MULTI-SCALE FUSION MAXIMUM ENTROPY SUBSPACE CLUSTERING FOR HYPERSPECTRAL BAND SELECTION

Haipeng Ma, Yulei Wang, Liru Jiang, Meiping Song, Chunyan Yu and Enyu Zhao

1. Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian 116026, China
2. Department of ground-air navigation, Air Force Communication Officer School, Dalian, China

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9884654

ABSTRACT

A novel multi-scale fusion maximum entropy subspace clustering (MFMESC) for hyperspectral image (HSI) band selection is proposed in this paper. Subspace clustering is combined as a self-expression layer with stacked convolutional autoencoder, so that subspace clustering working in linear subspaces can deal with complicated HSI data with nonlinear characteristics. Multiple fully-connected linear layers are inserted between the encoder layers and their corresponding decoder layers to promote learning more favorable representations for subspace clustering. A multi-scale fusion module is designed to guide the fusion of multi-scale information extracted from different layers to learn a more discriminative self-expression coefficient matrix. Furthermore, the maximum entropy regularization is introduced in the subspace clustering to promote the connectivity within each subspace. Experimental results demonstrate the superiority of the proposed model against state of-the-art methods.

Index Terms— hyperspectral band selection, maximum entropy regularization, subspace clustering, multi-scale fusion, stacked convolutional autoencoder

1. INTRODUCTION

Hyperspectral image (HSI) captures the spectral and spatial information of the target scene in hundreds of narrow and continuous spectral bands, thus providing an enormous amount of information about the region of interest. HSI has achieved great success in various application fields such as environmental detection and medical diagnosis. However, the high dimensionality and strong correlation of bands have brought the problems of data redundancy, heavy burden of computation and storage, and the curse of dimensionality. Therefore, dimensionality reduction has become an important technology in hyperspectral image processing [1].

There are two main measures for HSI dimensionality reduction: feature extraction and band selection (BS). Band selection is to select the most representative subset of bands from the original band set. Compared with feature extraction, band selection maintains the physical significance of the data, which makes it a research hotspot in the field of HSI processing.

According to the availability of label information, band selection (BS) methods can be categorized into supervised, unsupervised and semi-supervised BS methods. Considering the high cost of labelling data and the difficulty of obtaining sufficient labels, unsupervised BS is more flexible and applicable for HSI without using any prior knowledge. As a popular technology, self-representation uses the self-expressive property of data and various regularization terms to fulfill unsupervised BS, which has attracted much attention. Self-representation-based subspace clustering (SSC) models have made remarkable achievements in unsupervised BS [2], [3].

However, many BS methods based on subspace clustering only consider linear subspace, which is not suitable for the typical nonlinear structure of HSI, making it unable to achieve excellent performance. In addition, the subspace clustering-based BS methods commonly ignore the spatial information of HSI, which is also important for further process.

In recent years, deep learning-based models have been widely applied in band selection. Zeng et al. [4] proposed deep subspace clustering (DSC) for HSI band selection, and improved the SSC by embedding the self-representation into the deep convolutional autoencoder to learn the nonlinear spectral-spatial relationship. However, DSC model does not consider the low-level and high-level information of the input data to obtain more favorable representations for subspace clustering, and ignores the important multi-scale information embedded in deep autoencoder. Besides, the adopted regularization terms by DSC model ignore the connectivity within each subspace, which compromises the subsequent spectral clustering to some extent.

In this paper, a multi-scale fusion maximum entropy subspace clustering (MFMESC) is proposed for HSI band selection. Different from the existing DSC method, to learn more informative and discriminative subspace clustering representations, multiple fully-connected linear layers are inserted between the encoder layers and their corresponding decoder layers to generate multiple sets of self-expressive

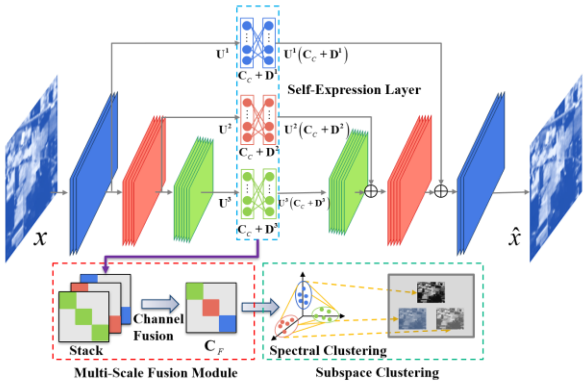


Fig. 1. Flowchart of proposed band selection method

and informative representations at different levels of the encoder. The multi-scale fusion module is introduced to fuse the multi-scale information extracted from different layers in stacked convolutional autoencoder to learn a more discriminative self-expression coefficient matrix, which is achieved by stacking the coefficient matrix extracted from different self-expression layers and then applying a convolutional kernel on the stacked coefficient matrix to fuse its channel. In addition, the maximum entropy regularization is introduced to strengthen the connectivity within each subspace, in which its elements corresponding to the same subspace are uniformly and densely distributed, benefiting the subsequent spectral clustering.

2. METHOD

In this section, the proposed MFMESC method for hyperspectral band selection is described. As shown in Fig. 1, the proposed MFMESC method is divided into four parts: feature extraction, self-expression, multi-scale fusion and subspace clustering. The feature extraction step extracts the inherent spatial information of HSI data through stacked convolutional autoencoder. The self-expression step embeds the self-expression model into stacked convolutional autoencoder to imitate the "self-expression" property. Multi-scale fusion is framed to generate the coefficient matrix of each convolutional layer in stacked convolutional autoencoder and then fuses them with the convolutional kernel. The subspace clustering selects band subset via spectral clustering.

2.1. Deep subspace clustering for band selection

Given HSI samples $\mathrm{X}=\left\{x_i \middle| x_i \in \mathbb{R}^{w\times h}\right\}_{i=1}^{N}$, where $w\times h$ and N denote the number of pixels and spectral bands, respectively. The goal of BS is to select a band subset with the size of n ($n<N$), which contains the maximum information with minimum redundancy.

The convolutional autoencoder is used in DSC model. The encoder is denoted as $u=E\left(x;\theta_e\right)$, where u is latent representation. The decoder can be defined as $\hat{x}=D\left(u;\theta_d\right)$. In addition, DSC model embeds self-expression model into convolutional autoencoder to achieve self-expression property. The latent representation u is first unfolded into a d-dimensional vector z. Besides, assuming that N band images are located in a union of n affine subspace S, where $S=S_1\cup S_2\cup\ldots\cup S_n$ denote n subspaces with dimensions $d_1,d_2,\ldots,d_n$ in the full space $\mathbb{R}^n$, and satisfy with $d=\sum_{i=1}^{n} d_i$. Therefore, the complete cost function of DSC model is expressed as:

$$\mathcal{L}(\mathrm{C})=\frac{1}{2}\left\|\mathrm{X}-\hat{\mathrm{X}}\right\|_2^2+\frac{\alpha}{2}\left\|\mathrm{U}-\mathrm{UC}\right\|_2^2+\frac{\lambda}{2}\left\|\mathrm{C}\right\|_2^2,\ \text{s.t. } \operatorname{diag}(\mathrm{C})=0 \quad (1)$$

where, $\left\|\mathrm{U}-\mathrm{UC}\right\|_2^2$ is a self-expression term. $\left\|\mathrm{C}\right\|_2^2$ is regarded as an additional network layer using back propagation, using l_2-norm regularization. X is the tensor form of the input band images, $\hat{\mathrm{X}}$ is the reconstructed band images. $\mathrm{U}=\left[\mu_1,\mu_2,\ldots,\mu_N\right]^T$ is the latent matrix. $\mathrm{C}\in\mathbb{R}^{N\times N}$ is the coefficient matrix. α and λ are two balancing coefficients.

2.2. Multi-scale fusion module

Inspired by providing increasingly complex input data representations based on different layers of the encoder, the feature learning process can be promoted by learning the low-level and high-level information of the input data through multi-level representation learning, so as to generate multiple sets of representations that satisfy the self-expressiveness property [5]. $\mathrm{C}_C\in\mathbb{R}^{N\times N}$ is defined as the consistency matrix to capture the relational information between the encoder layers and $\left\{\mathrm{D}^l\right\}_{l=1}^{L}\in\mathbb{R}^{N\times N}$ as distinctive matrices to produce the unique information of the individual layers. Promote learning of self-expressive representations through the following loss function

$$\mathcal{L}_{exp}=\sum_{l=1}^{L}\left\|\mathrm{U}^l-\mathrm{U}^l\left(\mathrm{C}_C+\mathrm{D}^l\right)\right\|_2^2 \quad (2)$$

The self-expression loss $\mathcal{L}_{exp}$ is used to promote learning self-expressive feature representations at different levels of the encoder. For the distinctive matrices, Frobenius norm is used to ensure the connectivity of the affinity graph associated with each fully connected layer. For the consistency matrix, l_1-norm is employed to generate a sparse representation of the data. Therefore, we add the following regular terms

$$\mathcal{L}_{C_C}=\left\|\mathrm{C}_C\right\|_1 \quad (3)$$

$$\mathcal{L}_D=\sum_{l=1}^{L}\left\|\mathrm{D}^l\right\|_F^2 \quad (4)$$

The multi-scale fusion module is used to fuse the consistency matrix and distinctive matrices of each convolution layer in the encoder. Stacked matrix C_S is obtained by stacking the consistency matrix and discrimination matrix along the channel dimension. A convolutional kernel k is used to integrate the channels of C_S, $\mathrm{C}_F \in \mathbb{R}^{N\times N} = k \otimes \mathrm{C}_S$, where $\otimes$ means convolution operation [6]. The multi-scale fusion module obtains a more distinctive self-expression coefficient matrix C_F.

2.3. Maximum entropy regularization

The element of C_F denoted by $C_{F_{i,j}}$ could be thought of as the similarity degree between data samples i and j. By applying maximum entropy regularization on the self-expression coefficient matrix, as well as the fact $\max H(\mathrm{C}) = \min - H(\mathrm{C})$, the loss function driving the learning of the self-expression coefficient matrix can be written as

$$\mathcal{L}_{C_F} = \sum_{i=1}^{N}\sum_{j=1}^{N} C_{F_{i,j}} \ln C_{F_{i,j}}, \text{ s.t. } C_{F_{i,j}} \geq 0 \tag{5}$$

When $C_{F_{i,j}} = 0$, the value of the corresponding summand $C_{F_{i,j}} \ln C_{F_{i,j}}$ is 0. Intuitively, when a sample is represented only by samples from the same subspace, minimizing Eq. (5) will force the connection strengths between samples belonging to the same subspace to be equal. Therefore, the previous constraint $\mathrm{diag}(\mathrm{C}) = 0$ is not required.

2.4. Multi-scale fusion maximum entropy subspace clustering for band selection

The overall cost function of MFMESC model is written as

$$\begin{aligned}\mathcal{L}(\mathrm{C}_F) = \frac{1}{2}\left\|\mathrm{X} - \hat{\mathrm{X}}\right\|_2^2 + \frac{\alpha_1}{2}\sum_{l=1}^{L}\left\|\mathrm{U}^l - \mathrm{U}^l\left(\mathrm{C}_C + \mathrm{D}^l\right)\right\|_2^2 + \lambda_1 \left\|\mathrm{C}_C\right\|_1 \\ + \lambda_2 \sum_{l=1}^{L}\left\|\mathrm{D}^l\right\|_F^2 + \lambda_3 \sum_{i=1}^{N}\sum_{j=1}^{N} C_{F_{i,j}} \ln C_{F_{i,j}}, \text{ s.t. } C_{F_{i,j}} \geq 0\end{aligned} \tag{6}$$

where $\alpha_1, \lambda_1, \lambda_2, \lambda_3 > 0$ are hyperparameters to balance the contribution of different losses. Adam gradient method is used to train the network. Standard back propagation is used to update parameters. Once C_F is obtained, we can create a symmetric affinity matrix W in the following form

$$\mathrm{W} = \frac{1}{2}\left(\left|\mathrm{C}_F\right| + \left|\mathrm{C}_F\right|^T\right) \tag{7}$$

which shows the pairwise relations between bands. According to the affinity matrix, we use spectral clustering to get the clustering results that segments all spectral bands into n clusters. Based on the clustering results, the average band in each class is used as the cluster center. Calculating the distance between the cluster center and each band. The selected band subset can be further determined by selecting these bands closest to their cluster centers.

3. EXPERIMENTS AND RESULTS

In this section, to demonstrate the effectiveness of our proposed method, we conduct experiments on real HSI data set and compare with existing band selection algorithms.

3.1. Data description

The Indian Pines data set was taken from the Multispectral Image Data Analysis System group at Purdue University. It was acquired by the AVIRIS sensor from JPL to record a scene from Northwest Indiana on June 12, 1992, covering an area of 6 miles west of West Lafayette, Indiana. It mainly consists of 220 spectral bands with the size of 145 × 145 pixels from 0.4 to 2.5 μm. The data has 20 m spatial resolutions and 10 nm spectral resolutions. There are 16 different classes of land cover objects of interest and 10249 pixels are labeled in the scene. We remove 20 spectral bands in 104–108, 150–163 and 200 with heavy noises due to water absorption, and finally get 200 bands in the experiment.

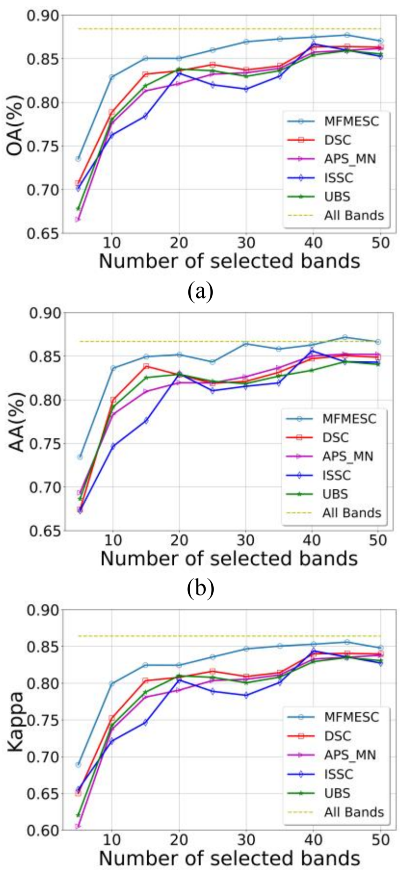
3.2. Experimental setup

To verify the superiority of the proposed MFMESC, we consider four other unsupervised hyperspectral band selection methods for comparison. The comparison methods are as follows: uniform band selection (UBS) [1], improved sparse spectral clustering (ISSC) [2], adaptive subspace partition strategy with minimum noise (ASPS_MN) [3] and deep subspace clustering (DSC) [4]. Support vector machine (SVM) with radial basis function kernel as the classifier is selected to prove the performance of different band selection methods. The parameters are selected via cross-validation. We utilize overall accuracy (OA), average accuracy (AA) and Kappa coefficient (Kappa) as quantitative assessments. In order to make the experiment more effective, we also add all bands to the experiment for comparison. During the experiment, 10% of labeled samples per class are randomly selected as training data and the rest are employed for testing. To ensure the fairness of random sampling, we repeat all experiments ten times and average results are reported.

We have employed an autoencoder model consisting of three stacked convolutional encoder layers with 10, 20 and 30 filters of sizes 5 × 5, 3 × 3, and 3 × 3, respectively. The parameters used in the experiments are as follows: $\alpha_1 = 1.0\times10^{-1}$, $\lambda_1 = 1.0\times10^{-2}$, $\lambda_2 = 1.0\times10^{-2}$, $\lambda_3 = 1.0\times10^{-2}$, the learning rate is set to 1.0×10^{-4}. For multi-scale fusion module, k is set to the convolutional kernel with 3 × 3 size.

3.3. Analysis of classification results

To evaluate the performance of MFMESC, we vary the number of selected bands n in the range of [5, 50] with a step size of 5, and compared the mean values of OA, AA, and Kappa on the Indian Pines dataset with competitors, as shown in Fig. 2. We also report the classification results when using all the spectral bands for reference. It can be clearly shown that the OA, AA and kappa of the proposed MFMESC method outperforms that of the other four methods with different numbers of selected bands. With the increase of n, the OA, AA and Kappa curves of MFMESC method are gradually rise and approach the whole band curve, and the performance becomes saturated when n is large. Only when $n = 45$, the value of AA is higher than the whole band. To analyze the effect of band selection method more intuitively, we also compared the performance of OA, AA and Kappa on Indian Pines using 30 bands, as shown in Table 1. Compared with other algorithms, MFMESC has higher classification accuracy.



(a) (b) (c)

Fig. 2. Classification results: OA, AA, and Kappa curves of Indian Pines by SVM classifier, respectively. (a) OA. (b) AA. (c) Kappa.

Table 1. Performance comparison of different methods in 30 bands of Indian Pines dataset

	OA	AA	Kappa
UBS	82.98±1.37%	81.86±2.67%	80.06±1.58%
ISSC	81.52±1.15%	81.55±3.80%	78.35±1.36%
APS_MN	83.39±1.27%	82.65±2.18%	80.53±1.49%
DSC	83.72±1.51%	82.11±3.97%	80.91±1.80%
MFMESC	**86.95±1.09%**	**86.43±2.87%**	**84.68±1.27%**

4. CONCLUSION

In this paper, a novel band selection method based on multi-scale fusion maximum entropy subspace clustering is proposed. The basic idea is to combine the subspace clustering as a self-expression layer into the stacked convolutional autoencoder, enabling it be trained end to end. To learn more informative representations for subspace clustering, the input HSI data is transformed into multi-level representations on the union of subspaces by leveraging information at different levels of the encoder. A multi-scale fusion module is devised to learn a more discriminative self-expression coefficient matrix. Maximum entropy regularizer is employed to strengthen the connectivity within each subspace, in which its elements corresponding to the same subspace are uniformly and densely distributed, so as to select a more discriminative bands subset through spectral clustering. Experiments on a benchmark dataset demonstrate that our method outperforms than other state-of-the-art band selection methods in classification accuracies.

5. ACKNOWLEDGEMENTS

This work is supported by the National Nature Science Foundation of China (61801075), China Postdoctoral Science Foundation (2020M670723) and the Fundamental Research Funds for the Central Universities (3132022232).

REFERENCES

[1] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.

[2] W. Sun, L. Zhang, B. Du, W. Li and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J.* Sel. *Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784-2797, June 2015.

[3] Q. Wang, Q. Li and X. Li, "Hyperspectral Band Selection via Adaptive Subspace Partition Strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940-4950, Dec. 2019.

[4] M. Zeng, Y. Cai, Z. Cai, X. Liu, P. Hu and J. Ku, "Unsupervised Hyperspectral Image Band Selection Based on Deep Subspace Clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1889-1893, Dec. 2019.

[5] M. Kheirandishfard, F. Zohrizadeh and F. Kamangar, "Multi-Level Representation Learning for Deep Subspace Clustering," 2020 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 2028-2037.

[6] Z. Dang, C. Deng, X. Yang and H. Huang, "Multi-Scale Fusion Subspace Clustering Using Similarity Constraint," 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 6657-6666.

[7] Z. Peng, Y. Jia, H. Liu, J. Hou and Q. Zhang, "Maximum Entropy Subspace Clustering Network," *IEEE Transactions on Circuits and Systems for Video Technology* (2021).