

Low-Rank and Sparse Representation with Adaptive Neighborhood Regularization for Hyperspectral Image Classification

Zhaohui XUE, Xiangyu NIE

School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

Abstract: Low-Rank and Sparse Representation (LRSR) method has gained popularity in Hyperspectral Image (HSI) processing. However, existing LRSR models rarely exploited spectral-spatial classification of HSI. In this paper, we proposed a novel Low-Rank and Sparse Representation with Adaptive Neighborhood Regularization (LRSR-ANR) method for HSI classification. In the proposed method, we first represent the hyperspectral data via LRSR since it combines both sparsity and low-rankness to maintain global and local data structures simultaneously. The LRSR is optimized by using a mixed Gauss-Seidel and Jacobian Alternating Direction Method of Multipliers (M-ADMM), which converges faster than ADMM. Then to incorporate the spatial information, an ANR scheme is designed by combining Euclidean and Cosine distance metrics to reduce the mixed pixels within a neighborhood. Lastly, the predicted labels are determined by jointly considering the homogeneous pixels in the classification rule of the minimum reconstruction error. Experimental results based on three popular hyperspectral images demonstrate that the proposed method outperforms other related methods in terms of classification accuracy and generalization performance.

Key words: Hyperspectral Image (HSI); spectral-spatial classification; Low-Rank and Sparse Representation (LRSR); Adaptive Neighborhood Regularization (ANR)

Citation: Zhaohui XUE, Xiangyu NIE. Low-Rank and Sparse Representation with Adaptive Neighborhood Regularization for Hyperspectral Image Classification [J]. Journal of Geodesy and Geoinformation Science, 2022, 5 (1): 73-90. DOI:10.11947/j.JGGS.2022.0108.

1 Introduction

In the past decades, HSI classification has been greatly promoted by advanced machine learning and pattern recognition methods^[1-4]. Especially in deep learning, which can automatically learn data high-level features in a hierarchical manner, various excellent HSI classification methods are continuously proposed, including CNN-based classification methods^[5-7], deep feature fusion/aggregation^[8-9], combining attention model^[10-11], etc. However, the optimization of complex parameters is time-consuming and a very tough task, making the designed deep learning framework unexplainable, and a large number of labeled samples are needed for training.

Representation-based classification is of great

interest due to the emerging paradigm of compressed sensing^[12]. Typical prototypes of representation-based classification methods include Sparse Representation (SR)^[13], Collaborative Representation (CR)^[14], and Low-Rank Representation (LRR)^[15]. SR was firstly introduced into HSI classification scenario by Chen et al.^[16], who then proposed the prototype of Joint Sparse Representation (JSR) method to incorporate spectral and spatial information. Earlier developments of SR-based methods focus on exploiting morphological component analysis^[17-18], total variation^[19], graph embedding^[20-21], graph regularization^[22-23], dictionary learning^[24-25]. Advanced progresses focus on JSR incorporated with nonlocal weighting^[26-28], maximum likelihood estimation^[29], kernel^[30-31], multi-feature/multi-scale^[32-33], manifold learning^[34-35]. However, SR-based methods

Received date: 2021-06-25; accepted date: 2021-12-06

Foundation support: National Natural Foundation of China (No. 41971279); Fundamental Research Funds of the Central Universities (No. B200202012)

First author: Zhaohui XUE (19 —), male, PhD, youth professor, and his interests include hyperspectral image classification, time-series image

analysis, pattern recognition, and machine learning.

E-mail: zhaohui.xue@hhu.edu.cn

have a common drawback of ignoring the global data structure of HSI.

LRR seeks the lowest rank representation among all the candidates that can represent the data samples as linear combinations of the bases in a given dictionary^[15]. Existing studies exploited LRR for multi-kernel/multi-task learning^[36], dimensionality reduction or band selection^[37-39], graph-based semi-supervised learning^[40-41], kernel-based LRR^[42-44], LRR with various regularizations^[45-46], spectral-spatial LRR^[47-48]. Although LRR has great ability to capture the global data structure of HSI, it lacks effective description of the local data structure.

Recently, LRSR is more advocated since it can combine both sparsity and low rankness to maintain global and local data structures simultaneously. Various explorations of LRSR on HSI classification can be categorized into spectral-spatial classification^[49-51], graph-based semi-supervised learning^[52], graph-based discriminant analysis^[53-57], feature fusion of hyperspectral and LiDAR/multispectral data^[58-59]. Although LRSR has obtained promising classification performance, it's still worth exploring on better LRSR-based models for spectral-spatial classification. Previous related work mainly focused on isolated spectral or spatial feature extraction based on some third-part feature extractor, and then conducted feature fusion or classification based on LRSR models. Whereas, the spatial regularizer is rarely exploited in existing LRSR models for spectral-spatial classification of HSI.

To address the above issue, this paper presents a LRSR-ANR model for HSI classification, where a mixed Gauss-Seidel and Jacobian Alternating Direction Method of Multipliers is used to solve the problem defined in our model, and ANR is designed to incorporate the spatial information by combining Euclidean and Cosine distance metrics. ANR aims to reduce the mixed pixels within a neighborhood when determining the predicted labels, which jointly considers the homogeneous pixels in the minimum reconstruction error rule.

It is worth noting that some existing studies explored spatial information via neighborhood regulari-

zation in LRSR. However, there is a major conceptual difference. In Literature [48], the influence of heterogeneous pixels in the neighborhood is constrained by calculating the spatial distance of the pixel in the coordinate. This strategy overemphasizes the nearest pixels around the center pixel, which is unfriendly to pixels at the edge of the class. In Literature [50], the spatial information is introduced by using a neighborhood weighting scheme, which is actually a spatial filtering strategy. In addition, neighborhood regularization through superpixels is also a common method^[43]. However, this method needs to compress the hyperspectral data into three bands, which will cause the loss of spectral information. In contrast, the proposed ANR scheme is quite different. On the one hand, ANR scheme constructs neighborhood samples under the premise of ensuring the integrity of the spectral information while screening out the heterogeneous pixels. On the other hand, ANR scheme enhances the traditional minimum reconstruction rule. In this context, we can summarize the main contribution and novelty of our work as follows:

① A novel low-rank and sparse representation with adaptive neighborhood regularization method is proposed for HSI classification, which is unique in the literature.

② A novel adaptive neighborhood regularization scheme is designed by combining Euclidean and Cosine distance metrics to reduce the mixed pixels within a neighborhood, which enhances the traditional pixel-wise labeling strategy based on minimum reconstruction rule and can be flexibly equipped into other representation-based classification models.

2 Proposed Method

Let $\mathbf{A} \in \mathcal{R}^{d \times n}$ be a hyperspectral image data set; d is the number of bands and n is the number of pixels. Let $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_c\} \in \mathcal{R}^{d \times m}$ be a dictionary by randomly selecting a certain percentage of pixels from each class in \mathbf{A} ; m is the number of atoms in \mathbf{B} ; and c represents the number of classes. It is worth noting that the dictionary is directly composed

of training samples, and dictionary learning is not performed. Let $\mathbf{Y} \in \mathcal{R}^{d \times k}$ be a test set, where $\mathbf{Y}_i = \{y_{i1}, y_{i2}, \dots, y_{iW_s}\}$ represents the neighborhood pixels for the i -th test sample y_{i1} , and W_s is the window size (i.e., $W_s=t \times t$).

A graphical illustration of the proposed method is shown in Fig.1. Firstly, an M-ADMM algorithm is

used to solve the LRSR problem. Secondly, the original HSI is reconstructed using the different sub-dictionary relative to the different class. Finally, the classification map is obtained according to the minimum reconstruction error regularized by an adaptive neighborhood.

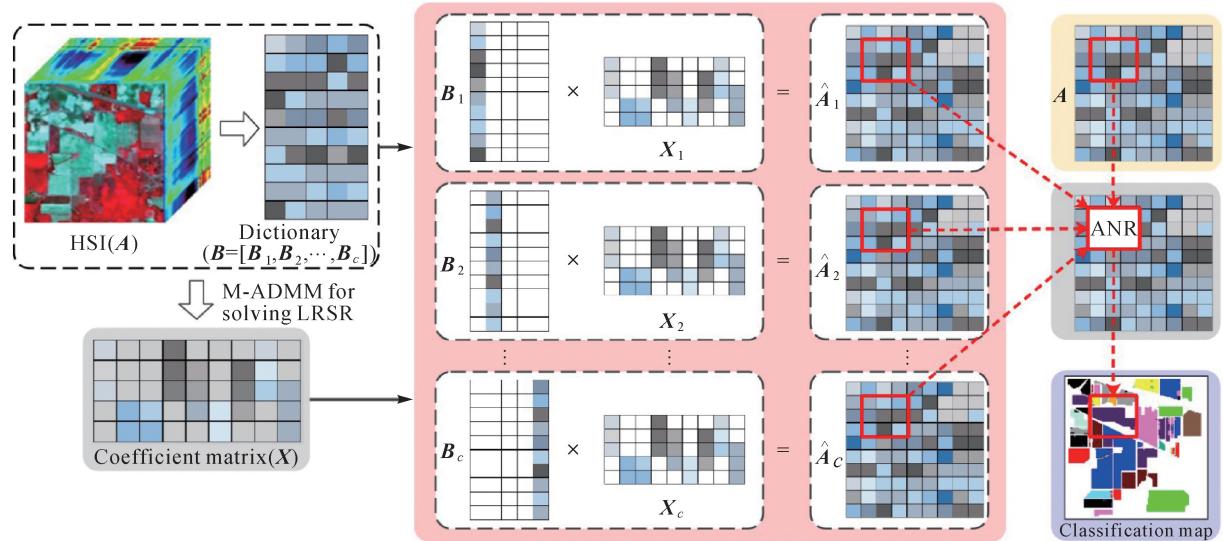


Fig.1 A graphical illustration of the proposed method for hyperspectral image classification

2.1 LRSR

In LRSR, the objective function is formulated as

$$\min_X \text{rank}(\mathbf{X}) + \alpha \|\mathbf{X}\|_0 \quad \text{s.t. } \mathbf{A} = \mathbf{B}\mathbf{X} \quad (1)$$

Where α is a positive regularization parameter balancing sparsity and low-rankness. Meanwhile, $\|\cdot\|_0$ is the number of all nonzero elements of the input matrix.

However, Eq. (1) is non-convex due to the low-rank operation and l_0 -norm. To solve non-convex issue, the equation can be reformulated as

$$\min_X \|\mathbf{X}\|_* + \alpha \|\mathbf{X}\|_1 \quad \text{s.t. } \mathbf{A} = \mathbf{B}\mathbf{X} \quad (2)$$

Actually, the hyperspectral image is usually noisy in real application. Therefore, a robust LRSR model considering noise reduction takes the form

$$\min_{\mathbf{X}, \mathbf{E}} \|\mathbf{X}\|_* + \alpha \|\mathbf{X}\|_1 + \beta \|\mathbf{E}\|_{2,1} \quad \text{s.t. } \mathbf{A} = \mathbf{B}\mathbf{X} + \mathbf{E} \quad (3)$$

Where $\|\cdot\|_*$ indicates the nuclear norm defined as the sum of all singular values of \mathbf{X} ; $\|\cdot\|_1$ represents the l_1 -norm which defines the sum of the absolute value of all elements; and $\|\cdot\|_{2,1}$ indi-

cates the sum of l_2 -norm along columns of the input matrix.

There are many ways to solve the LRSR model defined in Eq. (3). For instance, ADMM^[60], Augmented Lagrange Multiplier (ALM)^[61], Linearized Alternating Direction Method with Adaptive Penalty (LADMAP)^[62], and M-ADMM^[63].

2.2 M-ADMM for solving LRSR

Eq. (3) is a constraint convex optimization problem which can be solved by using the M-ADMM method, and the detailed analysis process can refer to Literature [64]. The specific solution process is given as follows.

By introducing two assistant variables \mathbf{Z} and \mathbf{J} , the objective function in Eq. (3) is converted to the equal optimization problems as

$$\begin{aligned} & \min_{\mathbf{X}, \mathbf{E}} \|\mathbf{X}\|_* + \alpha \|\mathbf{X}\|_1 + \beta \|\mathbf{E}\|_{2,1} \\ & \text{s.t. } \mathbf{A} = \mathbf{B}\mathbf{X} + \mathbf{E}, \mathbf{X} = \mathbf{Z}, \text{ and } \mathbf{X} = \mathbf{J} \end{aligned} \quad (4)$$

The augmented Lagrange function of Eq. (4) takes the form

$$\begin{aligned} L(\mathbf{X}, \mathbf{Z}, \mathbf{J}, \mathbf{E}, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3, \mu) = & \|\mathbf{X}\|_* + \alpha\|\mathbf{X}\|_1 + \beta\|\mathbf{E}\|_{2,1} + \\ & \langle \mathbf{Q}_1, \mathbf{A} - \mathbf{B}\mathbf{X} - \mathbf{E} \rangle + \langle \mathbf{Q}_2, \mathbf{X} - \mathbf{Z} \rangle + \langle \mathbf{Q}_3, \mathbf{X} - \mathbf{J} \rangle + \\ & \frac{\mu}{2} (\|\mathbf{A} - \mathbf{B}\mathbf{X} - \mathbf{E}\|_F^2 + \|\mathbf{X} - \mathbf{Z}\|_F^2 + \|\mathbf{X} - \mathbf{J}\|_F^2) \quad (5) \end{aligned}$$

Where \mathbf{Q}_1 , \mathbf{Q}_2 , and \mathbf{Q}_3 are the Lagrange multipliers; $\langle \mathbf{a}, \mathbf{b} \rangle = \text{trace}(\mathbf{a}^T \mathbf{b})$; and $\mu > 0$ is a penalty parameter. In this context, M-ADMM can be applied to solve Eq. (5) with four optimization variables, and the detailed updating rules are presented as follows (at iteration k)

① Fix \mathbf{X} , \mathbf{J} , and \mathbf{E} , update \mathbf{Z}

$$\begin{aligned} \mathbf{Z}^{k+1} = & \min_{\mathbf{Z}} \frac{1}{\mu} \|\mathbf{Z}\|_* + \frac{1}{2} \left\| \mathbf{Z}^k - \mathbf{X}^k - \frac{\mathbf{Q}_2^k}{\mu^k} \right\|_F^2 = \\ & \mathbf{U} \mathbf{S} \frac{1}{\mu^k} (\mathbf{\Sigma}) \mathbf{V}^T \quad (6) \end{aligned}$$

Where $\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ is the singular value decomposition (SVD) of $\mathbf{X}^k + \frac{\mathbf{Q}_2^k}{\mu^k}$; and $S_a(b) = \text{sgn}(b) \max(|b| - a, 0)$ is a soft-thresholding operator.

② Fix \mathbf{X} , \mathbf{Z} , and \mathbf{E} , update \mathbf{J}

$$\begin{aligned} \mathbf{J}^{k+1} = & \min_{\mathbf{J}} \frac{\alpha}{\mu^k} \|\mathbf{J}\|_1 + \frac{1}{2} \left\| \mathbf{J}^k - \mathbf{X}^k - \frac{\mathbf{Q}_3^k}{\mu^k} \right\|_F^2 = \\ & \Theta_{\frac{\alpha}{\mu^k}} \left(\mathbf{X}^k + \frac{\mathbf{Q}_3^k}{\mu^k} \right) \quad (7) \end{aligned}$$

Where $\Theta_a(b) = \max(b - a, 0) + \min(b + a, 0)$ is another soft-thresholding operator.

③ Fix \mathbf{X} , \mathbf{Z} , and \mathbf{J} , update \mathbf{E}

$$\mathbf{E}^{k+1} = \min_{\mathbf{E}} \frac{\beta}{\mu^k} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \left\| \mathbf{E}^k - \mathbf{A} + \mathbf{B}\mathbf{X}^k - \frac{\mathbf{Q}_1^k}{\mu^k} \right\|_F^2 \quad (8)$$

To solve the minimization problem in Eq. (8), we denote $\mathbf{G} = \mathbf{A} - \mathbf{B}\mathbf{X}^k + \frac{\mathbf{Q}_1^k}{\mu^k}$, and the optimal solution for the i -th column of \mathbf{E} is

$$\mathbf{E}^{k+1}(:, i) = \begin{cases} \|\mathbf{g}_i\|_2 - \frac{\beta}{\mu^k} & \frac{\beta}{\mu^k} < \|\mathbf{g}_i\|_2 \\ \frac{\beta}{\mu^k} \mathbf{g}_i & \frac{\beta}{\mu^k} \geq \|\mathbf{g}_i\|_2 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Where \mathbf{g}_i is the i -th column of \mathbf{G} .

④ Fix \mathbf{Z} , \mathbf{J} , and \mathbf{E} , update \mathbf{X}

$$\begin{aligned} \mathbf{X}^{k+1} = & (\mathbf{B}^T \mathbf{B} + 2\mathbf{I})^{-1} \left[\mathbf{B}^T \left(\frac{\mathbf{Q}_1^k}{\mu^k} - \mathbf{E}^{k+1} \right) + \mathbf{B}^T \mathbf{A} - \right. \\ & \left. \frac{(\mathbf{Q}_2^k + \mathbf{Q}_3^k)}{\mu^k} + \mathbf{Z}^{k+1} + \mathbf{J}^{k+1} \right] \quad (10) \end{aligned}$$

Where \mathbf{I} is an identity matrix.

The pseudo code of using M-ADMM method to solve LRSR is outlined in Algorithm 1.

Algorithm 1: M-ADMM for Solving LRSR.

Input: Data set \mathbf{A} , dictionary \mathbf{B} , regularization parameters α and β

Output: Coefficient matrix \mathbf{X} and noise matrix \mathbf{E}

Initialize: $\mathbf{X}^0 = \mathbf{Z}^0 = \mathbf{J}^0 = \mathbf{0}$, $\mathbf{E}^0 = \mathbf{0}$, $\mathbf{Q}_1^0 = \mathbf{Q}_2^0 = \mathbf{Q}_3^0 = \mathbf{0}$, $\mu^0 = 10^{-4}$, $\mu_{\max} = 10^6$, $\lambda = 1.1$ (λ is a growth rate of μ for each iteration), $k = 0$

While not converged do

 Update the variable \mathbf{Z} according to Eq. (6)

 Update the variable \mathbf{J} according to Eq. (7)

 Update the variable \mathbf{E} according to Eq. (8)

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

 Update the Lagrange multipliers \mathbf{Q}_1 , \mathbf{Q}_2 , and \mathbf{Q}_3

$$\mathbf{Q}_1^{k+1} = \mathbf{Q}_1^k + \mu^k (\mathbf{A} - \mathbf{B}\mathbf{X}^{k+1} - \mathbf{E}^{k+1})$$

$$\mathbf{Q}_2^{k+1} = \mathbf{Q}_2^k + \mu^k (\mathbf{X}^{k+1} - \mathbf{Z}^{k+1})$$

$$\mathbf{Q}_3^{k+1} = \mathbf{Q}_3^k + \mu^k (\mathbf{X}^{k+1} - \mathbf{J}^{k+1})$$

 Update the parameter μ

$$\mu^{k+1} = \min(\lambda\mu^k, \mu_{\max})$$

 Update k : $k = k + 1$

End while

2.3 Adaptive neighborhood regularization

As we all know, spatial information plays an extremely important role in HSI classification. However, the original LRSR only utilize the spectral information. Therefore, we combine the spatial information by designing an ANR scheme, which improves the decision rule of LRSR based on the minimum reconstruction error. A graphical illustration of the proposed ANR scheme can be seen in Fig. 2. Taking 5×5 window as an example, \mathbf{Y}_i , S_{ij} , S , and \mathbf{Y}'_i represent a neighborhood sample set, the similarity between the center pixel and its neighboring pixels, the similarity threshold, and the neighborhood sample set after screening, respectively.

To incorporate the spatial information, a neighborhood sample set $\mathbf{Y}_i = \{y_{i1}, y_{i2}, \dots, y_{iW_S}\}$ is established with i -th test sample y_{i1} as the center pixel. In order to facilitate the expression of the equations below, we put the center pixel y_{i1} in the first place, and the remaining pixels are sorted in the order of column first. For example, y_{i2} and y_{i3} are the first

pixel and the second pixel of the first column, respectively. Actually, the heterogeneous samples within a window will affect the decision accuracy based on the minimum reconstruction error rule. Therefore, we need to screen the neighboring pixels.

The similarities between the central pixel and its neighboring pixels are measured by combining Euclidean and Cosine distance metrics

$$S_{ij} = \frac{\langle y_{il}, y_{ij} \rangle}{|y_{il}| * |y_{ij}|} \times \exp\left(-\text{dist}\left(\frac{y_{il}}{|y_{il}|}, \frac{y_{ij}}{|y_{ij}|}\right)\right) \quad (11)$$

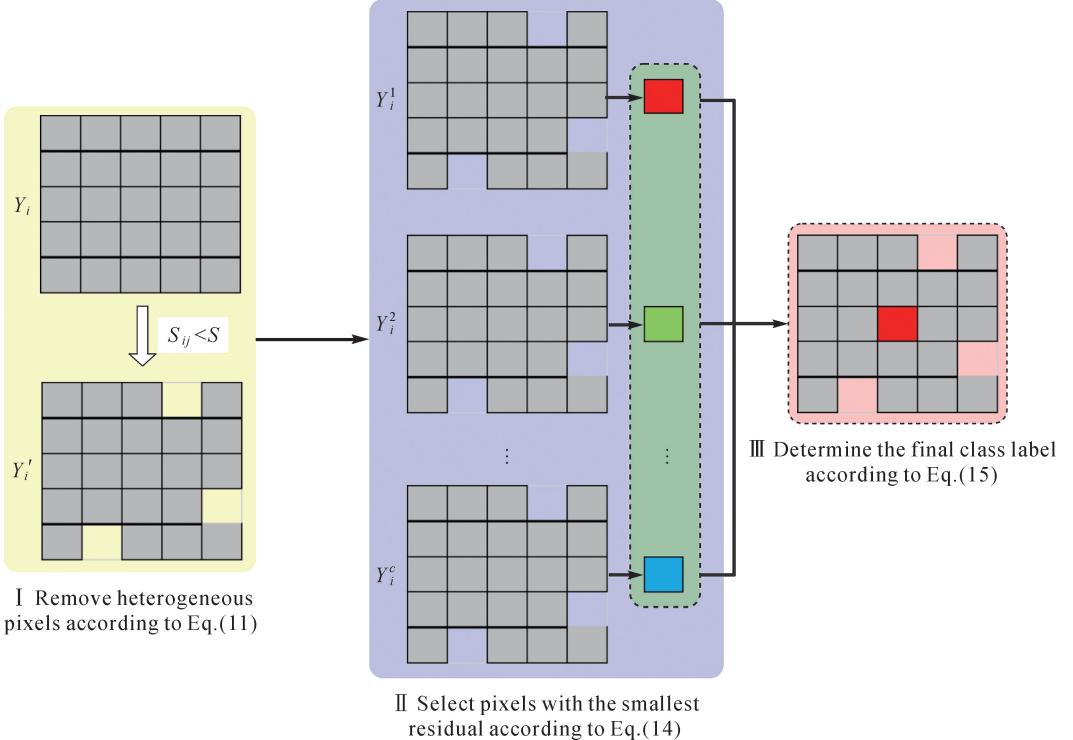


Fig.2 A graphical illustration of the proposed ANR scheme

Where $j=2, 3, \dots, W_s$; $\langle a, b \rangle$ means the inner product between a and b ; $\text{dist}(\cdot)$ stands for the Euclidean distance; and $\exp(\cdot)$ represents the exponential function. In this context, S_{ij} is designed to measure the similarity between y_{il} and y_{ij} . Obviously, a larger value of S_{ij} indicates a higher similarity. A threshold is determined to screen those heterogeneous samples. Suppose the optimal threshold of S_{ij} is S , when $S_{ij} < S$, we will remove y_{ij} from \mathbf{Y}_i . After screening, the remaining samples are re-sorted as $\mathbf{Y}'_i = \{y'_{i1}, y'_{i2}, \dots, y'_{iW'_s}\}$, where $W'_s \leq W_s$.

Since $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_c\}$, the equation $\mathbf{A} = \mathbf{BX} + \mathbf{E}$ can be expanded as

$$\mathbf{A} = [\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_c] [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c]^T + \mathbf{E} \quad (12)$$

Where \mathbf{X}_l is the coefficient matrix relative to sub-dictionary \mathbf{B}_l ($l = 1, 2, \dots, c$). The reconstructed data relative to different sub-dictionary can be

formulated as

$$\mathbf{A}_l = [0, \dots, \mathbf{B}_l, \dots, 0] [0, \dots, \mathbf{X}_l, \dots, 0]^T \quad (13)$$

Let r_i^l represent the smallest residual obtained by comparing the l -th reconstructed sample y_{il}^l with its original value y'_{il} , which can be formulated as

$$r_i^l = \min \left\{ \left\| \frac{y_{il}^l - y'_{il}}{|y_{il}^l| - |y'_{il}|} \right\|_1, \left\| \frac{y_{i2}^l - y'_{i2}}{|y_{i2}^l| - |y'_{i2}|} \right\|_1, \dots, \left\| \frac{y_{iW_s^l}^l - y'_{iW_s^l}}{|y_{iW_s^l}^l| - |y'_{iW_s^l}|} \right\|_1 \right\} \quad (14)$$

Finally, the smaller the value of r_i^l , the higher the likelihood of belonging to this class. Therefore, the class label of y'_{il} can be determined by the minimum of r_i^l .

$$\text{class}(y'_{il}) = \arg \min_{l=1,2,\dots,c} r_i^l \quad (15)$$

The pseudo code of the proposed LRSR-ANR method is outlined in Algorithm 2.

Algorithm 2: Low-Rank and Sparse Representation with Adaptive Neighborhood Regularization (LRSR-ANR).

Input: Data set \mathbf{A} , dictionary \mathbf{B} , parameters α, β, S, W_S

Output: Class labels for \mathbf{A}

Step 1: Select a certain percentage of training samples from each class in the ground-truth map to construct the dictionary.

Step 2: Solve the LRSR problem using Algorithm 1.

Step 3: Generate the neighborhood sample set \mathbf{Y}_i with i -th test sample y_{ii} as the center pixel.

Step 4: Calculate similarities according to Eq. (11) and remove heterogeneous pixels.

Step 5: Select pixels with the smallest residual according to Eq. (14).

Step 6: Determine the final class label according to Eq. (15).

3 Experimental Results

3.1 Hyperspectral data sets

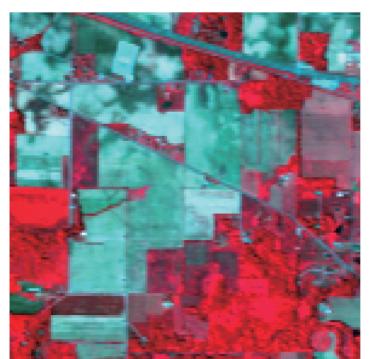
To verify the performance of the proposed LRSR-ANR, three popular hyperspectral datasets are used to conduct experimental. (For more detailed information about the data sets, please see in http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.)

① The first hyperspectral data set is Indian Pines, which was gathered by Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in north-western Indiana in

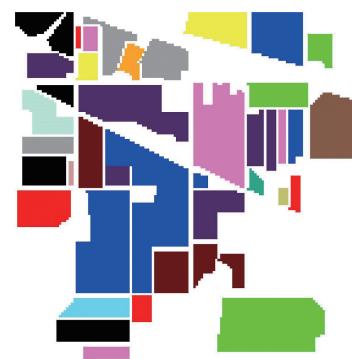
1992. The data set is consisting of 145×145 pixels and 224 spectral reflectance bands in the wavelength ranging from $0.4 \mu\text{m}$ to $2.5 \mu\text{m}$ with a spatial resolution of 20 m. There are 200 radiance channels used in the experiments after removing 20 noisy and water absorbed bands. The false color composite and the ground-truth map are shown in Fig. 3. A total of 10 249 samples containing 16 classes are available as reported in Tab.1.

Tab.1 Number of training and test samples used for the Indian Pines data set (10% for training)

Class	Name	Training	Test	Total
1	Alfalfa	5	41	46
2	Corn-notill	143	1285	1428
3	Corn-mintill	83	747	830
4	Corn	24	213	237
5	Grass-pasture	49	434	483
6	Grass-trees	73	657	730
7	Grass-pasture-mowed	3	25	28
8	Hay-windrowed	48	430	478
9	Oats	2	18	20
10	Soybean-notill	98	874	972
11	Soybean-mintill	246	2209	2455
12	Soybean-clean	60	533	593
13	Wheat	21	184	205
14	Woods	127	1138	1265
15	Buildings-grass-trees-drives	39	347	386
16	Stone-steel-towers	10	83	93
Total		1031	9218	10 249



(a) False color composite image (R: 50, G: 27, and B: 17)



(b) Ground-truth map

Fig.3 The Indian Pines

② The second hyperspectral data set is Pavia University, which was acquired by the Reflective Optics Spectrographic Imagine System (RODIS) sensor over the urban area of the University of Pavia, northern Italy, on July 8, 2002. The data set is consisting of 610×340 pixels and 115 spectral re-

flectance bands in the wavelength ranging from $0.43 \mu\text{m}$ to $0.86 \mu\text{m}$ with a spatial resolution of 1.3 m. There are 103 radiance channels used in the experiments after removing 12 noisy bands. The false color composite and the ground-truth map are shown in Fig.4. A total of 42 776 samples containing 9 clas-

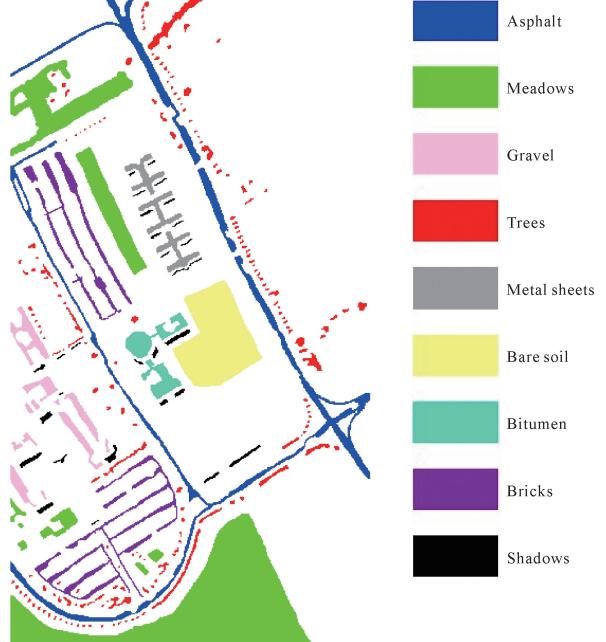
ses are available as reported in Tab.2.

③ The third hyperspectral data set is Salinas, which was obtained by AVIRIS sensor over Salinas Valley, CA, USA. The data set is consisting of 512×217 pixels and 224 spectral reflectance bands in the wavelength ranging from $0.4 \mu\text{m}$ to $2.5 \mu\text{m}$ with a

spatial resolution of 3.7 m . There are 204 spectral bands have been retained after discarding the 20 water absorption bands. The false color composite and the ground-truth map are shown in Fig.5. A total of 54 129 samples containing 16 classes are available as reported in Tab.3.



(a) False color composite image (R: 102, G: 56, and B: 31)

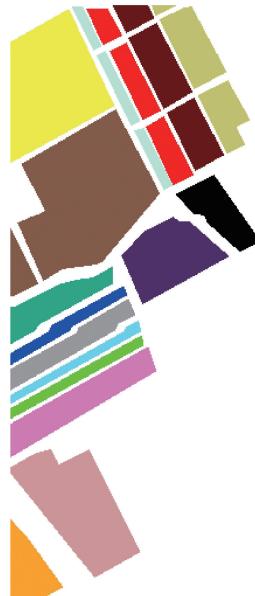


(b) Ground-truth map

Fig.4 The Pavia University



(a) False color composite image(R:50, G:27, and B:17)



(b) Ground-truth map

Fig.5 The Salinas

Tab.2 Number of training and test samples used for the Pavia University data set (2% for training)

Class	Name	Training	Test	Total
1	Asphalt	133	6498	6631
2	Meadows	373	18 276	18 649
3	Gravel	42	2057	2099
4	Trees	62	3002	3064
5	Metal sheets	27	1318	1345
6	Bare soil	101	4928	5029
7	Bitumen	27	1303	1330
8	Bricks	74	3608	3682
9	Shadows	19	928	947
Total		858	41 918	42 776

Tab.3 Number of training and test samples used for the Salinas data set (2% for training)

Class	Name	Training	Test	Total
1	Brocoli_green_weeds_1	41	1968	2009
2	Brocoli_green_weeds_2	75	3651	3726
3	Fallow	40	1936	1976
4	Fallow_rough_plow	28	1366	1394
5	Fallow_smooth	54	2624	2678
6	Stubble	80	3879	3959
7	Celery	72	3507	3579
8	Grapes_untrained	226	11 045	11 271
9	Soil_vinyard_develop	125	6078	6203
10	Corn_senesced_green_weeds	66	3212	3278
11	Lettuce_romaine_4wk	22	1046	1068
12	Lettuce_romaine_5wk	39	1888	1927
13	Lettuce_romaine_6wk	19	897	916
14	Lettuce_romaine_7wk	22	1048	1070
15	Vinyard_untrained	146	7122	7268
16	Stone-steel-towers	37	1770	1807
Total		1092	53 037	54 129

3.2 Experimental settings

The comparison methods, corresponding parameter settings, and evaluation indexes are introduced as follows.

① Comparison methods: We select eight methods for comparison including Support Vector Machines (SVM)^[65], SVM with Composite Kernel (SVMCK)^[66], SR^[13], Joint Sparse Representation (JSR)^[16], Low-Rank Representation (LRR)^[15], Low-Rank and Sparse Representation (LRSR)^[53], Joint Low-Rank and Sparse Representation (JLRSR)^[49], Fast Random Projection Low Rank approximation and Sparsity representation (FRPLRaS)^[51].

② Parameter settings: The number of training samples are set to 10%, 2%, and 2% for Indian Pines, Pavia University, and Salinas, respectively. More details of the training and test samples can be seen in Tabs. 1—3. The parameters of SVM and SVMCK are obtained by cross-validation^[65-66]. The parameters for SR, JSR, LRR, and LRSR are determined according to the experimental settings in Literatures [13], [15]—[16] and [53]. For JLRSR and FRPLRaS, the parameters are set following the original article. As for the proposed method, we adopt a one-by-one optimization strategy to obtain the optimal parameters for α , β , W_S , and S . The parameter settings of different methods for the three hyperspectral data sets are listed in Tab.4.

Tab.4 The parameter settings of different methods for the three hyperspectral data sets

Methods	Indian Pines	Pavia University	Salinas
SR	$T=5$	$T=10$	$T=10$
JSR	$\{T=20, W_S=5\times 5\}$	$\{T=25, W_S=11\times 11\}$	$\{T=20, W_S=11\times 11\}$
LRR	$\beta=20$	$\beta=0.2$	$\beta=20$
LRSR	$\{\alpha=1, \beta=20\}$	$\{\alpha=1, \beta=0.2\}$	$\{\alpha=1, \beta=20\}$
JLRSR	$\{\alpha=1, \beta=20, W_S=7\times 7\}$	$\{\alpha=1, \beta=0.2, W_S=15\times 15\}$	$\{\alpha=1, \beta=20, W_S=13\times 13\}$
FRPLRaS	$\{\text{Rank}=30, T=20, W_S=5\times 5\}$	$\{\text{Rank}=30, T=30, W_S=11\times 11\}$	$\{\text{Rank}=30, T=30, W_S=9\times 9\}$
Ours	$\{\alpha=1, \beta=20, W_S=7\times 7, S=0.9\}$	$\{\alpha=1, \beta=0.2, W_S=19\times 19, S=0.725\}$	$\{\alpha=1, \beta=20, W_S=13\times 13, S=0.85\}$

Note: T is sparsity level parameter for SR; W_S is the window size; α is a penalty parameter for SR; β is a penalty parameter for LRR; and S is a threshold for screening samples for our method.

③ Evaluation indexes: Three quantitative metrics including Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient (κ) are used to evaluate the classification performance of different methods. It's worth noting that we conduct

ten independent runs, where the training set for each run was obtained by random sampling from each class with the remaining samples used for test. The averaged results and standard deviations of all experiments are listed in the report.

④ **Running platform:** All the implementations were carried out using Matlab R2020a in a desktop PC equipped with an Intel Core i9-9900K CPU (at 3.6 GHz) and 64 GB of RAM. We have carefully optimized the involved parameters for all considered methods in order for them to obtain the best results, which ensures a fair comparison.

3.3 Parameter sensitiveness

For the proposed method, there are four parameters including α , β , W_s , and S need to be adjusted for different data sets. Although more parameters are involved, they can be easily optimized. To obtain the optimal value for each parameter, we optimize the four parameters one by one.

First of all, we optimize the parameter α by fixing the other parameters to $\{\beta=20, W_s=5\times5, S=0.85\}$, $\{\beta=20, W_s=5\times5, S=0.7\}$, and $\{\beta=20, W_s=5\times5, S=0.85\}$ for the Indian Pines, the Pavia University, and the Salinas, respectively. As shown in Fig.6(a), the optimal value for parameter α is 1 for the three data sets.

Secondly, in order to optimize β , the other parameters are fixed to $\{\alpha=1, W_s=5\times5, S=0.85\}$, $\{\alpha=1, W_s=5\times5, S=0.7\}$, and $\{\alpha=1, W_s=5\times5, S=0.85\}$ for the three data sets respectively. The evolution of OA as a function of parameter β is shown in Fig.6(b), where the optimal values of parameter β are 20, 0.2, 20 for Indian Pines, Pavia University, and Salinas, respectively.

Next, we test the impact of W_s on classification accuracy by fixing the other parameters to $\{\alpha=1, \beta=20, S=0.85\}$, $\{\alpha=1, \beta=0.2, S=0.7\}$, and $\{\alpha=1, \beta=20, S=0.85\}$ for the three data sets respectively. As shown in Fig.6(c), the optimal window sizes are 7×7 , 19×19 , and 13×13 respectively for the three data sets. Note that larger window size leads to higher accuracy for Pavia University according to the results. However, the classification map may be over-smoothed on between-class boundaries. Therefore, we limit the maximum value of W_s to 19.

Finally, we can optimize S since the optimal values for the other parameters have been obtained. According to Fig. 7, the optimal values of S are 0.90, 0.725, and 0.85 for the three data sets re-

spectively.

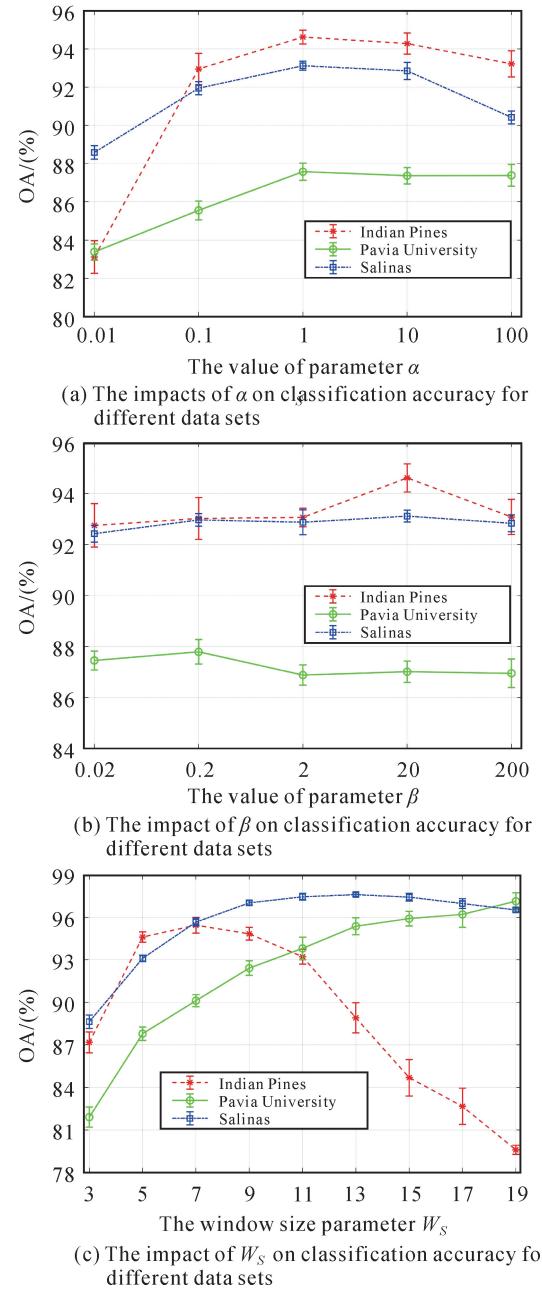


Fig.6 The impacts of α , β , W_s on classification accuracy for different data sets (Error bars indicate the standard deviations)

From the experimental results of the above parameter adjustments, the optimal values of the four parameters for the Pavia University data set are obviously different from the other two data sets. There are three main reasons for this observation. Firstly, the Pavia University data set and the other two data sets are collected from different sensors, and the number of bands is less than that of Indian Pines

and Salinas. Secondly, the land surface objects are different. The Pavia University data set contains 9 types of objects and the other two data sets contain 16 types. Finally, the spatial resolution is different. The spatial resolution of the Pavia University data set is higher than those of the other two datasets. The Indian Pines and Salinas data sets are gathered by the same sensor type. Both of these two data sets contain 16 types of objects which are mainly crops. These similarities make the optimal solutions of the three parameters α , β , and S for the two data sets almost the same. However, the window size parameter is determined by the spatial resolution and the homogeneity of the scene. Therefore, the optimal value of parameter W_s in Salinas is clearly greater than that of Indian Pines. In addition, we notice that the influence of parameter α is greater than the influence of parameter β , which proves that the sparsity level constraint plays a more important role in LRSR model for HSI classification.

3.4 Classification results

To validate the superiority of our method, we compare the proposed method with the other eight considered methods, and report the classification accuracies and maps obtained by different methods for the three data sets.

① Indian Pines: As listed in Tab.5, the proposed method (LRSR-ANR) obtains the highest OA of 96.29%, which is around 0.09%~28.59% higher than those of the other methods. For AA, LRSR-ANR yields the highest accuracy of 96.29%, which is around 0.58%~31.35% higher than those of the other methods. For κ , LRSR-ANR obtains the highest accuracy of 95.77%, which is around 0.1%~33.15% higher than those of the other methods. As for class specific accuracies, LRSR-ANR obtains higher accuracies for 5 classes. It is interesting to note that SR, LRR, and LRSR perform not very well in this scene. Whereas, JSR provides competitive results. In addition, SVMCK obviously outperforms SVM. By contrast, the results based on JLSR and FRPLRaS are outstanding, and their overall accuracy is higher than those of JSR and SVMCK. The superiority of LRSR-ANR can be visually

inspected in Fig.8, where the proposed method obtains a more accurate and smooth classification map, especially for the classes Hay-windrowed and Oats.

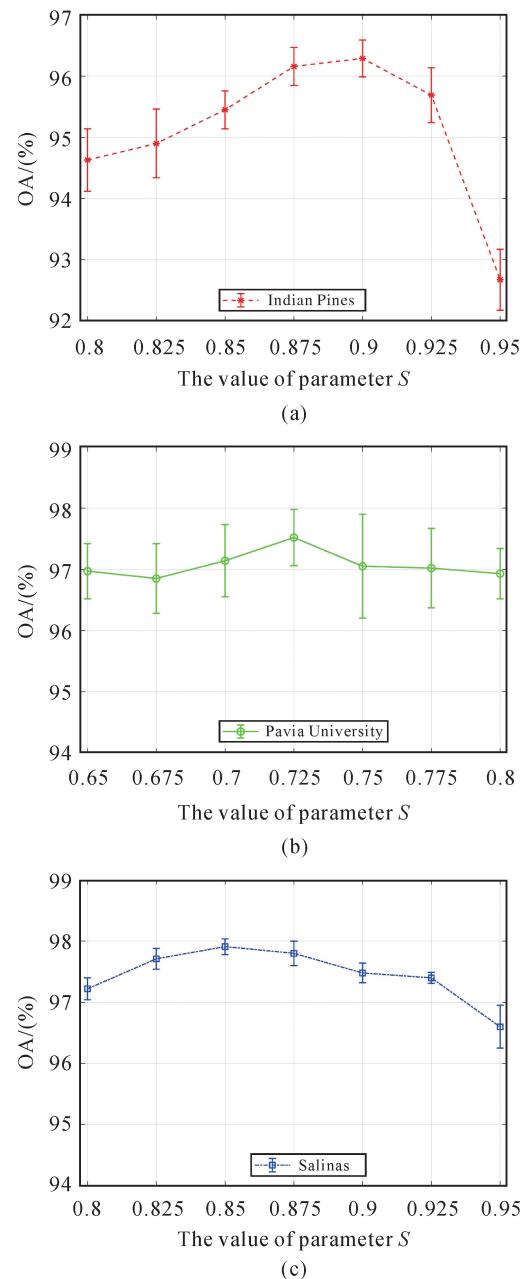


Fig.7 The impact of S on classification accuracy for different data sets (Error bars indicate the standard deviations)

② Pavia University: As reported in Tab. 6, LRSR-ANR significantly outperforms the other methods. For OA, our method obtains the highest accuracy of 97.52%, which is around 0.74%~26.72% higher than those of the other methods. For AA, the proposed method obtains an accuracy of 95.88%, which is around 2.47%~31.97% higher

than those of the other methods. For κ , the performance improvement is around $0.8\% \sim 34.67\%$ compared to the performance improvement of the other methods. As for class specific accuracies, our method achieves higher accuracy for 5 classes. Again, SR, LRR, and LRSR perform not well in

this scene. Whereas, SVMCK, JLRSR, and FRPLRaS exhibit competitive performance. The corresponding classification maps are shown in Fig. 9, where LRSR-ANR provides a more accurate and smooth classification map and can be verified in classifying the classes Meadows and Bitumen.

Tab.5 Classification accuracy of different methods for the Indian Pines data set (10% for training) (%)

Class	SVM	SVMCK	SR	JSR	LRR	LRSR	JLRSR	FRPLRaS	LRSR-ANR
1	11.46 ± 12.44	82.62 ± 10.10	42.44 ± 11.16	87.80 ± 12.44	70.24 ± 15.70	65.61 ± 16.52	99.76 ± 0.73	89.76 ± 11.33	98.54 ± 2.06
2	76.93 ± 1.87	90.17 ± 2.31	53.81 ± 2.35	88.41 ± 2.15	54.60 ± 2.77	54.84 ± 2.48	90.85 ± 1.60	94.76 ± 1.81	95.23 ± 0.67
3	64.70 ± 2.63	95.15 ± 1.54	51.27 ± 2.43	84.47 ± 3.82	39.80 ± 2.87	40.48 ± 3.30	92.99 ± 2.39	96.31 ± 1.43	93.19 ± 2.21
4	58.87 ± 8.19	83.66 ± 4.34	37.09 ± 5.13	79.67 ± 6.06	48.22 ± 4.87	48.78 ± 4.28	98.87 ± 2.04	92.68 ± 4.14	97.93 ± 1.72
5	88.62 ± 1.59	93.57 ± 2.57	82.67 ± 2.15	94.51 ± 2.96	75.12 ± 3.53	75.97 ± 3.10	92.17 ± 3.26	96.90 ± 1.47	94.98 ± 1.87
6	94.89 ± 2.03	99.23 ± 0.38	92.91 ± 2.18	98.63 ± 0.89	85.66 ± 3.37	86.15 ± 3.22	97.79 ± 0.55	99.15 ± 0.76	99.04 ± 0.23
7	6.40 ± 11.81	49.20 ± 29.21	73.20 ± 5.98	96.00 ± 3.77	88.40 ± 3.50	89.20 ± 3.79	100.00 ± 0.00	29.20 ± 34.18	94.40 ± 8.26
8	99.12 ± 0.92	99.75 ± 0.25	96.35 ± 2.22	99.40 ± 0.79	92.23 ± 2.78	91.70 ± 2.83	99.60 ± 0.79	99.91 ± 0.15	100.00 ± 0.00
9	0.00 ± 0.00	55.56 ± 9.63	22.78 ± 15.27	81.11 ± 13.15	42.22 ± 10.86	44.44 ± 8.28	90.56 ± 15.33	33.33 ± 32.77	94.44 ± 6.93
10	68.19 ± 3.47	90.28 ± 1.87	66.31 ± 1.94	93.86 ± 1.67	57.64 ± 1.99	57.88 ± 2.32	94.45 ± 1.51	93.38 ± 1.65	94.05 ± 1.45
11	86.01 ± 1.97	96.43 ± 1.25	71.22 ± 1.97	95.55 ± 0.80	71.65 ± 1.59	71.32 ± 2.28	96.17 ± 1.80	97.06 ± 0.90	97.18 ± 1.09
12	78.67 ± 3.38	89.65 ± 3.53	42.94 ± 1.92	84.89 ± 3.37	43.21 ± 3.38	43.00 ± 4.80	89.14 ± 4.20	95.43 ± 1.94	91.14 ± 1.90
13	95.11 ± 3.10	98.61 ± 1.16	92.66 ± 2.99	99.29 ± 1.06	94.84 ± 2.18	94.51 ± 1.84	98.70 ± 1.44	98.10 ± 1.40	97.61 ± 1.80
14	96.92 ± 0.75	99.16 ± 0.33	90.44 ± 1.65	99.02 ± 0.44	94.46 ± 1.04	93.99 ± 0.91	99.89 ± 0.08	99.22 ± 0.69	99.79 ± 0.33
15	56.71 ± 4.70	96.41 ± 2.50	34.38 ± 3.26	78.65 ± 5.07	37.61 ± 3.30	37.38 ± 3.25	93.92 ± 2.04	94.09 ± 3.67	92.62 ± 2.60
16	85.71 ± 5.50	85.41 ± 6.16	88.57 ± 2.98	98.21 ± 2.53	87.11 ± 5.24	88.43 ± 5.63	96.51 ± 2.44	92.74 ± 9.40	98.43 ± 1.28
OA	81.35 ± 0.71	94.41 ± 0.50	68.90 ± 0.67	92.69 ± 0.45	67.40 ± 0.65	67.43 ± 0.69	95.19 ± 0.46	96.20 ± 0.46	96.29 ± 0.30
AA	66.77 ± 1.33	87.80 ± 3.19	64.94 ± 1.60	91.22 ± 1.43	67.43 ± 1.12	67.73 ± 1.31	95.71 ± 0.88	87.63 ± 2.74	96.29 ± 1.84
$\kappa \times 100$	78.58 ± 0.81	93.63 ± 0.57	64.48 ± 0.76	91.65 ± 0.52	62.62 ± 0.73	62.66 ± 0.78	94.52 ± 0.52	95.67 ± 0.52	95.77 ± 0.66
Time/s	32	32	27	65	143	164	235	175	195

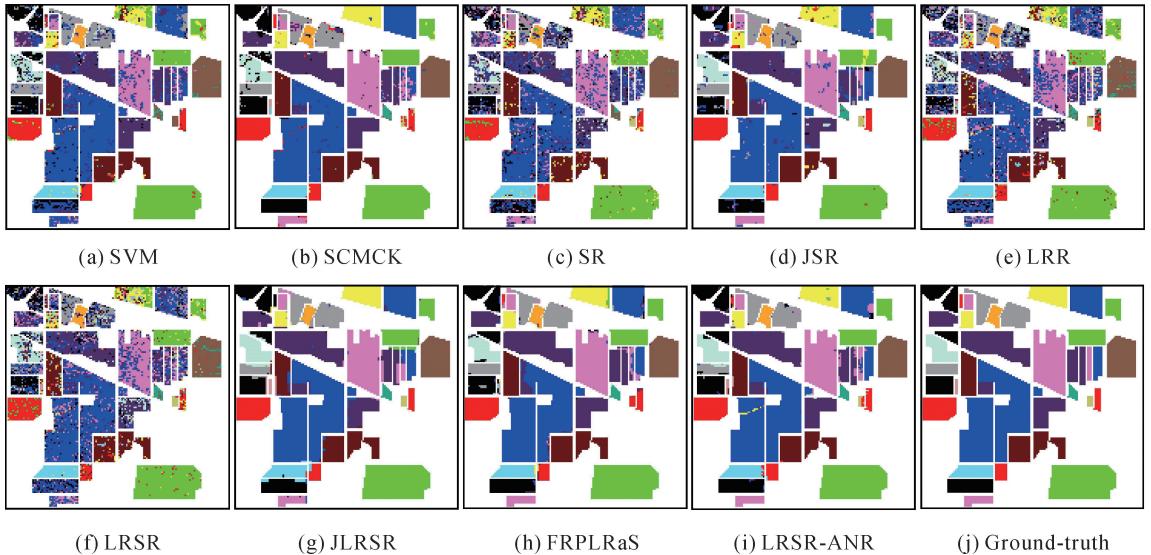


Fig.8 Classification maps obtained by different methods for the Indian Pines data set

③ Salinas: Tab.7 lists the classification accuracies obtained by different methods. And in Tab.7, LRSR-ANR achieves the best accuracies. For OA, LRSR-ANR yields an accuracy of 97.91% with the performance improvement of 1.3% ~ 16.61%

compared to the other methods. For AA, the improvement is around 0.31% ~ 9.4%. For κ , around 1.41% ~ 18.42% improvement can be seen by comparing our method with the other counterparts. As for class specific accuracies, the proposed method

obtains the highest accuracies for 3 classes. Fig. 10 exhibits the classification maps, where significant differences can be seen in the classification of

classes Grapes_untrained and Vinyard_untrained, and LRSR-ANR obtains more accurate and smooth results.

Tab.6 Classification accuracy of different methods for the Pavia University data set (2% for training) (%)

Class	SVM	SVMCK	SR	JSR	LRR	LRSR	JLRSR	FRPLRaS	LRSR-ANR
1	90.97±2.60	95.87±0.60	83.70±5.00	88.03±4.94	62.32±1.29	59.63±3.62	91.63±2.69	96.07±0.55	97.14±1.03
2	97.18±0.79	99.32±0.41	89.17±1.66	97.82±1.09	83.11±1.88	81.51±1.10	99.35±0.54	97.73±0.57	99.84±0.15
3	69.65±5.27	96.12±2.39	50.28±9.24	69.23±13.53	58.70±4.58	59.45±3.54	87.21±4.35	79.37±4.23	96.22±2.07
4	88.91±3.25	99.36±0.16	82.42±1.62	87.28±1.61	89.59±0.78	88.67±1.77	89.83±3.46	93.41±2.76	93.20±2.86
5	98.36±0.76	73.76±9.21	94.23±3.49	100.00±0.00	99.36±0.09	99.35±0.04	99.67±0.76	99.34±0.92	99.92±0.04
6	81.88±2.87	95.81±1.25	47.33±4.89	75.12±7.36	43.20±2.48	43.04±1.72	85.77±3.61	88.18±1.95	93.42±4.16
7	72.59±9.26	96.66±1.16	30.42±7.14	46.83±16.28	59.69±4.52	58.79±5.57	95.33±3.62	94.32±1.52	98.97±1.01
8	88.65±2.39	97.84±0.90	24.63±3.27	56.70±13.16	58.24±1.61	57.07±1.96	90.64±3.84	92.33±2.46	96.40±1.32
9	99.52±0.34	80.02±13.59	73.02±3.64	81.76±3.19	84.98±0.93	82.46±1.45	86.25±0.93	99.89±0.16	88.45±2.06
OA	91.06±0.48	96.78±0.52	73.43±0.88	86.07±2.08	72.15±0.80	70.80±1.08	94.13±0.76	94.67±0.49	97.52±0.46
AA	87.52±1.41	92.75±1.97	63.91±1.71	78.09±3.88	71.02±0.70	70.00±1.18	91.74±1.17	93.41±0.77	95.88±0.67
$\kappa \times 100$	88.06±0.64	95.73±0.69	64.26±1.14	81.27±2.82	63.48±0.88	61.86±1.32	92.15±1.02	92.91±0.67	96.53±0.79
Time/s	100	104	125	455	409	468	977	347	726

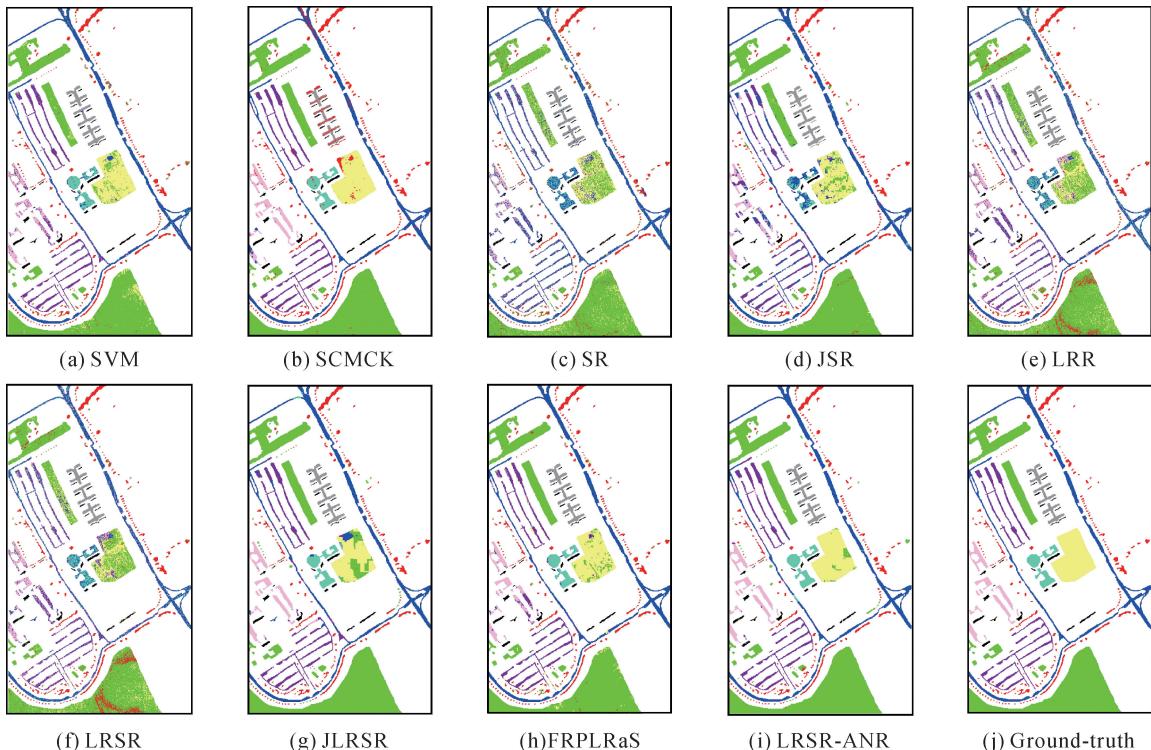


Fig.9 Classification maps obtained by different methods for the Pavia University data set

3.5 Generalization performance

To test the generalization performance of the proposed method, we apply all considered methods to the three data sets with different percentage of training samples. As shown in Fig.11, LRSR-ANR obtains the highest OAs in most of the cases for the three data sets, illustrating a better generalization performance. For Indian Pines, LRSR-ANR

performs better when the percentage of training samples is larger than 5%. For Pavia University and Salinas, LRSR-ANR always performs better under different percentage of training samples. For example, when using 0.1% training samples (i.e., only 42 samples for Pavia University and 54 samples for Salinas), the OAs obtained by LRSR-ANR are more stable and higher than 80%, demonstrating powerful

generalization performance of the proposed method.

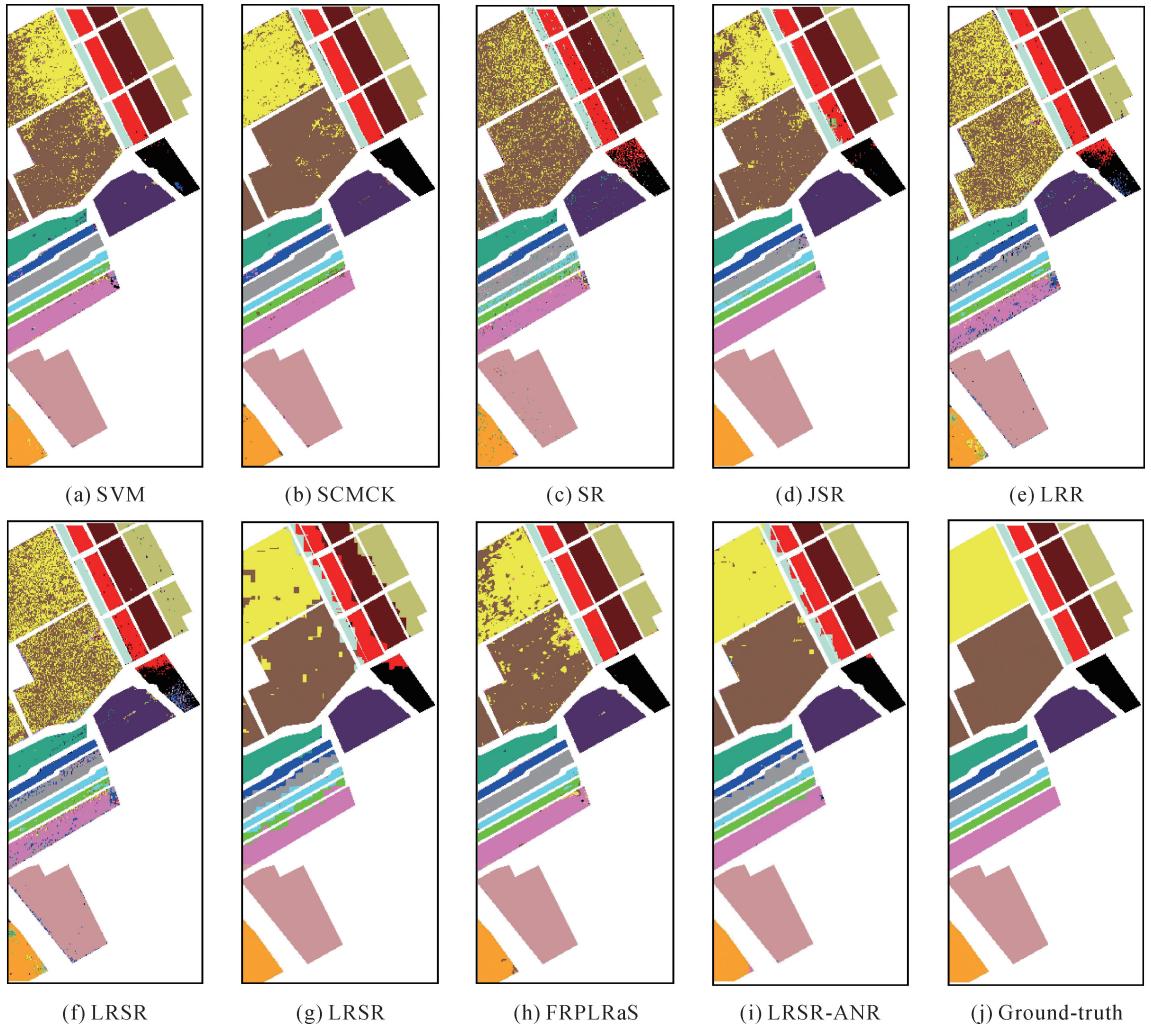
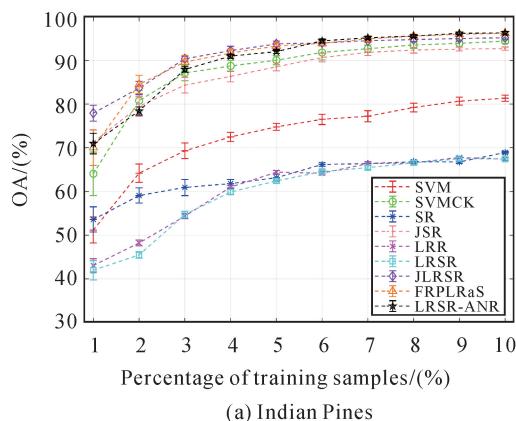


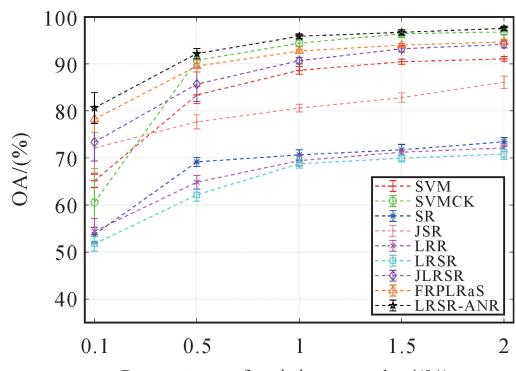
Fig.10 Classification maps obtained by different methods for the Salinas data set

Tab.7 Classification accuracy of different methods for the Salinas data set (2% for training) (%)

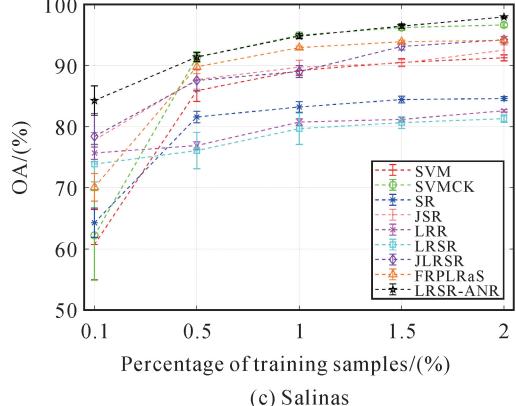
Class	SVM	SVMCK	SR	JSR	LRR	LRSR	JLRSR	FRPLRaS	LRSR-ANR
1	98.20± 0.91	99.08± 0.34	98.99± 0.36	99.99± 0.01	98.41± 0.38	97.48± 0.73	99.71± 0.16	99.67± 0.12	100.00± 0.00
2	99.46± 0.23	98.67± 0.52	98.19± 0.52	99.98± 0.02	97.23± 0.61	97.50± 0.57	99.95± 0.06	99.27± 0.22	99.95± 0.15
3	96.00± 2.95	98.78± 0.69	76.90± 3.31	96.59± 0.93	67.01± 3.88	67.92± 1.70	96.23± 2.62	99.83± 0.39	99.10± 2.09
4	98.37± 1.17	98.27± 1.94	92.88± 2.18	99.68± 0.23	98.55± 0.28	99.36± 0.06	72.03± 4.88	99.37± 0.81	81.23± 4.29
5	98.01± 0.52	97.80± 0.50	95.98± 0.87	95.01± 0.69	96.52± 1.22	97.67± 0.51	94.24± 1.87	99.29± 0.21	96.52± 1.04
6	99.52± 0.37	98.46± 1.56	98.80± 0.43	100.00± 0.00	99.44± 0.11	99.50± 0.07	96.4± 0.76	99.82± 0.13	99.93± 0.05
7	99.55± 0.25	98.97± 0.39	97.73± 0.52	100.00± 0.00	97.96± 0.28	97.42± 0.67	97.88± 0.86	99.34± 0.32	99.77± 0.11
8	88.87± 1.18	95.87± 0.74	84.88± 1.61	98.20± 0.50	64.12± 3.25	64.87± 2.29	94.55± 1.96	91.01± 0.70	98.22± 0.41
9	99.43± 0.38	99.22± 0.36	98.45± 0.51	99.99± 0.01	96.82± 0.23	97.05± 0.23	100.0± 0.00	99.64± 0.31	99.95± 0.03
10	90.65± 2.36	94.86± 2.76	91.31± 1.00	98.26± 0.53	76.21± 1.46	71.90± 1.19	92.91± 1.50	95.52± 1.03	97.69± 0.98
11	93.67± 2.78	94.59± 4.41	91.88± 4.39	99.87± 0.19	93.59± 1.11	93.71± 0.33	93.41± 2.09	95.46± 2.68	96.31± 1.78
12	99.52± 0.43	99.84± 0.09	83.04± 5.93	89.43± 7.05	91.06± 1.83	86.57± 3.90	88.87± 2.58	99.57± 0.28	92.86± 1.60
13	97.51± 1.96	94.44± 5.14	80.62± 5.11	87.82± 4.24	97.88± 0.26	97.41± 0.84	88.72± 5.19	97.94± 1.42	97.16± 1.79
14	90.94± 6.61	90.13± 4.30	83.34± 2.41	92.62± 3.53	87.14± 1.21	87.00± 0.97	85.22± 6.30	93.33± 4.38	92.66± 4.10
15	63.79± 3.29	91.51± 1.78	37.89± 3.76	56.07± 6.47	53.86± 4.33	53.08± 1.05	87.87± 3.25	76.33± 1.83	97.34± 0.97
16	96.05± 1.44	97.17± 2.22	94.85± 1.03	99.73± 0.27	88.71± 4.74	89.01± 3.28	99.01± 1.69	97.47± 0.69	99.21± 0.88
OA	91.24± 0.50	96.61± 0.35	84.59± 0.30	92.50± 0.86	81.58± 0.21	81.30± 0.48	94.21± 0.54	94.10± 0.28	97.91± 0.13
AA	94.35± 0.66	96.43± 0.70	87.86± 0.52	94.58± 0.53	87.78± 0.49	87.34± 0.32	92.94± 0.45	96.43± 0.29	96.74± 0.37
$\kappa \times 100$	90.22± 0.56	96.22± 0.39	82.77± 0.33	91.61± 0.97	79.52± 0.24	79.21± 0.53	93.55± 0.61	93.43± 0.32	97.63± 0.21
Time/s	277	280	262	1274	768	889	2217	580	1690



(a) Indian Pines



(b) Pavia University



(c) Salinas

Fig.11 OA obtained by different methods as a function of different percentage of training samples

3.6 Statistical test

To test the statistical differences between the proposed LRSR-ANR and other methods on the three data sets, we calculated the z -score values of OA and Kappa. The critical value of z -score is 1.96 at a confidence level of 0.95 and corresponding results are listed in Tab.8.

According to Tab.8, all test results for the Salinas are significant with 95% confidence. However, for Pavia University, the statistical difference between LRSR-ANR and SVMCK is not significant. Similarly, for Indian Pines, the statistical difference between LRSR-ANR and FRPLRaS does not reach the 95% confidence level. Comparing the results in Tabs.5, 6 and 7, we find that the classification performance of SVMCK and LRSR-ANR are unstable on the Pavia University data set. In addition, LRSR-ANR and FRPLRaS achieved similar and excellent classification results on the Indian Pines data set. These reasons lead to small statistical differences between the three methods.

3.7 Ablation analysis

To evaluate the effects of ANR scheme, we compare some variants of LRSR model including: ① LRSR without any window (the baseline method of LRSR); ② LRSR with fixed window (LRSR-no-adaptive); ③ LRSR-ANR using Euclidean distance (LRSR-ANR-Euc); ④ LRSR-ANR using Cosine distance (LRSR-ANR-Cos); And ⑤ the proposed method LRSR-ANR jointly using Euclidean and Cosine distance. Note that we have carefully optimized the involved threshold in different variants of LRSR model as listed in Tab.9.

Tab.8 Statistical test between methods in terms of κ z-score and OA z-score

Between-method	Indian Pines		Pavia University		Salinas	
	κ (z-score)	OA (z-score)	κ (z-score)	OA (z-score)	κ (z-score)	OA (z-score)
SVM/LRSR-ANR	16.45	19.38	11.3	9.72	12.39	12.91
SVMCK/ LRSR-ANR	2.45	3.22	0.76	1.06	3.18	3.48
SR/ LRSR-ANR	31.06	37.31	26.78	24.26	37.99	40.73
JSR/ LRSR-ANR	4.91	6.65	5.36	5.37	6.06	6.22
LRR/ LRSR-ANR	33.68	40.35	34.33	27.49	56.78	66.12
LRSR/ LRSR-ANR	32.4	38.35	25.18	22.76	32.31	33.4
JLRSR/LRSR-ANR	1.48	2.01	3.39	3.81	6.32	6.66
FRPLRaS /LRSR-ANR	0.12	0.16	3.50	4.24	10.97	12.34

Tab.9 The optimal values of S in different variants of LRSR model

Data Sets	LRSR-ANR-Euc	LRSR-ANR-Cos	LRSR-ANR (Ours)
Indian Pines	$S = 0.925$	$S = 0.995$	$S = 0.900$
Pavia University	$S = 0.815$	$S = 0.935$	$S = 0.725$
Salinas	$S = 0.905$	$S = 0.990$	$S = 0.850$

Tab. 10 reports the accuracies obtained by different variants for the three data sets. According to the results, LRSR obtains the lowest accuracies in

different image scenes due to the lack of considering the spatial information. LRSR-no-adaptive performs better than LRSR since it considers the spatial information. By comparing three variants of LRSR-ANR model, we can see that Cosine distance is better than Euclidean distance. However, the proposed method is the best by jointly considering the two distance metrics. This test verifies the effectiveness of ANR in the proposed model.

Tab.10 A comparison of different variants of LRSR model in terms of classification accuracy (%)

Data Sets	Metrics	LRSR	LRSR-no-adaptive	LRSR-ANR-Euc	LRSR-ANR-Cos	LRSR-ANR (Ours)
Indian Pines	OA	67.43 ± 0.69	94.13 ± 0.41	95.56 ± 0.41	95.87 ± 0.29	96.29 ± 0.30
	AA	67.73 ± 1.31	92.66 ± 1.36	94.93 ± 0.50	95.41 ± 0.49	96.29 ± 1.84
	$\kappa \times 100$	62.66 ± 0.78	93.31 ± 0.47	95.59 ± 0.72	95.65 ± 0.33	95.77 ± 0.66
Pavia University	OA	70.80 ± 1.08	92.01 ± 0.64	96.30 ± 0.66	96.57 ± 0.66	97.52 ± 0.46
	AA	70.00 ± 1.18	84.95 ± 0.82	95.23 ± 0.99	95.48 ± 0.94	95.88 ± 0.67
	$\kappa \times 100$	61.86 ± 1.32	89.39 ± 0.87	94.42 ± 0.52	94.61 ± 0.54	96.53 ± 0.79
Salinas	OA	81.30 ± 0.48	96.74 ± 0.16	97.04 ± 0.15	96.97 ± 0.24	97.91 ± 0.13
	AA	87.34 ± 0.32	94.47 ± 0.39	96.43 ± 0.24	96.47 ± 0.28	96.74 ± 0.37
	$\kappa \times 100$	79.21 ± 0.53	96.44 ± 0.17	96.19 ± 0.20	96.12 ± 0.33	97.63 ± 0.21

3.8 Comparison with deep learning methods

In this part, we compare the proposed LRSR-ANR with four deep learning methods, including 2D-CNN, 3D-CNN^[5], SSUN^[6], and SSRN^[7]. The experimental results based on 10 independent runs are listed in Tab.11. From the experimental results, the related methods based on deep learning can achieve excellent performance when the training samples are

sufficient. For example, in the Indian Pines data set based on 10% of the training samples, the SSRN achieves an overall classification accuracy of 97.50%, which is 1.21% higher than the LRSR-ANR. However, the performance of our method is still the best on the Pavia University and Salinas datasets based on 2% training samples.

Tab.11 Classification accuracy of different methods (%)

Data Sets	Metrics	2D-CNN	3D-CNN	SSUN	SSRN	LRSR-ANR
Indian Pines	OA	92.38 ± 1.05	94.53 ± 0.64	94.93 ± 1.14	97.50 ± 0.35	96.29 ± 0.30
	AA	90.25 ± 1.83	92.12 ± 2.38	93.85 ± 2.18	96.43 ± 1.41	96.29 ± 1.84
	$\kappa \times 100$	91.31 ± 1.19	93.76 ± 0.73	94.23 ± 1.30	97.15 ± 0.40	95.77 ± 0.66
Pavia University	OA	93.93 ± 0.61	96.22 ± 0.76	97.28 ± 0.50	96.76 ± 1.09	97.52 ± 0.46
	AA	91.77 ± 0.84	94.33 ± 1.05	95.43 ± 1.01	96.47 ± 0.89	95.88 ± 0.67
	$\kappa \times 100$	91.93 ± 0.80	94.18 ± 1.00	96.39 ± 0.66	95.69 ± 1.48	96.53 ± 0.79
Salinas	OA	91.91 ± 0.61	94.98 ± 0.70	97.69 ± 0.24	97.33 ± 0.47	97.91 ± 0.13
	AA	93.54 ± 0.48	96.22 ± 1.45	98.62 ± 0.16	98.61 ± 0.32	96.74 ± 0.37
	$\kappa \times 100$	91.01 ± 0.67	94.41 ± 0.77	97.42 ± 0.26	97.02 ± 0.52	97.63 ± 0.21

4 Conclusion

This paper proposed a novel LRSR-ANR method for HSI classification. In our method, LRSR is used to

maintain the global and local data structure of HSI and accelerate the convergence speed through the M-ADMM algorithm. In particular, an ANR scheme is designed to combine spatial information and eliminate

heterogeneous samples in the neighborhood, which enhances the traditional pixel-wise labeling strategy based on minimum reconstruction rule. Experimental results verify the superiority of the proposed model compared to other related methods. For classification accuracy, LRSR-ANR obtains an OA of 96.29% for Indian Pines with 0.09% ~ 28.59% improvements, an OA of 97.52% for Pavia University with 0.74% ~ 26.72% improvements, and an OA of 97.91% for Salinas with 1.3% ~ 16.61% improvements. As for the generalization performance, LRSR-ANR performs better in most of the cases under different percentage of training samples. Ablation verifies the effectiveness of the proposed ANR scheme.

Although our experimental results are encouraging, further work on additional scenes and comparison methods should be conducted in the future. We are also planning to further study the method from three perspectives: ① Exploring superpixel or other more effective adaptive neighborhood modeling methods; ② Extending our model to low-rank and JSR, with the aim of further improving the classification performance; ③ Addressing the imbalanced classification problem based on the proposed method.

References

- [1] GU Yanfeng, CHANUSSTOT J, JIA Xiuping, et al. Multiple kernel learning for hyperspectral image classification: a review [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(11) : 6547-6565.
- [2] GHAMISI P, YOKOYA N, LI Jun, et al. Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art[J]. IEEE Geoscience and Remote Sensing Magazine, 2017, 5(4) : 37-78.
- [3] HE Lin, LI Jun, LIU Chenying, et al. Recent advances on spectral-spatial hyperspectral image classification: an overview and new guidelines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(3) : 1579-1597.
- [4] GHAMISI P, MAGGIORI E, LI Shutao, 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 [J]. IEEE Geoscience and Remote Sensing Magazine, 2018, 6(3) : 10-43.
- [5] CHEN Yushi, JIANG Hanlu, LI Chunyang, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(10) : 6232-6251.
- [6] XU Yonghao, ZHANG Liangpei, DU Bo, et al. Spectral-spatial unified networks for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(10) : 5893-5909.
- [7] ZHONG Zilong, LI Jonathan, LUO Zhiming, et al. Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(2) : 847-858.
- [8] SONG Weiwei, LI Shutao, FANG Leyuan, et al. Hyperspectral image classification with deep feature fusion network [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(6) : 3173-3184.
- [9] XU Kejie, HUANG Hong, DENG Peifang, et al. Deep feature aggregation framework driven by graph convolutional network for scene classification in remote sensing[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(4) : 1-15.
- [10] XUE Zhaojun, ZHANG Mengxue, LIU Yifeng, et al. Attention-based second-order pooling network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11) : 9600-9615.
- [11] ZHU Minghao, JIAO Licheng, LIU Fang, et al. Residual spectral-spatial attention network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1) : 449-462.
- [12] LI Wei, DU Qian. A survey on representation-based classification and detection in hyperspectral remote sensing imagery [J]. Pattern Recognition Letters, 2016, 83(11) : 115-123.
- [13] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31 (2) : 210-227.
- [14] ZHANG Lei, YANG Meng, FENG Xiangchu. Sparse representation or collaborative representation: which helps face recognition? [C] // Proceedings of 2011 International Conference on Computer Vision (ICCV). Barcelona: IEEE, 2011: 471-478.
- [15] LIU Guangcan, LIN Zhouchen, YAN Shuicheng, et al. Robust recovery of subspace structures by low-rank representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1) : 171-184.
- [16] CHEN Y, NASRABADI N M, TRAN T D. Hyperspectral image classification using dictionary-based sparse representation [J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(10) : 3973-3985.
- [17] XUE Zhaojun, LI Jun, CHENG Liang, et al. Spectral-spatial classification of hyperspectral data via morphological component analysis-based image separation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53 (1) : 70-84.
- [18] XU Xiang, LI Jun, HUANG Xin, et al. Multiple morphological component analysis based decomposition for remote sensing

- image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(5) : 3083-3102.
- [19] DU Peijun, XUE Zhaohui, LI Jun, et al. Learning discriminative sparse representations for hyperspectral image classification [J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 9(6) : 1089-1104.
- [20] XUE Zhaohui, DU Peijun, LI Jun, et al. Simultaneous sparse graph embedding for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(11) : 6114-6133.
- [21] DUAN Yule, HUANG Hong, WANG Tao. Semisupervised feature extraction of hyperspectral image using nonlinear geodesic sparse hypergraphs [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 51(4) : 1-5.
- [22] LIU Jianjun, WU Zebin, SUN Le, et al. Hyperspectral image classification using kernel sparse representation and semilocal spatial graph regularization [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(8) : 1320-1324.
- [23] XUE Zhaohui, DU Peijun, LI Jun, et al. Sparse graph regularization for hyperspectral remote sensing image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(4) : 2351-2366.
- [24] XUE Zhaohui, DU Peijun, SU Hongjun, et al. Discriminative sparse representation for hyperspectral image classification: a semi-supervised perspective [J]. Remote Sensing, 2017, 9(4) : 386.
- [25] SUN Xiaoxia, QU Qing, NASRABADI N M, et al. Structured priors for sparse-representation-based hyperspectral image classification [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(7) : 1235-1239.
- [26] TU Bing, HUANG Siyuan, FANG Leyuan, et al. Hyperspectral image classification via weighted joint nearest neighbor and sparse representation [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(11) : 4063-4075.
- [27] PENG Jiangtao, SUN Weiwei, DU Qian. Self-paced joint sparse representation for the classification of hyperspectral images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2) : 1183-1194.
- [28] LI Dan, KONG Fanqiang, WANG Qiang. Hyperspectral image classification via nonlocal joint kernel sparse representation based on local covariance [J]. Signal Processing, 2021, 180: 107865.
- [29] PENG Jiangtao, LI Luoqing, TANG Yuanyan. Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(6) : 1790-1802.
- [30] YANG Weidong, PENG Jiangtao, SUN Weiwei, et al. Log-euclidean kernel-based joint sparse representation for hyperspectral image classification [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(12) : 5023-5034.
- [31] WANG Mingwei, JIA Zitong, LUO Jianwei, et al. A Hyper-spectral image classification method based on weight wavelet kernel joint sparse representation ensemble and β -whale optimization algorithm [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 2535-2550.
- [32] LIANG Miaoqiao, JIAO Licheng, XU Chundong. Deep feature-based multitask joint sparse representation for hyperspectral image classification [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(9) : 1578-1582.
- [33] ZHANG Chengkun, HAN Min. Multi-feature hyperspectral image classification with $L_2, 1$ norm constrained joint sparse representation [J]. International Journal of Remote Sensing, 2021, 42(12) : 4785-4804.
- [34] DUAN Yule, HUANG Hong, LI Zhengying, et al. Local manifold-based sparse discriminant learning for feature extraction of hyperspectral image [J]. IEEE Transactions on Cybernetics, 2021, 51(8) : 4021-4034.
- [35] DUAN Yule, HUANG Hong, TANG Yuxiao. Local constraint-based sparse manifold hypergraph learning for dimensionality reduction of hyperspectral image [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1) : 613-628.
- [36] GU Yanfeng, WANG Qianwang, WANG Hong, et al. Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(6) : 2739-2751.
- [37] AN Jinliang, ZHANG Xiangrong, ZHOU Huiyu, et al. Tensor-based low-rank graph with multimanifold regularization for dimensionality reduction of hyperspectral images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8) : 4731-4746.
- [38] SUN Weiwei, DU Qian. Graph-regularized fast and robust principal component analysis for hyperspectral band selection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(6) : 3185-3195.
- [39] ZHAI Han, ZHANG Hongyan, ZHANG Liangpei, et al. Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(3) : 1723-1740.
- [40] REN Shougang, WAN Sheng, GU Xingjian, et al. Semi-Supervised hyperspectral image classification using local low-rank representation [J]. Remote Sensing Letters, 2019, 10(2) : 195-204.
- [41] FENG Zhixi, YANG Shuyuan, WANG Min, et al. Learning dual geometric low-rank structure for semisupervised hyperspectral image classification [J]. IEEE Transactions on Cybernetics, 2021, 51(1) : 346-358.
- [42] LIU Qian, WU Zebin, SUN Le, et al. Kernel low-rank representation based on local similarity for hyperspectral image classification [J]. IEEE Journal of Selected Topics in Applied

- Earth Observations and Remote Sensing, 2019, 12 (6) : 1920-1932.
- [43] ZHAN Tianming, LU Zhenyu, WAN Minghua, et al. Multi-scale superpixel kernel-based low-rank representation for hyperspectral image classification [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(9) : 1642-1646.
- [44] SU Hongjun, YAO Wenjing, WU Zhaoyue, et al. Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171 : 238-252.
- [45] WANG Yuebin, MEI Jie, ZHANG Liqiang, et al. Self-supervised low-rank representation (SSLRR) for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(10) : 5658-5672.
- [46] WANG Qi, HE Xiang, LI Xuelong. Locality and structure regularized low rank representation for hyperspectral image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2) : 911-923.
- [47] MEI Shaohui, BI Qianqian, JI Jingyu, et al. Hyperspectral image classification by exploring low-rank property in spectral or/and spatial domain [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10 (6) : 2910-2921.
- [48] MEI Shaohui, HOU Junhui, CHEN Jie, et al. Simultaneous spatial and spectral low-rank representation of hyperspectral images for classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(5) : 2872-2886.
- [49] ZHANG Mengmeng, LI Wei, DU Qian. Joint low rank and sparse representation-based hyperspectral image classification [C] // Proceedings of 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). Los Angeles, CA: IEEE, 2016: 1-4.
- [50] PAN Lei, LI Hengchao, MENG Hua, et al. Hyperspectral image classification via low-rank and sparse representation with spectral consistency constraint [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(11) : 2117-2121.
- [51] CHEN Jie, CAO Hongju, CHEN Shuhan, et al. A fast low rank approximation and sparsity representation approach to hyperspectral anomaly detection [C] // Proceedings of 2020 IEEE International Geoscience and Remote Sensing Symposium. Waikoloa, HI: IEEE, 2020: 2432-2435.
- [52] DE MORSIER F, BORGEAUD M, GASS V, et al. Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54 (6) : 3410-3420.
- [53] LI Wei, LIU Jiabin, DU Qian. Sparse and low-rank graph for discriminant analysis of hyperspectral imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54 (7) : 4094-4105.
- [54] PAN Lei, LI Hengchao, DENG Yangjun, et al. Hyperspectral dimensionality reduction by tensor sparse and low-rank graph-based discriminant analysis [J]. Remote Sensing, 2017, 9 (5) : 452.
- [55] PAN Lei, LI Hengchao, LI Wei, et al. Discriminant analysis of hyperspectral imagery using fast kernel sparse and low-rank graph [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(11) : 6085-6098.
- [56] SUN Weiwei, YANG Gang, DU Bo, et al. A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55 (7) : 4032-4046.
- [57] SUN Weiwei, PENG Jiangtao, YANG Gang, et al. Fast and latent low-rank subspace clustering for hyperspectral band selection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(6) : 3906-3915.
- [58] RASTI B, GHAMISI P, PLAZA J, et al. Fusion of hyperspectral and LiDAR data using sparse and low-rank component analysis [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(11) : 6354-6365.
- [59] GAO Lianru, HONG Danfeng, YAO Jing, et al. Spectral superresolution of multispectral imagery with joint sparse and low-rank learning [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(3) : 2269-2280.
- [60] BOYD S, PARikh N, CHU E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. Foundations and Trends® in Machine Learning, 2011, 3(1) : 1-122.
- [61] LEE J, CHOE Y. Low rank matrix recovery via augmented lagrange mutiplier with nonconvex minimization [C] // Proceedings of 2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP). France: IEEE, 2016.
- [62] LIN Zhouchen, LIU Risheng, SU Zhixun. Linearized alternating direction method with adaptive penalty for low-rank representation [C] // Proceedings of 24th International Conference on Neural Information Processing Systems. Granada, Spain: Curran Associates Inc., 2011: 612-620.
- [63] LU Canyi, FENG Jiashi, YAN Shuicheng, et al. A unified alternating direction method of multipliers by majorization minimization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3) : 527-541.
- [64] CANDÈS E J, LI Xiaodong, MA Yi, et al. Robust principal component analysis? [J]. Journal of the ACM, 2011, 58 (3) : 1-31.
- [65] MELGANI F, BRUZZONE L. Classification of hyperspectral remote sensing images with support vector machines [J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42 (8) : 1778-1790.
- [66] CAMPS-VALS G, GOMEZ-CHOVA L, MUÑOZ-MARI J, et al. Composite kernels for hyperspectral image classification [J]. IEEE Geoscience and Remote Sensing Letters, 2006, 3(1) : 93-97.

© Mar 2022. This work is published under
<https://creativecommons.org/licenses/by-nc-nd/4.0/> (the
“License”). Notwithstanding the ProQuest Terms and
Conditions, you may use this content in accordance with the
terms of the License.