2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS) | 978-1-7281-3114-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/InGARSS48198.2020.9358920

A 3D-DEEP CNN BASED FEATURE EXTRACTION AND HYPERSPECTRAL IMAGE CLASSIFICATION

MURALI KANTHI[1], *T. HITENDRA SARMA*[2], *and C. SHOBHA BINDU*[1]

[1] Department of Computer Science and Engineering
JNTUA College of Engineering Anantapur, Ananthapuramu, A.P., India.
Email:{murali.kanthi, shobabindhu}@gmail.com
[2] Department of Computer Science and Engineering
Srinivasa Ramanujan Institute of Technology, Anantapur. A.P., India
Email:hitendrasarma@ieee.org

ABSTRACT

Hyperspectral image consists of huge spectral and special information. Deep learning models, such as deep convolutional neural networks (CNNs) being widely used for HSI classification. Most of the approaches are based on 2D CNN. Whereas, the HSI classification performance depends on both spatial and spectral information. This paper proposes a new 3D-Deep Feature Extraction CNN model for the HSI classification which uses both spectral and spatial information. Here the HSI data is divided into 3D patches and fed into the proposed model for deep feature extractions. Experimental results show that the performance of HSI classification is improved significantly with the proposed model. The experimental results on the publicly available HSI datasets, *viz.,* Indian Pines(IP), Pavia University scene(PU) and Salinas scene(SA), are compared with the contemporary models. The current results indicates that the proposed model provides comparatively better results than the state-of-the-art methods.

Index Terms— Hyperspectral Image (HSI), Classification, Convolutional Neural Networks (CNN), Spectral-Spatial, 3D-CNN

1. INTRODUCTION

Remote sensing plays a fundamental role for collecting rich set of information as a Hyperspectral images (HSI) for various applications such as monitoring the surface of earth, management of the environment, agriculture, security etc [1].

In the last two decades, many supervised classifiers have been proposed for HSI classification. However, Supervised methods face challenges in the classification of HSI data due to sufficient number of training samples are not available for the more number of spectral bands. Then the classification accuracy is unsatisfactory with this type of data. So the process of extracting hidden features is more important to improve the classification accuracy of the HSI data.

Deep learning based models have shown more and more attractive in different fields in many research areas, such as image classification, natural language processing, speech recognition, and remote sensing [2]. In recent years, many deep learning models have been examined for HSI classification. Convolution Neural Network (CNN) is one of the most commonly used network model for image classification in which features can be extracted from the hidden layer in the network with less amount of parameters, thus it has been a more preferable deep learning model in the current trend of HSI analysis. Recent studies with encouraging results are motivating many researchers to develop more deep CNNs models to cater the HSI classification problem [3].

The main objective of this paper is to propose a novel 3D deep feature extraction CNN structure for HSI classification. The current results indicates that the proposed model provides comparatively better results than the state-of-the-art methods in terms of classification accuracy.

The remainder of this paper is organized as follows: Section 2 presents the literature review corresponding to the proposed work, Section 3 initially describes the data preprocessing, and then presents the proposed model framework. Section 4 describes the experimental study and results of the proposed method. Section 5 describes the results and discussions. Finally, the conclusions are presented in Section 6.

2. LITERATURE SURVEY

In the literature, three types of CNN models have been found for HSI data classification based on whether they perform spectral feature analysis, spatial feature analysis, or spectral-spatial feature analysis [1]

Vectorized CNN (VCNN) proposed by Charmisha et al., [4] presents the effect of dimensionality reduction (DR) to achieve improved classification accuracy. Through DR technique for HSI classification using a VCNN, comparable classification accuracy was obtained using the reduced feature di-

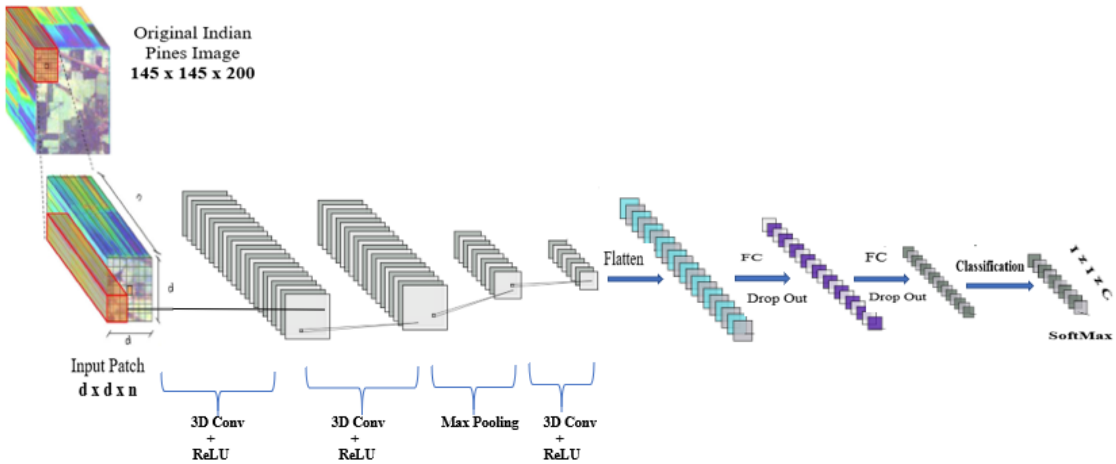

Fig. 1. Proposed 3D-DFE-CNN Model Architecture

mension and a lesser number of VCNN trainable parameters.

Huang Y et al., have applied deep convolutional neural networks to classify hyperspectral images directly in spectral domain. The architecture of the classifier contains sevaral layers that are implemented on each spectral signature to discriminate against others. The proposed model has overfitting problem caused by limited training samples and the spatial correlation was not considered [5]. Yushi Chen et al., have proposed an approach which employs several convolutional and pooling layers to extract deep features from HSIs. These features are useful for image classification and target detection. To avoid overfitting problem in class data modeling few strategies are investigated such as L2 regularization and dropout mechanism in the model. Finally they proposed a 3-D CNN-based FE model with combined regularization and a virtual sample enhanced method to extract effective spectral–spatial features of hyperspectral imagery to improve the performance of the model [6].

Several 3D convolutional neural network (3D-CNN) frameworks are proposed for accurate HSI classification. In these method the spectral–spatial features are extracted without doing any preprocessing on the HSI data cube [7]. Xiaoguang Mei et al., proposed to extract spectral-spatial features based on ARNN and ACNN for HSI classification by incorporating attention mechanism to the model. This can grasp various interspectral correlations in the continuous spectrum domain and focus on similar spatial features between neighboring pixels in spatial dimension by adding attention weights [8].

As most recent studies have suggested that spectral–spatial FE methods provide good improvement in terms of classification performance [9], the current work is motivated in the same direction and developed a new 3D CNN which is explained in the subsequent sections.

3. PROPOSED 3D-DFE-CNN

The proposed 3D-DFE-CNN model is shown in Figure 1 which consists of an input layer, three convolution layers with ReLU as nonlinear activation function, one max-pooling layer, and two fully-connected layers. The last softmax layer is the output layer for obtaining the desired labels for the input data. The following sub sections provide the details of data preprocessing strategy and the architecture.

3.1. Data Preprocessing

In general, every CNN model expects the normalized image before processing, However, a hyperspectral image consists of rich set of spatial and spectral information. To take the advantage of both the spatial and the spectral information of the image. The pre-processing of hyperspectral image is required to select the proper bands to reduce the dimensions and computational complexity. In this regard, PCA (principal component analysis) is used to perform an initial dimentionality reduction (DR) in order to remove the spectral correlation and redundancy. After pre-processing, The original hyperspectral image of size $D \times D \times N$, where D is the width and height of the input volume and N is the total number of bands, is divided into patches of size $d \times d \times n$.

In the proposed 3D-DFE-CNN model the normalized input volume neighborhood window of center around each pixel vector is given to the network for deep feature extraction.

3.2. CNN Architecture Details

After pre-processing the normalized hyperspectral image is taken into consideration for classification and is split into 3-D patches, these patches are grouped in batches of size b and sent to the CNN. Then, the $d \times d \times n$ patches are sent as

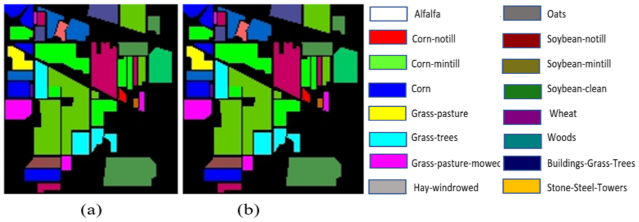

Fig. 2. Indian Pines dataset. (a) Ground truth (b) Classification map with respective classes.

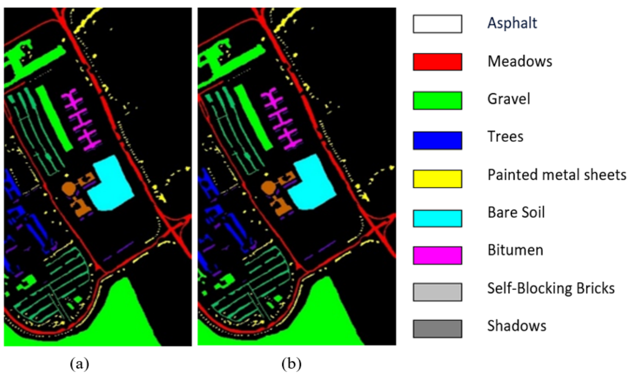

Fig. 3. Pavia University dataset. (a) Ground truth (b) Classification map with respective classes.

input to the first convolution layer (C1) of the model which is composed by $k1$ filters with stride is equals to 1, and without padding.

After applying the ReLU function, the $k1$ feature maps generated by C1 are sent to the second convolution layer (C2) of the model which is composed by $k2$ filters, with the same stride as in the first convolution layer and no padding. After applying the ReLU function, the $k2$ feature maps generated by C2 are sent to the first MaxPool layer (MP), with a kernel, stride of 2, and padding. The resulting output is sent to the last convolution layer (C3) with $k3$ filters, with the same stride as in the second convolution and no padding. Again, apply the ReLU function, and there is no second maxpool layer, so the resulting output volume is reshaped to send it to fully-connected layers.

Three fully-connected layers are implemented with $fc1$, $fc2$ and $fc3$ nodes. The first fully-connected layer compute their output as $z^{fc} = g(w^{fc}z^{fc-1} + b^{fc})$, where w^{fc} are their weight matrices, b^{fc} are their bias vectors, z^{fc-1} is the output of the previous layer and the activation function $g(.)$ is ReLU. Finally, the last resulting matrix z^{fc2} is sent to $fc3$, which computes the outputs of the network with a softmax function as $z^{fc3} = w^{fc2}z^{fc2} + b^{fc2}$, where z^{fc3} contains the desired labels for the original $d \times d \times n$ input volume.

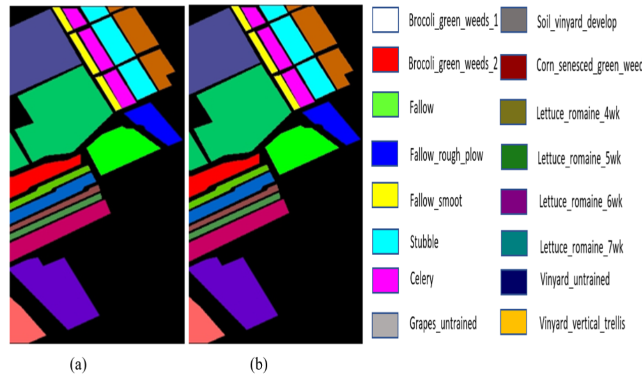

Fig. 4. Salinas dataset. (a) Ground truth (b) Classification map with respective classes.

4. EXPERIMENTAL STUDY

4.1. Data Description

To evaluate the performance of the proposed model, three publicly available benchmark HSI datasets Indian Pines (IP), Pavia University scene (PU) and Salinas scene (SA) are used in the experiments. For all the data, we randomly select 80% of each categories in dataset as train dataset and the remaining as test dataset.

4.2. Experimental Setup

Our experimental setup-platformm is a PC equipped with an Intel Core i7, 16 GB memory and Nvidia GPU. In order to evaluate the efficacy of the proposed model, we compared it with different deep learning HSI classification methods such as SVM, 2D-CNN, and 3D-CNN, Overall accuracy (OA), average accuracy (AA) and kappa statistic (K) are adopted to evaluate the classification performance of each model.

5. RESULTS AND DISCUSSION

The proposed method is compared with the state-of-the-art methods such as SVM, 2D-CNN, and 3D-CNN in terms of HSI classification accuracy. Overall accuracy (OA) is calculated to evaluate the performance of the model. In this section, detailed results are shown and discussed.

In the proposed architecture of the model, two convolutional layers, two ReLU layers, and one max-pooling layer are used. The Batch Normalization layer is used to speed up and optimizes the entire network. Dropout layer is added after every fully-connected layer with ratio of 0.4 to avoid overfitting. Then the softmax layer is the output layer of the network, which delivers the corresponding class labels. After this model, the input image will be converted into a vector. The batch size was 256, and the number of training epochs was 200 with learning rate 0.001. The window size 25×25

Table 1. Class wise accuracy for Indian Pines, Pavia University and Salinas datasets in comparison with the state-of-art methods.

Indian Pines					Pavia University					Salinas				
Class No.	**SVM**	**2D-CNN**	**3D-CNN**	**3D-DFE-CNN**	**Class No.**	**SVM**	**2D-CNN**	**3D-CNN**	**3D-DFE-CNN**	**Class No.**	**SVM**	**2D-CNN**	**3D-CNN**	**3D-DFE-CNN**
1	62.05	75.38	96.92	**98.94**	1	94.29	98.01	**100**	**100**	1	99.63	99.45	**100**	**100**
2	81.45	91.54	98.91	**99.92**	2	97.49	99.41	**100**	**100**	2	99.91	99.51	**100**	**100**
3	70.55	86.95	98.84	**99.23**	3	80.84	93.90	99.35	**99.86**	3	99.68	99.62	**100**	**100**
4	72.93	88.56	97.71	**98.86**	4	94.21	98.14	99.74	**99.87**	4	99.31	99.89	99.86	**99.92**
5	93.17	86.05	99.32	**99.84**	5	99.22	99.57	**100**	**100**	5	99.35	99.88	99.95	**99.97**
6	97.32	96.13	99.74	**99.96**	6	90.91	98.08	**100**	**100**	6	99.80	99.78	**100**	**100**
7	84.35	82.61	93.04	**98.46**	7	87.35	89.72	99.98	**100**	7	99.54	99.64	**100**	**100**
8	98.32	97.88	**100**	**100**	8	87.47	98.28	99.74	**99.84**	8	90.51	95.60	99.97	**99.99**
9	51.76	65.88	**100**	**100**	9	99.86	98.87	99.60	**99.72**	9	99.92	99.54	**100**	**100**
10	77.87	89.85	99.15	**99.87**						10	97.71	98.45	99.99	**100**
11	85.10	95.28	99.23	**99.**						11	98.88	98.73	**100**	**100**
12	79.09	88.65	97.86	**99.16**						12	99.79	99.58	99.99	**100**
13	98.39	97.82	99.89	**99.96**						13	98.88	99.13	99.98	**100**
14	95.59	98.40	99.59	**99.89**						14	97.65	97.53	99.98	**99.98**
15	61.28	89.21	98.48	**99.24**						15	70.54	95.01	99.95	**99.97**
16	87.60	82.53	95.70	**98.85**						16	99.18	97.00	99.94	**99.98**
OA	84.48	92.69	99.08	**99.94**	**OA**	94.10	98.27	99.92	**99.96**	**OA**	93.67	97.94	99.98	**99.99**
AA	81.05	88.29	98.40	**99.95**	**AA**	92.40	97.11	99.82	**99.92**	**AA**	96.89	98.65	99.98	**99.98**
Kappa	82.26	91.65	98.95	**99.94**	**Kappa**	92.17	97.71	99.89	**99.97**	**Kappa**	92.94	97.71	99.98	**99.99**

is chosen as input to the model. The classification results of IP, PU and SA datasets are listed in the Table 1.

Table 1 provides the comparative study on the classification accuracies of individual methods. From the results, it is shown that the classification accuracy of the proposed model performs better in terms of average accuracy (AA), kappa (K) coefficient and overall accuracy (OA) based on results on three datasets.

The classification maps are provided for the three scenes and compared with ground truths which are shown in Figure 2, Figure 3, and Figure 4 respectively. Some areas in the classification maps produced by the proposed method are less noisy compared to those of state-of-art methods.

6. CONCLUSION

In this paper, a novel 3D-Deep CNN based feature extraction framework framework is proposed for HSI classification that takes full advantage of both spectral-spatial information contained within HSI data. The proposed classification method is implemented on three publicly available benchmark datasets and compared with three state-of-the-art deep learning-based HSI classification methods. The experimental studies shown that the designed framework achieved the better overall accuracy on all the three datasets.

7. REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[2] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[3] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, 2017.

[4] K. Charmisha, V. Sowmya, and K. Soman, "Dimensionally reduced features for hyperspectral image classification using deep learning," in *International Conference on Communications and Cyber Physical Engineering 2018*. Springer, 2018, pp. 171–179.

[5] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, 2015.

[6] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[7] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.

[8] X. Mei, E. Pan, Y. Ma, X. Dai, J. Huang, F. Fan, Q. Du, H. Zheng, and J. Ma, "Spectral-spatial attention networks for hyperspectral image classification," *Remote Sensing*, vol. 11, no. 8, p. 963, 2019.

[9] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279–317, 2019.