuracy. This observation further illustrates that a higher average correlation between the augmented atom and its corresponding subdictionary atoms correlates with increased correctness of the atom. Therefore, removing low-correlation augmented atoms can effectively ensure the reliability of the resulting ID.

D. Subspace Feature Analysis

In this section, our objective is to validate the representation ability of the proposed SPCLSR-DID model for subspace features and the separability of these features within the subspace.

1) *Representation Ability*: We take the Indian Pines dataset as an example to verify the representation ability of SPCLSR-DID for subspace features. As depicted in Fig. 16, the restored HSI data for each subdictionary are first compressed by the principal component analysis (PCA) method, and then, the first principal component feature is displayed via a heatmap. Comparing Fig. 16(b) and (c), it can be clearly found that the contours of features for SPCLSR are more precise and complete than those of LRSR, especially for *Corn-notill* and *Corn-mintill*. Subsequently, as shown in Fig. 16(c), we can see that the overall smoothness and integrity of the features for SPCLSR-DID are further improved, especially for *Soybean-notill* and *Woods*. Moreover, compared

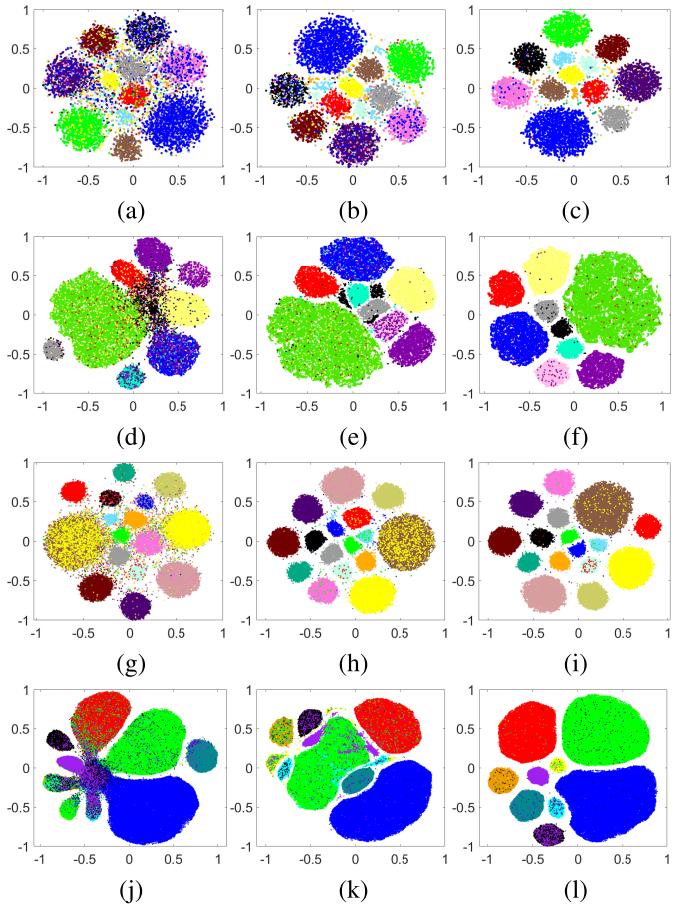


Fig. 17. Subspace feature separability for different methods in the four datasets. (a) LRSR, (b) SPCLSR, and (c) SPCLSR-DID for Indian Pines. (d) LRSR, (e) SPCLSR, and (f) SPCLSR-DID for the University of Pavia. (g) LRSR, (h) SPCLSR, and (i) SPCLSR-DID for Salinas. (j) LRSR, (k) SPCLSR, and (l) SPCLSR-DID for WHU-LongKou.

with LRSR and SPCLSR, the features of SPCLSR-DID have the sharpest contours. The reason is that the completeness of the dictionary is significantly enhanced after augmentation; thence, the representation ability of the model has also improved accordingly.

2) *Feature Separability*: The t-distributed stochastic neighbor embedding (t-SNE) method is employed to transform high-dimensional subspace features into a two-dimensional space, aiming to depict the separability between features. As depicted in Fig. 17, owing to the effectiveness of the prior structural constraint, the separability of subspace features in the SPCLSR method is significantly better than the basic LRSR. In comparison, it can also be clearly seen that the feature separability of SPCLSR-DID is significantly superior to SPCLSR, and this improvement is attributed to DID learning. In SPCLSR-DID, the augmented atoms obtained by the DID learning can be dedicated to constructing a more optimal structural prior constraint, thus indirectly enhancing the separability of subspace features.

E. Convergence Analysis

In this section, we investigate the optimization performance of ADMM by analyzing the evolution of the fitting residual with respect to the number of iterations.

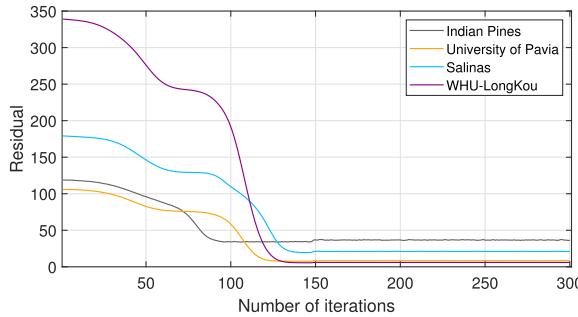


Fig. 18. Evolution of fitting residual as a function of the number of iterations for four datasets.

According to (7), in the process of solving the proposed SPCLSR-DID model by ADMM, three equality constraints are considered. Then, in the iterative optimization process, the fitting residual of the model (denoted as res) can be calculated as follows:

$$res = \sqrt{\|res_1\|_F^2 + \|res_2\|_F^2 + \|res_3\|_F^2} \quad (29)$$

where res_1 , res_2 , and res_3 represent $\mathbf{Y} - [\mathbf{D}, \bar{\mathbf{D}}]\mathbf{X} - \mathbf{E}$, $\mathbf{X} - \mathbf{X}_1$, and $\mathbf{X} - \mathbf{X}_2$, respectively. The evolution of the fitting residuals with the number of iterations is shown in Fig. 18, where we set the maximum number of iterations to 300 for four datasets. It can be observed that with the increase in the number of iterations, the fitting residuals sharply decrease and tend to converge around 150 iterations. As described in Algorithm 1, we set the maximum number of iterations to 200, ensuring that the ADMM algorithm achieves optimal optimization performance for four datasets.

V. CONCLUSION

In this article, a novel SPCLSR-DID model is proposed for HSI classification. On the basis of simultaneously maintaining the global and local data structures, this method further explores the intrinsic spectral characteristic and spatial structure information of HSI through the prior structural constraint and has a stronger subspace segmentation ability than the original LRSR model. Moreover, the presented DID scheme constructs a controllable ID by introducing augmented atoms to improve the representation ability of the dictionary effectively, and the way of integrating context information for classification ensures the accuracy of the final result and the smoothness of the map. For classification performance, experimental results on four HSI datasets show that our method can achieve better accuracy than related comparison methods. From the result of the ablation analysis, each part of SPCLSR-IDL jointly affects the final classification result, which verifies the effectiveness of the proposed method. In addition, the analysis of subspace features verifies that the imposed prior structural constraint enhances the subspace segmentation performance of LRSR, and the experiment on augmented atoms shows that the atoms obtained by DID learning are extremely reliable.

Although we have obtained encouraging experimental results, there are still some unresolved issues, and we are also planning to further study the method from two perspectives:

1) attempt to extend the idea of prior structural constraint into the tensor-based LRSR models to further explore the 3-D spatial structure of HSI and 2) integrate metric learning methods to learn discriminative augmented atoms more efficiently.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe for making the Airborne Visible/Infrared Imaging Spectrometer Indian Pines hyperspectral dataset available to the community, Prof. P. Gamba for providing the Reflective Optics Spectrographic Imaging System data over Pavia, Italy, and Prof. Y. F. Zhong for providing the WHU-LongKou hyperspectral dataset. They would also like to thank Dr. B. Tu for sharing the code of correlation coefficient and joint sparse representation (CCJSR), Dr. S. Yang for superpixel-guided discriminative low-rank representation (SP-DLRR), and Dr. H. Liu for global-local balanced low-rank approximation (GLB-LRA), respectively.

REFERENCES

- [1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
- [2] M. A. Moharram and D. M. Sundaram, "Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions," *Neurocomputing*, vol. 536, pp. 90–113, Jun. 2023.
- [3] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021.
- [4] N. Wambugu et al., "Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102603.
- [5] D. Datta, P. K. Mallick, A. K. Bhoi, M. F. Ijaz, J. Shafi, and J. Choi, "Hyperspectral image classification: Potentials, challenges, and future directions," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–36, Apr. 2022.
- [6] J. Wang, M. Bretz, M. A. A. Dewan, and M. A. Delavar, "Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects," *Sci. Total Environ.*, vol. 822, May 2022, Art. no. 153559.
- [7] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022.
- [8] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 4, pp. 213–247, Dec. 2022.
- [9] M. E. Paoletti, O. Mogollon-Gutierrez, S. Moreno-Álvarez, J. C. Sancho, and J. M. Haut, "A comprehensive survey of imbalance correction techniques for hyperspectral data classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5297–5314, 2023.
- [10] J. Peng et al., "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 10–43, Mar. 2022.
- [11] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
- [12] 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, Jan. 2013.
- [13] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [14] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, "Discriminative transfer subspace learning via low-rank and sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 850–863, Feb. 2016.
- [15] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

- [16] J. Wang, L. Jiao, H. Liu, S. Yang, and F. Liu, "Hyperspectral image classification by spatial-spectral derivative-aided kernel joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2485–2500, Jun. 2015.
- [17] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.
- [18] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.
- [19] Z. Xue, P. Du, J. Li, and H. Su, "Sparse graph regularization for robust crop mapping using hyperspectral remotely sensed imagery with very few in situ data," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 1–15, Feb. 2017.
- [20] Z. Xue, S. Yang, and L. Zhang, "Weighted sparse graph regularization for spectral-spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1630–1634, Sep. 2021.
- [21] W. Fu, S. Li, L. Fang, and J. A. Benediktsson, "Contextual online dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1336–1347, Mar. 2018.
- [22] T. Dundar and T. Ince, "Sparse representation-based hyperspectral image classification using multiscale superpixels and guided filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 246–250, Feb. 2019.
- [23] X. Tu, X. Shen, P. Fu, T. Wang, Q. Sun, and Z. Ji, "Discriminant sub-dictionary learning with adaptive multiscale superpixel representation for hyperspectral image classification," *Neurocomputing*, vol. 409, pp. 131–145, Oct. 2020.
- [24] Y. Duan, H. Huang, Z. Li, and Y. Tang, "Local manifold-based sparse discriminant learning for feature extraction of hyperspectral image," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 4021–4034, Aug. 2021.
- [25] Y. Duan, H. Huang, and T. Wang, "Semisupervised feature extraction of hyperspectral image using nonlinear geodesic sparse hypergraphs," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515115.
- [26] W. Li, J. Liu, and Q. Du, "Sparse and low-rank graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4094–4105, Jul. 2016.
- [27] S. Mei, J. Hou, J. Chen, L.-P. Chau, and Q. Du, "Simultaneous spatial and spectral low-rank representation of hyperspectral images for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2872–2886, May 2018.
- [28] H. Su, W. Yao, Z. Wu, P. Zheng, and Q. Du, "Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 238–252, Jan. 2021.
- [29] S. Yang, J. Hou, Y. Jia, S. Mei, and Q. Du, "Superpixel-guided discriminative low-rank representation of hyperspectral images for classification," *IEEE Trans. Image Process.*, vol. 30, pp. 8823–8835, 2021.
- [30] H. Sun, M. Liu, K. Zheng, D. Yang, J. Li, and L. Gao, "Hyperspectral image denoising via low-rank representation and CNN denoiser," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 716–728, 2022.
- [31] J. Huang, T.-Z. Huang, L.-J. Deng, and X.-L. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2419–2438, Apr. 2019.
- [32] L. Wu, J. Huang, and M.-S. Guo, "Multidimensional low-rank representation for sparse hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [33] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.
- [34] W. Sun, J. Peng, G. Yang, and Q. Du, "Fast and latent low-rank subspace clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3906–3915, Jun. 2020.
- [35] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.
- [36] X. He, J. Wu, Q. Ling, Z. Li, Z. Lin, and S. Zhou, "Anomaly detection for hyperspectral imagery via tensor low-rank approximation with multiple subspace learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509917.
- [37] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2021.
- [38] Q. Ran et al., "Anomaly detection for hyperspectral images based on improved low-rank and sparse representation and joint Gaussian mixture distribution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6339–6352, 2021.
- [39] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, "Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 973–988, Mar. 2019.
- [40] B. Zhao, M. O. Ulfarsson, J. R. Steinsson, and J. Chanussot, "Hyperspectral image denoising using spectral-spatial transform-based sparse and low-rank representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522125.
- [41] M. Wang et al., "Tensor decompositions for hyperspectral data processing in remote sensing: A comprehensive review," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 1, pp. 26–72, Mar. 2023.
- [42] W. Zhang et al., "An improved feature set for hyperspectral image classification: Harmonic analysis optimized by multiscale guided filter," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3903–3916, 2020.
- [43] X. Song et al., "Unsupervised robust projection learning by low-rank and sparse decomposition for hyperspectral feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [44] Y. Ding, Y. Chong, and S. Pan, "Sparse and low-rank representation with key connectivity for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5609–5622, 2020.
- [45] F. Feng, Y. Zhang, J. Zhang, and B. Liu, "Low-rank constrained attention-enhanced multiple spatial-spectral feature fusion for small sample hyperspectral image classification," *Remote Sens.*, vol. 15, no. 2, p. 304, Jan. 2023.
- [46] K. Tang, R. Liu, Z. Su, and J. Zhang, "Structure-constrained low-rank representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2167–2179, Dec. 2014.
- [47] Y. Liu, X. Li, C. Liu, and H. Liu, "Structure-constrained low-rank and partial sparse representation with sample selection for image classification," *Pattern Recognit.*, vol. 59, pp. 5–13, Nov. 2016.
- [48] L. Pan, H.-C. Li, H. Meng, W. Li, Q. Du, and W. J. Emery, "Hyperspectral image classification via low-rank and sparse representation with spectral consistency constraint," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2117–2121, Nov. 2017.
- [49] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.
- [50] J. Huang, K. Liu, and X. Li, "Locality constrained low rank representation and automatic dictionary learning for hyperspectral anomaly detection," *Remote Sens.*, vol. 14, no. 6, p. 1327, Mar. 2022.
- [51] H. Su, B. Zhao, Q. Du, P. Du, and Z. Xue, "Multifeature dictionary learning for collaborative representation classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2467–2484, Apr. 2018.
- [52] Z. Xue, X. Nie, and M. Zhang, "Incremental dictionary learning-driven tensor low-rank and sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5544019.
- [53] Y. Gao, H. Su, H. Lu, and Q. Du, "Self-balancing dictionary learning for relaxed collaborative representation of hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539918.
- [54] Z. Xue and X. Nie, "Low-rank and sparse representation with adaptive neighborhood regularization for hyperspectral image classification," *J. Geodesy Geoinf. Sci.*, vol. 5, no. 1, pp. 73–90, Mar. 2022.
- [55] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, and T.-Y. Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 734–749, Jan. 2020.
- [56] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [57] P. Ghamisi et al., "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.
- [58] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.

- [59] M. Imani and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges," *Inf. Fusion*, vol. 59, pp. 59–83, Jul. 2020.
- [60] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [61] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.
- [62] B. Tu, X. Zhang, X. Kang, G. Zhang, J. Wang, and J. Wu, "Hyper-spectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 340–344, Mar. 2018.
- [63] H. Liu, Y. Jia, J. Hou, and Q. Zhang, "Global-local balanced low-rank approximation of hyperspectral images for classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2013–2024, Apr. 2022.



Xiangyu Nie received the B.S. degree in geomatics engineering from the Suzhou University of Science and Technology, Suzhou, China, in 2019, and the M.E. degree in remote sensing from Hohai University, Nanjing, China, in 2022, where he is currently pursuing the Ph.D. degree in surveying and mapping.

His research interests include low-rank and sparse representation for hyperspectral image classification.



Zhaohui Xue (Member, IEEE) received the B.S. degree in geomatics engineering from Shandong Agricultural University, Tai'an, China, in 2009, the M.E. degree in remote sensing from the China University of Mining and Technology, Beijing, China, in 2012, and the Ph.D. degree in cartography and geographic information systems from Nanjing University, Nanjing, China, in 2015.

He is currently a Full Professor (Ph.D. Supervisor) with the College of Geography and Remote Sensing, Hohai University, Nanjing. He has authored more than 60 scientific papers, including more than 30 Science Citation Index (SCI) papers. His research interests include hyperspectral image classification, time-series image analysis, pattern recognition, and machine learning.

Dr. Xue has been an Editorial Board Member of the National Remote Sensing Bulletin since 2020. He was a recipient of the National Scholarship for Doctoral Graduate Students granted by the Ministry of Education of the People's Republic of China in 2014. He was awarded the Best Reviewer for the IEEE Geoscience and Remote Sensing Society. He has been a Reviewer for more than ten famous remote sensing journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Remote Sensing of Environment*, and *ISPRS Journal of Photogrammetry and Remote Sensing*.



Cong Lin received the Ph.D. degree in cartography and geographic information systems from Nanjing University, Nanjing, China, in 2021.

He is currently with Nanjing Research Institute of Surveying, Mapping and Geotechnical Investigation Company Ltd., Nanjing. His research interests include remote sensing image transfer learning and scene classification.



Ling Zhang received the B.S. degree in geomatics engineering from Shandong Agricultural University, Tai'an, China, in 2010, the M.E. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Beijing, China, in 2013, and the Ph.D. degree in photogrammetry and remote sensing from Hohai University, Nanjing, China, in 2022.

She is currently a Lecturer with the School of Naval Architecture and Ocean Engineering, Jiangsu Maritime Institute, Nanjing. Her research interests mainly focus on machine-learning-driven soil moisture estimation using multisource remote sensing.



Hongjun Su (Senior Member, IEEE) received the Ph.D. degree in cartography and geography information systems from the Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing, China, in 2011.

He is currently a Full Professor with the College of Geography and Remote Sensing, Hohai University, Nanjing. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmixing.

Dr. Su received the 2016 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. He is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.