Dimension Reduction Using Supervised LDA and Unsupervised PCA and 3D CNN for Classification of Hyper-spectral Image

Tasmia Jannat[1*],Abu Sayeed[2]

1. Rajshahi University of Engineering & Technology. Kazla, Rajshahi-6204, Bangladesh, Email: jannat22tasmia@gmail.com

2. Assistant Professor, Rajshahi University of Engineering & Technology. Kazla, Rajshahi-6204, Bangladesh, Email: abusayeed.cse@gmail.com

Keywords:

- Hyperspectral Image;
- Feature extraction;
- Dimensionality reduction;
- SVM;
- PCA;
- LDA;
- 3D-CNN

Abstract: Hyperspectral Image (HSI) classification is widely used now-a-days because it contains huge information about a scene. The main problem of dealing with hyperspectral image is "Hughes Phenomenon" or the "Curse of Dimensionality" which is due to the high number of dimensions. To overcome this, dimensionality reduction approaches need to be performed as a preprocessing step. In this paper, we used supervised LDA and unsupervised PCA for dimension reduction. Then for classification we used 3D CNN. Performance of classification depends on both spatial and spectral information for which 3D CNN is a good choice. But still it's not popular because of its computational complexity. But dimension reduction makes things easier by removing irrelevant spectral or information. So we propose a method that has applied these feature extraction methods on the dataset Indian Pine individually and then classified the dataset using 3D Convolutional Neural Network. The results are compared with Support Vector Machine and other Convolutional Neural network such as 2D CNN and 3D CNN. The experimental results show our method is best among these three with the accuracy of 99%.

1. Introduction

A hyperspectral image is a collection of several hundreds even thousands of reflectance bands of different wavelengths over the same spatial area, captured by a satellite sensor. If a RGB image is considered as a small book of 3 pages, then hyperspectral image is a book of hundred and more pages. Here every wavelength is presented as a page of the book. Due to its huge spectral range it's possible to identify the smallest difference between object, and therefore hyperspectral image is gaining more importance in different fields.

At the same time it has been a major challenge to process this huge amount of data effectively. If the training samples are limited, a reduction in the classification accuracy of the test data is often observed due to the poor generalization of the training results and this effect is known as

Submitted: xx Month, xxxx | Revised: xx Month, xxxx | Accepted: xx Month, xxxx

the Hughes phenomenon [1]. On the other hand, since the hyperspectral data are sensed in very close and contiguous spectral bands, some bands are highly correlated and can be removed [2]. This obstacle can be overcome by feature selection and feature extraction approaches [3].

Feature selection techniques refer to the selection of correlated spectral bands from the original HSI dataset [4] and Feature extraction methods involve transforming the data. The Principal Component Analysis (PCA) [5] and Linear Discriminant Analysis (LDA) [6] are two widely used unsupervised and supervised feature extraction methods respectively. One drawback of the LDA method is that only *G-1* features are generated in this approach, where *G* is the number of groups or the class labels. Unlike the LDA, the PCA does not focus on the individual class reparability while extracting features rather focus on variance of the data.

The CNN has shown very promising performance in visual information processing, such as image classification [7], object detection [8]. X. Kang proposed a model dual-path network (DPN) by combining the residual network and dense convolutional network [9]. Zhong et al. [10] has proposed the spectral–spatial residual network (SSRN).

From the literature we see using just 2-D-CNN or 3-D-CNN had few shortcomings such as very complex model or missing channel relationship information etc. Dimensionality reduction technique along with SVM is also not effective. This is the motivation of our work here. We propose a model using PCA and LDA as preprocessing step and 3D CNN for classification.

2. Background Work

2.1 PCA

The Principal Component Analysis [5] is a dimensionality reduction technique. It is mainly used in transforming a large set of value into small ones, containing most of the information of the original dataset, making the original hyperspectral $M \times N \times D$ to $M \times N \times B$ where $D > B$. PCA generates a low dimensional space representing as much of the variance [11].

2.2 LDA

Linear Discriminant Analysis (LDA) [12] also known as Fisher’s Discriminate Analysis, discriminates the samples in the training dataset by their class value. In LDA a new feature space is created. In that new space training space, data are projected in such a way that maximizes the class reparability.

2.3 3D CNN

3D-CNN [13] works as same way as 2D-CNN. The sole distinction is in 2D CNN, kernel moves in two directions. In 3D CNN, kernel moves in three directions. CNN or 3D-CNN operates in 3 layers.

1) *Convolution Layer:* Convolutional layers are the key building blocks of CNN. It is the initial layer to extract features from an input image. It applies a filter to an input to create a feature map that summarizes the presence of detected features in the input. It is a mathematical process that takes two inputs which are image matrix and a filter or kernel.
2) *Pooling Layer:* Pooling layers section would cut back the number of parameters. Pooling, or down-sampling, is done on the activation maps created during convolution layer. Pooling works same as convolution process.

Tasmia Jannat[1]*, Abu Sayeed; (jannat22tasmia@gmail.com)

3) *Fully Connected Layer:* In fully connected layer our matrix is flattened into vector and feed into a fully connected layer like a neural network. With the fully connected layers, we combined these features along to create a model. Then we have an activation function such as softmax or sigmoid to classify the outputs.

3. Experimental Analysis

3.1 Dataset Description

The Indian Pines image was captured by AVIRIS sensor over the Indian Pines field in North-western Indiana. It is comprised of 145×145 pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.5 μ m. The ground truth available is designated into sixteen classes and is not all mutually exclusive. By removing bands covering the region of water absorption: [104-108], [150-163], we have reduced the number of bands to 200 from 224.

3.2 Experimental Setup

In the process of implementing dimensionality reduction technique individually, many irrelevant bands were excluded. In PCA as a preprocessing step, only a few such as 16, 32 and 64 bands were selected manually. We chose these numbers because cumulative variance is nearly 99% when the number of component is above 14. In case of LDA 16 features have been computed. After reducing the dimension, main work of classification begins.

For better classification the HSI data cube is divided into small overlapping 3-D-patches of size $S \times S \times B$ covering the $S \times S$ window and all B spectral bands. In our case value of $S=25$ Total $(M-S+1) \times (N-S+1)$ number of 3D patches are created. The 3-D convolution is done by convolving a 3-D kernel with the 3-D-data. Two 3D CNN layers have been applied with the filter size 8 and 16 respectively. 3D CNN is effective in both spectral and spatial feature representation. Then to reduce the size of the tensor and speed up calculations Maxpooling layer has been applied. Then we used batched normalization for adjusting and scaling the activations and speeding up the learning. As we have used 3D CNN, the output of pooling layer will be four dimensional .Then the flatten layer prepares data for fully connected layer by turning 4-dimensional matrix to a vector. The final layer uses the softmax activation function instead of ReLU. It returns probabilities of the object in the image belonging to the different classes. Here we have used Adam optimizer and 100 epochs to train the data.

All experiments have been conducted on kaggle. Kaggle provides free access to NVIDIA TESLA P100 GPUs. Optimal learning rate is 0.001 based on the classification outcomes.

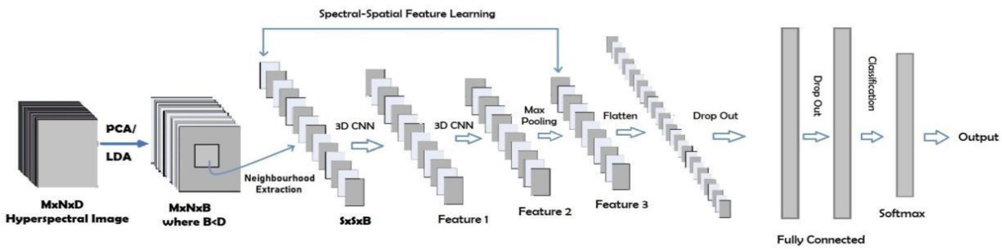

Fig.1. Proposed model of 3-D convolutions for HSI classification.

TABLE 1

Model summary of proposed architecture with window size 25 × 25.

Layer(type)	Output Shape	Parameter
Input_1(InputLayer)	(None; 25; 25; 16; 1)	0
conv3d_1(Conv3D)	(None; 23; 23; 12; 8)	368
conv3d_2(Conv3D)	(None, 21, 21, 10, 16)	3472
max_pooling3d_1 (MaxPooling3)	(None, 10, 10, 13, 16)	0
batch_normalization_1	(None, 10, 10, 13, 16)	64
flatten_1 (Flatten)	(None, 20800)	0
dense_1 (Dense)	(None, 256)	5325056
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 16)	2064
Total params:5,363,920 Trainable params:5,363,888 Non-trainable parameters: 32		

3.3 Classification Result

We have used the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa) to evaluate our classification performance.

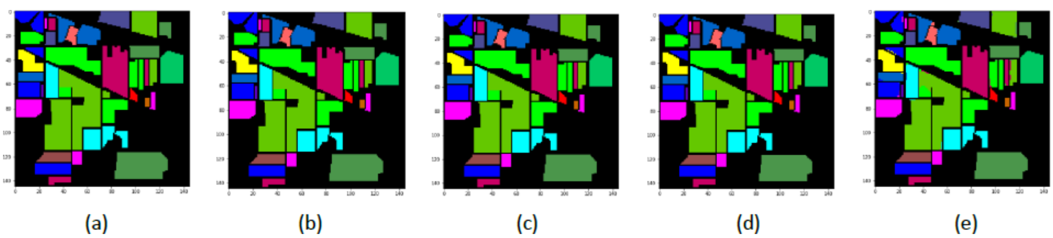

Fig.2. Classification map for Indian Pine. (a) Ground truth. (b)–(e) Predicted classification maps for PCA+3D-CNN(k=16), PCA+3D-CNN (k=32), PCA+3DCNN(k=64),LDA+3D-CNN respectively

	precision	recall	f1-score	support
Alfalfa	1.00	1.00	1.00	14
Corn-notill	1.00	1.00	1.00	428
Corn-mintill	1.00	1.00	1.00	249
Corn	1.00	1.00	1.00	71
Grass-pasture	1.00	0.99	0.99	145
Grass-trees	1.00	1.00	1.00	219
Grass-pasture-mowed	1.00	1.00	1.00	8
Hay-windrowed	1.00	1.00	1.00	143
Oats	1.00	1.00	1.00	6
Soybean-notill	1.00	1.00	1.00	292
Soybean-mintill	1.00	1.00	1.00	737
Soybean-clean	1.00	1.00	1.00	178
Wheat	1.00	1.00	1.00	61
Woods	1.00	1.00	1.00	380
Buildings-Grass-Trees-Drives	1.00	1.00	1.00	116
Stone-Steel-Towers	1.00	1.00	1.00	28
accuracy			1.00	3075
macro avg	1.00	1.00	1.00	3075
weighted avg	1.00	1.00	1.00	3075

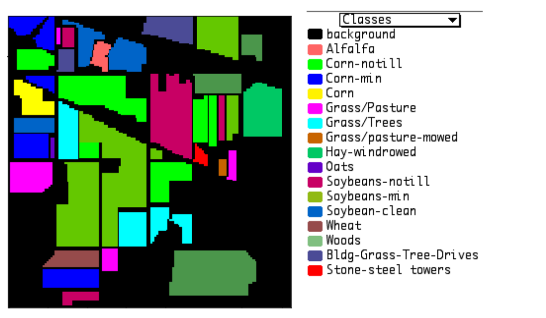

Fig.3. Classification Report which consists of the *Classwise Accuracy, Accuracy Precision, Recall, F1 Score*, and *Support for* Test ratio =0.3 ,Window size =25, pca component, K=32 and 16 classes of IP

Tasmia Jannat[1]*, Abu Sayeed; (jannat22tasmia@gmail.com)

The results of the proposed PCA+3D CNN and LDA+3D CNN model are compared with the most widely used supervised methods, such as SVM [14], 2-D-CNN [15], 3-D-CNN [16].

Table 2 shows impact of test ratio. When we train with bigger portion of the dataset, accuracies are comparatively better. We see accuracy falls with 80% test and only 20% train.

Table 3 shows the results in terms of the OA, AA and Kappa for different methods. It can be see that our proposed model outperforms them all. Both PCA+3D CNN and LDA+3D CNN work better than SVM, 2D-CNN and 3D-CNN. It is evident that 3-D or 2-D convolution alone is not able to represent the highly discriminative feature compared to our model.

TABLE 2
Impact of test ratio on classification accuracies of different method (In percentages) on IP

Method	Indian Pine dataset (test ratio=0.3, validation split=0.2)			Indian Pine dataset (test ratio=0.8, validation split=0.2)		
	OA	Kappa	AA	OA	Kappa	AA
PCA+3DCNN (k=16)	99.80	99.77	99.81	97.94	97.65	97.81
PCA+3DCNN (k=32)	99.90	99.88	99.82	97.74	97.42	98.03
PCA+3DCNN (k=64)	99.71	99.67	99.59	97.47	97.12	97.62
LDA+3dCNN	99.97	99.96	99.98	98.27	98.02	98.66

TABLE 3
Comparative Classification accuracies of different method (In percentages) on IP

Method	Indian Pine Dataset		
	OA (%)	Kappa (%)	AA (%)
SVM	85.30	83.10	79.03
2D-CNN	89.48	87.96	86.14
3D-CNN	91.10	89.98	91.58
PCA+3DCNN (k=16)	99.80	99.77	99.81
PCA+3DCNN (k=32)	99.90	99.88	99.82
PCA+3DCNN (k=64)	99.71	99.67	99.59
LDA+3dCNN	99.97	99.96	99.98

4. Conclusion

This paper has introduction 3D Convolutional Neural Network for classification with PCA and LDA as preprocessing step for reducing dimension. Our model works on both spatial and spectral features. We have used the Indian Pines dataset under experiments to reduce its dimensionality and check whether the classification accuracy gets better or not. In supervised LDA we used 16 bands (one less than the available class in the dataset) and in unsupervised PCA we used 16, 32, 64 bands respectively to check accuracy. In all cases, our model is computationally efficient than the 3D-CNN model. We got 99.8±0.1% accuracy in PCA and 3D CNN combination. In case of supervised LDA as preprocessing step, accuracy is 99.96±0.01%. We also compared our results with Support Vector Machine and 2D-CNN. Our model gets better result in both cases. We hope this may be a contribution to the fast-growing research interest in neural network in the field of hyperspectral image. In future we want to continue our research with other dataset.

References

[1] G. Hughes, “On the mean accuracy of statistical pattern recognizers,” IEEE Transactions on Information Theory, vol. 14, no. 1, pp. 55–63,1968.
[2] M. A. Hossain, X. Jia, and M. Pickering, “Subspace detection using a mutual information measure for hyperspectral image classification,” IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 2, pp. 424–428, 2014
[3] X. Jia, B. Kuo, and M. M. Crawford, “Feature mining for hyperspectral image classification,” Proceedings of the IEEE, vol. 101, no. 3, pp. 676–697, 2013.
[4] X. Bai, Z. Guo, Y. Wang, Z. Zhang, and J. Zhou, “Semisupervised hyperspectral band selection via spectral–spatial hypergraph model,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2774–2783, 2015.
[5] H. Hotelling, “Analysis of a complex of statistical variables into principal components.,” Journal of Educational Psychology, vol. 24, no. 7, p. 417–441, 1933.
[6] K. Fukunaga, Introduction to statistical pattern recognition, Academic Press, 1990.
[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.
[8] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems, pp. 91–99, 2015.
[9] X. Kang, B. Zhuo, and P. Duan, “Dual-path network-based hyperspectral image classification,” IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 3, pp. 447–451, 2018.
[10] Z. Zhong, J. Li, Z. Luo, and M. Chapman, “Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework,” IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 2, pp. 847–858, 2017.
[11] L. Van der Maaten, “An introduction to dimensionality reduction using matlab,” Report, vol. 1201, no. 07-07, p. 62, 2007.
[12] R. A. Fisher et al., “138: The use of multiple measurements in taxonomic problems.,” 1936.
[13] S. Ji, W. Xu, M. Yang, and K. Yu, “3d convolutional neural networks for human action recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 1, pp. 221–231, 2012.
[14] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” IEEE Transactions on geoscience and remote sensing, vol. 42, no. 8, pp. 1778–1790, 2004.
[15] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, “Deep supervised learning for hyperspectral data classification through convolutional neural networks,” in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4959–4962, IEEE, 2015.
[16] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, “3-d deep learning approach for remote sensing image classification,” IEEE Transactions on geoscience and remote sensing, vol. 56, no. 8, pp. 4420–4434.

Tasmia Jannat[1]*, Abu Sayeed; (jannat22tasmia@gmail.com)