Hyper-graph Regularized Subspace Clustering With Skip Connections for Band Selection of Hyperspectral Image

Meng Zeng[1], Bin Ning[1]*, Qiong Gu[1], Chunyang Hu[1], and Shuijia Li[2]

[1] School of Computer Engineering, Hubei University of Arts and Science
Xiangyang, Hubei, China, 441053
zengmeng@cug.edu.cn, {ningbin2000, qionggu, huchunyang}@hbuas.edu.cn
[2] School of Computer Science, China University of Geosciences
Wuhan, Hubei, China, 430074
shuijiali@cug.edu.cn

Abstract. The Hughes phenomenon of Hyperspectral images (HSIs) with the hundreds of continuous narrow bands makes the computational cost of HSIs processing high. Band selection is an effective way to solve such a problem and a lot of band selection methods have been proposed in recent years. In this paper, a novel hyper-graph regularized subspace clustering with skip connections (HRSC-SC) is proposed for band selection of hyperspectral image, which is a clustering-based band selection method. The networks combine subspace clustering into the convolutional auto-encoder by thinking of it as a self-expressive layer. To make full use of the historical feature maps obtained from the networks and tackle the problem of gradient vanishing caused by multiple nonlinear transformations, the symmetrical skip connections are added to the networks to pass image details from encoder to decoder. Furthermore, the hyper-graph regularization is presented to consider the manifold structure reflecting geometric information within data, which accurately describes the multivariate relationship between data points and makes the results of clustering more accurate so that select the most representative band subset. The proposed HRSC-SC band selection method is compared with the existing robust band selection algorithms on Indian Pines, Salinas-A, and Pavia University HSIs, showing that the results of the proposed method outperform the current state-of-the-art band selection methods. Especially, the overall accuracy of the clustering is the best on three real HSIs compared to other methods when the band selection number is 25, reaching 82.62%, 92.48%, and 96,5% respectively.

Keywords: Band selection, hyper-graph regularization, skip connections, subspace clustering, hyperspectral image

1. Introduction

Hyperspectral images (HSIs) with hundreds of narrow bands containing abundant spatial and spectral information so that it can identify the region of interest. Due to the high redundancy, the so-called Hughes phenomenon happens on HSIs frequently and increases computation complexity. Band selection (BS) is an effective way to reduce the dimensionality of HSIs, aiming to select the significant bands with the most information from the original data set as a band subset. The details of band selection are shown in Fig. 1. The

---

* Corresponding author

BS methods do not destroy the physical properties of the HSIs, which is different from the feature extraction methods that transform the physical characteristics of HSIs. Therefore, the BS methods are easier to explain than the feature extraction methods. BS methods can be classed as supervised and unsupervised fashions [1]. The supervised methods have to apply the prior knowledge, and the unsupervised methods are the opposite. Considering the difficulty to get the labeled samples of HSIs in reality, the unsupervised methods have better application prospects and attracted more attention in recent years.
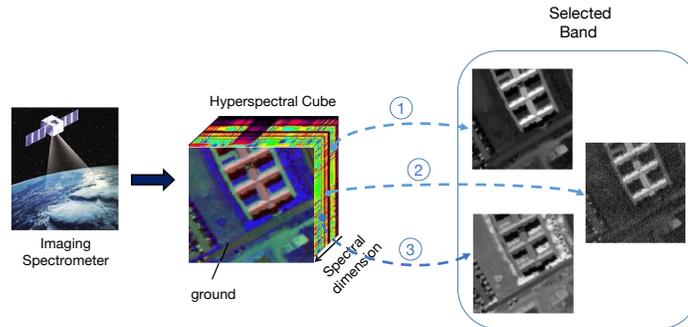

Fig. 1. Details of band selection.

The unsupervised BS methods can be divided into three categories: the searching-based methods, the ranking-based methods, and the clustering-based methods. The searching-based methods, such as multi-objective optimization based band selection (MOBS) [2], optimize the given metric using the heuristic searching, but such methods cost much time on heuristic searching. The ranking-based methods select band according to the importance of the bands, which allocate a rank for each band, such as maximum-variance principal component analysis (MVPCA) [3], sparsity-based band selection (SpaBS) [4], and Laplacian score (Lap-score)[5]. The clustering-based methods cluster the bands into several categories according to the assumption that all bands can be separated based on the similarity of the bands, and we can select the band subset from the categories, i.e., sparse non-negative matrix factorization clustering (SNMF) [6], improved sparse subspace clustering (ISSC) [7], deep subspace clustering for Band Selection (DSCBS) [8], etc. The clustering-based methods make full consideration the interaction between bands so that the clustering-based methods get great success over the past years. In this paper, we use a cluster-based method for band selection, which is based on the deep subspace clustering (DSC) [9] network. The proposed method makes full use of the globally nonlinear spectral-spatial relationship, improving the performance of band selection.

However, although DSC has a good effect on general images, it is difficult to achieve better results for complex HSIs. For example, the average clustering accuracy of DSC can reach 98% on the ORL data set [9], but only 75% on the Indian Pines data set and 70% on the Pavia University data set [10]. Therefore, skip connections and hyper-graph learning are introduced to optimize the network and improve the accuracy of BS.

The skip connections [11], also known as residual connections, can skip one or more layers in different layers to build extra connections between nodes. These skip connections can solve the problem of gradient vanishing, recover the original image and make full use of the historical feature maps obtained from the networks by pass image details from the encoder to the decoder in auto-encoder networks. The graph learning [12,13] is an important topic and has been widely used in image processing, which reflects the geometric information of data by learning the manifold structure. However, the simple graph model can only represent the simple relationships between data so it can not obtain robust results in complex high-order relationships of images. To overcome the obstacles, hypergraph learning [14] has been introduced to describe the multivariate relationship between simples of complex HSIs. To better compare the pros and cons of different methods, we summarize the above methods, as shown in table 2.

In this paper, a novel hyper-graph regularized subspace clustering with skip connections (HRSC-SC) for band selection of hyperspectral image is introduced, which is the clustering-based band selection method. The subspace clustering is used to combine the convolutional auto-encoder by thinking of it as a self-expressive layer. The symmetrical skip connections are added to the convolutional auto-encoder (CAE) [15] for HSIs clustering, which pass image details from encoder to decoder. In addition, the hypergraph regularized is introduced to describe the multivariate relationship between simples of complex HSIs. The main contributions of this paper are summarized as follows:

1. We propose a hyper-graph regularized subspace clustering with skip connections (HRSC-SC) for band selection of the hyperspectral image.
2. The symmetrical skip connections are added into the convolutional auto-encoder with a self-expressive layer to make full use of the historical feature maps obtained from the networks and tackle the problem of gradient vanishing caused by multiple nonlinear transformations, which pass image details from encoder to decoder and produce more beneficial representation for better clustering.
3. The hyper-graph regularized is introduced to describe the multivariate relationship between simples of complex HSIs that can fully consider the manifold structure reflecting geometric information within data, making the modeling of images more accurate. Three HSI data sets are utilized to evaluate the performance and efficiency of the proposed band selection algorithm. Experimental results show that the proposed HRSC-SC band selection method has state-of-the-art performance, outperforming the current robust band selection methods.

To make the subsequent expression clearer, we construct table 1 to summarize the acronyms in this paper. The rest of the paper is structured as follows. In Section 2, we briefly review the sparse subspace clustering and the convolutional auto-encoder. Then, the details of the proposed HRSC-SC for band selection are introduced in section 3. In section 4, we evaluate the HRSC-SC for three well-known HSI data sets. Finally, conclude with a summary and discuss the future research directions in section 5.

2.   Previous Work

2.1.   Sparse Subspace Clustering

The subspace clustering methods consist of two steps: first, evaluate the affinity for each pair of simples to build an affinity matrix, which is the most crucial step and determines

Table 1. Acronyms in the paper.

Full name	Acronyms
Average Accuracy	AA
Alternating Direction Method of Multipliers	ADMM
Auto-Encoders	AE
Band Selection	BS
Convolutional Auto-Encoder	CAE
Deep Subspace Clustering	DSC
Deep Subspace Clustering for Band Selection	DSCBS
Hyper-graph Regularized Subspace Clustering with Skip Connections	HRSC-SC
Hyperspectral Images	HSIs
Improved Sparse Subspace Clustering	ISSC
Laplacian score	Lap-score
Multi-objective Optimization based Band Selection	MOBS
Maximum-variance Principal Component Analysis	MVPCA
Overall Accuracy	OA
Self-expressive	SE
Sparse Non-negative Matrix Factorization	SNMF
Sparsity-based Band Selection	Spa-BS
Sparse Subspace Clustering	SSC

Table 2. Comparison of band selection methods.

Methods	Based	Technology	Band interaction	Spectral-spatial
MOBS [2]	searching	heuristic searching	no	not consider
MVPCA [3]	ranking	maximum-variance	no	not consider
SpaBS [4]	ranking	sparse representation	no	not consider
Lap-score [5]	ranking	Laplacian score	no	not consider
SNMF [6]	clustering	sparse non-negative matrix factorization	yes	not consider
ISSC [7]	clustering	sparse subspace clustering	yes	not consider
DSCBS [8]	clustering	deep auto-encoder+subspace clustering	yes	consider
HRSC-SC	clustering	Hyper-graph+Skip Connections+DSC	yes	full consider

the results of clustering; second, utilize spectral clustering [16] or normalized cuts [17] method by using the affinity matrix. Here, we briefly introduce the SSC [18] method. Let the data set be the size of $M \times N$, where M and N denote the dimension of the feature and the number of data points respectively. Assuming that all the data points lie in a union of t affine subspaces $S = S_1, S_2, \cdots S_t$, where the t subspace has dimensions $\{d_i\}_{i=1}^{t}$ and $d_1 + d_2 + \cdots d_t = M$. Based on the hypothesis above, the optimization equation can be represented as:

$$\min_{\mathbf{C}} \|\mathbf{C}\|_1 , s.t. \mathbf{Y} = \mathbf{YC} + \mathbf{N}, diag\left(\mathbf{C}\right) = 0 \qquad (1)$$

where $\mathbf{C}$ is the coefficient matrix. $\mathbf{Y}$ represents the data points. $\mathbf{N}$ denotes the error matrix. 1 stands for the l_1-norm regularization, which can be 0, 1, 2 in other subspace clustering methods. To avoid trivial solution, the $diag\left(\mathbf{C}\right)$ is constrained to be 0. The ADMM [19,20] method is applied to optimize Eq.1. Then, the similarity graph is constructed by

the coefficient matrix $\mathbf{C}$, the affinity matrix $\mathbf{M}$ can be written as:

$$\mathbf{M} = |\mathbf{C}| + |\mathbf{C}|^T \tag{2}$$

In the end, according to the similarity graph, the results of the clustering can be obtained by the spectral clustering method.

2.2. Convolutional Auto-encoder

The AE consists of the symmetrical encoder and decoder structure, which can convert the data points into the latent space representation. In recent years, the AE methods have been widely used and achieved state-of-the-art performance in learning data deep representation, e.g. variational auto-encoder [21], sparse auto-encoder [22], and denoising auto-encoder [23].

The structure of the CAE is similar to the AE, which consists of the symmetrical convolutional and deconvolutional layers. Here, we can define the convolutional layer as $\varphi = E(x; \alpha_e)$, in which φ denotes the latent representation (or bottleneck) to reveal the intrinsic information of the input data, x and α_e stand for data points and parameters respectively. Analogously, the deconvolutional can be defined as $\hat{x} = D(\varphi; \alpha_d)$, where $\hat{x}$ is the reconstruction of input data and α_d represents the parameters of the decoder. Then, we can define the loss function as:

$$\mathcal{L}(\alpha_e; \alpha_d) = \frac{1}{2} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|_F^2 \tag{3}$$

3.   Proposed Method

3.1.   Convolutional Auto-encoder with Self-expressive layer

The main structure of the HRSC-SC is to insert a SE layer into a Convolutional CAE [23] to learn a representation for HSI. The SE layer between the encoder and decoder of the CAE to imitate the "self-expressive" property of the traditional subspace clustering. Therefore, the SE layer is as similar as SSC [18], which is defined as:

$$\hat{\mathbf{Z}} = \mathbf{C}\mathbf{Z} \tag{4}$$

where $\hat{\mathbf{Z}}$ denotes the reconstruction of $\mathbf{Z}$, which is the input of the SE layer. $\mathbf{C}$ is the coefficient matrix obtained from the SE layer. The loss function of the SE layer as follows:

$$\mathcal{L}_{SE} = \frac{\lambda_1}{2} \left\| \mathbf{Z} - \hat{\mathbf{Z}} \right\|_F^2 + \frac{\lambda_2}{2} \|\mathbf{C}\|_F^2 \tag{5}$$

There are two parts in Eq.5: the first part is the reconstruction error; the second part uses an F-norm regularization to constrain $\mathbf{C}$. Here, we have not to use the condition of $diag(\mathbf{C}) = 0$ like the traditional SSC because using F-norm without diagonal constraints will not lead to trivial solutions [24].
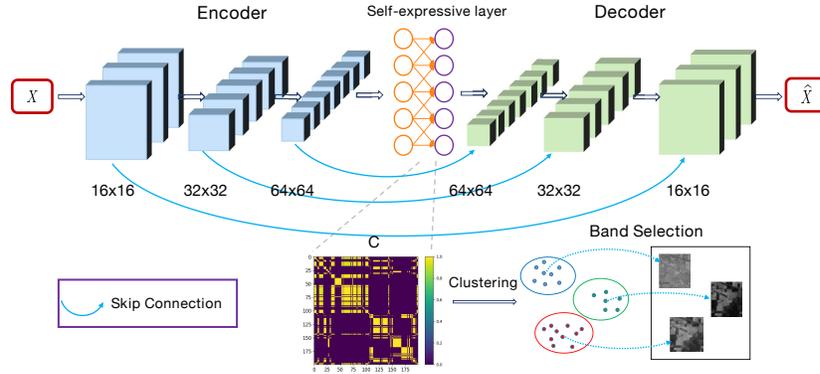
Fig. 2. Overall architecture of the HRSC-SC.

As shown in Fig. 2, giving the input HSI $\mathbf{X} = \{x_i\}_{i=1}^m$ and it can be coded through the encoder to obtain the latent representations $\mathbf{Z}$, the function of the encoder is defined as $\mathbf{Z} = f_\beta(\mathbf{X})$. Symmetrically, the refactored $\mathbf{X}$ can be defined as $\hat{\mathbf{X}} = g_\gamma(\mathbf{Z})$. Each layer of the CAE followed by a batch normalization and a ReLU activation. The loss function of the CAE and the CAE with SE layer are defined as Eq.6 and Eq.7, respectively:

$$\mathcal{L}_{AE} = \frac{1}{2} \left\| \mathbf{X} - \hat{\mathbf{X}} \right\|_F^2 \tag{6}$$

$$\mathcal{L}_{AES}(\beta;\gamma) = \mathcal{L}_{AE} + \mathcal{L}_{SE} \tag{7}$$

3.2. Skip Connections

Using the CAE with SE layer for HSI clustering can solve the problem of the nonlinear subspaces. However, all nodes in the SE layer are connected by linear weights that have no bias and non-linear activations so that N^2 parameters exist in the SE layer, making the SE layer the focus of training and overwhelming the optimize of CAE networks. Besides, the multiple layers networks may cause the problem of vanishing gradient.

To deal with the problems above, the symmetrical skip connections are introduced into the CAE. As shown in Fig. 2, skip shortcuts connect the convolution feature map and their deconvolution feature map in a symmetrical manner [25]. The gradients can back-propagate to the corresponding encoder layers directly without going through the SE layer. Ideally, using skip connections can make the network's train from scratch easier. Suppose x_j represents the j-th encoder layer mapping and $\tilde{x}_j$ is the corresponding j-th underlying decoder mapping, the skip connection mapping can fit:

$$\mathcal{H}(x_j) = \tilde{x}_j - x_j \tag{8}$$

According to Eq.8, the j-th underlying decoder mapping can be:

$$\tilde{x}_j = \mathcal{H}(x_j) + x_j \tag{9}$$

3.3.   Hyper-graph Learning

According to the graph learning methods [12], the potential geometry of high dimensional data points can be retained by the neighbor graph of the original data [13]. Therefore, we can use the theory to consider the manifold structure reflecting geometric information within data. Different from the traditional graph learning that can only explain the simple relationships between data points, the hyper-graph explains the complex relationships by connecting three or more vertices to describe the multivariate relationships of data accurately, achieving outstanding performance in recent works[26]. Therefore, we introduce the hyper-graph regularized into the subspace clustering networks to fully consider the multivariate relationships of complex HSIs. The hyper-graph structure can be represented as Fig. 3.
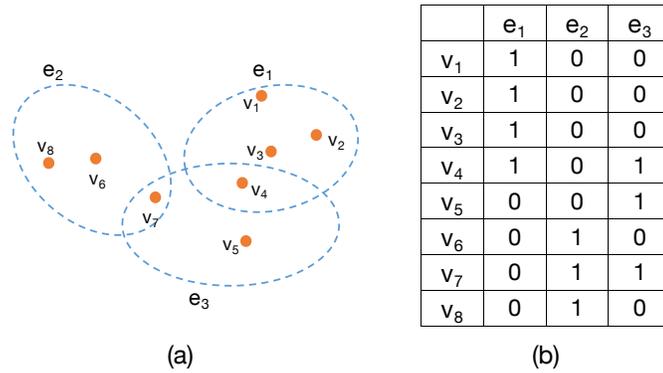
	e_1	e_2	e_3
v_1	1	0	0
v_2	1	0	0
v_3	1	0	0
v_4	1	0	1
v_5	0	0	1
v_6	0	1	0
v_7	0	1	1
v_8	0	1	0

(a)   (b)

Fig. 3. (a) Illustration of the hyper-graph. (b) Corresponding to (a), $\mathbf{V} = \{v_i\}_{i=1}^{m}$ is a vertex and $\mathbf{E} = \{e_j\}_{j=1}^{n}$ denote a hyper-edge, set $(v_i, e_j) = 1$ if the vertex i on the hyper-edge or $(v_i, e_j) = 0$ when the vertex i is not on the hyper-edge.

Assuming there is a hyper-graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$, in which $\mathbf{V} = \{v_i | i = 1, 2, \cdots, m\}$ denotes the multiple non-empty vertex set, $\mathbf{E} = \{e_j | j = 1, 2, \cdots n\}$ is non-empty hyper-edge subsets and $\mathbf{W}$ represents the weight matrix. We can define the $w(e)$ as weight of each hyper-edge. The incident matrix ζ is:

$$\zeta(v, e) = \begin{cases} 1, & if \ v \in e \\ 0, & if \ v \notin e \end{cases} \tag{10}$$

Then, the degree of a vertex v and the hyper-graph e can be defined as Eq.11 and Eq.12 respectively:

$$d(v) = \sum_{\{e \in E | v \in V\}} w(e) = \sum_{e \in E} w(e) \zeta(v, e) \tag{11}$$

$$\phi\left(e\right) = |e| = \sum_{v \in V} \zeta\left(v, e\right) \tag{12}$$

Finally, the hyper-graph Laplacian matrix $\mathbf{L}_h$ can be written as:

$$\mathbf{L}_h = \mathbf{D}_v - \zeta \mathbf{W} \mathbf{D}_e^{-1} \zeta^T \tag{13}$$

where D_v denotes the degree matrix of vertex and D_e is the degree matrix of hyper-edge, both of them are the diagonal matrices.

3.4.   Training and Band Selection

Algorithm 1 Pseudocode of HRSCNet

Input: image data set $\mathbf{X}$; number of clusters: k; hyper-parameters: $\lambda_1, \lambda_2, \lambda_3$.
Output: Clustering results.
1 Preprocess image data set;
2 Initialize $\mathbf{C}$, β and γ of the HRSCNet in Eq.14 ;
3 Compute the hyper-graph Laplacian matrix L_h according
   to Eq.13;
4 **while** maximum iteration is met **do**
5   Calculate the output of the encoder: $\mathbf{Z} = f_\beta\left(\mathbf{X}\right)$;
6   Calculate the output of the SE layer: $\hat{\mathbf{Z}} = \mathbf{C}\mathbf{Z}$;
7   Calculate the output of the decoder: $\hat{\mathbf{X}} = g_\gamma\left(\mathbf{Z}\right)$;
8   Calculate the loss according to Eq. 14;
9   Update $\mathbf{C}$, β and γ using Adam optimizer;
10 **end**
11 Construct affinity matrix according to Eq.15;
12 Calculate clustering results using spectral clustering;
13 Return clustering results.

There are four parts included in the HRSC-SC: the CAE, the SE layer, the skip connections, and the hyper-graph regularization. Therefore, the loss function of the HRSC-SC contains the CAE loss L_{AE} in Eq.6, the SE layer loss L_{SE} in Eq.5, and the hyper-graph regularized loss $\mathcal{L}_h = \frac{\lambda_3}{2} Tr\left(\mathbf{C}^T \mathbf{L}_h \mathbf{C}\right)$, expressed as:

$$\mathcal{L}\left(\mathbf{C}; \beta; \gamma\right) = \mathcal{L}_{AE} + \mathcal{L}_{SE} + \mathcal{L}_h$$
$$= \frac{1}{2}\left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_F^2 + \frac{\lambda_1}{2}\left\|\mathbf{Z} - \hat{\mathbf{Z}}\right\|_F^2 + \frac{\lambda_2}{2}\left\|\mathbf{C}\right\|_F^2 + \frac{\lambda_3}{2}Tr\left(\mathbf{C}^T \mathbf{L}_h \mathbf{C}\right) \tag{14}$$

where λ_1, λ_2 and λ_3 in $\mathcal{L}_{SE}$ and $\mathcal{L}_h$ denote the balancing parameters. The Adam [20] gradient descent method can be utilized to optimize Eq.14. It is worth noticing that we can train the HRSC-SC from scratch because of the skip connections technique, which is different from some existing deep subspace clustering methods that require pre-training.

According to the Fig. 2, we can get the coefficient matrix $\mathbf{C}$ after training the networks. Then, using $\mathbf{C}$ to construct a symmetric matrix $\mathbf{M}$:

$$\mathbf{M} = |\mathbf{C}| + |\mathbf{C}|^T \tag{15}$$

It can be seen in Fig. 2, the subspace clustering method is used to get the clustering results by cluster M into k classes. Then, the average of bands in each class is deemed as a cluster center, and we calculate the distance from each band to the cluster center. Finally, the selected band is the band closest to the cluster center in each category.

To express the HRSC-SC more clearly, the pseudocode of the proposed algorithm is shown in Algorithm 1.

4.   Results

4.1.   Data Set and Experimental Settings

We utilize three widely used images data sets for experiments: Indian Pines, Salinas-A, and Pavia University data sets [3]. For the convenience of experiment, the subscence located are used on Indian Pines and Pavia University data sets at $[50 \sim 120, 50 \sim 120]$ and $[200 \sim 300, 100 \sim 200]$, respectively. The details of the three data sets are shown in Table 3.

Table 3. Summary of Indian Pines, Salinas-A and Pavia University data sets.

Data sets	Indian Pines	Salinas-A	Pavia University
Pixels	70×70	86×83	100×100
Bands	200	204	103
Sensor	AVIRIS	AVIRIS	ROSIS

In the HRSC-SC network, the layers of encoder and decoder are all set to 3 and the encoder and the decoder have a symmetrical structure. Therefore, if the channels of the encoder in three layers are set to 16, 32, 64, respectively, the channels of the decoder will be set to 64, 32, 16. The learning rate is set to 1.0×10^{-4} and the hyper-parameters λ_1, λ_2 and λ_3 are all set to 1.0. For a more intuitive description, the hyper-parameters setting of all comparison BS methods is shown in Table 4. All methods except the MOBS method run in Python 3.6 and the MOBS method run in Matlab 2016. All code of the methods is run on the Intel Core i5 3.10GHz.

4.2.   Comparison of Performance

In this experiment, the number of the selected bands is changed in the range of 5 to 30, and the interval is set as 5. The SVM classifier [27,28] is used for all band selection methods to evaluate the OA, AA, and kappa coefficient (Kappa) of the different methods. To ensure the fairness of the experiment, we select 5% of labeled samples from each data set as the training set, and others as the testing set. To better test the performance of the algorithm, we evaluate all methods for 10 independent rums. The seven well-known band selection methods, Lap-score [5], SpaBS [4], ISSC [7], SNMF [6], MVPCA [3], DSCBS [8] and MOBS [2], are used as comparison methods to compare with the proposed HRSC-SC method on Indian Pines, Salinas-A, and Pavia University data sets.

---

[3] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes
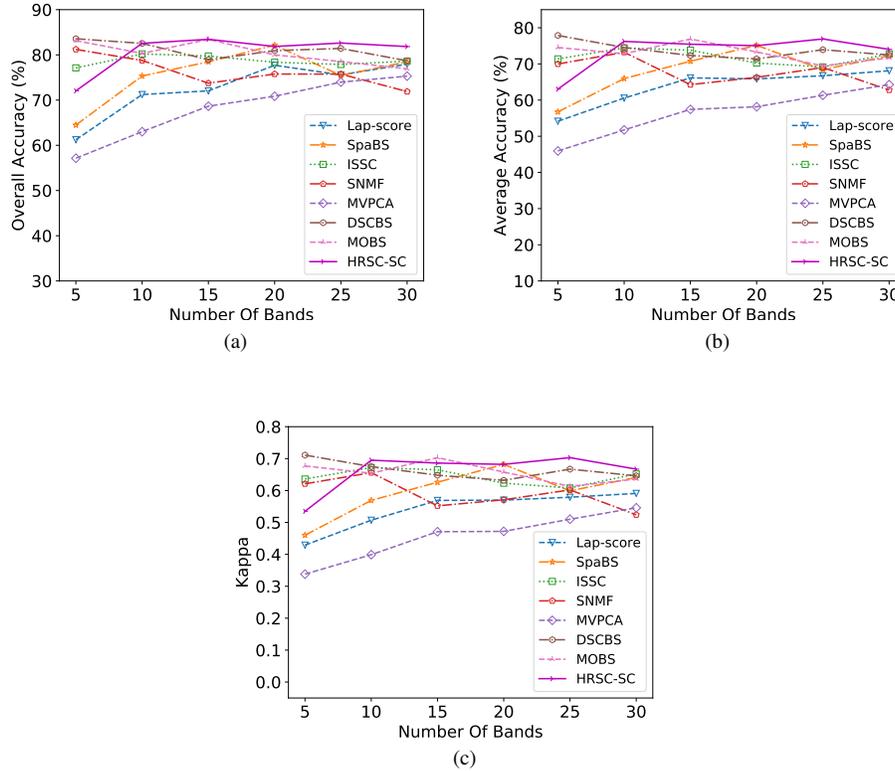
Fig. 4. Performance comparison of different BS methods with different band subset sizes on Indian Pines data set:(a) OA, (b) AA, and (c) Kappa.

The performance comparison of different BS methods with different band subset sizes is shown in Fig. 4 to Fig 6. Overall, it can be seen that the proposed HRSC-SC method achieves the best band selection results on three data sets. As shown in Fig. 4, the OA, AA, and Kappa of HRSC-SC are lower than some methods, but it achieves the best performance than other sever methods when the selected bands are 10 to 30. In addition, the HRSC-SC, SpaBS, ISSC, and MOBS show the best performance when selecting 15 bands, and show the downward trend after 15, that is the accuracy increases first and decreases with the selected bands increasing. The above phenomenon is called Hughes, which is mentioned in section 1. However, the HRSC-SC also achieves the best results than others, which shows the superiority of the HRSC-SC. It is noticed that the ranking-based methods, i.e. MVPCA, SpaBS, and Lap-score, achieve relatively poor accuracy than other approaches because of the increased chances of the misestimating band due to the noise of HSI.

In Fig. 5, similar to Indian Pines, the HRSC-SC shows the best performance when selecting more than 10 bands. Different from Indian Pines, the HRSC-SC has no obvious Hughes phenomenon on Salinas-A. But other methods, like SpaBS, ISSC, and MOBS,
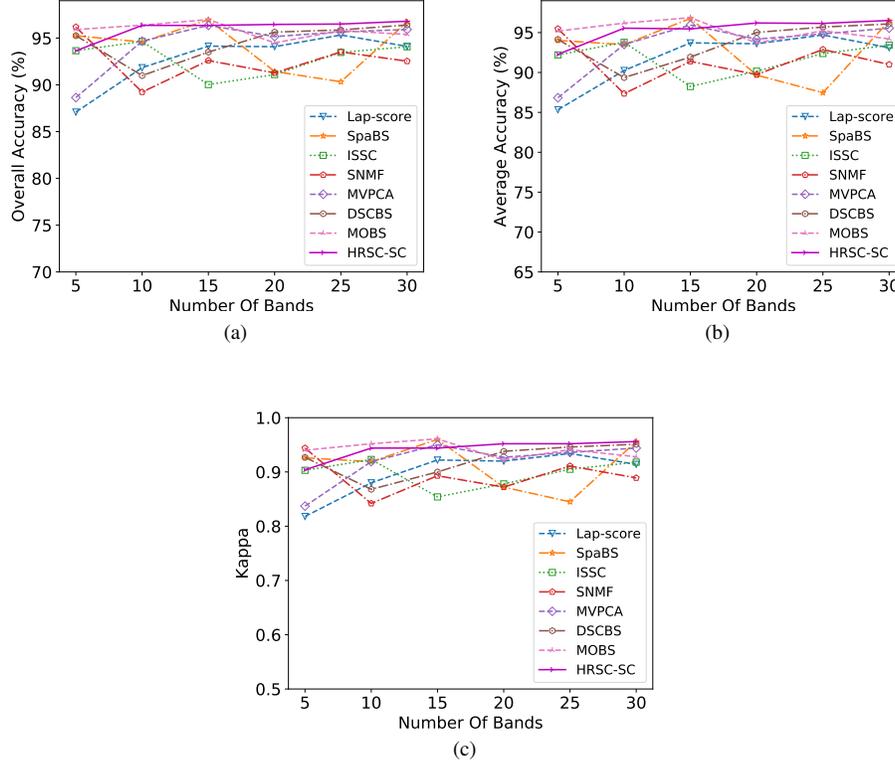
Fig. 5. Performance comparison of different BS methods with different band subset sizes on Salinas-A data set:(a) OA, (b) AA, and (c) Kappa.

also have the Hughes phenomenon. Therefore, HRSC-SC achieves significant results and outperforms all other methods.

In Fig. 6, the HRSC-SC shows great performance when the selected bands are 5 to 15. Then, the results of HRSC-SC are slightly inferior to the DSCBS method. However, the overall performance comparison result is similar to Indian Pines and Salinas-A data sets. The clustering-based methods can fully consider the mutual relations between different bands, especially the proposed method, which added the skip connections and the hyper-graph regularization to consider the deeper relationship of data. Consequently, the proposed clustering-based method can achieve better results than other methods.

4.3.   Analysis of the Selected Bands

We can analyze the selected bands according to the results of the tables and figures. To ensure the results more obvious, 15 bands are selected for all selected bands experiments on three data sets. Table 5 shows the selected bands of all methods when the number of selected bands is 15 on three data sets. Relatively, Fig. 7, Fig. 8, and Fig. 9 show the
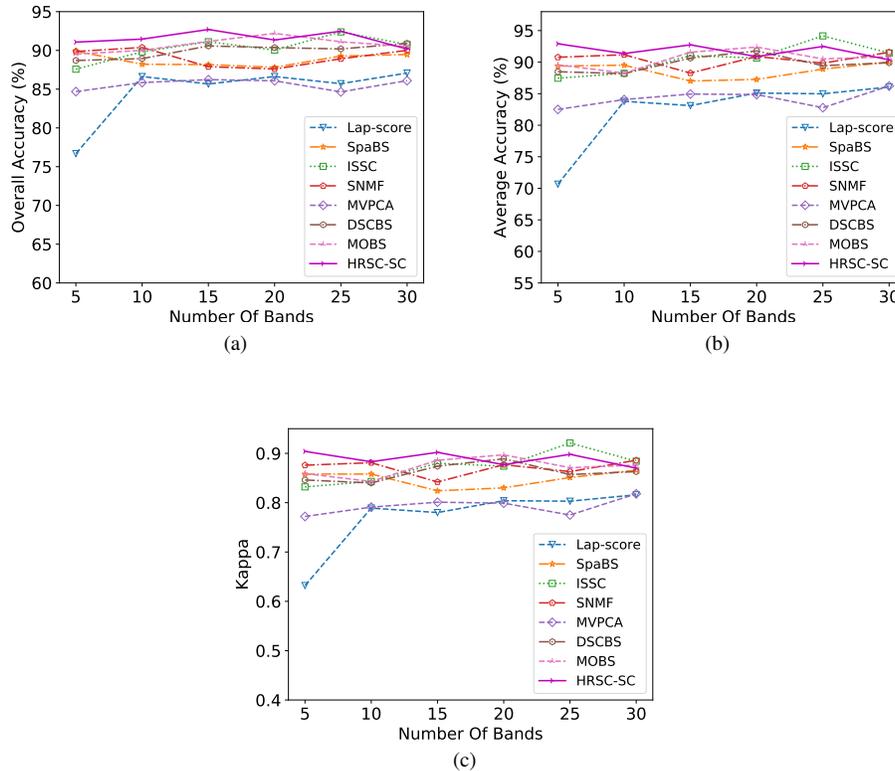
Fig. 6. Performance comparison of different BS methods with different band subset sizes on Pavia University data set:(a) OA, (b) AA, and (c) Kappa.

locations of the selected bands (above) on the spectrum, and the entropy curve (below) of three data sets.

In Fig. 7, the value of entropy is relatively average. For HRSC-SC, the selected bands have uniform distribution and no continuous bands. As for other algorithms, they all have continuous bands, especially MVPCA and Lap-score. The selected bands of the proposed HRSC-SC methods are most evenly distributed.

In Fig. 8, the value of entropy is similar to Indian Pines, it can be seen that the entropy curve has a relatively uniform distribution, which illustrates that better results can be obtained when the band selection is more uniform. In contrast, the MVPCA and the Lap-score methods select a lot of continuous bands. The other comparison algorithms select the relatively evenly bands. However, all of them select some continuous bands, which affect the final classification results. Therefore, the HRSC-SC method achieves the best performance.

Fig. 9 shows the selected bands and the entropy curve of Pavia University. As shown in this figure, the value of entropy is smooth initially and gradually increases, and reaches stability in the end. According to the distribution of entropy, choose server bands in num-

Table 4. Hyper-parameters setting of all comparison BS methods.

Methods	Hyper-paramenters
Lap-score	-
SpaBS	$\lambda = 1e2$
ISSC	$\lambda = 1e5$
SNMF	$maxiter = 100$
MVPCA	-
DSCBS	$\alpha = 1.0, \lambda = 1e - 3$
MOBS	$maxiter = 100, NP = 100$
HRSC-SC	$\lambda_1 = \lambda_2 = \lambda_3 = 1.0$
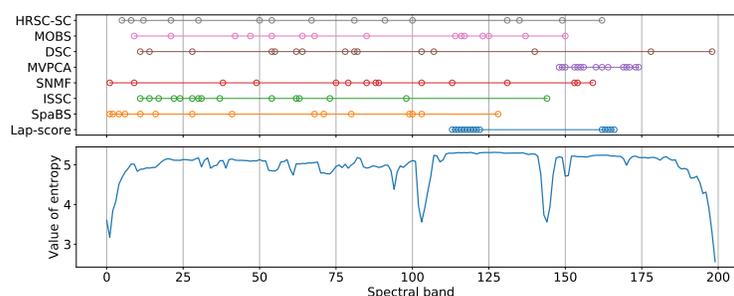
Fig. 7. The best 15 bands of Indian Pines data set selected by different BS methods (above) and the entropy value of each band (below).

bers 0 to 40, and approximately the average selection of bands after number 40 can get good results. Compare with the algorithm, HRSC-SC selects server bands and approximately average selects bands after band number 40. The performance of other methods is the same as using Indian Pines and Salinas-A. Therefore, HRSC-SC can get the best results for band selection on three data sets.

4.4. Impact of Epochs

As shown in Fig. 10, we plot the accuracy of the HRSC-SC and the training losses on different epochs to analyze the convergence on three data sets. The selected bands are set to 20. In Fig. 10 (a), we set the number of iterations from 0 to 200. It can be seen that the lower accuracy at the beginning, and the accuracy increases to maximum when the epoch increases close to 100, then fluctuate within a certain range the loss stabilizes. We set the number of iterations from 0 to 100 on the Salinas-A data set. According to the Fig.10 (b), the accuracy increases with the losses decrease, and then the accuracy reaches the maximum value when the epoch increases close to 50. When the epoch increases close to 100, the accuracy fluctuates within a certain range and the loss tends to be 0 while the accuracy achieve stable. As for Pavia University, we set the number of iterations from 0 to 300. As shown in Fig. 10 (c), before the number of iterations is 100, the accuracy fluctuates greatly. When it is close to 150 times, it tends to stabilize and reach a larger value, and the loss tends to be 0. Combining the above results, to get the best accuracy
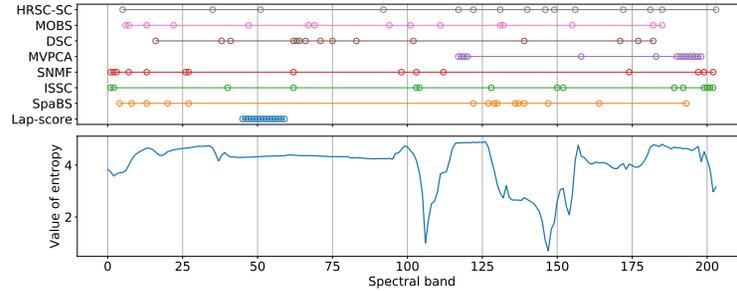
Fig. 8. The best 15 bands of Salinas-A data set selected by different BS methods (above) and the entropy value of each band (below).
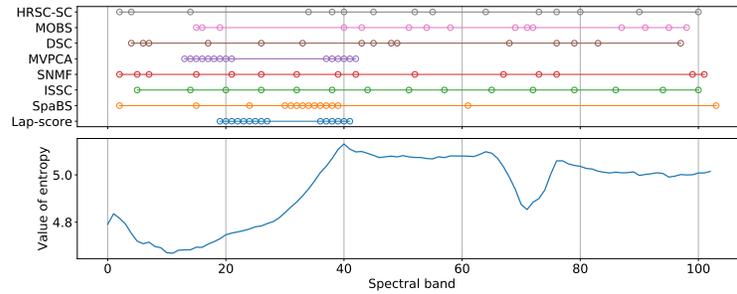

Fig. 9. The best 15 bands of Pavia University data set selected by different BS methods (above) and the entropy value of each band (below).

and cost minimal time, the number of the iterations for Indian Pines, Salinas-A, and Pavia University data sets are set to 100, 50, and 150, respectively.

4.5. Visualization of Affinity Matrix

The affinity matrices of three data set is shown in Fig. 11. According to the theory of the affinity matrix in section 3, the affinity matrix has a symmetrical and sparse block diagonal structure. The block of the affinity matrix can be used to cluster, which is the crucial element for subspace clustering. As shown in Fig. 11, the block is independent relatively and the affinity matrix obtained from the three data sets has good symmetry so that the affinity matrices achieve robust performance.

5.   Conclusions

In this paper, a hyper-graph regularized subspace clustering with skip connections (HRSC-SC) was proposed for band selection of the hyperspectral image. The networks combine subspace clustering into the convolutional auto-encoder by thinking of it as a self-expressive layer. Moreover, the symmetrical skip connections are added to the networks

Table 5. The best 15 bands of three data sets (Indian Pines, Salinas-A, Pavia University) selected by different band selection methods.

Data sets	Methods	Selected bands
Indian Pines	Lap-score	[45,46,47,48,49,50,51,52,53,54,55,56,57,58,59]
	SpaBS	[4,8,13,20,27,122,127,129,130,136,137,139,147,164,193]
	ISSC	[1,2,40,62,103,104,128,150,152,189,192,199,200,201,202]
	SNMF	[1,2,3,7,13,26,27,62,98,103,112,174,197,199,202]
	MVPCA	[117,118,119,120,158,183,190,191,192,193,194,195,196,197,198]
	DSCBS	[16,38,41,62,63,64,66,71,75,83,102,139,171,177,182]
	MOBS	[6,7,13,22,47,67,69,94,101,111,131,132,155,182,185]
	HRSC-SC	[5,35,51,92,117,122,131,140,146,149,156,172,181,185,203]
Salinas-A	Lap-score	[45,46,47,48,49,50,51,52,53,54,55,56,57,58,59]
	SpaBS	[4,8,13,20,27,122,127,129,130,136,137,139,147,164,193]
	ISSC	[1,2,40,62,103,104,128,150,152,189,192,199,200,201,202]
	SNMF	[1,2,3,7,13,26,27,62,98,103,112,174,197,199,202]
	MVPCA	[117,118,119,120,158,183,190,191,192,193,194,195,196,197,198]
	DSCBS	[16,38,41,62,63,64,66,71,75,83,102,139,171,177,182]
	MOBS	[6,7,13,22,47,67,69,94,101,111,131,132,155,182,185]
	HRSC-SC	[5,35,51,92,117,122,131,140,146,149,156,172,181,185,203]
Pavia University	Lap-score	[19,20,21,22,23,24,25,26,27,36,37,38,39,40,41]
	SpaBS	[2,15,24,30,31,32,33,34,35,36,37,38,39,61,103]
	ISSC	[5,14,20,26,32,38,44,51,57,65,72,79,86,94,100]
	SNMF	[2,5,7,15,21,26,32,39,42,52,67,73,76,99,101]
	MVPCA	[13,14,15,16,17,18,19,20,21,37,38,39,40,41,42]
	DSCBS	[4,6,7,17,26,33,43,45,48,49,68,76,79,83,97]
	MOBS	[15,16,19,40,43,51,54,58,69,71,72,87,91,95,98]
	HRSC-SC	[2,4,14,34,38,40,45,52,55,64,73,76,80,90,100]

to pass image details from encoder to decoder, which can make full use of the historical feature maps obtained from the networks and tackle the problem of gradient vanishing caused by multiple nonlinear transformations. Besides, we introduce the hyper-graph regularized to consider the manifold structure reflecting geometric information within data to accurately describe the multivariate relationship between data points and make the results of HSIs clustering more accurate. We execute the experiments on three HSI data sets for the proposed HRSC-SC method to show that the proposed method has state-of-the-art performance.

However, there are still the following problems that need to be improved in future work. First, due to the training characteristics of deep learning, the running time of the proposed HRSC-SC method spends much, which is an important topic to research in future works. Second, the proposed HRSC-SC method shows good performance for the band selection of HSIs, which can be used to study the clustering of HSIs in the future. Third, due to the limitation of training time and equipment, we only intercepted part of the three datasets for experiments. In future work, we will try to use the entire dataset for testing to verify the effectiveness of the proposed algorithm. Finally, a self-supervised network structure is used for learning in this paper, in future work, other methods, such as generative adversarial networks, will be introduced to optimize the network structure.
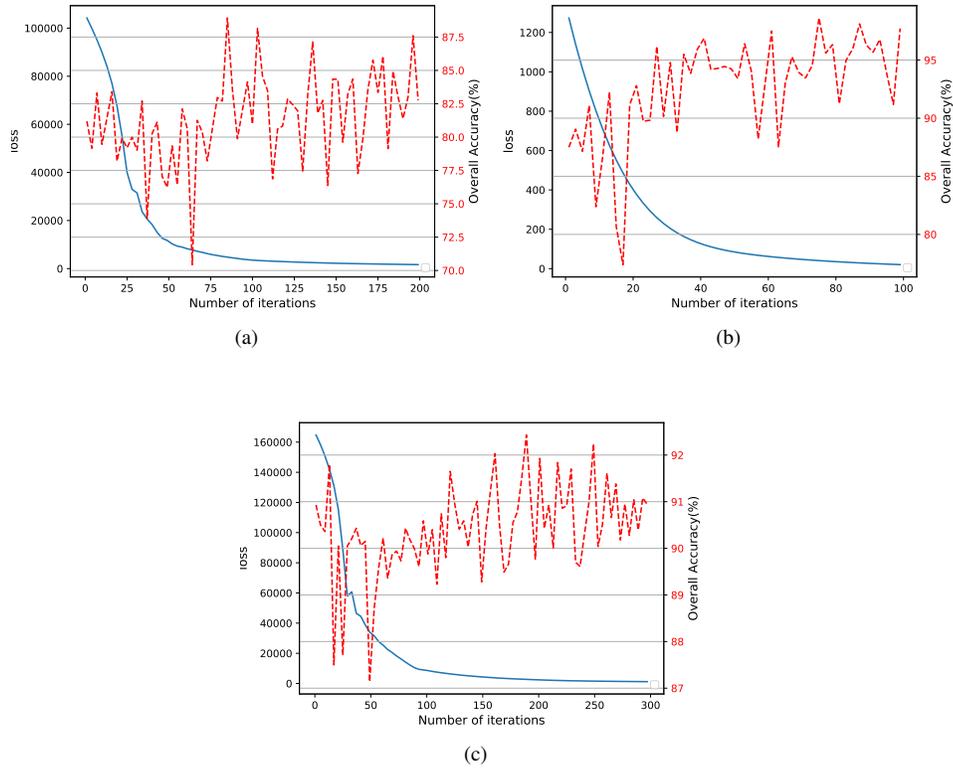
Fig. 10. Loss curve (blue full line) and accuracy curve (red dashed line) for the proposed HRSC-SC approach on three data sets. (a) Indian Pines, (b) Salinas-A, and (c) Pavia University.
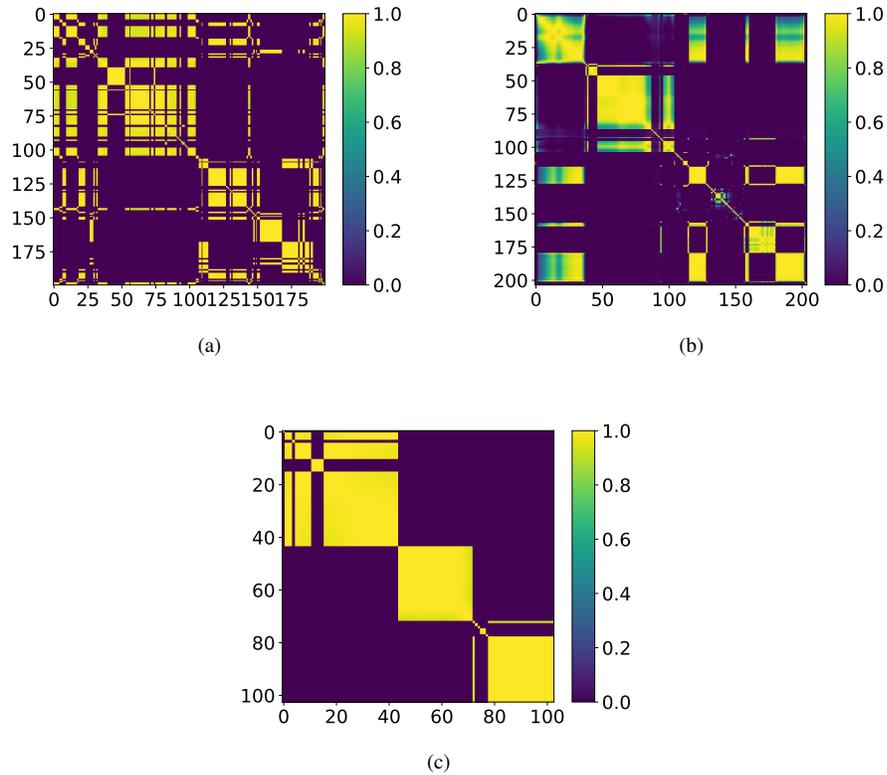
(a)

(b)

(c)

Fig. 11. Visualization of Affinity Matrix on three data sets. (a) Indian Pines, (b) Salinas-A, and (c) Pavia University.

References

1. Mateus Habermann, Vincent Fremont, and Elcio Shiguemori. Unsupervised hyperspectral band selection using clustering and single-layer neural network. *Revue Francaise de Photogramme-trie et de Teledetection*, 2018, 09 2018.
2. P. Hu, X. Liu, Y. Cai, and Z. Cai. Band selection of hyperspectral images using multiobjective optimization-based sparse self-representation. *IEEE Geoscience and Remote Sensing Letters*, 16(3):452–456, 2019.
3. Chein-I Chang, Qian Du, Tzu-Lung Sun, and M. L. G. Althouse. 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, 1999.
4. K. Sun, X. Geng, and L. Ji. A new sparsity-based band selection method for target detection of hyperspectral image. *IEEE Geoscience and Remote Sensing Letters*, 12(2):329–333, 2015.
5. Xiaofei He, Deng Cai, and Partha Niyogi. Laplacian score for feature selection. In *NIPS*, 2005.
6. Ji-ming Li and Yun-tao Qian. Clustering-based hyperspectral band selection using sparse non-negative matrix factorization. *Journal of Zhejiang University - Science C*, 12:542–549, 07 2011.

7. W. Sun, L. Zhang, B. Du, W. Li, and Y. Mark Lai. 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, 2015.

8. M. Zeng, Y. Cai, Z. Cai, X. Liu, P. Hu, and J. Ku. Unsupervised hyperspectral image band selection based on deep subspace clustering. *IEEE Geoscience and Remote Sensing Letters*, 16(12):1889–1893, 2019.

9. Pan Ji, Tong Zhang, Hongdong Li, Mathieu Salzmann, and Ian Reid. Deep subspace clustering networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

10. M. Zeng, Y. Cai, X. Liu, Z. Cai, and X. Li. Spectral-spatial clustering of hyperspectral image based on laplacian regularized deep subspace clustering. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019.

11. Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image denoising using very deep fully convolutional encoder-decoder networks with symmetric skip connections. *CoRR*, abs/1603.09056, 2016.

12. Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 585–591. MIT Press, 2002.

13. D. Cai, X. He, J. Han, and T. S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1548–1560, 2011.

14. X. Ma, W. Liu, S. Li, D. Tao, and Y. Zhou. Hypergraph p-laplacian regularization for remotely sensed image recognition. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3):1585–1595, 2019.

15. Y. Dai, B. Xu, S. Yan, and J. Xu. Study of cardiac arrhythmia classification based on convolutional neural network. *Computer Science and Information Systems*, 17(2):445–458, 2020.

16. Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Adv. Neural Inf. Process. Syst*, 14, 04 2002.

17. Jianbo Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

18. E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.

19. J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Püschel. D-admm: A communication-efficient distributed algorithm for separable optimization. *IEEE Transactions on Signal Processing*, 61(10):2718–2723, 2013.

20. Lu Canyi, Jiashi Feng, Zhouchen Lin, Tao Mei, and Shuicheng Yan. Subspace clustering by block diagonal representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP:1–1, 01 2018.

21. Qi Liu, Miltiadis Allamanis, Marc Brockschmidt, and Alexander Gaunt. Constrained graph variational autoencoders for molecule design. 05 2018.

22. J. A. Dabin, A. M. Haimovich, J. Mauger, and A. Dong. Blind source separation with l1 regularized sparse autoencoder. In *2020 29th Wireless and Optical Communications Conference (WOCC)*, pages 1–5, 2020.

23. G. Spigler. Denoising autoencoders for overgeneralization in neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):998–1004, 2020.

24. Pan Ji, M. Salzmann, and Hongdong Li. Efficient dense subspace clustering. In *IEEE Winter Conference on Applications of Computer Vision*, pages 461–468, 2014.

25. Xiao Jiao Mao, Chunhua Shen, and Yu Bin Yang. Image denoising using very deep fully convolutional encoder-decoder networks with symmetric skip connections. *CoRR*, abs/1603.09056, 2016.

26. F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao. Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image. *IEEE Transactions on Cybernetics*, 49(7):2406–2419, 2019.
27. G. Cheng and X. Tong. Fuzzy clustering multiple kernel support vector machine. In *2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, pages 7–12, 2018.
28. S. Rogic and L. Kascelan. Class balancing in customer segments classification using support vector machine rule extraction and ensemble learning. *Computer Science and Information Systems*, 18(3):893–925, 2021.

Meng Zeng received the B.Sc. from the Hubei University of Arts and Science, Xiangyang, China, in 2017 and received the M.S. degree in computer science with the China University of Geosciences, Wuhan, China, in 2020. She is currently pursuing the Ph. D. degree in School of Computer Science and Engineering, Central South University, Changsha, China. Her current research interests include machine learning, hyperspectral image processing, data mining and storage.

Bin Ning received the Master degree in software engineering from Beijing University of Technology, Beijing, China, in 2005. He is currently a Professor with the School of Computer Engineering, Hubei University of Arts and Science, Xiangyang, China. His research interests include software engineering, data mining, and algorithm design. He is a member of the China Computer Federation.

Qiong Gu received the M.S. degree in Computer science and technology from China University of Geosciences, in 2006,and the Ph.D. degree in Geosciences Information Engineerin from China University of Geosciences, Wuhan, China, in 2009. She is a member of the China Computer Federation. She is currently a professor with the School of Computer engineering, Hubei University of Arts and Science, She is particularly interested in Data mining, Machine Learning, she has extended her interests include Net-Mediated Public Sentiment, and internet of things.

Chunyang Hu received the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2011. He is currently an associate professor with the school of computer engineering, Hubei University of Arts and Science, Xiang yang, China. He served as a visiting scholar at University of Massachusetts Lowell in 2017. His research interests include big data processing, machine learning, software defined network, and block chain. Dr. Hu is a member of China Computer Federation.

Shuijia Li received the B.Sc. from the Hubei University of Arts and Science, Xiangyang, China, in 2017. He is currently pursuing the Ph. D. degree in computer science with the China University of Geosciences, Wuhan, China. His current research interests include evolutionary computation, computational intelligence and its application.