Hyperspectral image classification based on a convolutional neural network and discontinuity preserving relaxation

Author:
Gao, Q; Lim, S

Publication details:
International Geoscience and Remote Sensing Symposium (IGARSS)
v. 2018-July
pp. 3591 - 3594
9781538671504 (ISBN)

Event details:
IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium
2018-07-22 - 2018-07-27

Publication Date:
2018-10-31

Publisher DOI:
https://doi.org/10.1109/igarss.2018.8517463

HYPERSPECTRAL IMAGE CLASSIFICATION BASED ON A CONVOLUTIONAL NEURAL NETWORK AND DISCONTINUITY PRESERVING RELAXATION

Qishuo Gao[1], Samsung Lim[1]

[1] School of Civil and Environmental Engineering, University of New South Wales

ABSTRACT

In this paper, we present a novel method for hyperspectral image classification to take advantage of the merits of a convolutional neural network (CNN) and the spatial contextual information of hyperspectral imagery (HSI). We built a novel network consisting of several convolutional, pooling and activation layers to extract the effective features and predict the class membership probability distribution vectors for HSI pixels. Furthermore, in order to fully exploit the spatial contextual information and improve the classification accuracy under the condition of limited training samples, a promising discontinuity preserving relaxation (DPR) algorithm is applied to process the probabilistic results obtained by the CNN work. The proposed method was tested on two widely-used hyperspectral data sets: the Indian Pines and University of Pavia data sets. Experiments revealed that the proposed method can provide competitive results compared to some state-of-the-art methods.

Index Terms— Hyperspectral image (HSI) classification, convolutional neural network (CNN), discontinuity preserving relaxation (DPR) method.

1. INTRODUCTION

Hyperspectral imagery (HSI) is gaining attention due to its composition of hundreds of spectral channels over the same site. As an important approach to understand the remotely sensed images, HSI classification has been an active research field in recent years. However, the increased number of dimension in spectral domain and the limited number of training samples pose different processing problems for HSI classification, hence should be tackled under specific operations. Many attempts have been made to develop techniques that can classify HSI in an efficient manner. In the early stage of HSI classification, the majority of the methods have focused on processing of the spectral information. With the advancement of image processing knowledge, new techniques such as Markov Random Fields (MRFs) [1] and morphological operators [2] that incorporate spatial contextual information into HSI classification have been developed.

Very recently, convolutional neural networks (CNNs) have been investigated and they have outperformed many approaches in various domains [3, 4]. Some CNN-based models have been proposed for HSI classification recently. In the report of Makantasis et al.[5], the authors have successfully applied a new network for HSI classification and both spectral and spatial information have been considered. In [6], a 3D-CNN model was proposed for high accuracy HSI classification. In the study of Lee et al. [7], the authors proposed a deeper CNN which adopted a fully convolution network particularly for the HSI classification task. Compared with the traditional hand-crafted methods, CNNs can extract the effective and representative features without the need of predefined parameters. However, under the condition of the limited training samples, CNNs may not maximize the accuracy.

On the other hand, it has been observed that the spatial smoothing over the HSI can help enhance the classification accuracies [8]. As a post-processing approach, relaxation-based methods can help remove the noisy points and improve the classification results. These methods are often referred to as probabilistic relaxation (PR) methods [9], which should be applied to the posterior probability distribution results obtained by other probabilistic methods. In [10], a promising PR technique namely “discontinuity preserving relaxation (DPR)” was proposed as a post-processing technique to improve the classification results by imposing spatial consistency for the neighboring pixels in the final classification maps.

In this paper, we developed a novel framework that integrates a CNN and DPR to fully exploit the spectral and spatial information simultaneously. At first, a proper network is built to compute the probability distribution vectors for test pixels, and then DPR is applied to the class membership probability distribution to derive the final labels. The DPR improves the classification results by exploiting the intrinsic correlations between the test pixel and its neighbors. The main advantage of the proposed method is the integration of a CNN and DPR where the CNN can extract the features automatically and ensure the accuracy for most pixels, and then the DPR can enhance the classification by smoothing the results. This paper is organized as follows. Section 2 layouts the principles of the proposed framework. Section 3 presents the experimental validation. Section 4 concludes the paper.

2. THE CONTEXTUAL OF PROPOSED FRAMEWORK

Fig. 1 shows the framework of the proposed method. The proposed framework consists of two components, one is the CNN construction and the other is the probabilistic relaxation. The details of the two components are explained in detail in the following sections.

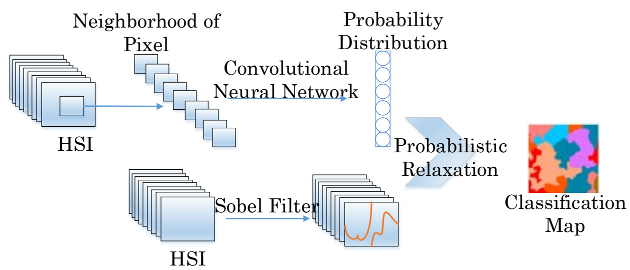

Fig.1. The flowchart of the proposed framework.

2.1. Convolutional Neural Network

CNNs extract the features via several non-linear functions. In this paper, we consider HSI classification with the so-called *directed acyclic graphs* (DAG) convolutional architecture in which the layers can share weights and the biases [11]. The basic layers in a CNN are introduced as follows:

The function of a convolution function can be defined as:

$$a = \sigma(fx + b) \quad (1)$$

where a is a feature map which can be obtained by convolving the input x with a weight filter f and biases b. $\sigma(\bullet)$ is a nonlinear activation function. The most frequently used activation function is rectified linear unit (ReLU). ReLU is defined as follows:

$$\sigma(x) = \max(0, x) \quad (2)$$

Pooling layers are also important in CNN construction. Max pooling is used throughout this paper. In the max pooling, a small $T \times 1$ patch is combined from the previous layer. The max pooling function can be defined as:

$$a_i = \max_{T\times 1}(a_j^{t\times 1} u(n,1)) \quad (3)$$

where $u(n,1)$ is a window function, a_i is the maximum value in the patch.

The network is trained through the back-propagation. Usually, the process can be treated as minimizing a defined loss function between the ground truth values (e.g. image labels) and the network output. Let $y_i = 1,...,c,...,C$ denote the target ground truth values, and let $P(y_i)$ denote the output class membership distribution with $i = 1,2,...,N$ as the number of training samples. The multi-class hinge loss throughout this paper is given by

$$L = \sum_{i=1}^{N}\sum_{c=1}^{C} \max(0, 1 - P(y_i = c)) \quad (4)$$

Finally, the test pixel is assigned to the label which minimizes the loss function:

$$\hat{y}_i = \arg\min_c L \quad (5)$$

2.2. The Construction of Convolutional Neural Network

The input of the whole framework is a three dimensional cube for each pixel, and the output is the corresponding class membership probability distribution vector. It can be observed from Fig. 1, for each pixel x_i, each convolutional layer of the proposed CNN has a $K \times K \times L$ format of input, where $K \times K$ is the kernel size and L is the number of spectral bands. Each convolutional layer for the proposed CNN has a three dimensional convolution where the third dimension is the number of kernels. The spatial size of the output is computed by $H'' = 1 + [\frac{H - H' + P}{S}]$, where H, H' and H'' represent the input size, kernel size and output size, respectively; P and S denote the padding and stride, respectively. For example, in Fig. 1, the input spatial format for the Pavia University data set is 7×7, the convolution kernel size for the first layer is 4×4 without pooling and stride, then the output spatial size of this layer is 4×4.

2.3. Discontinuity Preserving Relaxation

The DPR [10] can be applied to further locally smooth the homogenous areas in the original HSI. Let $p = [p_1,..., p_N] \in \mathbb{R}^{C\times N}$ (where N is the number of samples, and $1,...,c,...,C$ denote the class labels) be the class membership probability distribution obtained by the previous CNN step for all samples, $\theta_i = [\theta_{i,1},...,\theta_{i,c},...,\theta_{i,C}]^T$ be the final class membership probability distribution computed by the DPR method, and $\theta_{i,c}$ be the final probability for the $i-th$ test pixel belonging to class c. Let $\theta = [\theta_1,...,\theta_i,...,\theta_N] \in \mathbb{R}^{C\times N}$ be the distribution matrix for all the samples. θ can be computed by relaxing the following optimization function:

$$\min_{\theta}(1-\lambda)\|\theta - p\|^2 + \lambda\sum_i\sum_{j\in S_i}\delta_j\|\theta_j - \theta_i\|^2$$
$$subject\quad to: \quad \theta_i \geq 0 \quad 1^T\theta_i = 1 \quad (6)$$

where $0 \le \lambda \le 1$ is the weight parameter which controls the impacts of the two terms in (6), and S_i is the neighborhood of the test pixel x_i . λ measures the noisiness and smoothness levels of images. δ_j is the value at location $j \in S_i$ which is obtained by a Sobel filter:

$$\delta = \exp\left(-\sum_{i=1}^{L} sobel(I^{(i)})\right) \tag{7}$$

where $sobel()$ is the Sobel operator which produces two outputs: 0 or 1 corresponding to the existence and no existence of discontinuities in I . I denotes the original HSI cube.

During the DPR procedure, the discontinuities in the original images are detected by the Sobel filter at first, and then DPR is applied on the classification maps obtained by the previous CNN step without crossing the class boundaries. In that way, DPR smooths the spatially homogenous areas. In this paper, the objective function (6) is solved by the same way in [10]. After $\theta_i = [\theta_{i,1},...,\theta_{i,c},...,\theta_{i,C}]^T$ is recovered, x_i can be assigned to the class that has the maximum probability:

$$Class(x_i) = \underset{c=1,2,...,C}{\arg\max}\ \theta_{i,c} \tag{8}$$

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Datasets and Baselines

To verify the effectiveness of the proposed method, two data sets are applied in this paper: the Indian Pines and the University of Pavia data sets. The ground truth and false color images of the two data sets are shown in Fig. 2.

The performance of the proposed framework is compared with several spectral-spatial classifiers: SVM with extended multi-attribute profiles (EMP) features [12], SVM with DPR, and CNN without DPR (The CNN has the same architecture as the proposed framework). λ is set as 0.85, and S_i is set as the eight neighborhood pixels of the test pixel x_i . For fair comparison, the training samples are randomly selected for each data set and the remaining as the test set. For Indian Pines data set, 15% samples are selected as training set. 250 samples are selected for each labeled classes as training data set for the University of Pavia. There are three convolutional layers, one pooling layer, and one ReLU layer in the network. The details of the network structure are shown in Fig. 3. The input images are initially normalized into [-1 1]. The input sizes are set different for two data sets due to the different spatial resolutions and homogenous areas, and set as 5×5 and 7×7 for the Indian Pines and the University of Pavia, respectively. The learning rate for CNN models is set as 0.01; the number of epochs is set as 100 for the two data sets. The batch size is set as 10. To quantitatively validate the results of our proposed framework, overall accuracy (OA), average accuracy (AA) and the Kappa coefficient (k) are adopted as the metrics.

(a) (b)

Fig. 2. The (left) false color composite image bands (bands 50, 27, 17) and (right) ground truth for two data sets: (a) Indian Pines; (b) University of Pavia.

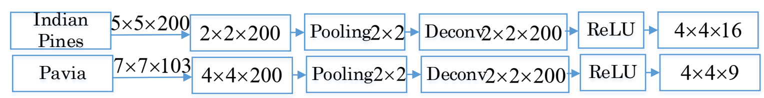

Fig. 3. The architecture of CNN for two data sets

3.2. Classification Results

Tables I-II show the classification performances for various classifiers. From the results, one can see that the classification performances of the CNN-based methods are better than those of other classifiers in terms of OA, AA and the k . Even with the limited training samples, the proposed method improves the classification results of the EMP-based and the original CNN. The OA of the proposed method is better than the CNN by 5.59% and 1.13% for two data sets, respectively. The improvement is realized by further exploitation of spatial information by the DPR. In addition, the performance enhancement compared to SVM-based methods is achieved by the integration of a properly built network and the efficient discontinuity preserving approach.

The classification maps are illustrated in Figs. 4-5. From the maps, one can figure out how DPR affects the classification results. The two SVM-based methods achieved the similar results for the two data sets and produced more noisy scatter points on the maps. One can observe from the maps that the SVM+DPR smooths more homogeneous areas. The CNN-based methods produce more detailed maps under the condition of same training samples, and the DPR algorithm help improve the performance by smoothing the homogeneous areas without blurring the boundaries. It should be noted that the proposed CNN may achieve a better performance with more training samples.

TABLE I.
CLASSIFICATION RESULTS (%) OF INDIAN PINES DATA SET

Accuracy	SVM+EMP	SVM+DPR	CNN	CNN+DPR
OA	87.03	87.81	89.14	94.73
AA	88.89	91.64	94.77	96.10
k	85.31	86.19	87.73	94.01

TABLE II.
CLASSIFICATION RESULTS (%) OF UNIVERSITY OF PAVIA DATA SET

Accuracy	SVM+EMP	SVM+DPR	CNN	CNN+DPR
OA	95.90	95.25	96.84	97.97
AA	97.07	96.17	96.93	97.92
k	94.63	93.75	95.84	97.32

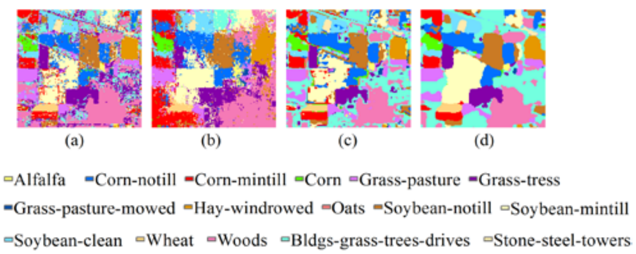

Fig. 4. Classification maps of Indian Pines data set for different classifiers: (a) SVM+EMP (b) SVM+DPR (c) CNN (d) CNN+DPR

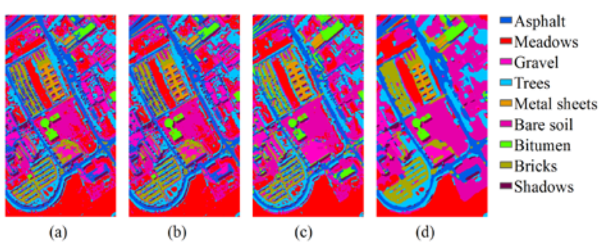

Fig. 5. Classification maps of the University of Pavia data set for different classifiers: (a) SVM+EMP (b) SVM+DPR (c) CNN (d) CNN+DPR

4. CONCLUSION

It is well known that CNNs can lead to an improved performance for image classification and the probabilistic relaxation strategy can incorporate the contextual information into a probabilistic classification results. Hence, we proposed a framework which integrates a CNN and probabilistic relaxation to leverage both spectral and spatial information for HSI classification. Firstly, the posterior probabilities for test pixels are obtained by the built network and then a discontinuity preserving relaxation algorithm is applied to derive the final pixel label vectors. The classification results on two popular data sets show that the proposed framework outperforms the state-of-the-art classifiers.

5. REFERENCES

[1] J. Li, J. M. Bioucas-Dias, and A. Plaza, “Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809-823, Aug. 2012.

[2] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, “Classification of hyperspectral data from urban areas based on extended morphological profiles,” *IEEE Trans. Geosci. Remote Sens.,* vol. 43, no. 3, pp. 480-491, Feb. 2005.

[3] A. Karpathy, G. Toderici, S. Shetty *et al.*, "Large-scale video classification with convolutional neural networks." pp. 1725-1732.

[4] S. Ren, K. He, R. Girshick *et al.*, "Faster R-CNN: Towards real-time object detection with region proposal networks." pp. 91-99.

[5] K. Makantasis, K. Karantzalos, A. Doulamis *et al.*, "Deep supervised learning for hyperspectral data classification through convolutional neural networks." pp. 4959-4962.

[6] Y. Chen, H. Jiang, C. Li *et al.*, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” *IEEE Trans. Geosci. Remote Sens.,* vol. 54, no. 10, pp. 6232-6251, Oct. 2016.

[7] H. Lee, and H. Kwon, “Going Deeper With Contextual CNN for Hyperspectral Image Classification,” *IEEE Trans. on Image Process.,* vol. 26, no. 10, pp. 4843-4855, Jul. 2017.

[8] Y. Wang, R. Niu, and X. Yu, “Anisotropic diffusion for hyperspectral imagery enhancement,” *IEEE Sens. J.,* vol. 10, no. 3, pp. 469-477, Feb. 2010.

[9] J. Richards, D. Landgrebe, and P. Swain, “Pixel labeling by supervised probabilistic relaxation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 2, pp. 188-191, Mar. 1981.

[10] J. Li, M. Khodadadzadeh, A. Plaza *et al.*, “A discontinuity preserving relaxation scheme for spectral–spatial hyperspectral image classification,” *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.,* vol. 9, no. 2, pp. 625-639, Feb. 2016.

[11] A. V. a. K. Lenc, "MatConvNet - Convolutional Neural Networks for MATLAB ".

[12] M. Fauvel, J. A. Benediktsson, J. Chanussot *et al.*, “Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles,” *IEEE Trans. Geosci. Remote Sens.,* vol. 46, no. 11, pp. 3804-3814, Nov. 2008.