Neural Networks 153 (2022) 292–302

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Graph regularized spatial–spectral subspace clustering for hyperspectral band selection

Jun Wang [a], Chang Tang [a,*], Xiao Zheng [b,*], Xinwang Liu [b], Wei Zhang [c], En Zhu [b]

[a] School of Computer Science, China University of Geosciences, Wuhan 430074, PR China
[b] School of Computer, National University of Defense Technology, Changsha 410073, PR China
[c] 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 250353, PR China

ARTICLE INFO

Article history:
Received 21 March 2022
Received in revised form 16 May 2022
Accepted 12 June 2022
Available online 16 June 2022

Keywords:
Clustering
Hyperspectral band selection
Feature learning
Similarity graph learning

ABSTRACT

Hyperspectral band selection, which aims to select a small number of bands to reduce data redundancy and noisy bands, has attracted widespread attention in recent years. Many effective clustering-based band selection methods have been proposed to accomplish the band selection task and have achieved satisfying performance. However, most of the previous methods reshape the original hyperspectral images (HSIs) into a set of stretched band vectors, which ignore the spatial information of HSIs and the difference between diverse regions. To address these issues, a graph regularized spatial–spectral subspace clustering method for hyperspectral band selection is proposed in this paper, referred to as GRSC. Specifically, the proposed method adopts superpixel segmentation to preserve the spatial information of HSIs by segmenting their first principal component into diverse homogeneous regions. Then the discriminative latent features are generated from each segmented region to represent the whole band, which can mitigate the effect of noise on the band selection. Finally, a self-representation subspace clustering model and an $l_{2,1}$-norm regularization are utilized to explore the spectral correlation among all bands. In addition, a similarity graph between region-aware latent features is adaptively learned to preserve the spatial structure of HSIs in the latent representation space. Extensive classification experimental results on three public datasets verify the effectiveness of GRSC over several state-of-the-art methods. The demo code of this work is publicly available at https://github.com/WangJun2023/GRSC.

1. Introduction

The hyperspectral sensor captures the observed surface information in tens to hundreds of spectral bands to obtain hyperspectral images. With advances in hyperspectral imaging technology, HSIs can distinguish the information that is not observed in visible and multispectral images. Hence, they contain abundant spectral–spatial information and have been widely used in many fields, e.g. environmental monitoring (Feng et al., 2021; Flores et al., 2021; Moroni, Lupo, Marra, & Cenedese, 2013), mineral exploration (Jakob, Zimmermann, & Gloaguen, 2017; Peyghambari & Zhang, 2021; Saralıoğlu, Görmüş, & Güngör, 2016), and so on. Nevertheless, tremendous hyperspectral bands also raise some issues. To be specific, the high dimensional features of HSIs cause the curse of dimensionality, which significantly increases the complexity of data processing. Furthermore, the strong correlation between adjacent bands makes it difficult to distinguish them from each other, affecting subsequent imaging processing tasks. Therefore, it is reasonable and necessary to conduct dimension reduction for HSIs.

During past decades, many representative algorithms have been proposed and have achieved satisfying performance for addressing hyperspectral dimension reduction, which can be roughly divided into feature extraction (Imani & Ghassemian, 2016; Rasti, Ulfarsson, & Sveinsson, 2016; Zabalza et al., 2016; Zhu et al., 2015) and band selection (Guo, Gunn, Damper, & Nelson, 2005; Tang, Liu, Zhu, Wang and Zomaya, 2021; Tang et al., 2020, 2021; Tschannerl et al., 2019; Wang et al., 2022). The former method mainly constructs a low dimensional feature space from which the new features are extracted to represent the whole HSIs, and the latter one only needs to select some representative bands from the original hyperspectral band set. Compared with the extracted features, the selected bands can fully retain the original physical meaning of hyperspectral data. Consequently, band selection is more commonly adopted for hyperspectral dimension reduction.

* Corresponding authors.
E-mail addresses: wang_jun@cug.edu.cn (J. Wang), tangchang@cug.edu.cn (C. Tang), zhengxiao@nudt.edu.cn (X. Zheng), xinwangliu@nudt.edu.cn (X. Liu), wzhang@qlu.edu.cn (W. Zhang), enzhu@nudt.edu.cn (E. Zhu).

https://doi.org/10.1016/j.neunet.2022.06.016
0893-6080/

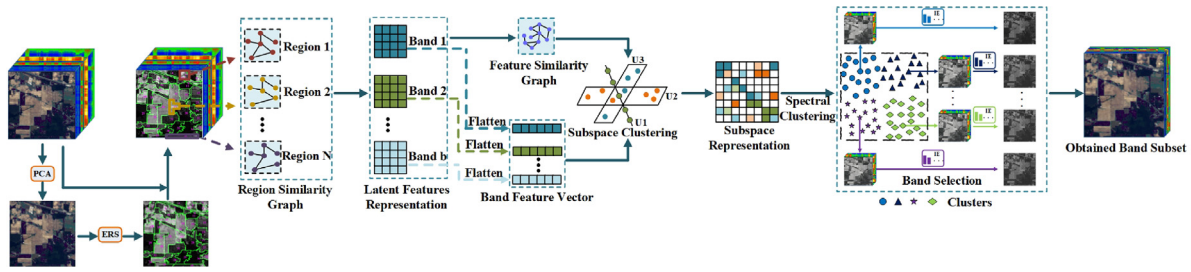


Fig. 1. Overview of the proposed GRSC. Firstly, the ERS is employed to segment the first principal component of HSIs into multiple homogeneous regions. For each segmented region, we can construct the corresponding similarity graph from which the latent features are extracted to represent each band. Based on the self-representation subspace clustering model, we reshape each band into a stretched band vector, which is employed as the dictionary to construct the coefficient representation matrix. In addition, the similarity graph between latent features is exploited to maintain the original spatial structure in the new representation space. Finally, the spectral clustering is conducted on the coefficient matrix to produce the multiple classes, and the band with the maximum IE is selected to form the optimal band subset.

For band selection, it can be conducted by using supervised (Yang, Du, Su, & Sheng, 2010; Zhang, He and Gao, 2019), semi-supervised (Cao et al., 2019; Feng, Jiao, Liu, Sun, & Zhang, 2014), and unsupervised manner (Du & Yang, 2008; Yuan, Zheng, & Lu, 2016) according to whether the label samples are used in the model. However, in most cases, it is expensive and laborious to acquire sample annotation to meet practical applications, so we mainly focus on the unsupervised band selection. In unsupervised band selection literature, many advanced methods have been reported and they can be generally categorized into four classes, i.e., clustering-based (Jia, Tang, Zhu, & Li, 2015), ranking-based (Datta, Ghosh, & Ghosh, 2015), searching-based (Zhang et al., 2019) and hybrid methods (Jia, Tang, Zhu, & Li, 2016). Clustering-based methods firstly partition all bands into different groups and then the final band subset is formed by selecting the representative bands from each group. Accordingly, the performance of this kind of method mainly depends on the calculation of similarity between hyperspectral bands. For ranking-based methods, the top-ranked bands are used to form the optimal band subset according to a specific metric. As a result, the determination of optimal metrics is the most important factor of ranking-based methods. Searching-based methods mainly adopt specific search criteria to obtain the desired number of feature bands. The common searching strategies are updated searching (Li, Zhu, Wan, & Feng, 2013), incremental searching (Du, 2003), and eliminating searching (Sun, Geng, & Ji, 2014). Hybrid band selection methods mainly combine others kinds of methods to select the representative band subset.

Among the clustering-based band selection methods, the self-representation subspace clustering approach has attracted widely attention and has achieved great success in recent years. Assuming that each data point can be represented by a combination of other data points in the same space, the self-representation subspace clustering method aims to find a reconstruction matrix to encode all data samples and then conducts spectral clustering on the reconstruction matrix to produce clustering results. Therefore, the performance of this kind of method on hyperspectral band selection mainly depends on the quality of the generated reconstructed matrix. To improve clustering performance, many advanced subspace clustering algorithms have been designed. For instance, Huang, Zhang, and Pižurica (2021) proposed to decompose the reconstructed matrix into a desirable coefficient matrix and a sparse matrix to adaptively capture the local and global structure of HSIs. To effectively capture the global data structure of HSIs, Zhu, Huang, Li, Tang, and Liang (2017) employed a coefficient representation with low rank to conduct band selection. Considering the high similarity property of adjacent bands, Zhai, Zhang, Zhang, and Li (2018) adopted the low-rank representation with Laplacian regularization to improve the clustering performance on the hyperspectral band selection. Furthermore, with advances in deep learning, many deep subspace-clustering algorithms have been proposed for band selection, such as Meng, Bin, Zhihua, and Qiong (2020), Xie et al. (2019) and Zeng et al. (2019).

Despite the satisfying performance in most existing subspace clustering methods (Tang et al., 2019), they still face several limitations. Firstly, the manner of stretching each band into a single feature vector ignores the spatial information of HSIs, which causes information loss in data processing, such as Wang, Zhang and Li (2020). Secondly, due to atmospheric conditions and other unavoidable factors, some original features of HSIs are contaminated by noise, and a mechanism to minimize the negative effect is needed. Thirdly, apart from the strong correlation between adjacent bands, the global information between non-adjacent bands should also be considered.

In this article, we take into account the above limitations and propose a graph regularized spatial–spectral subspace clustering method to address the hyperspectral band selection problem, as shown by Fig. 1. Specifically, the entropy rate superpixel segmentation (ERS) (Liu, Tuzel, Ramalingam, & Chellappa, 2011) is adopted to segment the first principal component of HSIs into diverse homogeneous regions, which can fully preserve the original spatial information of a certain scene. Instead of directly using the pixel-wise features, we generate the low dimensional latent features from different segmented regions to improve the representation capability of each band. Benefiting from the advantages of sparse coding (Yang et al., 0000) and entropy theory (Yang, Tan, & Chen, 2022), which motivate us to exploit them for improving the performance of the proposed method. Thus, a self-representation model with $l_{2,1}$-norm regularization is used to obtain a robust coefficient matrix on which the spectral clustering is conducted to get multiple band partitions, and the band with the maximum information entropy (IE) is simply selected in each partition group. In addition, with respect to the hardware implementation of the proposed method, some existing works can contribute to the feasibility of our subsequent experiments, such as Yang et al. (2019) and Yang, Wang, Deng, Azghadi, and Linares-Barranco (2021).

In a word, the main contributions of the proposed method are listed as follows:

(1) A self-representation subspace clustering model is proposed for hyperspectral band selection, which can capture both spatial and spectral information of original hyperspectral data. By exploiting the spatial information of HSIs and learning discriminative latent features, a robust coefficient representation matrix is generated for depicting the correlation of spectral bands.
(2) An adaptively similarity graph learning method is introduced to preserve the original spatial structure of HSIs for the new spectral self-representation space.

(3) An optimization algorithm is designed to solve the resultant model, and extensive classification experimental results reveal the effectiveness of GRSC when compared with other state-of-the-art ones.

The remainder of this article is structured as follows. Section 2 reviews the self-representation subspace clustering methods for hyperspectral band selection. In Sections 3 and 4, the proposed method and corresponding optimization solution are introduced, respectively. Extensive experiments are conducted in Section 5. Finally, in Section 6, the conclusion is given.

2. Prior works

In this section, most existing self-representation subspace clustering methods for band selection are reported. Before that, the notations and corresponding definitions throughout this paper are introduced, as shown in Table 1. Given hyperspectral image cube $\mathbf{H} \in \mathbb{R}^{w\times h\times b}$, where w, h and b denote the width, height and the number of contained bands of HSIs, respectively. Assuming that each band can be represented by a combination of other bands, the subspace clustering methods aim to find a coefficient matrix to encode the correlation of all spectral bands. For the previous subspace clustering methods, they often stretch the original 3-D hyperspectral image cube $\mathbf{H}$ into a set of 2-D slices $\mathbf{X} \in \mathbb{R}^{wh\times b}$ along the spectral band dimension, and the band selection problem can be formulated as follows:

$$\min_{\mathbf{Z}} \Phi(\mathbf{X}, \mathbf{XZ}) + \lambda\Psi(\mathbf{Z}), \tag{1}$$

where $\Phi(\cdot)$ and $\Psi(\cdot)$ represent the loss function and regularization term, respectively, and λ represents the trade-off parameter. For most existing subspace clustering methods, $\Phi(\cdot)$ is often represented in form of the Frobenius norm $\|\mathbf{X}\|_F^2$ to measure the error of reconstruction. As for the regularization term, $\Psi(\cdot)$ can be replaced by $\|\mathbf{Z}\|_F^2$ (Sun, Zhang, Du, Li, & Lai, 2015), $\|\mathbf{Z}\|_{2,1}$ (Sun, Zhang, Zhang, & Lai, 2016; Zhang, Wang, Jiang, & Zhou, 2021) and $\|\mathbf{Z}\|_*$ (Huang et al., 2021; Sun, Peng, Yang, & Du, 2020; Zhai et al., 2018).

Different regularization terms would lead to the diverse properties of the generated coefficient matrix $\mathbf{Z}$. For example, Zhu, Zuo, Zhang, Hu and Shiu (2015) adopted $\|\mathbf{Z}\|_{2,1}$ norm to induce a row sparse structure to the coefficient matrix $\mathbf{Z}$, which can detect the outliers and improve the performance on feature selection. In order to simultaneously take the sample and feature representation into consideration, Tang et al. (2018) proposed a dual self-representation method for feature selection. In addition, based on the assumption that different bands can be mapped into a unified low-dimensional latent space, Yu, Cen, Chang, and Li (2018) exploited $\|\mathbf{Z}\|_*$ to produce the low rankness in the coefficient matrix. Despite the great success in these methods, most of them adopt the original high-dimensional pixel-wise features to exploit the relationships among all bands in the new representation space, which would limit their performance once the features are grossly corrupted by noise. Thus, we propose GRSC to address this issue.

Based on the coefficient matrix $\mathbf{Z}$, the corresponding similarity matrix $\mathbf{C} = \frac{|\mathbf{Z}|+|\mathbf{Z}|^\top}{2}$ is constructed. And then the common way to generate the final clustering results is conducting standard spectral clustering on the similarity matrix $\mathbf{C}$. To be specific, the Laplacian matrix $\mathbf{L}$ is constructed based on the similarity matrix $\mathbf{C}$, i.e. $\mathbf{L} = diag(\mathbf{C}) - \mathbf{C}$. Assuming that the desired number of selected bands is K, the eigenvectors corresponding to the first K smallest eigenvalues of $\mathbf{L}$ are formed via eigenvalue decomposition, and the k-means clustering algorithm is adopted on it to produce multiple classes. Finally, the feature bands are selected from each class to form the optimal band subset.

Table 1
Notations and definitions.

Notations	Definitions
$\mathbf{H}$	3-D hyperspectral image cube
$\mathbf{X}$	2-D spectral matrix of $\mathbf{H}$
$\mathbf{Z}$	Coefficient matrix
$\mathbf{Z}^\top$	The transpose of $\mathbf{Z}$
$\mathbf{I}_d$	d dimension identity matrix
$\|\cdot\|$	Absolute value operator
Tr()	Trace operator
$\|\|\cdot\|\|_F^2$	Frobenius norm
$\|\|\cdot\|\|_{2,1}$	$l_{2,1}$-norm
$\|\|\cdot\|\|_*$	Nuclear norm
N	The number of segmented regions
K	The number of selected bands
$\lambda, \alpha, \beta, \gamma$	Regularization parameters

3. Proposed method

3.1. Region-wise latent features learning

For HSIs, the diverse regions of each band often reflect different pixel values, and the same regions have similar spectral properties. However, for most existing clustering-based band selection methods, a 2-D spectral matrix with stacking the pixel-wise features of the original HSIs is directly constructed to conduct subsequent data processing, which ignores the different spectral properties of diverse regions in each band. Furthermore, the stacked high dimensional pixel features will lead to the curse of dimensionality and significantly increase the computational burden for some clustering tasks. Therefore, to address these issues, the superpixel segmentation algorithm is firstly adopted to distinguish different scenes in HSIs. Considering that different spectral bands represent the same object or land cover, the first principal component of hyperspectral image cube $\mathbf{H}$ is generated by using principal component analysis (PCA). And then the entropy rate superpixel segmentation algorithm is used to segment the generated first principal component into multiple homogeneous regions. Mathematically, the problem can be reformulated as follows:

$$\mathbf{H} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_N\}\ s.t. \mathcal{R}_i \cap \mathcal{R}_j = \emptyset(\forall i, j, i \neq j), \tag{2}$$

where $\mathcal{R}_i$ and N denote the ith segmented region and the number of regions, respectively.

After obtaining all segmented regions of HSIs, the critical problem is how to represent each band. A direct way is to use the original averaged spectral values of each region as a new feature, and each band can be represented by the size of $N \times 1$ feature vector. Nevertheless, the information encoded in the averaged spectral values is limited and some pixel values are contaminated by noise, which significantly influence the clustering performance on the hyperspectral band selection. Therefore, the discriminative latent features are extracted from each region to replace the original pixel-wise features. To be specific, for each segmented region, the corresponding Laplacian matrix is constructed, from which the d dimension latent features are obtained as follows:

$$\min_{\mathbf{Y}_{(i)}} \sum_{i=1}^{N} \mathrm{Tr}(\mathbf{Y}_{(i)}^\top \mathbf{L}_{(i)} \mathbf{Y}_{(i)}) \quad s.t.\ \mathbf{Y}_{(i)}^\top \mathbf{Y}_{(i)} = \mathbf{I}_d, \tag{3}$$

where $\mathbf{L}_i = \mathbf{I} - \mathbf{D}_{(i)}^{-\frac{1}{2}} \mathbf{W}_{(i)} \mathbf{D}_{(i)}^{-\frac{1}{2}}$ and $\mathbf{Y}_{(i)} \in \mathbb{R}^{b\times d}$ represent the Laplacian matrix and latent feature matrix of ith region, respectively. $\mathbf{W}_{(i)}$ denotes the similarity matrix of ith region, which is constructed via k-nearest neighbor graph and Euclidean distance, and $\mathbf{D}_{(i)}$ denotes the diagonal matrix of $\mathbf{W}_{(i)}$. In this way, the original hyperspectral image cube $\mathbf{H}$ can be expressed in a new

form by using the low dimensional latent features, i.e. $\mathcal{X} \in \mathbb{R}^{b\times d\times N}$, and each band is described by a 2-D latent feature matrix with the size of $d \times N$.

3.2. Subspace clustering with graph regularization

Given new latent feature representation $\mathcal{X} \in \mathbb{R}^{b\times d\times N}$, the coefficient matrix $\mathbf{Z} \in \mathbb{R}^{b\times b}$ can be generated according to Eq. (1), i.e.:

$$\min_{\mathbf{Z}} \|\mathbf{F}^\top - \mathbf{F}^\top\mathbf{Z}\|_F^2 + \alpha\|\mathbf{Z}\|_{2,1} \quad s.t.\ 0 \le Z_{ij} \le 1, \tag{4}$$

where $\mathbf{F} \in \mathbb{R}^{b\times dN}$ denotes a 2-D latent feature matrix which is formed by stacking all latent features of $\mathcal{X}$. For regularization term, a $l_{2,1}$-norm is used on the coefficient matrix $\mathbf{Z}$ to constrain its row sparse structure.

However, as to Eq. (4), we can find that each band is represented with all concatenated latent features, i.e., a $1 \times dN$ feature vector, which only exploits the spectral correlation between different bands, while ignores the spatial structural information of HSIs. To this end, a similarity graph $\mathbf{S}$ is employed to preserve the prior information in the subspace clustering. Specifically, we regard each latent feature as a node of the graph, and the similarity among them is encoded into $\mathbf{S} \in \mathbb{R}^{Nd\times Nd}$ to record the spatial structure of HSIs, i.e.:

$$\min_{\mathbf{S}} \sum_{i,j=1}^{Nd} \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 S_{ij} + \gamma S_{ij}^2 \quad s.t.\ \mathbf{s}_i^\top\mathbf{1} = 1, 0 \le S_{ij} \le 1, \tag{5}$$

where $\mathbf{f}_i$ and $\mathbf{f}_j$ denote the ith and jth column of $\mathbf{F}$, respectively. For clustering-based band selection methods, one rational criterion is that the similarity between two bands should keep consistent in the original feature space and the new representation subspace, and the problem can be expressed as follows:

$$\min_{\mathbf{Z}} \sum_{i,j=1}^{Nd} \|\mathbf{f}_i^\top\mathbf{Z} - \mathbf{f}_j^\top\mathbf{Z}\|_2^2 S_{ij}. \tag{6}$$

To simplify the expression, Eq. (6) can be rewritten as:

$$\min_{\mathbf{Z}} \mathrm{Tr}(\mathbf{Z}^\top\mathbf{F}\mathbf{L}_S\mathbf{F}^\top\mathbf{Z}), \tag{7}$$

where $\mathbf{L}_S$ denotes the corresponding Laplacian matrix of $\mathbf{S}$, i.e., $\mathbf{L}_S = \mathbf{D}_S - \mathbf{S}$, $\mathbf{D}_S$ represents the diagonal matrix of $\mathbf{S}$.

3.3. Overall objective function

Formally, combining above items, the overall objective function of the proposed method is formulated in the following:

$$\begin{aligned}&\min_{\mathbf{Z},\mathbf{S}} \|\mathbf{F}^\top - \mathbf{F}^\top\mathbf{Z}\|_F^2 + \textstyle\sum_{i,j=1}^{Nd}(\|\mathbf{f}_i - \mathbf{f}_j\|_2^2 S_{ij} + \gamma S_{ij}^2)\\ &+\alpha\|\mathbf{Z}\|_{2,1} + \beta\mathrm{Tr}(\mathbf{Z}^\top\mathbf{F}\mathbf{L}_S\mathbf{F}^\top\mathbf{Z})\\ &s.t. 0 \le Z_{ij} \le 1, \mathbf{s}_i^\top\mathbf{1} = 1, 0 \le S_{ij} \le 1, diag(\mathbf{Z}) = 0.\end{aligned} \tag{8}$$

The first term of Eq. (8) depicts the property that each band can be linearly represented by the other relevant bands. Different from the traditional self-representation band selection methods, we employ the latent features to represent each band rather than the original high-dimensional pixel-wise features, which can mitigate the effect of noise on the feature representation of each band. To model the outliers, i.e., improve the robustness of the proposed method, l_{21}-norm regularization is used in the third term to induce a column sparse structure to the coefficient matrix $\mathbf{Z}$. The second term denotes the adaptively similarity graph learning, which means that a smaller distance between two bands should be assigned a larger similarity value. Considering that the invariance of feature space structure, a graph based manifold regularization is introduced in the fourth term to maintain the local geometric structure of HSIs, while the most existing band selection methods do not take it into consideration. Furthermore, by coupling the band correlation learning and adaptively similarity graph learning into a unified framework, not only the spectral correlation between different bands is exploited, but also the spatial structure of HSIs can be maintained in the subspace clustering. Noted that the low dimension latent features are utilized to represent the whole HSIs, which can significantly reduce the computation overhead for subsequent clustering tasks.

With the coefficient matrix $\mathbf{Z}$, a standard spectral clustering algorithm is adopted to generate multiple groups, and from which the band with the maximum IE is selected to form an optimal band subset.

4. Optimization algorithm

For Eq. (8), the proposed model contains two variables, i.e., the coefficient matrix $\mathbf{Z}$ and the similarity matrix $\mathbf{S}$. In order to optimize the resultant model, an iterative optimization algorithm is designed as follows.

4.1. Optimizing $\mathbf{Z}$

When $\mathbf{S}$ is fixed, the Eq. (8) can be simplified to:

$$\begin{aligned}&\min_{\mathbf{Z}} \|\mathbf{F}^\top - \mathbf{F}^\top\mathbf{Z}\|_F^2 + \alpha\|\mathbf{Z}\|_{2,1} + \beta\mathrm{Tr}(\mathbf{Z}^\top\mathbf{F}\mathbf{L}_S\mathbf{F}^\top\mathbf{Z})\\ &s.t. 0 \le Z_{ij} \le 1, diag(\mathbf{Z}) = 0.\end{aligned} \tag{9}$$

Noticed that the value of $\|\mathbf{Z}\|_{2,1}$ can be zero in theory, which makes the optimization problem become non-differentiable. Thus, $\|\mathbf{Z}\|_{2,1}$ is replaced by $\sum_{i=1}^{b}(\mathbf{z}_i^\top\mathbf{z}_i + \varepsilon)^{\frac{1}{2}}$ to avoid this condition, where $\mathbf{z}_i$ and ε denote the ith row of $\mathbf{Z}$ and a very small constant, respectively, and Eq. (9) can be transformed into:

$$\begin{aligned}&\min_{\mathbf{Z}} \mathrm{Tr}(\mathbf{Z}^{\top\mathbf{G}\mathbf{Z}}) - 2\mathrm{Tr}(\mathbf{Z}^\top\mathbf{F}\mathbf{F}^\top) + \alpha\textstyle\sum_{i=1}^{b}(\mathbf{z}_i^\top\mathbf{z}_i + \varepsilon)^{\frac{1}{2}}\\ &s.t.\ \mathbf{G} = \mathbf{F}\mathbf{F}^\top + \beta\mathbf{F}\mathbf{L}_S\mathbf{F}^\top, 0 \le Z_{ij} \le 1, diag(\mathbf{Z}) = 0.\end{aligned} \tag{10}$$

Taking the derivative of Eq. (10) w.r.t $\mathbf{Z}$ and setting it to zero, the following equation is obtained:

$$\mathbf{G}\mathbf{Z} - 2\mathbf{F}\mathbf{F}^\top + \alpha\mathbf{Q}\mathbf{Z} = 0 \quad s.t.\ Q_{ii} = \frac{1}{2}(\mathbf{z}_i^\top\mathbf{z}_i + \varepsilon)^{-\frac{1}{2}}. \tag{11}$$

For Eq. (11), $\mathbf{Q}$ is an unknown diagonal matrix and depends on $\mathbf{Z}$. To solve the problem, an iterative algorithm is designed. Specifically, when $\mathbf{Q}$ is fixed, the solution $\mathbf{Z}^*$ of Eq. (9) is obtained as follows:

$$\mathbf{Z}^* = (\mathbf{G} + \alpha\mathbf{Q})^{-1}\mathbf{F}\mathbf{F}^\top. \tag{12}$$

When $\mathbf{Z}$ is fixed, the solution of $\mathbf{Q}$ can be directly solved via Eq. (11).

4.2. Optimizing $\mathbf{S}$

When $\mathbf{Z}$ is fixed, the Eq. (8) is simplified to:

$$\begin{aligned}&\min_{\mathbf{S}} \textstyle\sum_{i,j=1}^{Nd}(\|\mathbf{f}_i - \mathbf{f}_j\|_2^2 S_{ij} + \gamma S_{ij}^2 + \beta\|\mathbf{f}_i^\top\mathbf{Z} - \mathbf{f}_j^\top\mathbf{Z}\|_2^2 S_{ij})\\ &s.t.\ \mathbf{s}_i^\top\mathbf{1} = 1, 0 \le S_{ij} \le 1.\end{aligned} \tag{13}$$

Since Eq. (13) is independent between different i, the solution of problem (13) is equivalent to:

$$\begin{aligned}&\min_{\mathbf{S}} \textstyle\sum_{j=1}^{Nd}(\|\mathbf{f}_i - \mathbf{f}_j\|_2^2 S_{ij} + \gamma S_{ij}^2 + \beta\|\mathbf{f}_i^\top\mathbf{Z} - \mathbf{f}_j^\top\mathbf{Z}\|_2^2 S_{ij})\\ &s.t.\ \mathbf{s}_i^\top\mathbf{1} = 1, 0 \le S_{ij} \le 1.\end{aligned} \tag{14}$$

Let auxiliary variable $m_{ij} = \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 + \beta\|\mathbf{f}_i^\top\mathbf{Z} - \mathbf{f}_j^\top\mathbf{Z}\|_2^2$, which denotes the ith row and jth column entry of $\mathbf{M}$, then Eq. (14) can be reformulated as:

$$\min_{\mathbf{s}} \|\mathbf{s}_i + \frac{1}{2\gamma}\mathbf{m}_i\|_2^2 \quad s.t.\ \mathbf{s}_i^\top\mathbf{1} = 1, 0 \le S_{ij} \le 1, \tag{15}$$

where $\mathbf{m}_i \in \mathbb{R}^{Nd\times 1}$ denotes the column vector of $\mathbf{M}$. According to the Section 2.4 in Nie, Wang, and Huang (2014), the optimal solution of $\mathbf{S}$ and the value of parameter γ can be efficiently obtained, i.e., the value of $\mathbf{S}$ is:

$$S_{i,j} = \begin{cases} \frac{m_{i,k+1} - m_{ij}}{k m_{i,k+1} - \sum_{v=1}^{k} m_{iv}} & j \le k, \\ 0 & j > k, \end{cases} \tag{16}$$

where k denotes the number of neighbors of data point $\mathbf{f}_i$. And the value of γ is determined as:

$$\gamma = \frac{1}{Nd}\sum_{i=1}^{Nd}\left(\frac{k}{2}\|\mathbf{f}_i - \mathbf{f}_{k+1}\|_2^2 - \frac{1}{2}\sum_{j=1}^{k}\|\mathbf{f}_i - \mathbf{f}_j\|_2^2\right). \tag{17}$$

In a word, the detailed of the proposed method can be summarized in Algorithm 1.

Algorithm 1 GRSC

Input: Hyperspectral data $\mathbf{H} \in \mathbb{R}^{w\times h\times b}$, the number of selected bands K, the latent feature dimension d, the number of neighbors k, the number of segmented regions N, regularization parameters α and β.
Output: The obtained band subset $\mathbf{H}^*$.
1: Segment $\mathbf{H}$ into N regions via PCA and ERS.
2: Calculate $\{\mathbf{Y}_{(i)}\}_{i=1}^N$ via Eq. (3).
3: Construct $\mathbf{F}$ by concatenating $\{\mathbf{Y}_{(i)}\}_{i=1}^N$.
4: Initialize $\mathbf{Z}$ and $\mathbf{S}$.
5: **while not converged do**
6: Update $\mathbf{Z}$ via Eq. (12).
7: Update $\mathbf{S}$ via Eq. (16).
8: **end while**
9: Conduct the spectral clustering algorithm on $\mathbf{Z}$ to generate the multiple classes.
10: Calculate the IE of each band and the band with the maximum IE is selected from each class.
11: **Return the band subset $\mathbf{H}^*$.**

4.3. Convergence analysis

Since it is difficult to directly obtain the optimal solution of each variable, an iterative algorithm is adopted to solve the problem (8). As each sub-problem is independent and the corresponding optimal solution can be obtained, the proposed method converges obviously. Now the detailed proof is shown as follows.

For optimizing $\mathbf{Z}$, when $\mathbf{Q}$ is fixed, it is easily to obtain that the problem (11) is equivalent to:

$$\begin{aligned} &\min_{\mathbf{Z}} \mathrm{Tr}(\mathbf{Z}^{\top\mathbf{GZ}}) - 2\mathrm{Tr}(\mathbf{Z}^\top\mathbf{F}\mathbf{F}^\top) + \alpha(\mathbf{Z}^{\top\mathbf{QZ}}) \\ &s.t.\ 0 \le Z_{ij} \le 1, diag(\mathbf{Z}) = 0. \end{aligned} \tag{18}$$

Suppose $\mathbf{Z}^t$ and $\mathbf{Z}^{t+1}$ denote the tth and $t+1$th iteration of $\mathbf{Z}$, respectively, we have:

$$\begin{aligned} &\mathrm{Tr}(\mathbf{Z}^{t+1\top}\mathbf{G}\mathbf{Z}^{t+1}) - 2\mathrm{Tr}(\mathbf{Z}^{t+1\top}\mathbf{F}\mathbf{F}^\top) + \alpha(\mathbf{Z}^{t+1\top}\mathbf{Q}\mathbf{Z}^{t+1}) \\ &\le \mathrm{Tr}(\mathbf{Z}^{t\top}\mathbf{G}\mathbf{Z}^t) - 2\mathrm{Tr}(\mathbf{Z}^{t\top}\mathbf{F}\mathbf{F}^\top) + \alpha(\mathbf{Z}^{t\top}\mathbf{Q}\mathbf{Z}^t). \end{aligned} \tag{19}$$

When we plug the value of $\mathbf{Q}$ into Eq. (19), the following inequality holds:

$$\begin{aligned} &\mathrm{Tr}(\mathbf{Z}^{t+1\top}\mathbf{G}\mathbf{Z}^{t+1}) - 2\mathrm{Tr}(\mathbf{Z}^{t+1\top}\mathbf{F}\mathbf{F}^\top) + \tfrac{\alpha}{2}\textstyle\sum_{i=1}^{b}\frac{\mathbf{z}_i^{t+1\top}\mathbf{z}_i^{t+1}+\varepsilon}{(\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon)^{\frac{1}{2}}} \\ &\le \mathrm{Tr}(\mathbf{Z}^{t\top}\mathbf{G}\mathbf{Z}^t) - 2\mathrm{Tr}(\mathbf{Z}^{t\top}\mathbf{F}\mathbf{F}^\top) + \tfrac{\alpha}{2}\textstyle\sum_{i=1}^{b}\frac{\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon}{(\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon)^{\frac{1}{2}}}. \end{aligned} \tag{20}$$

In addition, according to Lemma 3 in Nie, Zhu, and Li (2019), the following inequality can be derived:

$$\begin{aligned} &\alpha\textstyle\sum_{i=1}^{b}(\mathbf{z}_i^{t+1\top}\mathbf{z}_i^{t+1}+\varepsilon)^{\frac{1}{2}} - \tfrac{\alpha}{2}\textstyle\sum_{i=1}^{b}\frac{\mathbf{z}_i^{t+1\top}\mathbf{z}_i^{t+1}+\varepsilon}{(\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon)^{\frac{1}{2}}} \\ &\le \alpha\textstyle\sum_{i=1}^{b}(\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon)^{\frac{1}{2}} - \tfrac{\alpha}{2}\textstyle\sum_{i=1}^{b}\frac{\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon}{(\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon)^{\frac{1}{2}}}. \end{aligned} \tag{21}$$

Based on Eqs. (20) and (21), we can arrive at:

$$\begin{aligned} &\mathrm{Tr}(\mathbf{Z}^{t+1\top}\mathbf{G}\mathbf{Z}^{t+1} - 2\mathbf{Z}^{t+1\top}\mathbf{F}\mathbf{F}^\top) + \alpha\sum_{i=1}^{b}(\mathbf{z}_i^{t+1\top}\mathbf{z}_i^{t+1}+\varepsilon)^{\frac{1}{2}} \\ &\le \mathrm{Tr}(\mathbf{Z}^{t\top}\mathbf{G}\mathbf{Z}^t - 2\mathbf{Z}^{t\top}\mathbf{F}\mathbf{F}^\top) + \alpha\sum_{i=1}^{b}(\mathbf{z}_i^{t\top}\mathbf{z}_i^t+\varepsilon)^{\frac{1}{2}}. \end{aligned} \tag{22}$$

Furthermore, since Eq. (9) has the lower bound, the sub-problem of optimizing $\mathbf{Z}$ is convergent.

For optimizing $\mathbf{S}$, it can be easily checked that Eq. (15) is a convex function due to its second order of derivative w.r.t $\mathbf{s}_i$ is more than zero. Thus, this sub-problem is also convergent.

4.4. Time complexity analysis

The time complexity of GRSC mainly exists in three parts, i.e. superpixel segmentation, the learning of latent features, and the solution to problem (8). For hyperspectral image cube $\mathbf{H}$, the computational complexity of ERS is $\mathcal{O}(wh \times log(wh))$. As for the learning of latent features, i.e. Eq. (3), it costs $\mathcal{O}(Nb^3)$. To solve the problem (8), we divide it into two sub-problems. For updating $\mathbf{Z}$, it costs $\mathcal{O}(b^3t)$, where t denotes the number of iterations. And the time complexity of updating $\mathbf{S}$ is $\mathbf{O}(Ndkt)$. After optimization, to generate the optimal band subset on the coefficient matrix $\mathbf{Z}$, the extra complexity of spectral clustering is $\mathcal{O}(b^3 + bK^2)$. Thus, the overall time complexity of the proposed method is $\mathcal{O}\left(wh \times log(wh) + bK^2 + (N+1)b^3 + (b^3 + Ndk)t\right)$.

5. Experiments

5.1. Datasets

Indian Pines Scene was acquired by the AVIRIS sensor over the agricultural area of northwestern Indians in 1992. This dataset contains 220 spectral bands and 16 classes, and the size of each band is 145 × 145. In addition, 10 249 pixels are labeled in the corresponding ground truth. Due to water absorption, 200 spectral bands are exploited in the experiment by discarding 20 bands with the index of 104–108, 150–163, and 200. **KSC** was captured by the AVIRIS sensor in 1996. It consists of 224 spectral bands and each band contains 512 × 614 pixels. In the experiment, 176 spectral bands are selected by removing 48 noisy bands.

Botswana was collected by NASA EO-1 sensor over Okavango Delta, Botswana between 2001 and 2004. The image of size 1476 × 256 × 242 contains 14 classes. By removing 94 spectral bands due to water absorption, 145 bands are used in the experiment.

Table 2
The classification results of all competitors on three metrics.

Datasets	Classifier	Metrics	UBS	EFDPC	OPBS	OCF	ASPS-IE	FNGBS	SR-SSIM	GRSC
Indian Pines	KNN	OA	59.68 ± 0.56	58.59 ± 0.64	60.37 ± 0.95	61.89 ± 0.45	69.96 ± 0.65	69.46 ± 0.49	68.57 ± 0.62	**72.76 ± 0.24**
		AA	47.36 ± 0.71	44.31 ± 1.35	48.41 ± 0.66	49.16 ± 1.31	55.99 ± 1.44	56.15 ± 1.94	57.15 ± 1.78	**58.69 ± 1.22**
		Kappa	57.11 ± 0.58	56.00 ± 0.63	57.85 ± 0.95	59.31 ± 0.44	67.67 ± 0.67	67.08 ± 0.51	66.24 ± 0.65	**70.50 ± 0.24**
	SVM	OA	78.58 ± 0.62	69.09 ± 0.87	70.96 ± 0.91	77.32 ± 0.43	79.28 ± 0.41	80.51 ± 0.49	76.44 ± 0.51	**80.92 ± 0.45**
		AA	72.62 ± 1.85	55.82 ± 1.24	64.48 ± 1.53	72.45 ± 2.58	75.68 ± 1.75	77.26 ± 2.17	72.99 ± 1.62	**77.72 ± 2.77**
		Kappa	76.68 ± 0.65	66.52 ± 0.94	68.74 ± 0.92	75.35 ± 0.46	77.42 ± 0.44	78.71 ± 0.52	74.45 ± 0.54	**79.15 ± 0.47**
	LDA	OA	66.62 ± 0.78	63.85 ± 0.79	56.36 ± 0.68	66.45 ± 0.52	66.14 ± 0.90	70.20 ± 0.85	66.01 ± 0.70	**70.54 ± 1.18**
		AA	61.94 ± 2.29	57.39 ± 2.53	50.89 ± 2.01	60.86 ± 1.31	60.61 ± 1.53	66.86 ± 2.88	63.37 ± 0.67	**67.54 ± 1.62**
		Kappa	64.26 ± 0.81	61.54 ± 0.84	53.55 ± 0.68	64.07 ± 0.51	63.87 ± 0.91	68.03 ± 0.89	63.54 ± 0.69	**68.43 ± 1.20**
KSC	KNN	OA	85.12 ± 0.97	82.74 ± 0.79	59.71 ± 2.62	85.36 ± 0.49	84.97 ± 0.75	85.36 ± 0.70	82.30 ± 0.67	**85.75 ± 0.91**
		AA	77.71 ± 1.29	74.67 ± 1.14	47.45 ± 3.11	77.29 ± 0.64	78.02 ± 1.40	77.77 ± 1.19	73.87 ± 0.80	**78.66 ± 0.19**
		Kappa	83.73 ± 1.04	81.17 ± 0.84	57.08 ± 2.66	83.97 ± 0.53	83.57 ± 0.81	83.98 ± 0.75	80.70 ± 0.71	**84.41 ± 0.98**
	SVM	OA	80.13 ± 0.72	79.55 ± 0.95	55.90 ± 1.02	82.51 ± 1.11	81.47 ± 1.04	80.02 ± 0.81	80.18 ± 0.87	**83.08 ± 0.64**
		AA	68.98 ± 1.57	69.34 ± 1.39	36.64 ± 1.38	73.62 ± 1.99	70.75 ± 2.04	67.06 ± 1.89	70.10 ± 1.88	**74.81 ± 1.40**
		Kappa	78.36 ± 0.79	77.71 ± 1.01	52.74 ± 1.07	80.89 ± 1.20	79.78 ± 1.12	78.20 ± 0.88	78.39 ± 0.94	**81.52 ± 0.69**
	LDA	OA	88.95 ± 0.59	89.68 ± 0.57	61.59 ± 3.38	89.19 ± 0.70	86.89 ± 0.70	88.12 ± 0.69	88.03 ± 0.67	**89.71 ± 0.59**
		AA	84.06 ± 0.87	84.46 ± 1.11	48.44 ± 4.25	84.25 ± 1.04	80.86 ± 1.04	82.81 ± 0.75	81.81 ± 0.84	**84.47 ± 0.25**
		Kappa	87.87 ± 0.64	**88.66**±0.62	58.96 ± 3.49	88.14 ± 0.76	85.64 ± 0.75	86.97 ± 0.75	86.86 ± 0.73	88.60 ± 0.64
Botswana	KNN	OA	85.25 ± 0.88	83.39 ± 0.82	82.81 ± 0.43	84.07 ± 0.80	84.29 ± 0.81	84.82 ± 1.23	86.33 ± 0.81	**86.35 ± 0.79**
		AA	86.37 ± 1.08	84.36 ± 1.02	83.90 ± 0.82	85.22 ± 0.70	85.56 ± 0.82	85.89 ± 2.01	87.43 ± 0.68	**87.78 ± 0.12**
		Kappa	84.36 ± 0.92	82.43 ± 0.85	81.82 ± 0.45	83.13 ± 0.83	83.36 ± 0.84	83.91 ± 1.29	**85.49 ± 0.85**	85.44 ± 0.81
	SVM	OA	89.58 ± 0.51	88.20 ± 0.78	89.64 ± 0.79	88.79 ± 0.66	89.46 ± 0.84	89.54 ± 0.70	90.27 ± 0.64	**90.39 ± 0.44**
		AA	90.71 ± 0.50	89.31 ± 0.63	90.56 ± 0.75	89.88 ± 0.72	90.43 ± 0.66	90.66 ± 0.66	91.21 ± 0.68	**91.44 ± 0.30**
		Kappa	88.90 ± 0.54	87.44 ± 0.81	88.96 ± 0.83	88.06 ± 0.69	88.77 ± 0.88	88.85 ± 0.73	89.61 ± 0.67	**89.74 ± 0.46**
	LDA	OA	90.06 ± 0.71	87.99 ± 0.59	90.06 ± 0.97	88.53 ± 0.80	90.00 ± 0.65	90.22 ± 0.36	90.70 ± 0.65	**90.89 ± 0.72**
		AA	90.76 ± 1.04	89.10 ± 0.49	90.98 ± 1.09	89.55 ± 0.96	90.72 ± 0.93	90.84 ± 0.59	91.59 ± 0.73	**91.71 ± 0.48**
		Kappa	89.39 ± 0.75	87.22 ± 0.62	89.40 ± 1.02	87.79 ± 0.84	89.34 ± 0.68	89.56 ± 0.38	90.07 ± 0.68	**90.27 ± 0.75**

5.2. Compared methods

In order to demonstrate the superiority of GRSC, seven competitors are selected, including two clustering-based band selection methods ASPS-IE (Wang, Li, & Li, 2019) and OCF (Wang, Zhang, & Li, 2018), one ranking-based method SR-SSIM (Xu, Li, Hou, Wang, & Wei, 2021), three searching-based methods FNGBS (Wang, Li and Li, 2020), UBS (Chang, Du, Sun, & Althouse, 1999) and OPBS (Zhang, Li, Dou, & Zhao, 2018), and one hybrid-based method EFDPC (Jia et al., 2016).

5.3. Experimental setup

In most previous methods, the performance of band selection is often evaluated by conducting classification experiments. Thus, three classical classifiers are selected to experiment, i.e., KNN, SVM, and LDA classifier. Furthermore, to measure the classification performance on these classifiers, three metrics are employed, including overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa). Since all classifiers are supervised, we randomly select 10% of the hyperspectral data as the training set. To the best of our knowledge, the optimal number of selected bands remains an open problem. Thus, we set it to range from 5 to 50 via traversing. For all competitors, the optimal parameter values are set in the experiment. With respect to GRSC, it contains two parameters, i.e., α and β, they are uniformly tuned in the range of $\{10^{-6}, 10^{-4}, \ldots, 10^{4}, 10^{6}\}$. In addition, the latent feature dimension of GRSC is 5, and the number of neighbors is 15. To reduce the random experimental bias, we repeatedly experiment 10 times and the best values of all methods are reported. All experiments are conducted using Windows 10 with Intel Core i7-7700 3.60-GHz CPU, 24-GB RAM, and MATLAB 2020a.

The optimal number of superpixels is often difficult to be determined due to the complex correlation between different pixels in the literature of superpixel segmentation. Based on the first principal component of HSIs, Jiang, Ma, Wang, Chen, and Liu (2018) proposed to exploit the textural information of it to guide the segmentation of HSIs, and the following objective function is formed:

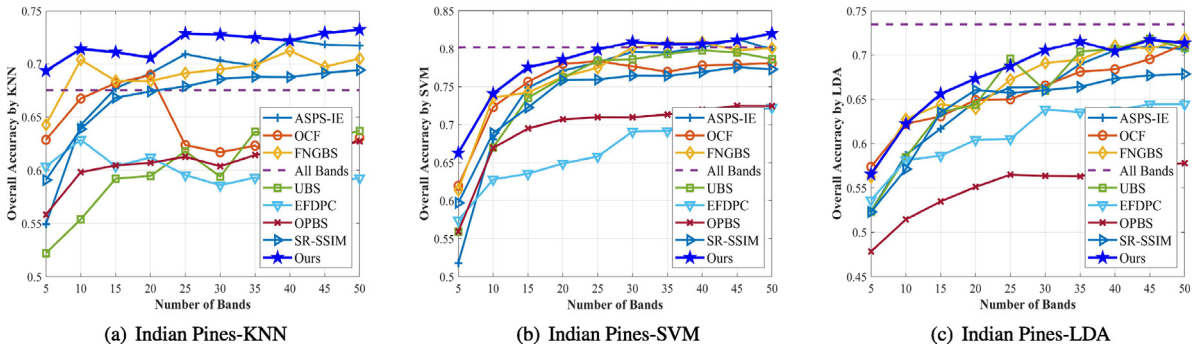
$$N = T \times \frac{N_z}{w \times h}, \tag{23}$$

where N_z represents the number of non-zero values at the edge of the first principal component of HSIs, and T denotes a constant. Similar to this work, the value of T is also set to 2000 in our experiments. In such a manner, the number of segmented regions of different HSIs can be adaptively determined.

5.4. Results analysis

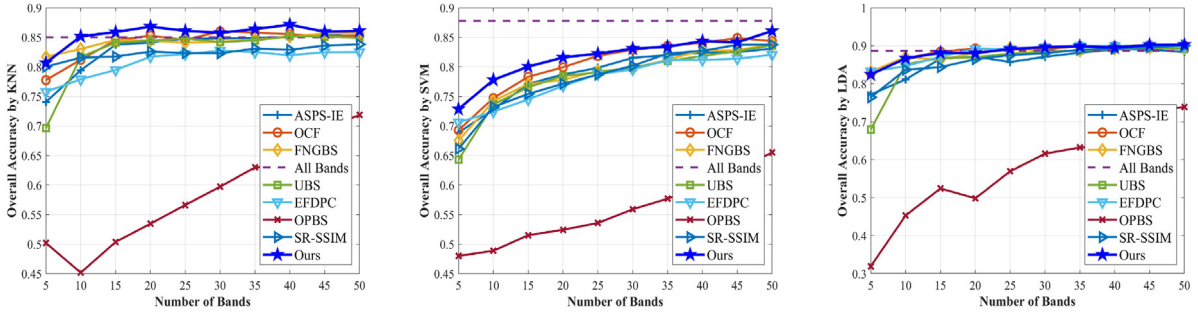
In Table 2, the averaged values and corresponding standard values of all algorithms on three classifiers are reported. As seen from Table 2, GRSC can achieve satisfying performance in most cases for all hyperspectral datasets, which demonstrates that GRSC is a promising method on the band selection. Furthermore, we also plot the OA curves for each dataset to verify the effectiveness of GRSC, as seen in Figs. 2–4. In addition, the results of all bands are also plotted as a baseline in all figures. Based on the above results, the following observations can be obtained:

(1) Our proposed method GRSC outperforms other competitors in terms of OA, AA, and Kappa on all hyperspectral datasets. From the results of all algorithms, we can find that the curves trend of each classifier is consistent with the KNN classifier on Indian Pines, KSC, and Botswana datasets. Furthermore, when more spectral bands are selected, the classification performance will increase, which is why the spectral signatures of objects with more bands become more distinguishable. However, for the clustering-based band selection method OCF, the classification performance becomes worse when K increases in Fig. 2(a), which can be attributed to the noisy bands are adopted and results in the high redundancy in the selected band subset. In addition, with the number of selected bands is set to 5, it should be noticed that GRSC can gain about 5% improvement on the Indian Pines dataset when compared with FNGBS, as seen in Fig. 2(a) and (b). As to others datasets, GRSC can also obtain satisfying performance



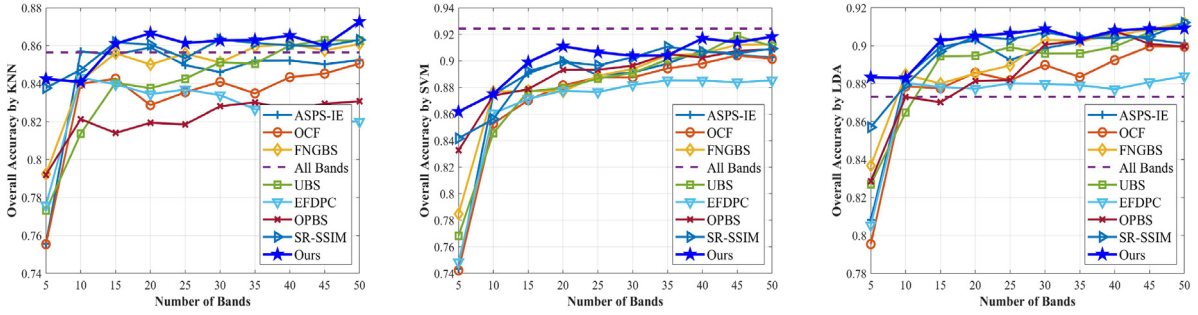
(a) Indian Pines-KNN (b) Indian Pines-SVM (c) Indian Pines-LDA

Fig. 2. The classification results in terms of OA of all competitors on the Indian Pines dataset.



(a) KSC-KNN (b) KSC-SVM (c) KSC-LDA

Fig. 3. The classification results in terms of OA of all competitors on the KSC dataset.



(a) Botswana-KNN (b) Botswana-SVM (c) Botswana-LDA

Fig. 4. The classification results in terms of OA of all competitors on the Botswana dataset.

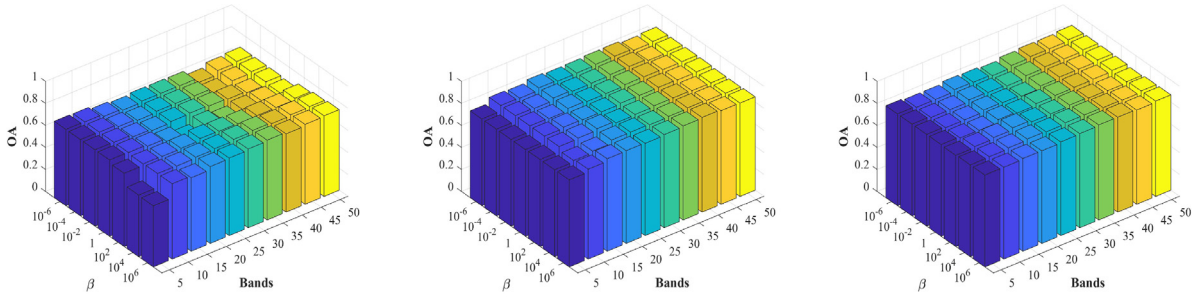
on the band selection. Overall, these results demonstrate the effectiveness of GRSC.

(2) Among the clustering-based band selection methods, GRSC can achieve the best results on three public hyperspectral datasets. Based on the assumption that the bands with contiguous wavelengths should be divided into the same cluster, OCF optimizes the objective function to obtain the optimal band subset. Similarly, ASPS-IE mainly generates the band subset based on the strong correlation between adjacent bands. However, the non-adjacent bands might also have high similarities, which conflicts with the above assumptions. Thus, the selected bands within the different clusters may have a significant correlation, resulting in the high redundancy in the optimal band subset. For GRSC, it takes into account the correlation among all bands, and thereby reduces the redundancy of the selected band subset. An intuitive comparison can be seen in all diagrams. For instance, the proposed method achieves 72.76% in terms of OA on the Indian Pines dataset with KNN classifier, while OCF and ASPS-IE obtain 61.89% and 69.96%, respectively. Similar results can also be seen in other datasets.

(3) The reason why the proposed method can obtain the superiority performance can be attributed to the utilization of spatial–spectral information and the latent features, which help to improve the distinguishable between different bands on the band selection. For other competitors, they do not take fully into account the prior information and yield sub-optimal solutions. As seen in Fig. 4(b) and (c), when 5 bands are selected, GRSC gains about 2% improvement in terms of OA compared with the second performer SR-SSIM. Therefore, these results verify the effectiveness of spatial–spectral information and latent features on the hyperspectral band selection.

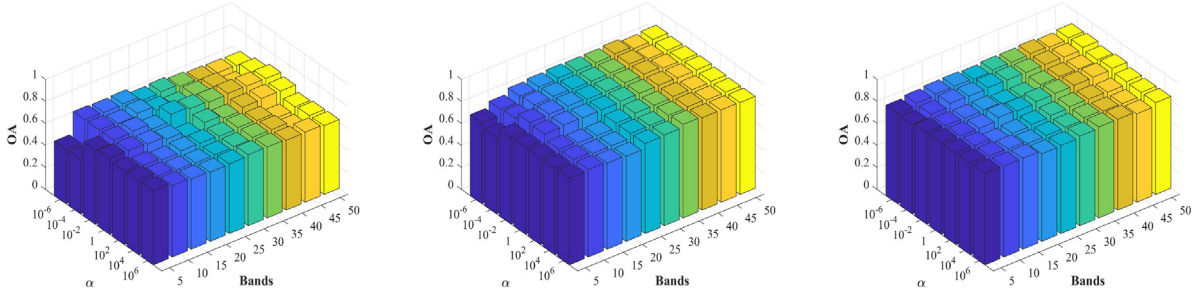
5.5. Parameter and convergence study

Based on Eq. (8), we can find that GRSC contains two parameters, i.e., α and β. To further study the effect of parameters on the proposed model, we fix one parameter and vary another parameter in the experiment, and the corresponding results with using KNN classifier are shown in Figs. 5 and 6. Specifically, we fix $\alpha = 1$ and $\beta = 1$, respectively, and vary the rest one to conduct



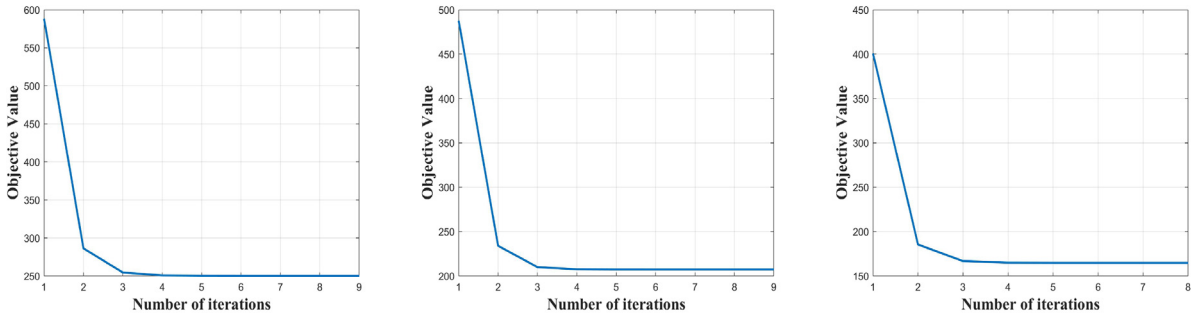
(a) Indian Pines (b) KSC (c) Botswana

Fig. 5. Parameter sensitivity analysis in terms of OA with using KNN classifier when α is fixed to 1.



(a) Indian Pines (b) KSC (c) Botswana

Fig. 6. Parameter sensitivity analysis in terms of OA with using KNN classifier when β is fixed to 1.



(a) Indian Pines (b) KSC (c) Botswana

Fig. 7. The objective function values of GRSC on three public hyperspectral datasets.

band selection. As seen in the figures, the proposed method is not sensitive to the two parameters, and it also can obtain stable results in a certain parameter range on all hyperspectral datasets.

As aforementioned, the proposed algorithm can converge in theory. To verify its convergence property, we plot the convergence curve of all datasets, as seen in Fig. 7. From the curves, we can observe that the proposed algorithm converges in several iterations.

5.6. Effect of different training samples sizes

As all we know, the size of training samples is also a key factor for the classification performance of classifiers. In order to further study the performance of all algorithms on hyperspectral band selection with different sizes of training samples, the corresponding results are reported in Fig. 8 within the training set of {1%, 5%, 10%, 20%, 50%}. In Fig. 8, it can be observed that GRSC still can achieve satisfying performance on all datasets, which demonstrates its superiority over other competitors.

5.7. Study on the different numbers of segmented regions

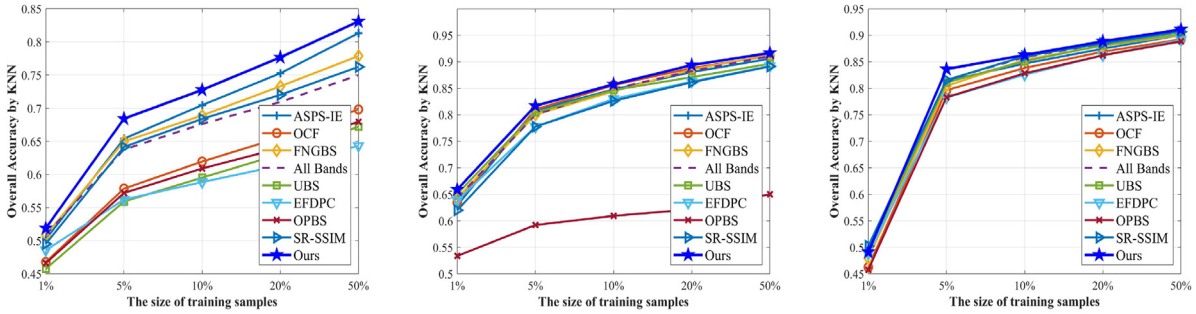
As aforementioned in Section 5.3, the number of segmented regions is determined via Eq. (23). To analyze the impact of different values of T on band selection, we set it to 500, 1000, 2000, and 3000 respectively, and the results in terms of OA on the KNN classifier are reported in Table 3. As seen in Table 3, although the results are not the best on some specific numbers of selected bands, GRSC can consistently achieve satisfying performance on all hyperspectral datasets when $T = 2000$. Thus, we think it is a reasonable value in the experiment.

6. Conclusion

In this paper, a graph regularized spatial–spectral subspace clustering method for hyperspectral band selection is proposed, dubbed GRSC, which takes into account the spectral correlation between different bands, as well as the spatial information of HSIs. In order to alleviate the impact of the noise and maintain the

Table 3
OA with different values of T on KNN classifier (%).

Datasets	T	N	5	10	15	20	25	30	35	40	45	50
Indian Pines	500	41	69.34	70.74	71.14	70.28	70.56	70.30	70.99	72.58	70.72	70.84
	1000	83	69.34	70.93	71.37	70.04	72.77	71.61	71.75	71.54	72.38	72.17
	2000	165	69.34	71.39	71.09	70.57	72.87	72.76	72.49	72.21	72.91	73.24
	3000	248	69.34	70.91	70.93	71.53	73.29	72.70	73.25	73.25	73.51	73.74
KSC	500	47	83.94	84.82	85.60	86.35	85.24	85.22	86.10	87.04	85.92	85.82
	1000	93	82.74	85.38	85.65	86.00	85.05	85.75	85.92	86.80	85.37	85.91
	2000	186	80.67	85.16	85.89	86.82	86.10	85.75	86.40	87.15	85.94	86.10
	3000	279	83.32	85.22	85.47	85.93	85.63	85.48	86.00	87.32	86.26	85.88
Botswana	500	45	83.66	85.42	86.91	86.85	85.69	86.53	86.80	86.95	86.43	87.51
	1000	90	83.67	85.28	86.23	86.30	85.84	86.58	86.84	86.75	86.13	87.35
	2000	180	84.25	84.08	86.11	86.66	86.14	86.28	86.30	86.53	86.01	87.27
	3000	270	83.40	84.19	85.40	86.31	85.73	86.09	86.12	86.69	86.12	87.60



(a) Indian Pines (b) KSC (c) Botswana

Fig. 8. The results with different sizes of training samples in terms of OA.

original spatial structure in the new self-representation space, the low dimensional latent features and a similarity graph between them are exploited in the proposed method. Finally, extensive experimental results demonstrate the effectiveness of GRSC on three public hyperspectral datasets.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work was supported by National Key R & D Program of China (No. 2020AAA0107100), and the National Natural Science Foundation of China (No. 62076228), and the National Natural Science Foundation of Shandong Province (No. ZR2021LZH001).

References

Cao, Xianghai, Wei, Cuicui, Ge, Yiming, Feng, Jie, Zhao, Jing, & Jiao, Licheng (2019). Semi-supervised hyperspectral band selection based on dynamic classifier selection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(4), 1289–1298.

Chang, Chein-I, Du, Qian, Sun, Tzu-Lung, & Althouse, M. L. G. (1999). A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(6), 2631–2641.

Datta, Aloke, Ghosh, Susmita, & Ghosh, Ashish (2015). Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(6), 2814–2823.

Du, Qian (2003). Band selection and its impact on target detection and classification in hyperspectral image analysis. In *IEEE workshop on advances in techniques for analysis of remotely sensed data, 2003* (pp. 374–377). IEEE.

Du, Qian, & Yang, He (2008). Similarity-based unsupervised band selection for hyperspectral image analysis. *IEEE Geoscience and Remote Sensing Letters*, *5*(4), 564–568.

Feng, Jie, Jiao, Licheng, Liu, Fang, Sun, Tao, & Zhang, Xiangrong (2014). Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(5), 2956–2969.

Feng, Wei, Quan, Yinghui, Dauphin, Gabriel, Li, Qiang, Gao, Lianru, Huang, Wenjiang, et al. (2021). Semi-supervised rotation forest based on ensemble margin theory for the classification of hyperspectral image with limited training data. *Information Sciences*, *575*, 611–638.

Flores, Hernan, Lorenz, Sandra, Jackisch, Robert, Tusa, Laura, Contreras, I. Cecilia, Zimmermann, Robert, et al. (2021). UAS-based hyperspectral environmental monitoring of acid mine drainage affected waters. *Minerals*, *11*(2), 182.

Guo, Baofeng, Gunn, Steve, Damper, Bob, & Nelson, James (2005). Adaptive band selection for hyperspectral image fusion using mutual information. In *2005 7th international conference on information fusion, Vol. 1* (p. 8). IEEE.

Huang, Shaoguang, Zhang, Hongyan, & Pižurica, Aleksandra (2021). A structural subspace clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–15.

Imani, Maryam, & Ghassemian, Hassan (2016). Binary coding based feature extraction in remote sensing high dimensional data. *Information Sciences*, *342*, 191–208.

Jakob, Sandra, Zimmermann, Robert, & Gloaguen, Richard (2017). 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 Sensing*, *9*(1), 88.

Jia, Sen, Tang, Guihua, Zhu, Jiasong, & Li, Qingquan (2015). A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(1), 88–102.

Jia, Sen, Tang, Guihua, Zhu, Jiasong, & Li, Qingquan (2016). A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(1), 88–102.

Jiang, Junjun, Ma, Jiayi, Wang, Zheng, Chen, Chen, & Liu, Xianming (2018). Hyperspectral image classification in the presence of noisy labels. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(2), 851–865.

Li, Shijin, Zhu, Yuelong, Wan, Dingsheng, & Feng, Jun (2013). Spectral similarity-preserving hyperspectral band selection. *Remote Sensing Letters*, *4*(10), 969–978.

Liu, Ming-Yu, Tuzel, Oncel, Ramalingam, Srikumar, & Chellappa, Rama (2011). Entropy rate superpixel segmentation. In *CVPR 2011* (pp. 2097–2104). IEEE.

Meng, ZENG, Bin, NING, Zhihua, CAI, & Qiong, GU (2020). Hyperspectral band selection based on deep adversarial subspace clustering. *Journal of Computer Applications*, *40*(2), 381.

Moroni, Monica, Lupo, Emanuela, Marra, Emanuela, & Cenedese, Antonio (2013). Hyperspectral image analysis in environmental monitoring: setup of a new tunable filter platform. *Procedia Environmental Sciences*, *19*, 885–894.

Nie, Feiping, Wang, Xiaoqian, & Huang, Heng (2014). Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 977–986).

Nie, Feiping, Zhu, Wei, & Li, Xuelong (2019). Structured graph optimization for unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*, *33*(3), 1210–1222.

Peyghambari, Sima, & Zhang, Yun (2021). Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. *Journal of Applied Remote Sensing*, *15*(3), Article 031501.

Rasti, Behnood, Ulfarsson, Magnus Orn, & Sveinsson, Johannes R. (2016). Hyperspectral feature extraction using total variation component analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(12), 6976–6985.

Saralıoğlu, Ekrem, Görmüş, Esra Tunç, & Güngör, Oğuz (2016). Mineral exploration with hyperspectral image fusion. In *2016 24th signal processing and communication application conference (SIU)* (pp. 1281–1284). IEEE.

Sun, Kang, Geng, Xiurui, & Ji, Luyan (2014). A band selection approach for small target detection based on CEM. *International Journal of Remote Sensing*, *35*(13), 4589–4600.

Sun, Weiwei, Peng, Jiangtao, Yang, Gang, & Du, Qian (2020). Fast and latent low-rank subspace clustering for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(6), 3906–3915.

Sun, Weiwei, Zhang, Liangpei, Du, Bo, Li, Weiyue, & Lai, Yenming Mark (2015). Band selection using improved sparse subspace clustering for hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(6), 2784–2797.

Sun, Weiwei, Zhang, Liangpei, Zhang, Lefei, & Lai, Yenming Mark (2016). A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(9), 4374–4388.

Tang, Chang, Liu, Xinwang, Li, Miaomiao, Wang, Pichao, Chen, Jiajia, Wang, Lizhe, et al. (2018). Robust unsupervised feature selection via dual self-representation and manifold regularization. *Knowledge-Based Systems*, *145*, 109–120.

Tang, Chang, Liu, Xinwang, Zhu, En, Wang, Lizhe, & Zomaya, Albert (2021). Hyperspectral band selection via spatial-spectral weighted region-wise multiple graph fusion-based spectral clustering. In *Proceedings of the thirtieth international joint conference on artificial intelligence organization, Montreal, QC, Canada* (pp. 19–27).

Tang, Chang, Liu, Xinwang, Zhu, Xinzhong, Xiong, Jian, Li, Miaomiao, Xia, Jingyuan, et al. (2020). Feature selective projection with low-rank embedding and dual Laplacian regularization. *IEEE Transactions on Knowledge and Data Engineering*, *32*(9), 1747–1760.

Tang, Chang, Zheng, Xiao, Liu, Xinwang, Zhang, Wei, Zhang, Jing, Xiong, Jian, et al. (2021). Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*.

Tang, Chang, Zhu, Xinzhong, Liu, Xinwang, Li, Miaomiao, Wang, Pichao, Zhang, Changqing, et al. (2019). Learning a joint affinity graph for multiview subspace clustering. *IEEE Transactions on Multimedia*, *21*(7), 1724–1736.

Tschannerl, Julius, Ren, Jinchang, Yuen, Peter, Sun, Genyun, Zhao, Huimin, Yang, Zhijing, et al. (2019). MIMR-DGSA: Unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm. *Information Fusion*, *51*, 189–200.

Wang, Qi, Li, Qiang, & Li, Xuelong (2019). Hyperspectral band selection via adaptive subspace partition strategy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(12), 4940–4950.

Wang, Qi, Li, Qiang, & Li, Xuelong (2020). A fast neighborhood grouping method for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(6), 5028–5039.

Wang, Jun, Tang, Chang, Li, Zhenglai, Liu, Xinwang, Zhang, Wei, Zhu, En, et al. (2022). Hyperspectral band selection via region-aware latent features fusion based clustering. *Information Fusion*, *79*, 162–173.

Wang, Qi, Zhang, Fahong, & Li, Xuelong (2018). Optimal clustering framework for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(10), 5910–5922.

Wang, Qi, Zhang, Fahong, & Li, Xuelong (2020). Hyperspectral band selection via optimal neighborhood reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(12), 8465–8476.

Xie, Weiying, Lei, Jie, Yang, Jian, Li, Yunsong, Du, Qian, & Li, Zhen (2019). Deep latent spectral representation learning-based hyperspectral band selection for target detection. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(3), 2015–2026.

Xu, Buyun, Li, Xihai, Hou, Weijun, Wang, Yiting, & Wei, Yiwei (2021). A similarity-based ranking method for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(11), 9585–9599.

Yang, Shuangming, Deng, Bin, Wang, Jiang, Li, Huiyan, Lu, Meili, Che, Yanqiu, et al. (2019). Scalable digital neuromorphic architecture for large-scale biophysically meaningful neural network with multi-compartment neurons. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(1), 148–162.

Yang, He, Du, Qian, Su, Hongjun, & Sheng, Yehua (2010). An efficient method for supervised hyperspectral band selection. *IEEE Geoscience and Remote Sensing Letters*, *8*(1), 138–142.

Yang, Shuangming, Gao, Tian, Wang, Jiang, Deng, Bin, Azghadi, Mostafa Rahimi, Lei, Tao, et al. (0000). SAM: A unified self-adaptive multicompartmental spiking neuron model for learning with working memory. Frontiers in Neuroscience, 467.

Yang, Shuangming, Tan, Jiangtong, & Chen, Badong (2022). Robust spike-based continual meta-learning improved by restricted minimum error entropy criterion. *Entropy*, *24*(4), 455.

Yang, Shuangming, Wang, Jiang, Deng, Bin, Azghadi, Mostafa Rahimi, & Linares-Barranco, Bernabe (2021). Neuromorphic context-dependent learning framework with fault-tolerant spike routing. *IEEE Transactions on Neural Networks and Learning Systems*.

Yu, Chunyan, Cen, Kun, Chang, Chein-I, & Li, Fang (2018). Unsupervised hyperspectral band selection method based on low-rank representation. In *International conference in communications, signal processing, and systems* (pp. 1053–1061). Springer.

Yuan, Yuan, Zheng, Xiangtao, & Lu, Xiaoqiang (2016). Discovering diverse subset for unsupervised hyperspectral band selection. *IEEE Transactions on Image Processing*, *26*(1), 51–64.

Zabalza, Jaime, Ren, Jinchang, Zheng, Jiangbin, Zhao, Huimin, Qing, Chunmei, Yang, Zhijing, et al. (2016). Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing*, *185*, 1–10.

Zeng, Meng, Cai, Yaoming, Cai, Zhihua, Liu, Xiaobo, Hu, Peng, & Ku, Junhua (2019). Unsupervised hyperspectral image band selection based on deep subspace clustering. *IEEE Geoscience and Remote Sensing Letters*, *16*(12), 1889–1893.

Zhai, Han, Zhang, Hongyan, Zhang, Liangpei, & Li, Pingxiang (2018). Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(3), 1723–1740.

Zhang, Peng, He, Haixia, & Gao, Lianru (2019). A nonlinear and explicit framework of supervised manifold-feature extraction for hyperspectral image classification. *Neurocomputing*, *337*, 315–324.

Zhang, Wenqiang, Li, Xiaorun, Dou, Yaxing, & Zhao, Liaoying (2018). A geometry-based band selection approach for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(8), 4318–4333.

Zhang, Aizhu, Ma, Ping, Liu, Sihan, Sun, Genyun, Huang, Hui, Zabalza, Jaime, et al. (2019). Hyperspectral band selection using crossover-based gravitational search algorithm. *IET Image Processing*, *13*(2), 280–286.

Zhang, Yongshan, Wang, Xinxin, Jiang, Xinwei, & Zhou, Yicong (2021). Marginalized graph self-representation for unsupervised hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*.

Zhu, Guokang, Huang, Yuancheng, Li, Shuying, Tang, Jun, & Liang, Dong (2017). Hyperspectral band selection via rank minimization. *IEEE Geoscience and Remote Sensing Letters*, *14*(12), 2320–2324.

Zhu, Zexuan, Jia, Sen, He, Shan, Sun, Yiwen, Ji, Zhen, & Shen, Linlin (2015). Three-dimensional gabor feature extraction for hyperspectral imagery classification using a memetic framework. *Information Sciences*, *298*, 274–287.

Zhu, Pengfei, Zuo, Wangmeng, Zhang, Lei, Hu, Qinghua, & Shiu, Simon CK (2015). Unsupervised feature selection by regularized self-representation. *Pattern Recognition*, *48*(2), 438–446.

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.

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, ACMM, 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.

Xiao Zheng received her master degree from the Tianjin Medical University, Tianjin, China. She is currently pursuing the Ph.D. degree with the National University of Defense Technology, China. Her recent research interests focus on machine learning and data mining.

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.

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.