



# Hyperspectral band selection via region-aware latent features fusion based clustering

Jun Wang<sup>a</sup>, Chang Tang<sup>a,b,\*</sup>, Zhenglai Li<sup>a</sup>, Xinwang Liu<sup>c</sup>, Wei Zhang<sup>d</sup>, En Zhu<sup>c</sup>, Lizhe Wang<sup>a</sup>

<sup>a</sup> School of Computer Science, China University of Geosciences, Wuhan 430074, PR China

<sup>b</sup> State Key Lab. for Novel Software Technology, Nanjing University, Nanjing 210023, PR China

<sup>c</sup> School of Computer, National University of Defense Technology, Changsha 410073, PR China

<sup>d</sup> Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputing Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250000, PR China

## ARTICLE INFO

### Keywords:

Hyperspectral band selection  
Clustering  
Latent feature learning  
Feature fusion

## ABSTRACT

Band selection is one of the most effective methods to reduce the band redundancy of hyperspectral images (HSIs). Most existing band selection methods tend to regard each band as a whole, and then explore the band redundancy with the pixel-wise features directly. However, since the regions of HSIs corresponding to different objects have diverse spectral properties and spatial structure, such above scheme limits the performance of hyperspectral band selection due to the lack of spatial information. To address above issues, a novel band selection method via region-aware latent features fusion based clustering (RLFFC) is proposed. Specifically, we employ the superpixel segmentation to segment HSIs into multiple regions so that the spatial information of HSIs can be fully preserved. In order to capture the priori information, we construct its corresponding Laplacian matrix from which a group of low dimensional latent features are generated to further enhance the separability among different bands. Then, a shared latent feature representation of HSIs is obtained by fusing region-aware latent features to effectively capture the band redundancy of HSIs. Finally, the  $k$ -means clustering algorithm is utilized to obtain the index of the selected bands from the shared latent feature representation. As a result, the spectral and spatial properties are well exploited in the proposed method. Extensive experiments on four public hyperspectral datasets show that the proposed method achieves superior performance when compared with other state-of-the-art ones.

## 1. Introduction

Hyperspectral sensor captures the spectral and spatial information of target scenes by collecting dozens to hundreds of continuous hyperspectral bands. Compared with RGB images, HSIs have richer information and higher resolution. With the continuous development and maturity of hyperspectral imaging technology and image classification technology, HSIs are widely used in various fields, such as salient target detection [1–3], medical imaging processing [4–6], mineral exploration [7–9], and so on. Nevertheless, there are also some problems with hyperspectral images. For instance, the large amount of hyperspectral bands make hyperspectral image processing more difficult. In addition, within a certain range, the spectral of adjacent bands are relatively similar, which increase the computational complexity of image classification and affect the final accuracy of band selection. Hence, it is extremely necessary to reduce the dimension of hyperspectral images so as to tackle the curse of dimensionality.

Hyperspectral dimensional reduction can be generally divided into two categories, i.e., feature extraction [10–13] and feature selection (band selection) [14–17]. For feature extraction, the hyperspectral data is mapped to a lower dimensional feature space by specific mathematical transformation criteria, then the new features are used to represent the original hyperspectral images. The typical feature extraction methods include principal component analysis [18], independent component analysis [19] and linear discriminant analysis [20]. In the procedure of feature extraction [21–23], it processes and optimizes the original data features. Although it is beneficial to obtain better results in some cases, through feature transformation, it changes the physical meaning of hyperspectral images and leads to the loss of some key information. For the latter, band selection is usually carried out in the original feature space, i.e., it only selects some representative feature bands to form a band subset from the whole hyperspectral images without changing the original data information. Therefore, this kind of methods

\* Corresponding author at: School of Computer Science, China University of Geosciences, Wuhan 430074, PR China.

E-mail addresses: [wang.jun@cug.edu.cn](mailto:wang.jun@cug.edu.cn) (J. Wang), [tangchang@cug.edu.cn](mailto:tangchang@cug.edu.cn) (C. Tang), [yuezhenguan@cug.edu.cn](mailto:yuezhenguan@cug.edu.cn) (Z. Li), [xinwangliu@nudt.edu.cn](mailto:xinwangliu@nudt.edu.cn) (X. Liu), [wzhang@qlu.edu.cn](mailto:wzhang@qlu.edu.cn) (W. Zhang), [enzhu@nudt.edu.cn](mailto:enzhu@nudt.edu.cn) (E. Zhu), [lizhe.wang@gmail.com](mailto:lizhe.wang@gmail.com) (L. Wang).

not only retain the physical meaning of original data, but also reduce the redundancy of original hyperspectral bands. Consequently, we also focus on band selection in this paper.

According to the availability of class labels, band selection can be further divided into two types, supervised band selection [24] and unsupervised band selection [25], which is very similar to feature selection [26–29]. For the former, it selects a band subset based on certain metric characteristics and the correlation of sample class labels. This kind of methods require training samples firstly, and then select the optimal band subset. However, under most conditions, it is difficult and laborious to obtain the sample labels, and there is no guarantee that the number of samples with labels can meet the training requirements. On the contrary, unsupervised methods only need to select band subset according to the importance of bands without label information. In addition, some evaluation metrics are proposed to measure the importance of a certain band, such as information divergence [30], minimum noise value [31], Euclidean distance [32] and so on. As a result, this kind of methods are preferable for us.

In unsupervised band selection, the clustering-based methods attract much attention in recent years. Overall, these methods show good performance, such as [33–35]. They stretch each band into a single feature vector, then select the hyperspectral bands according to their corresponding objective functions. In practice, for a certain band, different regions often correspond to different objects or land covers in which spectral features are diverse. High dimensional pixel-wise features of each band contain much redundant information which limit the performance of clustering on hyperspectral band selection. Therefore, it is inappropriate to regard each band as a single feature vector directly. In addition, the utilization of spatial information is very effective to improve the low-dimensional feature representation and clustering accuracy [36–39], while most existing clustering-based band selection methods do not give it sufficient attention. Except for the clustering-based methods, we find that the image fusion can also be utilized to reduce the band redundancy of HSIs, e.g., [40–44]. Inspired by the above methods, in order to retain the spectral and spatial information of HSIs maximumly, the superpixel segmentation is adopted to segment HSIs into multiple regions in this paper. For all segmented regions, since the high dimensional features of each band contain extensive redundant information which are obviously unfriendly in computational complexity, we propose to learn latent features to reduce the feature dimension. Moreover, the latent features are also beneficial to enhance the separability among bands.

Due to the diversity and complementarity of the information among different regions, when the latent feature matrix of a certain region is used alone to represent the whole HSIs, it will result in the loss of some key information. Furthermore, if the latent feature matrices of all segmented regions are simply spliced and then used as the feature representation of HSIs, it will not only ignore the potential correlation among different regions, but also will cause the curse of dimensionality as the data dimension increases. This does not meet our actual needs, nor does it make any sense in theory. Toward this end, we adopt a fusion strategy to generate a shared latent feature representation of HSIs.

Thus, based on the above descriptions, we propose a novel band selection method via region-aware latent features fusion based clustering (RLFFC) to address above limitations simultaneously in this paper. The overall flowchart of RLFFC is shown in Fig. 1. To be specific, RLFFC couples the learning and fusion of all latent features into a joint framework. The main contributions of this work can be summarized as follows.

- (1) A region-aware latent features learning method is proposed to fully capture the spatial and spectral information for hyperspectral band selection. By learning low dimensional latent features from each region, the differentiation among bands can be improved while reducing dimension.

- (2) Since the spectral reflectance in various regions is different, an effective feature fusion method to explore the complementary information of each region is proposed. In such a manner, the shared latent feature representation of HSIs is obtained and the proposed method as a framework can be applied to any clustering algorithm.
- (3) An iterative optimization algorithm is designed to solve the resultant model. Extensive experiments on four public hyperspectral datasets show that the proposed method can achieve more outstanding performance when compared with other representative methods.

In order to distinguish various variables throughout this paper, we use boldface uppercase letters and boldface lowercase letters to denote matrices and vectors respectively. In addition, scalars are represented in non-bold italic font. For a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ ,  $\text{Tr}(\mathbf{M})$  denotes the trace of  $\mathbf{M}$ ,  $M_{ij}$  represents  $i$ th row and  $j$ th column element of matrix  $\mathbf{M}$ , and  $\mathbf{M}^T$  denotes the transpose of itself. A special case is if  $\mathbf{M}$  is a tensor of third order, then  $\mathbf{M}^{(n)}$  is used to represent  $\mathbf{M}(:, :, n)$ .  $\mathbf{I}_n$  denotes an identity matrix with size  $n \times n$ . And  $m_i$  denotes the  $i$ th element of vector  $\mathbf{m}$ .

The rest of this paper is organized as follows. Section 2 introduces several typical unsupervised hyperspectral band selection methods, including clustering-based ones and ranking-based ones. Section 3 presents the detail of the proposed method. In Section 4, the optimization algorithm is discussed, including its convergence proof and computation complexity analysis. In Section 5, experiments on four public datasets and the corresponding results as well as analysis are presented. Finally, the conclusion is given in Section 6.

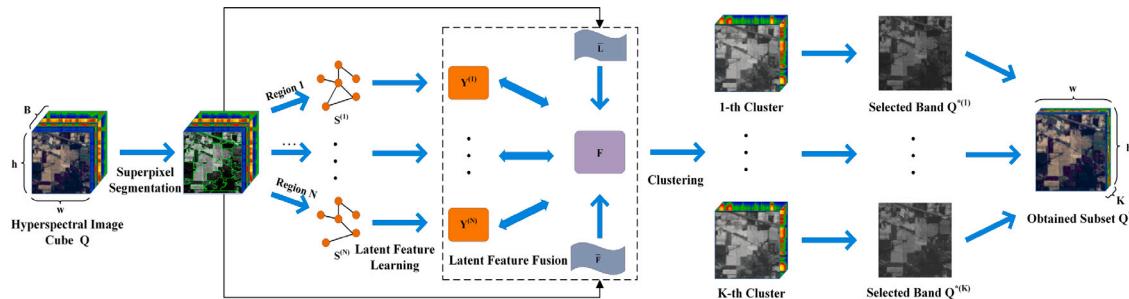
## 2. Related work

The typical unsupervised hyperspectral band selection methods can be roughly divided into two categories, i.e., clustering-based methods and ranking-based methods. Before introducing the proposed method, we give a short review about previous hyperspectral band selection methods.

### 2.1. Clustering-based methods

The clustering-based methods mainly partition all bands into multiple clusters based on a certain clustering algorithm, and then select the feature band subset from all clusters. Therefore, this kind of methods need to solve two problems, i.e., how to separate all bands into multiple clusters and how to select feature bands from each cluster. For the former, based on the fact that hyperspectral bands are ordered, Wang et al. [45] adopted a search algorithm to determine optimal unrelated hyperspectral subcube. Wang et al. [46] utilized dynamic programming to traverse all band combinations to determine the optimal partition of HSIs. In addition, some traditional clustering algorithms are used to partition all bands into multiple clusters, such as hierarchical clustering [47,48], spectral clustering [49,50] and  $k$ -means clustering [51,52]. As for the latter, Zhu et al. [53] selected the feature band that is closest to the clustering center from each class. Wang et al. [45] selected the bands with the maximum information entropy or the minimum noise value as the feature bands.

Overall, the above algorithms usually employ a group-wise strategy to select the representative bands from each cluster so as to avoid high correlation among hyperspectral subspace. Although these methods adopt many metrics to measure the importance of bands in each cluster, they still need to select the feature bands independently, which ignore the correlation among bands. Furthermore, they do not take into account the spatial information contained in HSIs, and they are also inaccurate to partition hyperspectral image cube only by relying on spectral information.



**Fig. 1.** Framework of RLFFC. Firstly, superpixel segmentation is adopted to segment hyperspectral image cube  $Q \in \mathbb{R}^{w \times h \times B}$  into multiple regions. And then we construct the similarity matrices  $\{S^{(i)}\}_{i=1}^N$  of all segmented regions by using  $k$ -nearest neighbor graph. Next latent feature matrices  $\{Y^{(i)}\}_{i=1}^N$  are obtained with the spectral embedding of the Laplacian matrices constructed by  $\{S^{(i)}\}_{i=1}^N$ , and the shared latent feature representation  $F \in \mathbb{R}^{B \times d}$  of HSIs is generated from  $\{Y^{(i)}\}_{i=1}^N$ . In addition, average latent feature matrix  $\bar{F}$  and average Laplacian matrix  $\bar{L}$  of all regions are introduced to constrain  $F$ . Finally, the  $k$ -means algorithm is utilized to obtain the index of the selected bands in  $F$ . In this way, the whole band selection is completed and the selected band subset  $Q^*$  is the final result of  $Q$  with dimension reduction.

## 2.2. Ranking-based methods

The ranking-based methods mainly evaluate the weight of all bands according to a certain measurement metric, and then select the optimal subset from the bands by weight ranking. In [54–57], they calculate the weight of each band to the whole dataset according to an objective function, and then sort these weight bands to select the feature band subset. The typical ranking standard is the amount of information in one band. In addition, many metrics are also proposed to measure the amount of information, such as information divergence [47], correlation [58], mutual information [59], etc. The following is a brief introduction of these methods.

The method based on the information divergence [47] mainly calculates bands and their corresponding K-L divergence between the gaussian distribution, and then measures the difference through divergence. Finally, the band with largest divergence is selected. The method based on correlation [58] mainly measures the correlation between bands and the whole dataset by an objective function, and the feature band subset is selected according to the relevant of bands. The method based on mutual information [59] calculates the mutual information between the band and its reference band so as to represent the information of the band. The selection of reference band mainly depends on estimation.

On the whole, these information-based band selection methods can often select bands with more information, but neglect the correlation among bands, which lead to the high information redundancy of the selected band subset and poor performance compared with other types of methods.

## 3. Proposed method

In this section, the motivation and implementation of the proposed method (RLFFC) are introduced.

### 3.1. Latent feature matrices construction

As mentioned in Section 1, hyperspectral band selection should be considered not only from spectral information but also from spatial information. However, most of previous algorithms cannot satisfy the two requirements simultaneously. In the classification of hyperspectral images literature, only using the spectral information cannot obtain the satisfactory performance for hyperspectral band selection due to the lack of spatial structure and texture information. Not only that, it also does not address the problem of synonyms spectrum or foreign objects with the same spectrum. Thus, in order to address above issues, we segment HSIs into multiple homogeneous regions to capture the spatial information via superpixel segmentation. Specifically, we depict the spatial structure of original data by distilling the first principal component of HSIs which is then used to capture the spatial information via

entropy rate segmentation (ERS). In this way, the multiple segmented regions of HSIs are generated when we extend the segmented maps into all bands. Broadly speaking, each band can be regarded as a data point and the similarity between them can be considered as the weight of edge in space. As a result, the data matrices of all segmented regions can be transformed to graphs. To the best of our knowledge, most existing clustering-based band selection methods tend to employ the high-dimensional pixel-wise features to calculate the similarity among bands for band selection, which are obviously not constructive for subsequent clustering. Toward this end, instead of using original high dimensional pixel-wise features of each band, we propose to learn discriminative latent features via spectral embedding [60–63] which can enhance the separability among regions while reducing the feature dimension. The specific objective function is formulated as follows:

$$\min_{Y^{(i)}} \text{Tr}(Y^{(i)\top} L^{(i)} Y^{(i)}) \quad s.t. \quad Y^{(i)\top} Y^{(i)} = I, \quad (1)$$

where  $\{Y^{(i)}\}_{i=1}^N \in \mathbb{R}^{B \times d}$  denotes latent feature matrix of each region,  $B$  is the number of bands contained in HSIs,  $d$  is the dimension of latent feature space, and  $N$  is the number of segmented regions.  $L^{(i)} = I - D^{(i)-\frac{1}{2}} W^{(i)} D^{(i)-\frac{1}{2}}$  is the normalized Laplacian matrix of region  $i$ .  $W^{(i)} \in \mathbb{R}^{B \times B}$  is the similarity matrix of all bands in region  $i$ , and  $D^{(i)}$  is a diagonal matrix with each element  $D_{jj} = \sum_{i=1}^B W_{ij}$ . In this paper, the method for constructing similarity matrices is using the  $k$ -nearest neighbor graph, and the similarities among bands are mainly defined by the Euclidean distance. Moreover, Eq. (1) can be easily solved via eigenvalue decomposition on  $L^{(i)}$ , i.e., the optimal solution  $Y^{(i)}$  is conducted by the eigenvectors of  $L^{(i)}$  corresponding to the first  $d$  smallest eigenvalues. In this way, we can conduct the latent feature matrices of all segmented regions.

### 3.2. Latent feature matrices fusion

With respect to the latent features of all segmented regions, the trivial method to obtain a unified representation matrix of HSIs is simply sum up them. Nevertheless, it will cause the curse of dimensionality as the latent feature dimension increases, which is obviously inappropriate to the large scale HSIs. Due to the high correlation between adjacent bands, hyperspectral bands inevitably contain redundant latent features apart from their specific features. As a result, we learn the shared latent feature representation matrix of HSIs by fusing regions-aware latent features to exploit the complementary features of each region while preserving the common features of all regions. Furthermore, it is also helpful to boost the clustering performance for hyperspectral band selection. Hence, we can model the following problem:

$$\begin{aligned} & \max_{F, R, \gamma} \text{Tr}(F^\top \sum_{i=1}^N \gamma_i Y^{(i)\top} R^{(i)}) \\ & s.t. \quad F^\top F = I_d, R^{(i)\top} R^{(i)} = I_d, \sum_{i=1}^N \gamma_i^2 = 1, \gamma_i \geq 0, \end{aligned} \quad (2)$$

where  $\mathbf{F}$  denotes the shared latent feature matrix with the size of  $B \times d$ . Considering that the weights of diverse regions are different, we introduce a weight factor  $\gamma$  to balance the latent features corresponding to each region, and  $\{\gamma_i\}_{i=1}^N$  denotes the contribution rate of each region to the HSIs.  $\{\mathbf{R}^{(i)}\}_{i=1}^N \in \mathbb{R}^{d \times d}$  denotes rotation matrices for different region-aware features.

However, from above objective function, we can find that the obtained representation matrix is greatly affected by the weight factor, which directly influences the final clustering performance. In order to improve the quality of the obtained shared latent feature matrix, we introduce some constraints to regularize it. Toward this end, the final objective function of the proposed method in this paper is shown below:

$$\begin{aligned} & \max_{\mathbf{F}, \mathbf{R}, \gamma} \text{Tr}(\mathbf{F}^\top \sum_{i=1}^N \gamma_i \mathbf{Y}^{(i)} \mathbf{R}^{(i)}) + \lambda \text{Tr}(\mathbf{F}^\top \bar{\mathbf{F}}) + \beta \text{Tr}(\mathbf{F}^\top \bar{\mathbf{L}} \mathbf{F}) \\ & \text{s.t. } \mathbf{F}^\top \mathbf{F} = \mathbf{I}_d, \mathbf{R}^{(i)^\top} \mathbf{R}^{(i)} = \mathbf{I}_d, \sum_{i=1}^N \gamma_i^2 = 1, \gamma_i \geq 0, \end{aligned} \quad (3)$$

where  $\lambda$  and  $\beta$  are both balance parameters. For all latent features derived from the similarity graphs of HSIs, they usually contain common underlying cluster structure except for their unique cluster structure. To exploit this advantage, we introduce an average latent feature matrix  $\bar{\mathbf{F}} \in \mathbb{R}^{B \times d}$  to prevent the obtained fusion features from being too far from latent features. In Eq. (2), the Laplacian matrix for each region can only describe the local similarity among bands in this region, so it neglects the global similarity in all regions. If two bands are highly similar in the original feature space, they should also be highly similar in the new latent feature space. Therefore, average Laplacian matrix  $\bar{\mathbf{L}}$  of all regions is utilized to capture the consistent information among bands. Now we give the solution for  $\bar{\mathbf{L}}$  as follows:

$$\bar{\mathbf{L}} = \frac{1}{N} \sum_{i=1}^N \mathbf{L}^{(i)}. \quad (4)$$

In addition, the specific solution for  $\bar{\mathbf{F}}$  is similarity to  $\mathbf{Y}^{(i)}$  in Eq. (1), so it is not introduced in detail.

Finally, the  $k$ -means algorithm is utilized to calculate the distance between bands and the clustering centers on the shared latent feature matrix, and the band which is closest to the clustering center is selected as the feature band in each cluster. In such a manner, the whole band selection is completed.

#### 4. Optimization algorithm

The proposed model mainly contains three variables, i.e.,  $\mathbf{F}$ ,  $\{\mathbf{R}^{(i)}\}_{i=1}^N$  and  $\gamma$ , it is difficult to solve them at one step. Therefore, in order to solve all problems mentioned in Eq. (3), we design an iterative optimization algorithm to obtain optimal solution. The specific processes of solving different variables are as follows.

##### 4.1. Optimization F

When  $\{\mathbf{R}^{(i)}\}_{i=1}^N$  and  $\{\gamma_i\}_{i=1}^N$  are fixed, Eq. (3) can be rewritten as:

$$\min_{\mathbf{F}} \text{Tr}(\mathbf{F}^\top \mathbf{G}) + \text{Tr}(\mathbf{F}^\top \mathbf{A} \mathbf{F}) \quad \text{s.t. } \mathbf{F}^\top \mathbf{F} = \mathbf{I}_d, \quad (5)$$

where  $\mathbf{G} = -\sum_{i=1}^N \gamma_i \mathbf{Y}^{(i)} \mathbf{R}^{(i)} - \lambda \bar{\mathbf{F}}$  and  $\mathbf{A} = -\beta \bar{\mathbf{L}}$ . This first-order optimization problem with orthogonal constraint can be solved by [64].

##### 4.2. Optimization $\{\mathbf{R}^{(i)}\}_{i=1}^N$

When  $\mathbf{F}$  and  $\{\gamma_i\}_{i=1}^N$  are fixed, the equivalent of Eq. (3) can be simplified to:

$$\max_{\mathbf{R}^{(i)}} \text{Tr}(\mathbf{R}^{(i)^\top} \mathbf{Z}) \quad \text{s.t. } \mathbf{Z} = \gamma_i \mathbf{Y}^{(i)^\top} \mathbf{F}, \quad (6)$$

the above objective function can be solved by Theorem 1.

**Theorem 1.** Suppose the matrix  $\mathbf{Z}$  is decomposed by singular value decomposition (SVD) as  $\mathbf{Z} = \mathbf{U} \Sigma \mathbf{V}^\top$ , then the optimal solution of  $\mathbf{R}^{(i)}$  is:

$$\mathbf{R}^{(i)} = \mathbf{U} \mathbf{V}^\top. \quad (7)$$

**Proof.** Note that  $\mathbf{Z} = \mathbf{U} \Sigma \mathbf{V}^\top$ , then the Eq. (6) can be transformed to:

$$\max_{\mathbf{R}^{(i)}} \text{Tr}(\mathbf{R}^{(i)^\top} \mathbf{U} \sum \mathbf{V}^\top) = \max_{\mathbf{R}^{(i)}} \text{Tr}(\mathbf{V}^\top \mathbf{R}^{(i)^\top} \mathbf{U} \sum). \quad (8)$$

Let  $\mathbf{X} = \mathbf{V}^\top \mathbf{R}^{(i)^\top} \mathbf{U}$ , then  $\mathbf{X} \mathbf{X}^\top = \mathbf{V}^\top \mathbf{R}^{(i)^\top} \mathbf{U} \mathbf{U}^\top \mathbf{R}^{(i)} \mathbf{V} = \mathbf{I}_d$ . Therefore,  $\text{Tr}(\mathbf{V}^\top \mathbf{R}^{(i)^\top} \mathbf{U} \sum) = \text{Tr}(\mathbf{X} \sum) = \text{Tr}(\mathbf{X}^\top \sum) \leq \frac{1}{2} (\text{Tr}(\mathbf{X}^\top \mathbf{X}) + \text{Tr}(\sum^\top \sum)) = \frac{1}{2} (d + \sum_{i=1}^d \omega_i)$ . It is obviously that the objective function has a maximum value when  $\mathbf{X} = \sum$ , i.e.,  $\text{Tr}(\mathbf{V}^\top \mathbf{R}^{(i)^\top} \mathbf{U} \sum) = d$ . So Eq. (7) is the optimal solution of Eq. (6).  $\square$

##### 4.3. Optimization $\gamma$

By fixing  $\{\mathbf{R}^{(i)}\}_{i=1}^N$  and  $\mathbf{F}$ , the optimization problem of Eq. (3) can be formulated as:

$$\max_{\gamma} \sum_{i=1}^N \gamma_i \sigma_i \quad \text{s.t. } \sum_{i=1}^N \gamma_i^2 = 1, \quad (9)$$

where  $\sigma_i = \text{Tr}(\mathbf{F}^\top \mathbf{Y}^{(i)} \mathbf{R}^{(i)})$ . According to the Cauchy inequality, we can obtain that  $\sum_{i=1}^N \gamma_i \sigma_i \leq \sqrt{\sum_{i=1}^N \gamma_i^2 \sum_{i=1}^N \sigma_i^2} = \sqrt{\sum_{i=1}^N \sigma_i^2}$ . When the equal sign holds,  $\frac{\gamma_1}{\sigma_1} = \frac{\gamma_2}{\sigma_2} = \dots = \frac{\gamma_N}{\sigma_N} = k$ , then we can get  $\sum_{i=1}^N \gamma_i^2 = k^2 \sum_{i=1}^N \sigma_i^2 = 1$  and  $k = \sqrt{\frac{1}{\sum_{i=1}^N \sigma_i^2}}$ . So  $\gamma_i = \frac{\sigma_i}{\sqrt{\sum_{i=1}^N \sigma_i^2}}$  is the optimal solution of Eq. (9).

In a nut shell, the details of RLFFC can be summarized in Algorithm 1, where  $obj^t$  represents the value of objective function in the  $t$ th iteration. The implementation code of the algorithm can be downloaded from <https://github.com/WangJun2023/RLFFC>.

---

#### Algorithm 1 Hyperspectral Band Selection via RLFFC

---

**Input:** Hyperspectral image cube  $\mathbf{Q} \in \mathbb{R}^{w \times h \times B}$ , the number of selected bands  $K$ , the latent feature dimension  $d$ , balance parameter  $\lambda$  and  $\beta$ .

**Output:** The selected feature band subset  $\mathbf{Q}^*$ .

- 1: Segment HSIs into multiple regions via superpixel segmentation.
  - 2: Calculate latent feature matrix  $\{\mathbf{Y}^{(i)}\}_{i=1}^N$  via Eq. (1).
  - 3: Calculate average Laplacian matrix  $\bar{\mathbf{L}}$  and average latent feature matrix  $\bar{\mathbf{F}}$  for all segmented regions via Eq. (10).
  - 4: Initialize  $\{\mathbf{R}^{(i)}\}_{i=1}^N = \mathbf{I}_d$ ,  $\{\gamma_i\}_{i=1}^N = \frac{1}{\sqrt{N}}$  and  $t = 1$ .
  - 5: **while** not converged **do**
  - 6: Update  $\mathbf{F}$  by solving Eq. (5).
  - 7: Update  $\{\mathbf{R}^{(i)}\}_{i=1}^N$  by solving Eq. (6).
  - 8: Update  $\gamma$  by solving Eq. (9).
  - 9:  $t = t + 1$ .
  - 10: **end while** if  $\frac{(obj^t - obj^{t-1})}{obj^{t-1}} \leq 10^{-6}$  or  $t \geq 100$ .
  - 11: Apply  $k$ -means algorithm on the shared latent feature representation matrix  $\mathbf{F}$  to get all clusters.
  - 12: Select the band which is closest to the clustering centering as the feature band from each cluster.
  - 13: **Return the selected feature band subset  $\mathbf{Q}^*$ .**
- 

##### 4.4. Convergence analysis

Since the optimal solution of each subproblem can be obtained in Algorithm 1, the proposed method is obviously convergence. Now we provide the detailed proof as follows.

**Theorem 2.** The proposed algorithm RLFFC can converge to the globally optimal value in a finite number of iterations.

**Proof.** Considering that for  $\forall i, j$ , the inequality we can get as follows.  
 $\text{Tr}((\gamma_i \mathbf{Y}^{(i)} \mathbf{R}^{(i)})^\top (\gamma_j \mathbf{Y}^{(j)} \mathbf{R}^{(j)})) \leq \text{Tr}[(\mathbf{Y}^{(i)} \mathbf{R}^{(i)})^\top (\mathbf{Y}^{(j)} \mathbf{R}^{(j)})] \leq \frac{1}{2}(\text{Tr}[(\mathbf{Y}^{(i)} \mathbf{R}^{(i)})^\top (\mathbf{Y}^{(i)} \mathbf{R}^{(i)})] + \text{Tr}[(\mathbf{Y}^{(j)} \mathbf{R}^{(j)})^\top (\mathbf{Y}^{(j)} \mathbf{R}^{(j)})]) = d$ . Therefore, we can obtain that  $\text{Tr}(\mathbf{F}^\top \sum_{i=1}^N \gamma_i \mathbf{Y}^{(i)} \mathbf{R}^{(i)}) \leq \frac{1}{2}(\text{Tr}(\mathbf{F}^\top \mathbf{F}) + \text{Tr}[(\sum_{i=1}^N \gamma_i \mathbf{Y}^{(i)} \mathbf{R}^{(i)})^\top (\sum_{i=1}^N \gamma_i \mathbf{Y}^{(i)} \mathbf{R}^{(i)})]) \leq \frac{1}{2}(d + dN^2)$ . In addition, the inequality  $\lambda \text{Tr}(\mathbf{F}^\top \bar{\mathbf{F}}) \leq \frac{\lambda}{2}(\text{Tr}(\mathbf{F}^\top \mathbf{F}) + \text{Tr}(\bar{\mathbf{F}}^\top \bar{\mathbf{F}})) = \lambda d$ ,  $\beta \text{Tr}(\mathbf{F}^\top \bar{\mathbf{L}} \mathbf{F}) = \beta \text{Tr}((\mathbf{F} \mathbf{F}^\top)^\top \bar{\mathbf{L}} \mathbf{F}) \leq \frac{\beta}{2}[\text{Tr}((\mathbf{F} \mathbf{F}^\top)^\top (\mathbf{F} \mathbf{F}^\top)) + \text{Tr}(\bar{\mathbf{L}}^\top \bar{\mathbf{L}})] = \frac{\beta}{2}(d + \sum_{i,j=1}^B \bar{L}_{ij}^2)$  can also be obtained. Consequently, Eq. (3) is upper bounded. Furthermore, the algorithm [64] used to solve  $\mathbf{F}$  can also be guaranteed that the value of Eq. (5) is monotonically increasing and globally convergent when other variables are fixed. In addition, for the solving of  $\{\mathbf{R}^{(i)}\}_{i=1}^N$  and  $\gamma$ , when one variable is optimized and other variables are fixed, the problems solved by each variable are convex problems and independent of each other, and the value of objective function is monotonically increasing. So the global optimal solution can be obtained for both subproblems. As a result, the convergence of objective function can be achieved after multiple iterations.  $\square$

#### 4.5. Computational complexity analysis

The computational complexity of RLFFC mainly lies in superpixel segmentation,  $\mathbf{F}$  and  $\{\mathbf{R}^{(i)}\}_{i=1}^N$ . For HSIs, the computational complexity of ERS is  $\mathcal{O}(P \log P)$ , where  $P$  is the number of pixels of each band. For updating  $\mathbf{F}$ , it costs  $\mathcal{O}(Bd + Bd^2 + d^3)$  in each iteration.  $d$  and  $B$  are the dimension of the latent feature space and the number of bands respectively. For updating  $\{\mathbf{R}^{(i)}\}_{i=1}^N$ , it costs  $\mathcal{O}(Bd^2 + Nd^3)$  in each iteration,  $N$  is the number of segmented regions. Finally, the computational complexity of  $k$ -means clustering is  $\mathcal{O}(BK^2)$ .  $K$  is the number of clusters, i.e., the number of selected bands. Therefore, the total computational complexity of RLFFC is  $\mathcal{O}(Bd + 2Bd^2 + BK^2 + (N + 1)d^3 + P \log P)$ .

### 5. Experiment

In this section, extensive experiments on four datasets are conducted to verify the efficiency of RLFFC. Firstly, several public datasets and all algorithms compared in experiment are introduced. Then we present the experimental settings. Finally, the results of all algorithms on different datasets with classifiers are analyzed in detail, which verify the superior performance of the proposed method. Furthermore, we also analyze the influence of all parameters on the experimental results and verify the convergence of RLFFC in the experiment.

#### 5.1. Datasets

**Indian Pines Scene** was collected by AVIRIS sensor in 1992. It is mainly consists of  $145 \times 145$  pixels and 220 spectral reflectance bands. The wavelengths of these bands are all between 0.4 and 2.5  $\mu\text{m}$ . It contains 16 categories in the available ground truth. We select 200 bands for experiment by removing 20 bands which are water absorption bands.

**Salinas Scene** was captured by AVIRIS sensor in California. It is mainly composed of  $512 \times 217$  pixels and 224 spectral bands and has 3.7 m pixels spatial resolution. We select 204 bands for experiment by removing 20 bands which are water absorption bands.

**Kennedy Space Center (KSC)** was collected by AVIRIS sensor in 1996. It is mainly consists of  $512 \times 614$  pixels and 224 spectral bands. We select 176 bands for experiment by removing 48 bands which are water absorption bands.

**Botswana** was captured by NASA EO-1 satellite sensors between 2001 and 2004. It is mainly consists of  $1476 \times 256$  pixels and 242 spectral reflectance bands. The wavelengths of these bands are all between 0.4 and 2.5  $\mu\text{m}$ . We select 145 bands for experiment by removing 97 bands which are water absorption bands.

#### 5.2. Compared methods

In order to verify the efficiency of RLFFC, several state-of-the-art algorithms are compared with it.

**ASPS\_MN, ASPS\_IE** [45] adopt a search algorithm to determine the optimal neighbor subspace, and the band with maximum information entropy or minimum noise value is selected from each subspace.

**TOF** [46] mainly uses dynamic programming to search an optimal segmentation of hyperspectral image cube by traversing all combinations of subcube, and the band with the highest score which is calculated by [65] is selected from each subcube.

**Uniform Band Selection (UBS)** [66] just partition HSIs into multiple subspace uniformly, and each segmentation point is selected as the feature band.

**FNGBS** [67] mainly uses a coarse–fine strategy to segment HSIs into several subspace, and then the band with the maximum product of local density and information entropy is selected from each subspace.

**ONR** [68] selects the optimal band combination which can minimize the error of reconstructing the original data.

In addition, with the purpose of fully reflecting the performance of all algorithms compared, we consider all bands as feature bands. It should be noted that the parameters of all algorithms are consistent with those in the original paper.

#### 5.3. Experimental setup

In order to evaluate the performance of RLFFC and other competitors, we choose three classifiers, including  $k$ -nearest neighborhood (KNN) classifier [69], support vector machine (SVM) classifier [70] and linear discriminant analysis (LDA) classifier [71]. In the experiment, the parameter  $k$  of KNN classifier is set 5, and gaussian kernel function is used uniformly in SVM. In addition, for Indian Pines, Salinas, KSC and Botswana dataset, the penalty coefficient of SVM is  $0.5 \times 10^4$ , 100,  $10^4$ ,  $10^4$  and the gamma of SVM is 0.5, 16, 16, 0.5. Three classifiers are all supervised classification, so we random choose 10% of the whole dataset as training samples and the rest as test samples. Since the optimal number of selected bands for each dataset is unknown, the number of selected bands to range from 5 to 50 each 5 interval is set. Besides, three metrics are adopted to measure the final results of classification, i.e., overall accuracy (OA), average accuracy (AA) and Kappa coefficient (Kappa). All the experiments are conducted 10 times individually to reduce the randomness of all results. In the experiment, all algorithms are implemented by MATLAB 2020a, and are conducted using the Intel Core i7-7700 3.60 GHz CPU with 24 GB memory.

Generally speaking, a hyperspectral image cube can be defined as  $\mathbf{Q} \in \mathbb{R}^{P \times B}$ ,  $P$  and  $B$  represent the number of pixels contained in each band and the number of bands respectively. In the segmentation of HSIs, one of the most important problem is to segment the target image thoroughly so that each region can be considered as homogeneous after segmentation [72]. Too large or too small of segmented regions of HSIs will affect the accuracy of final band selection. However, to the best of our knowledge, how to determine the number of superpixels is still an open problem. It is difficult to determine the unified number of superpixels  $N$  due to the difference between spatial features and spatial resolutions of HSIs. So its determination is often experimental. In order to alleviate the influence of this variable on the results and enhance the adaptive ability of the proposed method to new scenarios, we adopt a new manner to determine the number of superpixels based on the proposed method by [73,74]. Except for the texture information of HSIs, we think the spatial resolution of HSIs should also be taken into account. The hyperspectral datasets with low spatial resolution usually contain more spatial information, and the ones with high spatial resolution can reflect more contextual structure. Therefore, the distinction should be used to guide the determination of  $N$ , and we define the number of superpixels as follows:

$$N = 500 \times \frac{N_z}{P} + 40 \times \frac{N_z}{N_b \times \sqrt{res}}, \quad (10)$$

**Table 1**

OA, AA and Kappa of all algorithms on four public hyperspectral datasets. The best results are highlighted in bold in Table (%).

Datasets	Classifier	Metrics	UBS	TOF	ASPS_MN	ASPS_IE	FNGBS	ONR	RLFFC
Indian Pines	KNN	OA	59.68 ± 0.60	61.76 ± 0.67	71.24 ± 0.62	71.54 ± 0.57	69.48 ± 0.82	69.27 ± 0.45	<b>71.65 ± 0.68</b>
		AA	47.45 ± 0.45	49.93 ± 1.68	57.08 ± 1.50	56.89 ± 0.67	56.95 ± 1.66	56.93 ± 1.46	<b>60.50 ± 0.80</b>
		Kappa	57.10 ± 0.60	59.23 ± 0.67	68.95 ± 0.63	69.25 ± 0.47	67.15 ± 0.82	66.93 ± 0.46	<b>69.37 ± 0.69</b>
	SVM	OA	78.10 ± 0.50	77.55 ± 0.60	78.48 ± 0.47	79.38 ± 0.80	80.23 ± 0.66	79.38 ± 0.49	<b>80.87 ± 0.22</b>
		AA	73.12 ± 1.84	73.65 ± 1.79	74.42 ± 0.91	75.41 ± 0.97	77.87 ± 1.04	74.46 ± 0.84	<b>79.54 ± 0.50</b>
		Kappa	76.16 ± 0.51	75.58 ± 0.63	76.55 ± 0.50	77.51 ± 0.86	78.44 ± 0.68	77.54 ± 0.52	<b>79.09 ± 0.24</b>
	LDA	OA	65.92 ± 0.61	66.49 ± 0.74	68.32 ± 0.84	70.02 ± 0.87	69.40 ± 0.59	70.01 ± 0.63	<b>71.33 ± 0.21</b>
		AA	62.14 ± 1.65	60.47 ± 1.71	66.76 ± 0.68	66.71 ± 1.81	66.21 ± 1.41	66.61 ± 1.65	<b>71.32 ± 0.51</b>
		Kappa	63.58 ± 0.59	64.12 ± 0.77	66.13 ± 0.84	67.86 ± 0.89	67.19 ± 0.63	67.85 ± 0.65	<b>69.21 ± 0.21</b>
Salinas	KNN	OA	86.80 ± 0.19	89.08 ± 0.16	87.83 ± 0.17	89.28 ± 0.18	88.04 ± 0.21	89.65 ± 0.16	<b>90.14 ± 0.07</b>
		AA	92.10 ± 0.17	93.82 ± 0.10	93.10 ± 0.12	94.07 ± 0.21	92.90 ± 0.14	94.31 ± 0.19	<b>94.72 ± 0.06</b>
		Kappa	85.73 ± 0.20	88.16 ± 0.17	86.83 ± 0.18	88.38 ± 0.19	87.05 ± 0.22	88.77 ± 0.17	<b>89.29 ± 0.07</b>
	SVM	OA	92.45 ± 0.11	92.99 ± 0.11	92.50 ± 0.18	92.19 ± 0.13	92.58 ± 0.08	92.90 ± 0.07	<b>93.02 ± 0.05</b>
		AA	95.80 ± 0.15	96.27 ± 0.16	95.97 ± 0.21	95.82 ± 0.11	96.09 ± 0.12	96.35 ± 0.09	<b>96.55 ± 0.05</b>
		Kappa	91.76 ± 0.12	92.33 ± 0.11	91.81 ± 0.19	91.48 ± 0.14	91.90 ± 0.09	92.24 ± 0.08	<b>92.37 ± 0.06</b>
	LDA	OA	89.36 ± 0.18	89.54 ± 0.07	89.29 ± 0.11	88.36 ± 0.17	89.69 ± 0.18	90.07 ± 0.12	<b>90.38 ± 0.05</b>
		AA	93.19 ± 0.18	92.67 ± 0.08	93.13 ± 0.16	92.56 ± 0.16	93.48 ± 0.19	93.65 ± 0.13	<b>94.11 ± 0.07</b>
		Kappa	88.44 ± 0.19	88.63 ± 0.07	88.37 ± 0.11	87.37 ± 0.18	88.79 ± 0.19	89.21 ± 0.13	<b>89.53 ± 0.06</b>
KSC	KNN	OA	85.24 ± 0.87	85.50 ± 0.46	84.62 ± 0.77	84.63 ± 0.71	84.50 ± 0.86	85.37 ± 0.60	<b>87.15 ± 0.24</b>
		AA	78.38 ± 1.48	77.57 ± 1.09	77.30 ± 1.13	77.47 ± 1.39	77.04 ± 1.32	78.78 ± 0.70	<b>81.33 ± 0.28</b>
		Kappa	83.88 ± 0.93	84.12 ± 0.50	83.19 ± 0.83	83.21 ± 0.77	83.06 ± 0.93	84.00 ± 0.64	<b>85.92 ± 0.26</b>
	SVM	OA	79.76 ± 0.40	82.84 ± 1.18	80.81 ± 0.73	80.32 ± 0.70	80.07 ± 0.76	81.46 ± 0.89	<b>84.67 ± 0.58</b>
		AA	68.48 ± 1.19	73.55 ± 1.89	70.62 ± 1.89	68.99 ± 0.98	67.66 ± 1.79	72.34 ± 1.48	<b>76.76 ± 0.69</b>
		Kappa	77.96 ± 0.43	81.23 ± 1.28	79.07 ± 0.78	78.54 ± 0.75	78.26 ± 0.82	79.76 ± 0.97	<b>83.22 ± 0.63</b>
	LDA	OA	89.28 ± 0.56	89.17 ± 0.47	88.77 ± 0.38	86.86 ± 0.85	89.10 ± 0.69	89.00 ± 0.68	<b>90.89 ± 0.09</b>
		AA	84.42 ± 0.76	84.33 ± 0.80	83.85 ± 0.65	80.90 ± 1.38	84.11 ± 0.81	84.29 ± 0.91	<b>86.84 ± 0.31</b>
		Kappa	88.22 ± 0.61	88.11 ± 0.51	87.68 ± 0.41	85.60 ± 0.92	88.03 ± 0.74	87.92 ± 0.74	<b>89.98 ± 0.10</b>
Botswana	KNN	OA	85.04 ± 0.85	83.48 ± 0.68	86.23 ± 0.40	84.98 ± 0.71	85.24 ± 1.10	85.92 ± 0.57	<b>88.40 ± 0.29</b>
		AA	86.27 ± 0.96	84.71 ± 0.53	87.31 ± 0.56	86.33 ± 0.64	86.43 ± 1.20	87.02 ± 0.74	<b>89.58 ± 0.27</b>
		Kappa	84.15 ± 0.89	82.52 ± 0.70	85.38 ± 0.42	84.08 ± 0.74	84.35 ± 1.14	85.06 ± 0.60	<b>87.65 ± 0.30</b>
	SVM	OA	89.33 ± 0.96	88.78 ± 0.66	89.42 ± 0.56	89.21 ± 0.80	88.74 ± 0.46	90.64 ± 0.55	<b>92.55 ± 0.19</b>
		AA	90.41 ± 0.94	89.71 ± 0.61	90.41 ± 0.57	90.32 ± 0.77	89.98 ± 0.48	91.51 ± 0.75	<b>93.39 ± 0.20</b>
		Kappa	88.63 ± 1.01	88.05 ± 0.69	88.73 ± 0.59	88.51 ± 0.85	88.01 ± 0.48	90.01 ± 0.58	<b>92.03 ± 0.20</b>
	LDA	OA	90.16 ± 0.73	88.50 ± 0.86	90.55 ± 0.90	89.46 ± 0.52	90.26 ± 0.52	91.13 ± 0.52	<b>92.16 ± 0.25</b>
		AA	91.03 ± 0.83	89.41 ± 0.95	91.28 ± 1.09	90.20 ± 0.74	90.98 ± 0.52	91.72 ± 0.61	<b>92.96 ± 0.25</b>
		Kappa	89.51 ± 0.77	87.76 ± 0.90	89.91 ± 0.95	88.76 ± 0.55	89.60 ± 0.55	90.53 ± 0.55	<b>91.61 ± 0.27</b>

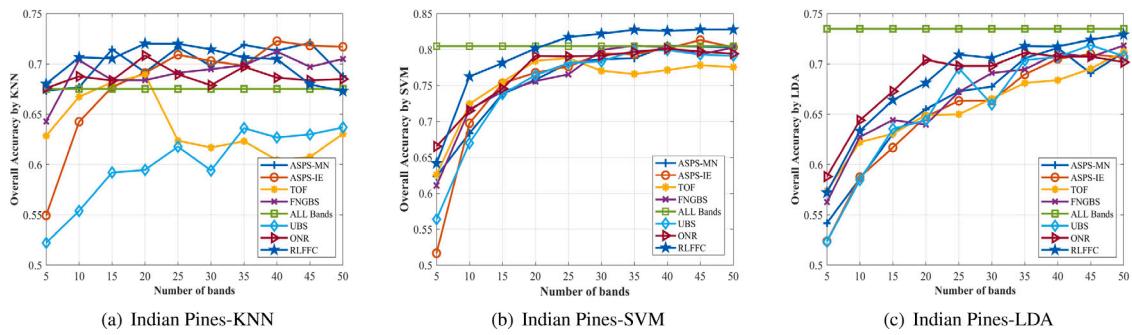


Fig. 2. OA for three classifiers by selecting different numbers of bands on the Indian Pines dataset.

where  $N_z$  represents the number of non-zero values at the edge of detected image.  $res$  represents spatial resolution of HSIs, and  $N_b$  is a fixed number. In all subsequent experiments, we set  $N_b = 10000$ . As a result, the number of superpixels  $N$  is set adaptively in this way.

#### 5.4. Results analysis

Table 1 shows the OA, AA and Kappa of all algorithms on four public datasets. In addition, the number of selected bands for four public datasets is set to 30 uniformly, and the data in Table 1 reflects the range of fluctuations. In order to clearly show the best result on each dataset, we highlight them in bold. As seen in Table 1, the proposed RLFFC

achieves outstanding performance and consistently presents the best result on four public datasets when compared with other algorithms. For example, the proposed method obtains 2.17%, 2.17% and 2.27% improvement in terms of OA, AA and Kappa than the second performer (ASPS\_MN) on the Botswana dataset with KNN classifier respectively. Therefore, it not only shows RLFFC has superior performance, but also demonstrates the effectiveness of HSIs segmentation. Overall, among all competitors, the results of RLFFC are more stable than other algorithms on four public datasets, which reflects its strong robustness. To fully compare the performance of RLFFC with other competitors, the results on four public datasets are analyzed in detail below.

Fig. 2 shows the performance of all competitors on the Indian Pines dataset. It can be clearly seen that the proposed method achieves satisfactory results compared with other algorithms. In Fig. 2(a), the

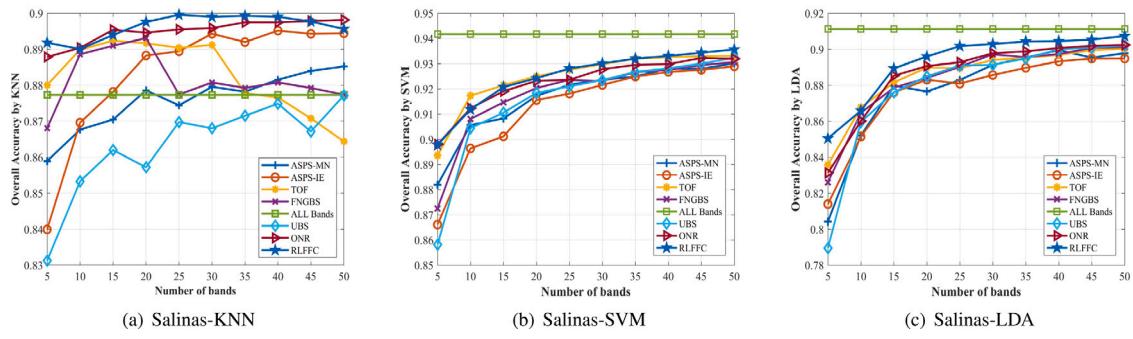


Fig. 3. OA for three classifiers by selecting different numbers of bands on the Salinas dataset.

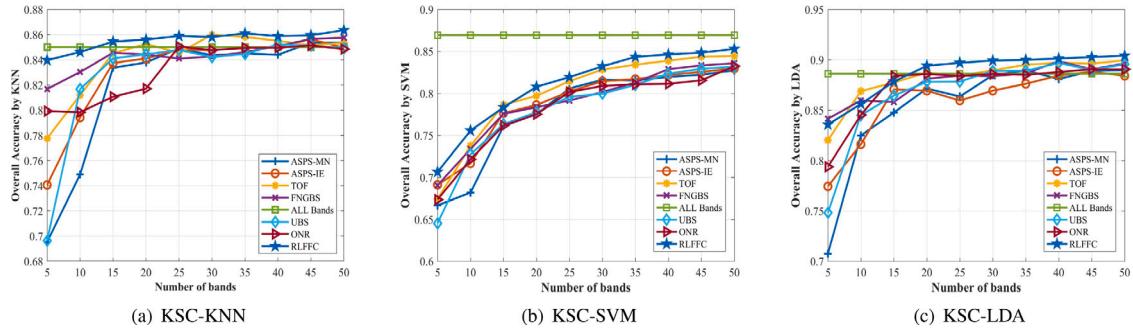


Fig. 4. OA for three classifiers by selecting different numbers of bands on the KSC dataset.

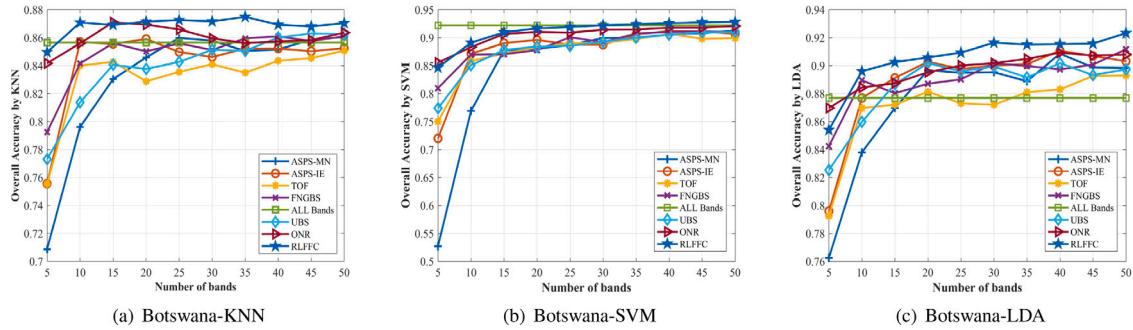


Fig. 5. OA for three classifiers by selecting different numbers of bands on the Botswana dataset.

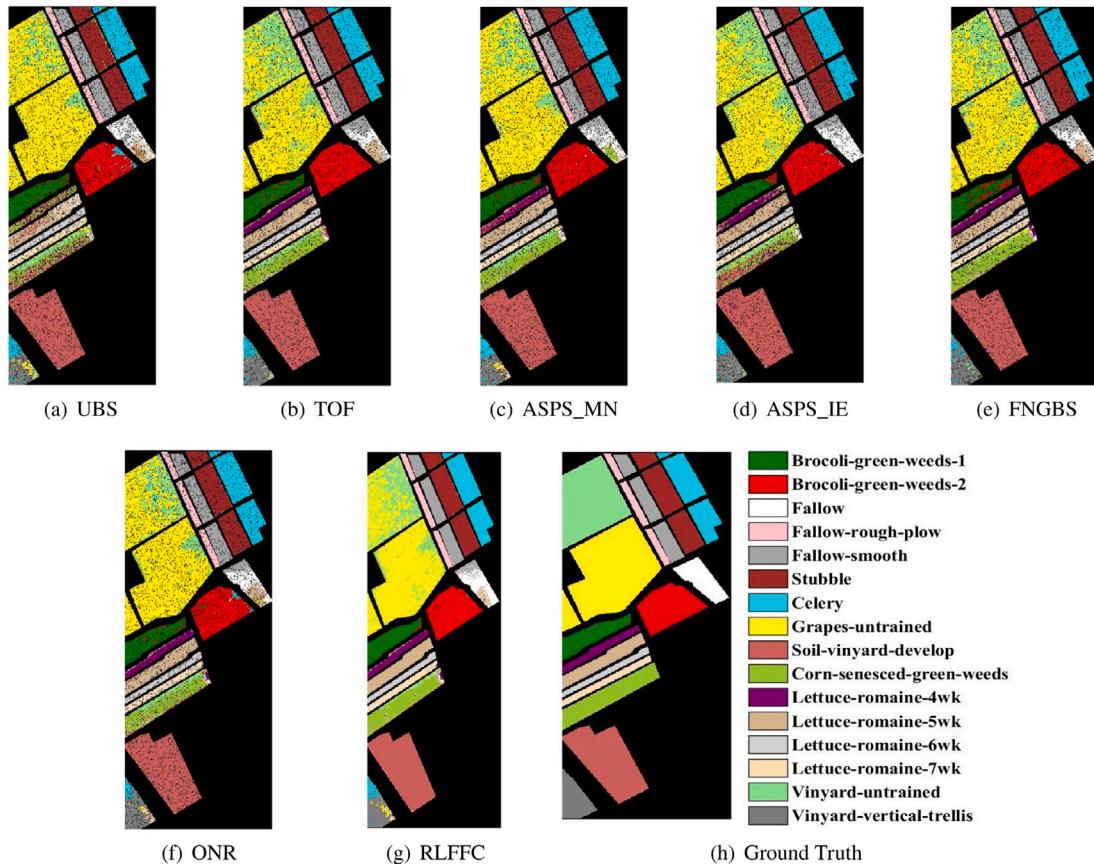
performance of RLFFC is optimal in most cases when the number of selected bands is less than 30, while some competitors obtain unstable performance, such as TOF and UBS. With respect to Fig. 2(b) and Fig. 2(c), the performance of RLFFC is also comparable with other competitors. For example, the OA of our method is about 3% higher than that of the second performer (TOF) when the number of selected bands is 10 in Fig. 2(b). Moreover, it should be noted that the proposed method always obtain outstanding performance when the number of selected bands is greater than 5 in Fig. 2(b). On the whole, the proposed method can achieve superior performance on the Indian Pines dataset.

For Salinas dataset in Fig. 3, although the advantages are not as obvious as those in Fig. 2, it is also comparable with other algorithms. Among all the competitors, we can find that our method can achieve satisfactory results in all classifiers. On the contrary, ASPS\\_IE obtains poor performance on three classifiers. In Figs. 3(b) and 3(c), the results of base line (ALL Bands) are always the highest. The reason is that the number of bands contained in this dataset is 204, and we only select less than a quarter of the bands as the feature bands. Nevertheless, the proposed method can also achieve more outstanding performance than the baseline. In Fig. 3(a), among all the algorithms, our method can obtain best result in most cases compared with the base line. Although all competitors achieve approximately results in SVM and

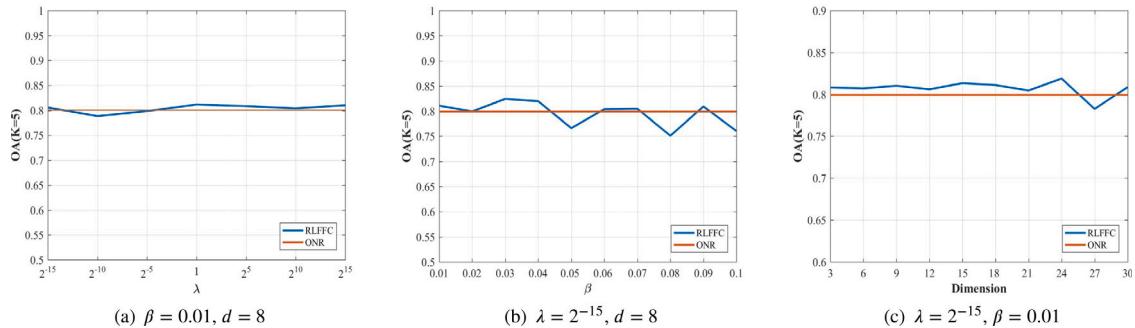
LDA classifier, our method can still perform better than other competitors. For example, it is always optimal when the number of selected bands exceeds 15 in LDA classifier. As a whole, it demonstrates the effectiveness of RLFFC in band selection.

Fig. 4 demonstrates the results of all algorithms on KSC dataset. In Fig. 4, we can observe that the performance of RLFFC is optimal in most cases. Moreover, when the number of selected bands is 5, the result of RLFFC is about 2% higher than the second performer (FNGBS) in Fig. 4(a). As to Figs. 4(b) and 4(c), ONR and TOF perform well, but RLFFC still shows better performance than them. For instance, when the number of selected bands exceeds 20, the performance of RLFFC is always superior to their. Overall, RLFFC also achieves satisfactory results on this dataset with three classifiers.

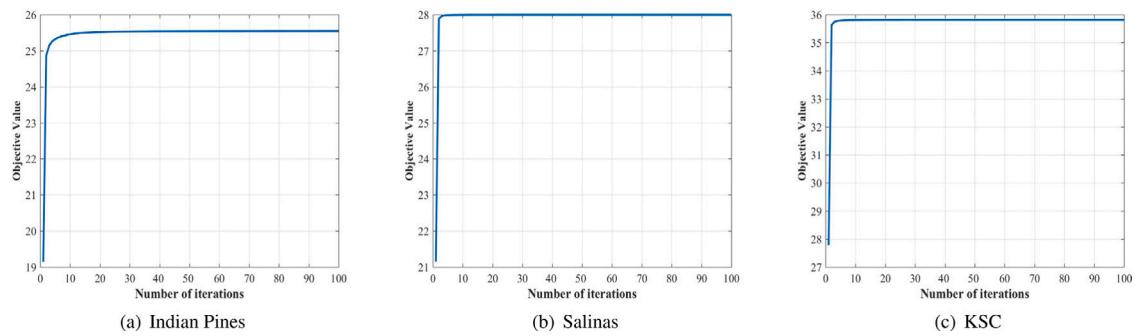
Fig. 5 shows the results of all algorithms on the Botswana dataset. According to the curve in Fig. 5, it can be seen that except when the number of selected bands is 5, the results of RLFFC are always optimal on three classifiers. On the contrary, the performance of TOF and ASPS\\_MN both are unsatisfactory. Furthermore, when the number of selected bands is 5, the result (70.9% and 52.7%) for ASPS\\_MN is significantly lower than other algorithms in Figs. 5(a) and 5(b). With respect to Fig. 5(c), the proposed method is superior to other algorithms. For example, RLFFC can obtain more than 1.5% improvement compared



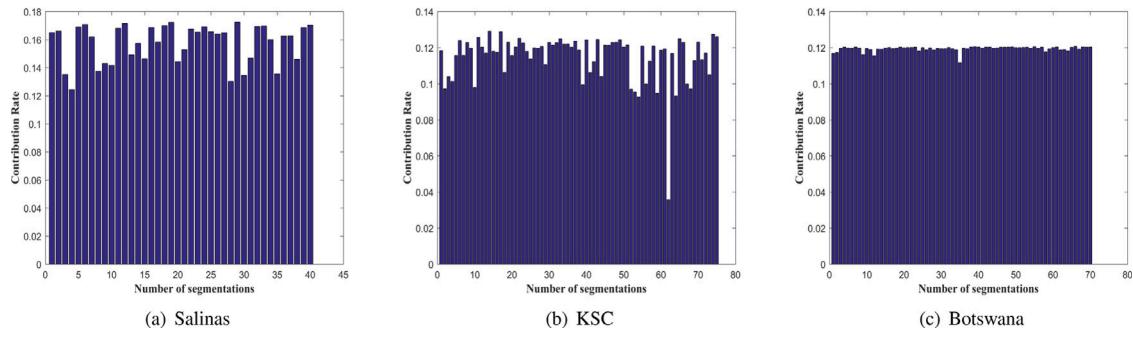
**Fig. 6.** The classification maps of all competitors on Salinas dataset.



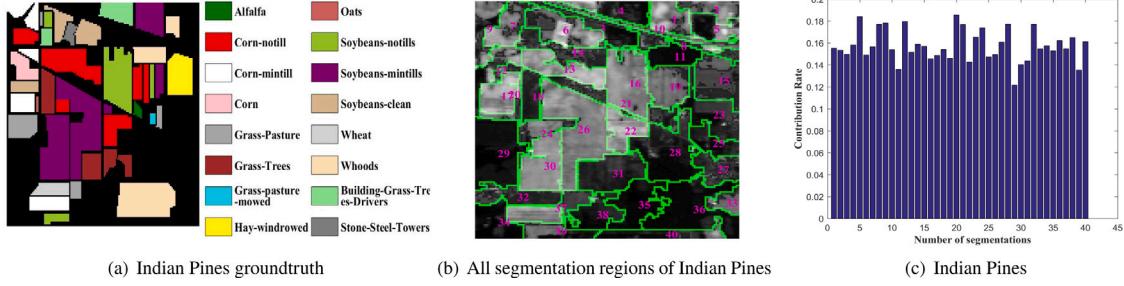
**Fig. 7.** OA for KNN classifier with different parameters  $\lambda$ ,  $\beta$  and  $d$  on the KSC dataset.



**Fig. 8.** Objective function values of Eq. (3) after multiple iterations on the Indian Pines, Salinas and KSC dataset (The Botswana dataset is similar and is omitted due to space limit).



**Fig. 9.** The contribution rate of each segmented region for the Salinas, KSC and Botswana dataset.



**Fig. 10.** The groundtruth map, all segmented regions and their contribution rate of the Indian Pines dataset.

with the second performer (ONR) when the number of selected bands is 30. Therefore, RLFFC can achieve outstanding performance in this dataset.

Furthermore, in order to clearly show the performance of the proposed method on hyperspectral band selection, the visual classification results of all competitors on the Salinas dataset are presented in Fig. 6. As seen in Fig. 6, the classification map of RLFFC is clearer than that of other algorithms, which verifies the effectiveness of it on band selection.

### 5.5. Parameter sensitivity analysis

According to Eq. (3), it can be seen that the proposed method has three parameters, namely  $\lambda$ ,  $\beta$  and  $d$ . In order to study the influence of three parameters on the final results, we set the number of selected bands is 5 and choose the ONR as the competitor on KSC dataset with KNN classifier. We fix two parameters every time and show the performance of RLFFC by changing the value of other parameters. Fig. 7 shows the sensitivity of RLFFC on the KSC dataset with KNN classifier. Because the results on other classifiers and datasets are similar to Fig. 7, we omit them here. In Figs. 7(a) and 7(b), it can be seen that the results fluctuate gently, which indicate that RLFFC is not very sensitive to parameter  $\lambda$  and  $\beta$ . Furthermore, it also reflect that RLFFC can achieve satisfactory results even under a wide range of parameters. Therefore, in the experiment, we set the range of two parameters as  $[2^{-15}, 2^{-14}, \dots, 2^{15}]$  and  $[0.01, 0.02, \dots, 0.1]$  respectively. In order to facilitate the selection of optimal parameters for different datasets, grid search is adopted to determine the optimal  $\lambda$  and  $\beta$  of each dataset in the experiment. As far as we know, how to determine the dimension of latent feature is an open problem, and there is no effective solution to determine the optimal feature dimension. So the strategy of grid search is adopted to determine its approximate range, as shown in Fig. 7(c). In Fig. 7(c), it can also be seen that RLFFC is not very sensitive to the latent feature dimension, i.e., parameter  $d$ . In order to reduce the influence of parameter  $d$  on the final results, in the experiment, the  $d$  of Indian Pines, Salinas, KSC and Botswana is set as 5, 6, 8 and 5 respectively.

### 5.6. Convergence validation

In Section 4, the convergence property of RLFFC has been proved in theory. Hence, we verify its convergence through experiments in this section. Fig. 8 demonstrates the objective function values of Eq. (3) after each iteration ( $\lambda = 2^{-15}$ ,  $\beta = 0.01$ ,  $d$  is the fixed value mentioned above). As shown in Fig. 8, it can be found that RLFFC can converge in less than 5 iterations on several datasets.

### 5.7. Contribution rate of each region validation

In this section, we verify the contribution rate of each segmentation region to HSIs is various in experiment, as shown in Figs. 9 and 10. For example, in Fig. 10(c), it can be seen that the contribution rate of the 29th region is 6% lower than that of the 32th region. The reason is that the 29th region in HSIs is the background while the 32th region is wheat in Figs. 10(a) and 10(b). Therefore, the above results not only demonstrate the various regions reflect different objects or land covers, but also show that it is reasonable to assign different contribution rate to all regions. As a result, it verifies the efficiency of the proposed method.

### 5.8. Discussion

In this section, we will discuss some phenomena that occurs in the experiment, so as to provide some constructive suggestions for subsequent work.

- (1) The reason why our method can achieve outstanding performance on hyperspectral band selection. First, the proposed method exploits the spectral and spatial properties of HSIs to explore the band redundancy simultaneously. The utilization of rich spatial information of HSIs can improve the quality of latent features and enhance the separability among bands. In this way, the proposed method can offer a strong discrimination when two bands with highly similarity are selected simultaneously. In order to evaluate the effectiveness of using the spatial information, we create a variant of RLFFC, i.e., we stretch each band into a single feature vector and then

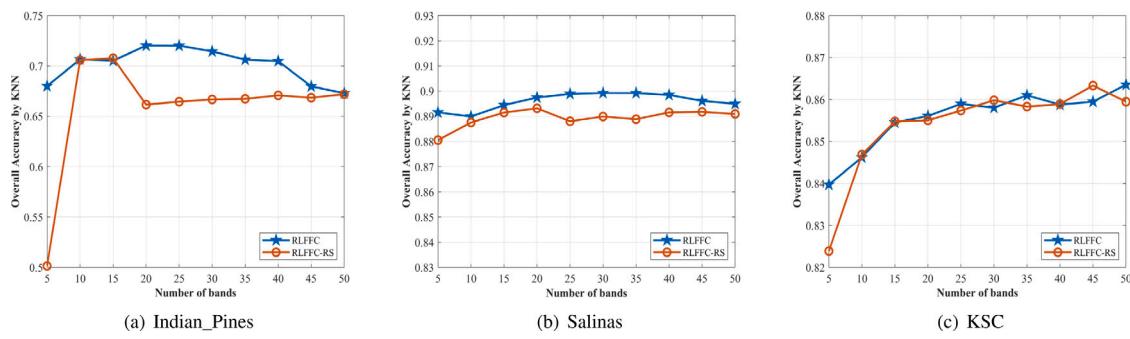


Fig. 11. Performance comparison of RLFFC and its variant on three hyperspectral datasets (The Botswana dataset is similar and is omitted due to space limit).

**Table 2**  
OA by selecting different parameters on four public datasets with using KNN classifier (%).

Datasets	$N_b$	$N$	5	10	15	20	25	30	35	40	45	50
Indian Pines	1000	46	68.48	69.17	71.03	71.06	70.58	70.78	70.86	69.69	66.74	65.01
	10000	40	68.00	70.66	70.53	72.02	72.01	71.45	70.63	70.48	67.98	67.30
	100000	40	68.00	70.66	70.53	72.02	72.01	71.45	70.63	70.48	67.98	67.30
Salinas	1000	145	88.58	89.14	89.60	89.71	89.48	89.63	89.27	89.08	89.23	89.02
	10000	40	89.16	88.99	89.45	89.75	89.89	89.93	89.92	89.85	89.61	89.50
	100000	25	88.88	88.88	89.54	89.86	89.86	89.44	89.39	89.39	89.20	89.18
KSC	1000	320	82.73	84.28	85.55	85.64	85.74	85.93	86.02	86.02	85.85	85.96
	10000	75	83.97	84.62	85.45	85.61	85.90	85.81	86.10	85.88	85.95	86.35
	100000	50	82.05	84.45	85.17	85.30	85.58	85.94	85.61	85.78	85.80	85.96
Botswana	1000	290	85.26	85.91	86.57	86.57	86.70	86.82	87.11	86.81	86.93	87.01
	10000	70	84.96	87.07	86.90	87.14	87.25	87.16	87.48	86.92	86.80	87.03
	100000	47	82.10	86.53	86.85	86.87	86.69	86.62	86.54	86.75	86.66	86.87

segment it into multiple regions randomly. In the experiment, the proposed method with superpixel segmentation is named as RLFFC, and the variant with randomly segmentation is named as RLFFC-RS. From Fig. 11, we can see that RLFFC outperforms its variant. This indicates that the utilization of spatial information is benefit for the band selection. Second, the optimization algorithm utilized in RLFFC can obtain the high quality latent feature representation of HSIs, which can improve the subsequent clustering performance.

(2) The influence of different numbers of segmented regions on hyperspectral band selection. As mentioned in Section 5.3, the optimal number of superpixels is hard to determine for different hyperspectral datasets, the determination of it is usually experimental. For all hyperspectral datasets utilized in the experiment, the number of their real categories is about 16. Thus, we empirically set the number of superpixels to two or three times of this value, i.e., the  $N_b$  is set to 10000 in Eq. (10). To further explore the impact of  $N_b$  on the final results, we set it to 1000, 10000, 100000 in the experiment respectively, as seen in Table 2. It should be noted that the number of superpixels is same when  $N_b$  is set to 10000 and 100000 for the Indian Pines dataset. From Table 2, we can find that the  $N_b$  is not sensitivity to the final results, and the proposed method can achieve satisfactory performance when it is set to 10000.

(3) The curse of dimensionality on the hyperspectral band selection. For all datasets utilized in the experiment, they contain more than 150 bands, and we only show the results for less than a third of them. Thus, the curse of dimensionality is not obvious in some datasets. Furthermore, as we can see from all experimental results, the performance of using all bands are worse in most cases, which indicate that the results have a descending process as the number of selected bands increases. By following the previous work, we do not select more bands to verify the effectiveness of the proposed method on hyperspectral band selection. Accordingly, the process of decline is not clearly shown in the experiment.

## 6. Conclusion

In this paper, a novel band selection method via region-aware latent features fusion based clustering (RLFFC) is presented. To be specific, the proposed method captures both the spatial and spectral information by employing the superpixel segmentation to segment HSIs into multiple regions firstly. And then the latent features are generated to enhance the separability among different bands from each segmented region. In order to capture the complementary information of each region, one fusion method is adopted to obtain a shared latent feature representation of HSIs. In particular, RLFFC can automatically determine the weight of each segmented region. With clustering on the shared latent feature matrix, the band closest to the clustering center is selected as the feature band. Extensive experiments on four public hyperspectral datasets show that the proposed method achieves more outstanding performance than other algorithms.

In the future, we will further consider how to optimize the learning and fusion of latent features. In addition, how to reduce the computational complexity of RLFFC is also a problem we need to consider.

## CRediT authorship contribution statement

**Jun Wang:** Writing, Coding. **Chang Tang:** Methodology, Supervision. **Zhenglai Li:** Coding. **Xinwang Liu:** Data curation. **Wei Zhang:** Reviewing and Editing. **En Zhu:** Investigation. **Lizhe Wang:** Investigation.

## Acknowledgments

This work was supported by National Key R & D Program of China (No. 2020AAA0107100), and the National Natural Science Foundation of China (No. 62076228, 61802356 and 61976196) and the Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education (Grant NO. GLAB2020ZR18) and the Fundamental Research Funds for the Central Universities, and the Opening Fund of State Key Lab. for Novel Software Technology, Nanjing University (No. KFKT2021B24).

## References

- [1] C. Tang, X. Zhu, X. Liu, P. Wang, Salient object detection via recurrently aggregating spatial attention weighted cross-level deep features, in: IEEE International Conference on Multimedia and Expo, 2019, pp. 1546–1551, <http://dx.doi.org/10.1109/ICME.2019.000267>.
- [2] Chinsu Lin, Shih-Yu Chen, Chia-Chun Chen, Chia-Huei Tai, Detecting newly grown tree leaves from unmanned-aerial-vehicle images using hyperspectral target detection techniques, *ISPRS J. Photogramm. Remote Sens.* 142 (2018) 174–189.
- [3] Changzhe Jiao, Chao Chen, Ronald G. McGarvey, Stephanie Bohlman, Licheng Jiao, Alina Zare, Multiple instance hybrid estimator for hyperspectral target characterization and sub-pixel target detection, *ISPRS J. Photogramm. Remote Sens.* 146 (2018) 235–250.
- [4] Mihaela Antonina Calin, Sorin Viorel Parasca, Dan Savastru, Dragos Manea, Hyperspectral imaging in the medical field: Present and future, *Appl. Spectrosc. Rev.* 49 (6) (2014) 435–447.
- [5] Baowei Fei, Hyperspectral imaging in medical applications, in: Data Handling in Science and Technology, volume 32, Elsevier, 2020, pp. 523–565.
- [6] Liu Zhi, David Zhang, Jing-qi Yan, Qing-Li Li, Qun-lin Tang, Classification of hyperspectral medical tongue images for tongue diagnosis, *Comput. Med. Imaging Graph.* 31 (8) (2007) 672–678.
- [7] Sandra Jakob, Robert Zimmermann, Richard Gloaguen, The need for accurate geometric and radiometric corrections of drone-borne hyperspectral data for mineral exploration: Mephysto — A toolbox for pre-processing drone-borne hyperspectral data, *Remote Sens.* 9 (1) (2017) 88.
- [8] Arindam Guha, Mineral exploration using hyperspectral data, in: Hyperspectral Remote Sensing, Elsevier, 2020, pp. 293–318.
- [9] Thais Andressa Carrino, Alvaro Penteado Crósta, Catarina Labouré Bemfica Toledo, Adalene Moreira Silva, Hyperspectral remote sensing applied to mineral exploration in southern Peru: A multiple data integration approach in the Chapi Chiara gold prospect, *Int. J. Appl. Earth Obs. Geoinformation* 64 (2018) 287–300.
- [10] Chang Tang, Xinwang Liu, Xinzhong Zhu, Jian Xiong, Miaomiao Li, Jingyuan Xia, Xiangke Wang, Lizhe Wang, Feature selective projection with low-rank embedding and dual Laplacian regularization, *IEEE Trans. Knowl. Data Eng.* (2019).
- [11] Jaime Zabalza, Jinchang Ren, Mingqiang Yang, Yi Zhang, Jun Wang, Stephen Marshall, Junwei Han, Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing, *ISPRS J. Photogramm. Remote Sens.* 93 (2014) 112–122.
- [12] Chang Tang, Meiru Bian, Xinwang Liu, Miaomiao Li, Hua Zhou, Pichao Wang, Hailin Yin, Unsupervised feature selection via latent representation learning and manifold regularization, *Neural Netw.* 117 (2019) 163–178.
- [13] Siwei Wang, Xinwang Liu, En Zhu, Chang Tang, Jiyuan Liu, Jingtao Hu, Jingyuan Xia, Jianping Yin, Multi-view clustering via late fusion alignment maximization, in: International Joint Conference on Artificial Intelligence, 2019, pp. 3778–3784.
- [14] Jie Feng, Jiantong Chen, Qigong Sun, Ronghua Shang, Xianghai Cao, Xiangrong Zhang, Licheng Jiao, Convolutional neural network based on bandwise-independent convolution and hard thresholding for hyperspectral band selection, *IEEE Trans. Cybern.* (2020) 1–15.
- [15] Yuan Yuan, Guokang Zhu, Qi Wang, Hyperspectral band selection by multitask sparsity pursuit, *IEEE Trans. Geosci. Remote Sens.* 53 (2) (2014) 631–644.
- [16] Jie Feng, Di Li, Jing Gu, Xianghai Cao, Ronghua Shang, Xiangrong Zhang, Licheng Jiao, Deep reinforcement learning for semisupervised hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* (2021).
- [17] Buyun Xu, Xihai Li, Weijun Hou, Yiting Wang, Yawei Wei, A similarity-based ranking method for hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* (2021).
- [18] Weiwei Sun, Qian Du, Graph-regularized fast and robust principal component analysis for hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* 56 (6) (2018) 3185–3195.
- [19] Masaki Nakanishi, Yijun Wang, Xiaogang Chen, Yu-Te Wang, Xiaorong Gao, Tzyy-Ping Jung, Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis, *IEEE Trans. Biomed. Eng.* 65 (1) (2017) 104–112.
- [20] Christelle Piantop Mbo'o, Kay Hameyer, Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection, *IEEE Trans. Ind. Appl.* 52 (5) (2016) 3861–3868.
- [21] Chang Tang, Xinwang Liu, Shan An, Pichao Wang, BR<sup>2</sup> net: Defocus blur detection via a bidirectional channel attention residual refining network, *IEEE Trans. Multimed.* 23 (2020) 624–635.
- [22] Chang Tang, LIU Xinwang, Xiao Zheng, Wanqing Li, Jian Xiong, Lizhe Wang, Albert Zomaya, Antonella Longo, DeFusionNET: Defocus blur detection via recurrently fusing and refining discriminative multi-scale deep features, *IEEE Trans. Pattern Anal. Mach. Intell.* (2020).
- [23] Nan Zhang, Shiliang Sun, Multiview graph restricted Boltzmann machines, *IEEE Trans. Cybern.* (2021).
- [24] Xianghai Cao, Tao Xiong, Licheng Jiao, Supervised band selection using local spatial information for hyperspectral image, *IEEE Geosci. Remote Sens. Lett.* 13 (3) (2016) 329–333.
- [25] Mingyang Zhang, Jingjing Ma, Maoguo Gong, Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization, *IEEE Geosci. Remote Sens. Lett.* 14 (5) (2017) 773–777.
- [26] Peng Zhou, Liang Du, Xuejun Li, Yi-Dong Shen, Yuhua Qian, Unsupervised feature selection with adaptive multiple graph learning, *Pattern Recognit.* 105 (2020) 107375.
- [27] Peng Zhou, Jiangyong Chen, Mingyu Fan, Liang Du, Yi-Dong Shen, Xuejun Li, Unsupervised feature selection for balanced clustering, *Knowl.-Based Syst.* 193 (2020) 105417.
- [28] Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, Jing Zhang, Jian Xiong, Lizhe Wang, Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection, *IEEE Trans. Knowl. Data Eng.* (2021).
- [29] Shiliang Sun, Xuli Sun, Qiuyang Liu, Multi-view Gaussian processes with posterior consistency, *Inform. Sci.* 547 (2021) 710–722.
- [30] Yao Liu, Tao Wu, Junjie Yang, Kezhu Tan, Shuwen Wang, Hyperspectral band selection for soybean classification based on information measure in FRS theory, *Biosyst. Eng.* 178 (2019) 219–232.
- [31] Weizhao Chen, Zhijing Yang, Jinchang Ren, Jiangzhong Cao, Nian Cai, Huimin Zhao, Peter Yuen, MIMN-DPP: Maximum-information and minimum-noise determinantal point processes for unsupervised hyperspectral band selection, *Pattern Recognit.* 102 (2020) 107213.
- [32] Leo Liberti, Carlile Lavor, Nelson Maculan, Antonio Mucherino, Euclidean distance geometry and applications, *SIAM Rev.* 56 (1) (2014) 3–69.
- [33] Weiwei Sun, Liangpei Zhang, Bo Du, Weiyue Li, Yenming Mark Lai, Band selection using improved sparse subspace clustering for hyperspectral imagery classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8 (6) (2015) 2784–2797.
- [34] Fahong Zhang, Qi Wang, Xuelong Li, Hyperspectral image band selection via global optimal clustering, in: IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2017, pp. 1–4.
- [35] Weiwei Sun, Jiangtao Peng, Gang Yang, Qian Du, Fast and latent low-rank subspace clustering for hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* 58 (6) (2020) 3906–3915.
- [36] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, James C. Tilton, Advances in spectral-spatial classification of hyperspectral images, *Proc. IEEE* 101 (3) (2012) 652–675.
- [37] Yuliya Tarabalka, Jón Atli Benediktsson, Jocelyn Chanussot, Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques, *IEEE Trans. Geosci. Remote Sens.* 47 (8) (2009) 2973–2987.
- [38] Leyuan Fang, Shutao Li, Xudong Kang, Jon Atli Benediktsson, Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation, *IEEE Trans. Geosci. Remote Sens.* 52 (12) (2014) 7738–7749.
- [39] Junjun Jiang, Jiayi Ma, Chen Chen, Zhongyuan Wang, Zhihua Cai, Lizhe Wang, SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery, *IEEE Trans. Geosci. Remote Sens.* 56 (8) (2018) 4581–4593.
- [40] Wenda Zhao, Zhijun Xu, Jian Zhao, Gradient entropy metric and p-Laplace diffusion constraint-based algorithm for noisy multispectral image fusion, *Inf. Fusion* 27 (2016) 138–149.
- [41] Wenda Zhao, Huimin Lu, Dong Wang, Multisensor image fusion and enhancement in spectral total variation domain, *IEEE Trans. Multimed.* 20 (4) (2017) 866–879.
- [42] Wenda Zhao, Dong Wang, Huchuan Lu, Multi-focus image fusion with a natural enhancement via a joint multi-level deeply supervised convolutional neural network, *IEEE Trans. Circuits Syst. Video Technol.* 29 (4) (2018) 1102–1115.
- [43] Fan Zhao, et al., Learning specific and general realm feature representations for image fusion, *IEEE Trans. Multimed.* (2020).
- [44] Fan Zhao, Wenda Zhao, Libo Yao, Yu Liu, Self-supervised feature adaption for infrared and visible image fusion, *Inf. Fusion* (2021).
- [45] Qi Wang, Qiang Li, Xuelong Li, Hyperspectral band selection via adaptive subspace partition strategy, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12 (2019) 4940–4950.
- [46] Qi Wang, Fahong Zhang, Xuelong Li, Optimal clustering framework for hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* 56 (10) (2018) 5910–5922.
- [47] Adolfo Martínez-Usó-Martínez-Usó, Filiberto Pla, José Martínez-Sotoca, Pedro García-Sevilla, Clustering-based hyperspectral band selection using information measures, *IEEE Trans. Geosci. Remote Sens.* 45 (12) (2007) 4158–4171.
- [48] Ibrahim Delibasoglu, Mufit Cetin, Hyperspectral band selection using structural information via hierarchical clustering, *J. Appl. Remote Sens.* 13 (1) (2019) 014526.
- [49] Weiwei Sun, Jiangtao Peng, Gang Yang, Qian Du, Correntropy-based sparse spectral clustering for hyperspectral band selection, *IEEE Geosci. Remote Sens. Lett.* 17 (3) (2019) 484–488.
- [50] Xiao Bai, Zhouxiao Guo, Yanyang Wang, Zhihong Zhang, Jun Zhou, Semisupervised hyperspectral band selection via spectral-spatial hypergraph model, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8 (6) (2015) 2774–2783.
- [51] Muhammad Ahmad, Dr Ihsan Ul Haq, Qaisar Mushtaq, Muhammad Sohaib, A new statistical approach for band clustering and band selection using K-means clustering, *Int. J. Eng. Technol.* 3 (6) (2011) 606–614.

- [52] A.C.S. Santos, H. Pedrini, A combination of k-means clustering and entropy filtering for band selection and classification in hyperspectral images, *Int. J. Remote Sens.* 37 (13) (2016) 3005–3020.
- [53] Guokang Zhu, Yuancheng Huang, Shuying Li, Jun Tang, Dong Liang, Hyperspectral band selection via rank minimization, *IEEE Geosci. Remote Sens. Lett.* 14 (12) (2017) 2320–2324.
- [54] G.P. Saradhi Varma, et al., Spatial residual clustering and entropy based ranking for hyperspectral band selection, *Eur. J. Remote Sens.* 53 (2020).
- [55] Chein-I Chang, Yi-Mei Kuo, Shuhua Chen, Chia-Chen Liang, Kenneth Yeonkong Ma, Peter Fuming Hu, Self-mutual information-based band selection for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2020).
- [56] Baofeng Guo, Steve R Gunn, Robert I Damper, James DB Nelson, Band selection for hyperspectral image classification using mutual information, *IEEE Geosci. Remote Sens. Lett.* 3 (4) (2006) 522–526.
- [57] Qian Du, He Yang, Similarity-based unsupervised band selection for hyperspectral image analysis, *IEEE Geosci. Remote Sens. Lett.* 5 (4) (2008) 564–568.
- [58] John A. Rice, *Mathematical Statistics and Data Analysis*, Cengage Learning, 2006.
- [59] Alexander Kraskov, Harald Stögbauer, Peter Grassberger, Estimating mutual information, *Phys. Rev. E* 69 (6) (2004) 066138.
- [60] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, Wen Gao, Late fusion incomplete multi-view clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (10) (2018) 2410–2423.
- [61] Zhanxuan Hu, Feiping Nie, Rong Wang, Xuelong Li, Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding, *Inf. Fusion* 55 (2020) 251–259.
- [62] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, En Zhu, Efficient and effective regularized incomplete multi-view clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* (2020).
- [63] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Guanghui Yue, Wei Zhang, En Zhu, Consensus graph learning for multi-view clustering, *IEEE Trans. Multimed.* (2021).
- [64] Bin Gao, Xin Liu, Xiaojun Chen, Ya-xiang Yuan, A new first-order algorithmic framework for optimization problems with orthogonality constraints, *SIAM J. Optim.* 28 (1) (2018) 302–332.
- [65] Sen Jia, Guihua Tang, Jiasong Zhu, Qingquan Li, A novel ranking-based clustering approach for hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* 54 (1) (2015) 88–102.
- [66] Chein-I Chang, Qian Du, Tzu-Lung Sun, Mark L.G. Althouse, A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 37 (6) (1999) 2631–2641.
- [67] Qi Wang, Qiang Li, Xuelong Li, A fast neighborhood grouping method for hyperspectral band selection, *IEEE Trans. Geosci. Remote Sens.* (2020).
- [68] Qi Wang, Fahong Zhang, Xuelong Li, Hyperspectral band selection via optimal neighborhood reconstruction, *IEEE Trans. Geosci. Remote Sens.* (2020).
- [69] Thomas Cover, Peter Hart, Nearest neighbor pattern classification, *IEEE Trans. Inform. Theory* 13 (1) (1967) 21–27.
- [70] Farid Melgani, Lorenzo Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, *IEEE Trans. Geosci. Remote Sens.* 42 (8) (2004) 1778–1790.
- [71] Tatyana V. Bandos, Lorenzo Bruzzone, Gustavo Camps-Valls, Classification of hyperspectral images with regularized linear discriminant analysis, *IEEE Trans. Geosci. Remote Sens.* 47 (3) (2009) 862–873.
- [72] Nikhil R. Pal, Sankar K. Pal, A review on image segmentation techniques, *Pattern Recognit.* 26 (9) (1993) 1277–1294.
- [73] Ming-Yu Liu, Oncel Tuzel, Srikanth Ramalingam, Rama Chellappa, Entropy rate superpixel segmentation, in: *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2011, pp. 2097–2104.
- [74] Junjun Jiang, Jiayi Ma, Zheng Wang, Chen Chen, Xianming Liu, Hyperspectral image classification in the presence of noisy labels, *IEEE Trans. Geosci. Remote Sens.* 57 (2) (2018) 851–865.



**Chang Tang** received his Ph.D. degree from Tianjin University, Tianjin, China in 2016. He joined the AMRL Lab of the University of Wollongong between Sep. 2014 and Sep. 2015. He is now a full professor at the School of Computer Science, China University of Geosciences, Wuhan, China. Dr. Tang has published 50+ peer-reviewed papers, including those in highly regarded journals and conferences such as IEEE T-PAMI, IEEE T-KDE, IEEE T-MM, IEEE T-HMS, IEEE SPL, AAAI, IJCAI, ICCV, CVPR, ACM, ICME, etc. He regularly served as the Technical Program Committees of top conferences such as NIPS, ICML, IJCAI, ICME, AAAI, ICCV, CVPR, etc. His current research interests include machine learning and computer vision.



**Zhenglai Li** received the B.E. degree from China University of Geosciences, Wuhan, China, in 2018. Currently, he is a now pursuing the master degree in China University of Geosciences, Wuhan, China. His research interests include multi-view clustering.



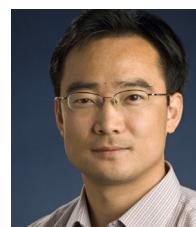
**Xinwang Liu** received his Ph.D. degree from National University of Defense Technology (NUDT), China. He is now full professor of School of Computer, NUDT. His current research interests include kernel learning and unsupervised feature learning. Dr. Liu has published 60+ peer-reviewed papers, including those in highly regarded journals and conferences such as IEEE T-PAMI, IEEE T-KDE, IEEE T-IP, IEEE T-NNLS, IEEE T-MM, IEEE T-IFS, NeurIPS, ICCV, CVPR, AAAI, IJCAI, etc. He is a senior member of IEEE. More information can be found at [xinwangliu.github.io](http://xinwangliu.github.io).



**Wei Zhang** received the B.E. degree from Zhejiang University in 2004, the M.S. degree from Liaoning University in 2008, and the Ph.D. degree from Shandong University of Science and Technology in 2018. He is currently an Associate Professor with the Shandong Computer Science Center (National Supercomputing Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences). His research interests include future generation network architectures, edge computing and edge intelligence.



**En Zhu** received his Ph.D. degree from National University of Defense Technology (NUDT), China. He is now Professor at School of Computer Science, NUDT, China. His main research interests are pattern recognition, image processing, machine vision and machine learning. Dr. Zhu has published 150+ peer-reviewed papers, including IEEE T-CSVT, IEEE T-NNLS, PR, AAAI, IJCAI, etc. He was awarded China National Excellence Doctoral Dissertation.



**Lizhe Wang** received the B.E. and M.E. degrees from Tsinghua University, Beijing, China, and the Dr.Eng. (Magna Cum Laude) degree from the University of Karlsruhe, Karlsruhe, Germany. He is currently a ChuTian Chair Professor with the School of Computer Science, China University of Geosciences, Beijing, and also a Professor with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing. His research interests include high-performance computing, eScience, and remote sensing image processing. Dr. Wang is a fellow of the IET and the British Computer Society. He serves as an Associate Editor of the IEEE Transactions on Parallel and Distributed Systems, the IEEE Transactions on Cloud Computing, and the IEEE Transactions on Sustainable Computing. He is a Fellow of IEEE.



**Jun Wang** received the B.E. degree in Information Management and Information System from Hubei University of Technology, Wuhan, China, in 2019. He is currently working toward the Master's degree at the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests are hyperspectral image processing and machine learning.