HYPERSPECTRAL BAND SELECTION BASED ON TERNARY WEIGHT CONVOLUTIONAL NEURAL NETWORK

Jie Feng, Di Li, Jiantong Chen, Xiangrong Zhang, Xu Tang, Xiande Wu

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education,
Xidian University, Xi'an 710071, China

ABSTRACT

In this paper, a novel ternary weight convolution neural network (TWCNN) is proposed for band selection of hyperspectral images. TWCNN constructs deep-wise convolution layer with 1×1 filters as the first layer of the network, which is used for band selection. In the deep-wise convolution layer, weights are constrained to -1, 0, or 1. -1 and 1 represent that the corresponding band is selected, while 0 indicates it's not. TWCNN constructs subsequent layers to extract features and classify for selected spectral bands. It combines band selection, feature extraction and classification into a unified optimization procedure, which makes it to achieve end-to-end band selection and classification. Furthermore, the constraints of the number of spectral bands is added to the cost function of TWCNN. The specific number of spectral bands can be selected. The experiment results show that the proposed model provides a competitive result to state-of-the-art methods.

Index Terms— hyperspectral images, band selection, ternary weight network, convolutional neural network, deep learning, classification

1. INTRODUCTION

The high spectral resolution of hyperspectral images (HSI) provides abundant information of land cover, which makes hyperspectral remote sensing have great potential applications in various fields. Although the high spectral resolution enriches the feature information, it also leads to data redundancy and "Hughes phenomenon". Data redundancy refers to the strong correlation among the spectral bands of HSI, especially the adjacent bands. Redundant information increases the computational cost of HSI processing. Hughes phenomenon, caused by high-dimensional spectral bands and limited training samples of HSI, deteriorates the classification performance of HSI. Therefore, dimensionality reduction of HSI is an important step in hyperspectral image processing.

There are two types of methods to reduce the dimensionality of HSI: feature extraction and feature selection. Feature selection is to select the most discriminative band subset from original spectral bands of HSI. Compared with feature extraction, feature selection observes the physical information of original spectral bands.

Feature selection algorithms can be divided into three categories: filter, wrapper and embedded. Filter-based methods, such as minimum-redundancy maximum-relevancy (mRMR) [1] and mutual information and clonal selection (MI-CSA) [2], consider an evaluation criterion independent of the classifier to search the band subset. The execution efficiency of filter-based methods is high, but the performance of these methods is limited. Wrapper-based methods take the performance of the classifier as the criterion to measure the quality of the feature subset directly, which take a huge computational cost. Embedded-based methods integrate the process of feature selection and training of the model, which means the feature selection step is completed automatically in the process of model training.

Recent years, deep learning-based models have made a great success in HSI classification, and it was quickly applied in the field of band selection. In [3], a deep feature selection model with a sparse regularized loss function was proposed. It is called elastic net. The main idea of elastic net is to add a sparse one-to-one linear layer between the input layer and the first hidden layer of multilayer perceptrons. Compared with shallow network based methods such as LOASS regression [4] and [5], deep networks extract nonlinear and abstract high-level features, and can also be directly applied to multi-classification tasks. Ying Zhan et al. proposed a convolutional neural network (CNN) based wrapper approach for HSI band selection, which is named as BSCNN+ [6]. Random search strategy based on distance density is used to search for band subset

This work was supported by the National Natural Science Foundation of China under Grant 61871306, by the National Natural Science Foundation of Shaanxi Province under Grant 2019JM-194, by the Fundamental Research Funds for the Central Universities under Grant JBX181707, by the Joint Fund of Equipment Research of Ministry of Education under Grant 6141A020333, by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, under Grant LSIT201803D.

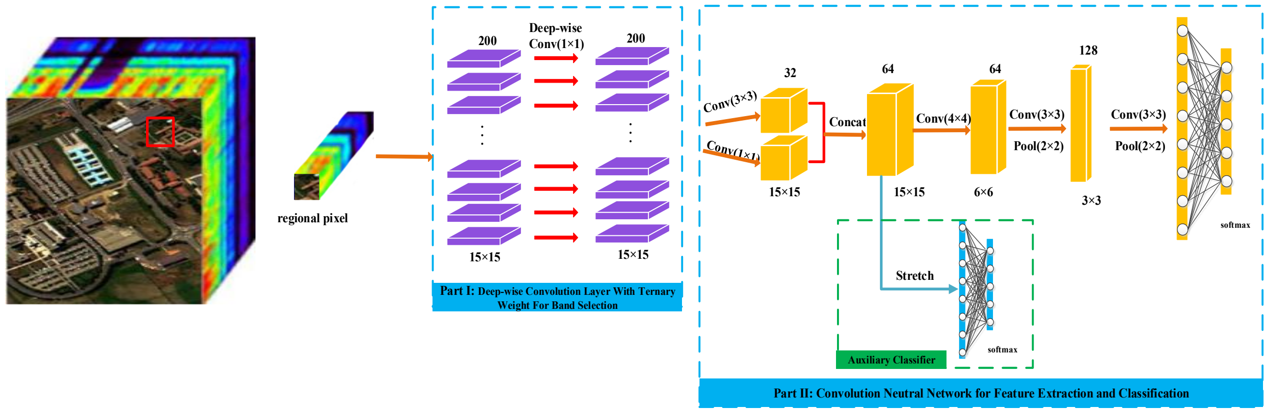

Fig. 1 Flowchart of the proposed TWCNN method

in BSCNN+, then the 1D-CNN trained by full-band data is used to find the discriminative band combinations for HSI classification.

In this paper, an embedded method based on ternary weight CNN (TWCNN) is proposed for HSI band selection. In TWCNN, ternary weight network is first applied in feature selection. Ternary weight network selects discriminative band subset via depth-wise convolution with ternary weights. To control the number of selected bands, a constraint term is added into the cost function in addition to the cost of classification errors. TWCNN completes band selection, feature extraction and classification in the process of network training. The proposed methods are applied on a well-known hyperspectral dataset, and the result shows that our method obtains more satisfactory results on HSI band selection.

2. THE PROPOSED TWCNN METHOD

In this section, the proposed TWCNN for hyperspectral band selection is described. As shown in Fig. 1, the proposed TWCNN method is divided into two parts: band selection part and classification part. The band selection part selects spectral bands via depth-wise convolution, where weights are ternarized. Based on selected spectral bands, the next part of the network is constituted by CNN with an auxiliary classifier.

2.1. Band selection via depth-wise convolution

In the band selection part, the first layer of TWCNN is a one-to-one layer which is named as deep-wise convolution layer [7]. To preserve spatial information, a pixel region with size W×H is set as the input of TWCNN network. Different to regular convolution layer, a convolution kernel of depth-wise convolution corresponds to one channel, and one channel is convoluted by only one convolution kernel. As is shown in fig. 2, the number of output channels is equal to the number of input channels after deep-wise convolution operation. Here, the kernel size of each convolution operation is 1×1 without padding, which means the output size of the first layer is same to the input size. So, for each spectral band $B_i \in \mathbb{R}^{W\times H\times 1}, i \in \{1,2,...,C\}$,where C is the total number of spectral bands, there is only one parameter $w_i \in \mathbb{R}^{1\times 1}, i \in \{1,2,...,C\}$. Deep-wise convolution operation with 1×1 filter is equivalent to multiplying a weight by the input for each band.

We constrain the weights W_i to -1, 0, or 1, where -1 and 1 represent that the corresponding band is selected, while 0 indicates it's not. Specifically, the weights of the first layer are triangulated by a threshold-based ternary function as follows:

$$W_i^t = f_t(W_i \mid \Delta) = \begin{cases} +1, & \text{if } W_i > \Delta \\ 0, & \text{if } |W_i| < \Delta \\ -1, & \text{if } W_i > -\Delta \end{cases} \qquad (1)$$

where $\Delta \in [0,1]$ is an positive threshold parameter.

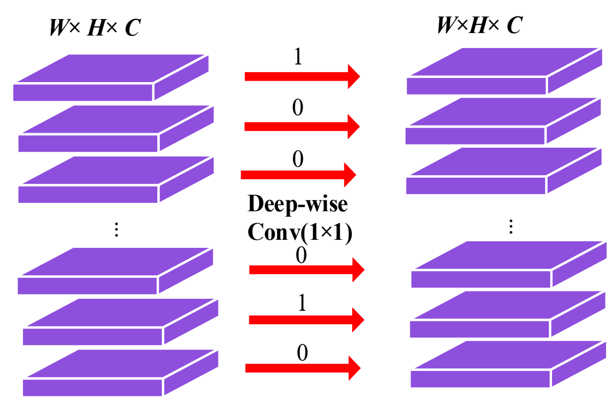

Fig. 2 Deep-wise convolution. W and H indicates the width and height of the input block; C is the total number of bands. Deep-wise convolution is a one-to-one operation, the number 0,1 or -1 on the red arrow shows the weight of every deep-wise convolution layer with 1×1 filter size.

2.2. Feature extraction and classification via convolutional neural network

The rest of the network is shown as part II in Fig. 1. In this part, we proposed a CNN to extract joint spectral–spatial features of HSI. The first layer of this part is similar to the Inception module [8], which consists of 1 × 1 and 3 × 3 convolutional filters. The 1 × 1 convolution aims at extract spectral information and 3 × 3 aims at extract the spatial information of HSI. After 1 × 1 and 3 × 3 convolution operation, output of these convolutions are cascaded. The rest of the network is stacked by several convolutional layers, maxpool layers and fully connected layers. Finally, the extracted deep features are fed into a fully connected layer with softmax activation. Batch normalization, dropout and *l*-2 norm are used in the network to speed up calculation and mitigate overfitting. Inspired by GoogleNet [8], we add an auxiliary classifier into the middle layer of the network to overcome the network gradient dispersion, The Auxiliary classifier is used only during the training phase but not in testing step.

2.3. Cost function of TWCNN

In TWCNN, a new cost function is defined to combine the band selection and classification. Here, a mini-batch update strategy is adopted, and cross entropy function C_O is used as the cost function to measure the classification error:

$$C_O = -\frac{1}{m}\sum_{i=1}^{m}[x_i \log(z_i) + (1-x_i)\log(1-z_i)] \tag{2}$$

$$C = C_o + \lambda_1 C_a \tag{3}$$

where m denotes the mini-batch size, x_i and z_i denote the truth label and predicted label in the i-th input of the mini-batch. C_o denotes the cost of the auxiliary classifier, and $\lambda_1 \in [0,1]$ is the weight of this auxiliary classifier.

To constrain the number of selected bands, we add a constraint term into the cost function *C*. Then, Equation (3) is modified to:

$$C = C_o + \lambda_1 C_a + \lambda_2 \frac{1}{m}\sum_{i=1}^{m}\left\|\sum_{i=1}^{N}|W_i^t| - n_b\right\|_2^2 \tag{4}$$

where the hyper parameter n_b denotes the number of selected bands, $|W_i^t|$ denotes the absolute value of ternarized weights of the first layer in the network, and $\lambda_2 \in [0,1]$ denotes the weight of this constraint term. With the training process of the network, the number of selected bands will converge to n_b gradually.

2.4. Training the TWCNN with stochastic gradient descent

Mini-batch stochastic gradient descent (SGD) is used to update the parameters of the proposed network. The derivative of the ternary function is zero almost everywhere, which makes it apparently incompatible with backpropagation. We follow a method similar to that in [9]. Ternary-valued weights are used during the forward and backward propagations, but not during the parameter update. During the process of parameters update, weights are updated as full precision weights by SGD.

3. EXPERIMENT

3.1. Dataset descriptions and experimental setup

The Indian Pines data set was collected by Airborne Visible Infrared Imaging Spectrometer sensors (AVIRIS) in 1992. The scene contains 16 different land-cover classes and the image size is 145×145. After removing 20 absorbent bands, there are 200 bands left. There are 10249 labeled samples in the dataset. The labeled samples of the dataset are normalized to [0, 1] and divided into test set and training set. Specifically, 5% of the labeled samples are randomly selected as the training set to train the network, and the remaining 95% of the labeled samples are used as the test set to evaluate the final performance of the network. The performance of all the methods is measured by overall accuracy (OA).

Four typical approach of HSI band selection are compared with our proposed method: three filter-based methods (MI-CSA [2], mRMR [1] and OPBS [10]), a wrapper-based method (BSCNN+ [6]). Support vector machine (SVM) with radial basis function (RBF) kernel is chosen as the classifier for all the three filter-based methods. The parameters C and γ of SVM classifier are determined with five-fold cross validation,

For TWCNN, the size of input spatial neighborhood window is set to 15×15, the weights λ_1 and λ_2 in cost function are set to 0.05 and 0.01, respectively. The threshold Δ of the ternary function is set to 0.5. In the training phase, the learning rate is initialized to 0.1, and then reduce the learning rate by a factor of 0.8 for every 100 iterations. Note that the weights of the first layer of the network are randomly initialized by the uniform distribution of (0,1), while other parameters are initialized by the normal distribution.

3.2. Classification result analysis

Fig.2 shows the mean overall accuracy of classification as the number of selected bands increases from 10 to 100. Each experiment was repeated five times to obtain the average values. It is shown that the accuracy of the proposed TWCNN outperforms that of the other four methods with different numbers of selected bands. When the number of bands increases, the accuracy of other four algorithms sharply increase, especially in the range of [10-40]. TWCNN gets an excellent OA (95.9%) even when the number of selected bands is very limited (10 bands). Compared with the other algorithms, the classification accuracy of TWCNN is much higher. There are two main reasons why the performance of our proposed method is better than others, one is that TWCNN extracts not only spectral but also spatial information, the other is that the bands selected by TWCNN is “tailor-made” for the classifier.

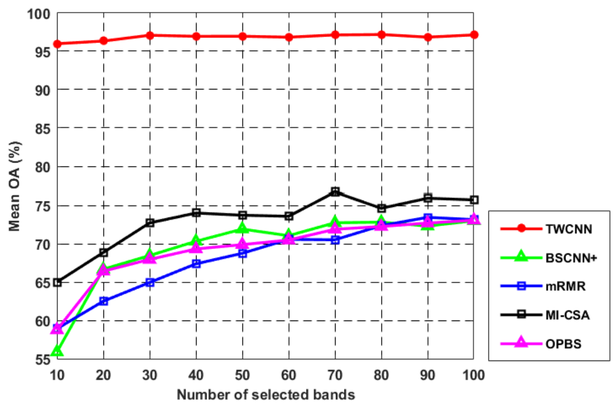

Fig. 3 Mean OA values of TWCN, BSCNN+, mRMR, MI-CSA and OPBS with different numbers of selected bands on the Indian Pines dataset.

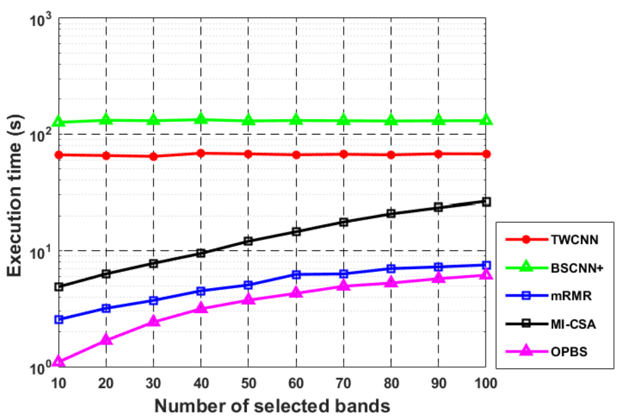

Fig. 4. Execution times of TWCN, BSCNN+, mRMR, MI-CSA and OPBS.

3.3. Time complexity analysis

Fig. 4 shows the execution time of the five algorithms. Here，the execute time includes time for training the model and evaluating the capability of the trained models. Note that the figure is plotted in exponential coordinates. As the figure illustrates, the two approaches based on CNN (TWCNN and BSCNN+) take more time than the others. This is because deep learning based method takes much time in updating the massive parameters of the networks. Compared with BSCNN+, TWCNN have lower time cost. That is because TWCNN completes band selection during the process of training the model, instead of random search which costs much computation.

4. CONCLUSION

A novel band selection method based on ternary weight convolutional neural network is proposed. Compared with traditional HSI band selection approaches, the proposed method selects more discriminative band subset by deep feature extraction ability of CNN. Compared with the existing deep learning based method, the proposed method achieves an end-to-end band selection by integrating band selection into the cost function of TWCNN. We validated our approach on an open hyperspectral dataset and obtained the state-of-art classification performance with only 5% of original spectral bands.

5. REFERENCES

[1] Hanchuan Peng, Fuhui Long and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226-1238, Aug. 2005.
[2] Feng, Jie, et al. "Hyperspectral band selection based on trivariate mutual information and clonal selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no.7, pp. 4092-4105, 2014.
[3] Li, Yifeng, Chih-Yu Chen, and Wyeth W. Wasserman. "Deep feature selection: theory and application to identify enhancers and promoters," *Journal of Computational Biology*, vol. 23, no.5, pp. 322-336, 2016.
[4] Tibshirani, Robert. "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267-288, 1996.
[5] M. Habermann, V. Fremont and E. H. Shiguemori, "Problem-based band selection for hyperspectral images," *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 1800-1803.
[6] Zhan, Ying, et al. "Hyperspectral Band Selection Based on Deep Convolutional Neural Network and Distance Density," *IEEE Geoscience and Remote Sensing Letters.*, vol. 14, no.12, pp. 2365-2369, 2017.
[7] Chollet, François. "Xception: Deep learning with depthwise separable convolutions," *arXiv preprint:* 1610-02357, 2017
[8] Szegedy, Christian, et al. "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
[9] Li, Fengfu, B. Zhang, and B. Liu. "Ternary Weight Networks." *arXiv preprint*: 1610-02357, 2016
[10] X. Geng, K. Sun, L. Ji and Y. Zhao, "A Fast Volume-Gradient-Based Band Selection Method for Hyperspectral Image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7111-7119, Nov. 2014.