International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Evaluation of deep learning Convolutional Neural Network for Crop Classification

Kavita Bhosle, Vijaya Musande

Abstract: ***In this paper, we have done exploratory experiments using deep learning convolutional neural network framework to classify crops into cotton, sugarcane and mulberry. In this contribution we have used Earth Observing-1 hyperion hyperspectral remote sensing data as the input. Structured data has been extracted from hyperspectral data using a remote sensing tool. An analytical assessment shows that convolutional neural network (CNN) gives more accuracy over classical support vector machine (SVM) and random forest methods. It has been observed that accuracy of SVM is 75 %, accuracy of random forest classification is 78 % and accuracy of CNN using Adam optimizer is 99.3 % and loss is 2.74 %. CNN using RMSProp also gives the same accuracy 99.3 % and the loss is 4.43 %. This identified crop information will be used for finding crop production and for deciding market strategies.***

Index Terms: Convolutional neural network, Hyperspectral remote sensing data, Random forest classifier, Support vector machine.

I. INTRODUCTION

This paper focuses on classification of crops using remote sensing images. As hyperspectral remote sensors have become available more recently, imagery from these sensors has been evaluated for crop identification and area estimation.[1][2] Using these high resolution images, land cover classification has been studied.[3][4] For spatial and spectral category of classification for hyperspectral imagery, 3D CNN has been implemented.[5][6] For crop discrimination, temporal data indices have been used by researcher.[7][8] Convolutional neural network architecture of deep learning has been used for crops and land cover classification from remote sensing image segment.[9] Very high resolution remote sensing images can be classified using CNN.[10] CNN is also used for scene and region based classification.[11] PCA plays a significant role to reduce the dimensions or number of bands of hyperspectral images.[12] Hyper spectral data has lots of information in hundreds of bands. In order to minimize the number of bands, PCA produces Eigen values and Eigen vectors. [13] CNN based flexible momentum with PCA and SVM has been used for hyperspectral data classification. [14] Optimizers have been evaluated for getting maximum accuracy using CNN. Stochastic Gradient Descent (SGD) has been recommended to use with nesterov for shallow networks. [15] Adam and RMSprop are used for deep learning networks. The Adam optimization algorithm is an extension to SGD to update network weights iteratively. [16][17]

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Kavita Bhosle*, Computer Science and Engineering, Department, Maharashtra Institute of Technology, Aurangabad, MH, India.
Dr. Vijaya Musande, Computer Science and Engineering Department, MGM's Jawaharlal Nehru Engineering College, Aurangabad, MH, India.

II. MATERIALS AND METHODS

A. Study Area

The study area lies between upper left corner latitude 20.310883, longitude 75.401182, upper right corner latitude 20.297426, longitude 75.472334, lower left corner latitude 19.377469, longitude 75.260141, lower right corner latitude 19.390566, longitude 75.189663 in Waregaon village, Aurangabad district of Maharashtra. Required data has been collected from Space borne Earth Observing -1 Hyperion, on Dec 24, 2015 in winter season, as weather is clear and non cloudy. In this season, the crop under study attains its middle stage of growth with enough foliage. Hence the crop gives clear reflectance. Same study area of Aurangabad district was used by researchers.

The 36 extracted principal components are given as an input to CNN, support vector machine and random forest classifier for crop classification as shown in Fig. 1. Accuracy assessment model has been implemented to compare the above mentioned classification methods. [18]

B. Preprocessing

Hyperspectral data is unstructured data. Many researchers have provided unstructured data as input to CNN. In this paper, this raw data has been preprocessed using ENVI tool. We have extracted features for the region of interest (ROI). These extracted features consist of atmospheric corrected 155 bands and their reflectance at each wavelength. As per previous study, crop can be identified by its reflectance values at each wavelength. Ground truth for each pixel has been collected by surveying the ROI.

Dimensionality reduction has been implemented using different methods. [19] Principal component analysis (PCA), a dimensionality minimization or reduction technique, has been used to minimize the number of bands in order to extract the useful information. Then cumulative variance for all principal components is calculated. It has been observed that out of 155 principal components, the first 36 yields more information. Cumulative variance for remaining components is constant.

C. Convolutional Neural Network (CNN)

60 % of the input data has been used to train the model, and 40 % data has been used for testing. We have implemented 2D CNN architecture with one convolution layer and one max pooling layer which is connected to the fully connected layer.

ReLU activation function has been used in convolution layer and softmax function has been used in fully connected layer. 36 channels and 1 X 1 filters have been used in the convolution layer. To achieve maximum possible accuracy, optimized parameters of CNN, such as learning rate, batch size, and optimizer have been set. Fine tuning search method has been used to set these parameters. [20][21]

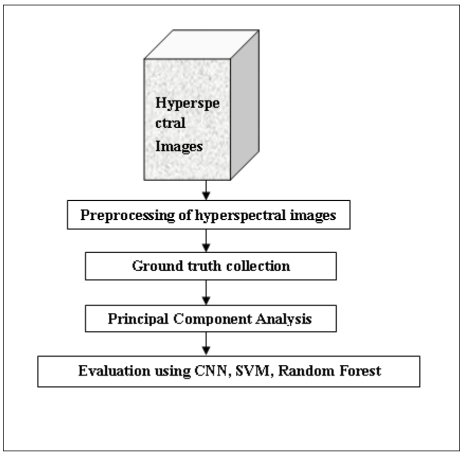

Fig 1 Flow of methods used

Given input k × k square neuron layer x convolved with m × m filter ω, convolutional layer produced output of size (k−m+1)×(k−m+1). Following equation has been used to calculate output of each convolutional layer.

$$F_{ab}^{k} = \sigma\left(\sum_{i=0}^{m-1}\sum_{j=0}^{m-1}\left(w_{ij}\, x_{(a+i)(b+j)}^{k-1}\right) + bias\right)$$

The max-pooling layer takes p x p region and given output a single value which is maximum for that region. If input layer is t x t layer, the output will be a (t/p) x (t/p) layer, as each p x p block is converted into a single value using the max function.

D. Optimizers

Stochastic Gradient Descent (SGD) is recommended to use SGD with Nesterov for shallow networks [15]. Momentum considers the past gradients to smooth out the steps of gradient descent. It can be applied with batch gradient descent. Adagrad makes small updates for frequent parameters and big updates for infrequent parameters and first published [16]. So it is good for dealing with sparse data. The main advantage of Adagrad is that we do not need to tune the learning rate manually. Default value of learning rate in Adagrad in most implementations is 0.01.
RMSProp uses a moving average of squared gradients to normalize the gradient itself. That has an effect of balancing the step size. It decreases the step for large gradient and increases the step for small gradient to avoid exploding and vanishing respectively [17]. Adam and RMSprop are used for deep learning networks. The Adam optimization algorithm is an extension to SGD to update network weights iteratively. It is adaptive moment estimation. We can say that *Adam is equivalent to RMSprop plus Momentum. It has been observed that it required r*elatively low memory. It works efficiently with little tuning of hyperparameters.

E. Learning Rate

The experiment designed by Girshick et al. shows that during fine tuning, some reducing value of learning rate optimizes the performance [21]. Nicholas Becherer et.al implemented optimization of CNN using fine tuning method [20]. In this study we have observed performance parameters of CNN by changing learning rate from 0 to 1.

After fine tuning, 0.01 learning rate given maximum accuracy and minimum loss.

F. Batch Size

CNN model is trained and tested using different batch size like 16, 32, 64, and 128. It is observed that with batch size 16, we obtained efficient result. Batch size 16 is efficient.

After fine tuning of learning rate, batch size and optimizer, we obtained maximum accuracy of 99.33 % and loss of 2.75%.

G. Support Vector Machine (SVM)

SVM has been implemented with polynomial kernel and without polynomial kernel. In processing high-resolution remote sensing images, Many researchers used classification algorithms, based on the support vector machine (SVM). [22][23] Algorithmic Implementation of Multiclass Kernel-based SVM was shown by M Crammer et. al. [24]

H. Random Forest Classifier

One of the ensemble learning method is random forests or random decision forests. It has been used for regression, classification and other tasks. At training time and predicting the class labels, it constructs a multitude of decision trees. [25][26]

III. RESULT AND DISCUSSIONS

As shown in Figure 1, PCA algorithm is implemented and evaluated on structured data obtained using ENVI remote sensing tool. It is observed that cumulative explained variance increased for first 36 principal components and then it remains constant as shown in Figure 2. CNN with 155 bands data without PCA has been evaluated for three crops. An accuracy of 53.45 % has been obtained. CNN is implemented by providing 36 principal components as an input. It has been observed that observed that accuracy increased up to 75.27 %.

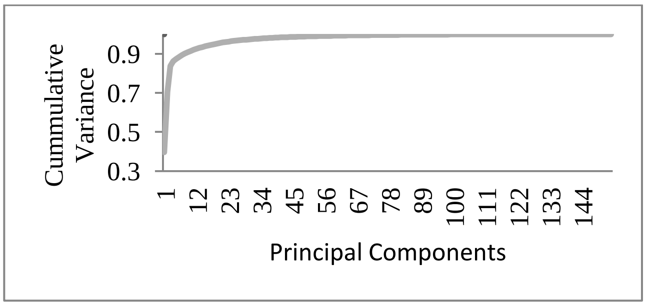

Fig 2 Cumulative Explained Variance using PCA

Retrieval Number: B2872078219/19©BEIESP
DOI: 10.35940/ijrte.B2872.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fine tuning search algorithm is implemented to obtain best learning rate, batch size and optimizer. It has been observed that CNN has given maximum accuracy of 99.3 % accuracy and 2.74 % loss using 0.01 learning rate, 16 batch size and Adam optimizer. RMSProp has given accuracy of 99.3 % with slightly more loss of 4.434 % by keeping all other parameter unchanged as shown in Table 1 and Table 2.

SVM with polynomial kernel and without polynomial kernel has given 75 % accuracy. Crop has been classified using random forest classifier with 78 % accuracy as shown in Table 3 and Figure 3.

Table- I: Comparison of the accuracy using different optimizers

Accuracy						
Optimizers Epochs	SGD	SGD, momentum=0.3	SGD momentum=0.3, nesterov=True	Adam	Adagrad	RMSprop
Learning rate= 0.01 Batch Size=16						
0	0.35	0.51	0.43	0.54	0.55	0.57
1	0.58	0.66	0.65	0.75	0.75	0.77
2	0.65	0.69	0.67	0.85	0.80	0.89
3	0.66	0.70	0.71	0.88	0.83	0.87
4	0.67	0.69	0.70	0.96	0.85	0.93

Table- II: Comparison of the loss using different optimizers

Loss						
Optimizers Epochs	SGD	SGD, momentum=0.3	SGD momentum=0.3, nesterov=True	Adam	Adagrad	RMSprop
Learning rate= 0.01 Batch Size=16						
0	1.3	1.0	1.1	1.1	1.0	1.2
1	0.9	0.9	0.8	0.5	0.6	0.5
2	0.8	0.8	0.8	0.3	0.5	0.3
3	0.8	0.8	0.7	0.2	0.4	0.3
4	0.7	0.7	0.7	0.2	0.4	0.2

Table- III: Comparison of the overall accuracy using different models

Methods	CNN without PCA	CNN with PCA	Fine tuned CNN with PCA	SVM with PCA	Random Forest Classifier with PCA
Overall Accuracy (%)	53.45	75.27	99.33	75.00	78.00

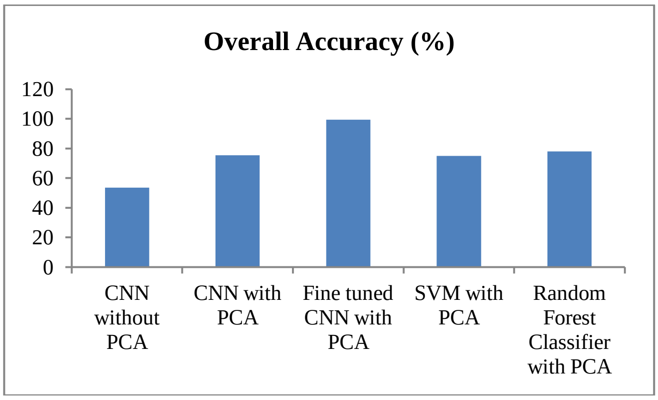

Fig 3 Overall accuracy of different methods

IV. CONCLUSION

This study demonstrated the accuracy of various methodologies for identification of cotton, sugarcane, and mulberry fields using hyperspectral data. CNN method has been compared with SVM and random forest classifiers and evaluated for crop identification. Results showed that cotton, sugarcane and mulberry fields could be accurately identified using CNN with PCA method. The first finding is that we can convert unstructured hyperspectral remote sensing images to structured data for evaluation of different methods. The second finding is that for hyperspectral data PCA played an important role for reducing dimensionality and for increasing accuracy. This research enhanced our understanding of crop identification and their role in agricultural production. This information will help in farm use planning with suitable crops and farming practices and that could lead to increased yield.

REFERENCES

1. John E. Ball, Derek T. Anderson, Chee Seng Chan, "Feature and Deep Learning in Remote Sensing Applications," *J. Appl. Remote Sens.* 11(4), 042601 (2018), doi: 10.1117/1.JRS.11.042601.
2. Chengming Zhang, Jiping Liu, Fan Yu, Shujing Wan, Yingjuan Han, Jing Wang, and Gang Wang, "Segmentation model based on convolutional neural networks for extracting vegetation from Gaofen-2 images", *J. Appl. Remote Sens.* 12(4), 042804 (2018), doi: 10.1117/1.JRS.12.042804
3. Vladimir Crnojevic, Predrag Lugonja, Branko Brkljač, Borislav Brunet, "Classification of small agricultural fields using combined Landsat-8 and RapidEye imagery: case study of northern Serbia". *Journal of Applied Remote Sensing* 083512-1 Vol. 8, 2014
4. Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data.*IEEE Geoscienc and Remote Sensing Letter*, VOL. 14, NO. 5, MAY 2017

5. Ying Li, Haokui Zhang, Qiang Shen. (2017). "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network". *Remote Sens*. 9, 67; doi:10.3390/rs9010067
6. Wenzhi Zhao, Zhou Guo, Jun Yue, Xiuyuan Zhang, and Liqun Luo.(2015). "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery". Taylor & Francis, *International Journal of Remote Sensing*, 2015. DOI:10.1080/2150704X.2015.1062157
7. Vijaya Musande & Anil Kumar & Karbhari Kale. (2012). "Cotton Crop Discrimination Using Fuzzy Classification Approach". *Springer J Indian Soc Remote Sens* 40(4):589–597.
8. Vijaya Musande , Anil Kumar, P.S. Roy, Karbhari Kale.(2015). "Evaluation of fuzzy-based classifiers for cotton crop identification". Taylor & Francis *Geocarto International*, 28:3, 243-257.
9. Bin Zhang, Yueyan Liu, Zuyu Zhang, Yonglin Shen, "Land use and land cover classification for rural residential areas in China using soft-probability cascading of multifeatures". *Journal of Applied Remote Sensing* 045010-1 Oct–Dec 2017, Vol. 11(4)
10. Tian Tian ·Lang Gao ·Weijing Song · Kim-Kwang Raymond Choo · Jijun He. (2017). "Feature extraction and classification of VHR images with attribute profiles and convolutional neural networks". *Springer Multimedia Tools Appl* https://doi.org/10.1007/s11042-017-5331-4
11. Jinying Zhong, Bin Yang, Guoyu Huang, Fei Zhong, Zhongze Chen (2016). "Remote Sensing Image Fusion with Convolutional Neural Network". *Springer Sens Imaging* DOI 10.1007/s11220-016-0135-6
12. G.F. Byrnep .F.crapperk.K.Mayo. (1980)."Monitoring land-cover change by principal component analysis of multitemporal landsat data". *Journal of Remote Sensing of Environment* Volume 10, Issue 3, Pages 175-184 https://doi.org/10.1016/0034-4257(80)90021-8
13. Wenzhi Zhao and Shihong Du.(2016). "Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach". *IEEE Transactions on Geoscience and Remote Sensing*, VOL. 54, NO. 8, 10.1109/TGRS.2016.2543748
14. Qi Yue & Caiwen Ma. (2018) . "Hyperspectral data classification based on flexible Momentum deep convolution neural network". *Springer Multimed Tools Appl* 77:4417–4429, DOI 10.1007/s11042-017-4734-6
15. Robbins H, Monro S 1951 A stochastic approximation method. *Ann Math Stat.* 22:400–407.
16. Duchi John, Hazan Elad, Singer Yoram (2011). "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". *Journal of Machine Learning Research* 12 (2011) 2121-2159
17. Tieleman et. al, Tijmen and Hinton, Geoffrey (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning
18. Dongmei Huang, Shoujue Xu, Jingqi Sun, Suling Liang, Wei Song, Zhenhua Wang, " Accuracy assessment model for classification result of remote sensing image based on spatial sampling". *Journal of Applied Remote Sensing* 046023-1 Oct–Dec 2017 • Vol. 11
19. S. Enayat Hosseini Aria, Massimo Menenti, Ben G. H. Gorte, "Spectral region identification versus individual channel selection in supervised dimensionality reduction of hyperspectral image data," *J. Appl. Remote Sens*. 11(4), 046010 (2017), doi: 10.1117/1.JRS.11.046010.
20. Nicholas Becherer, John Pecarina, Scott Nykl, Kenneth Hopkinson(2017). "Improving optimization of convolutional neural networks through parameter fine-tuning". Springer *Neural Computing and Applications*
21. Girshick R, Donahue J, Darrell T, Malik J (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation". Proceedings of the IEEE conference on *computer vision and pattern recognition*, pp 580–587
22. M. Ustuner, F. Balik Sanli, and B. Dixon, "Application of support vector machines for land use classification using high-resolution rapid eye images: a sensitivity analysis," *Eur. J. Remote Sens*. 48, 403–422 (2015).
23. X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens*. 51(1), 257–272 (2013).
24. Crammer, Koby & Singer, Yoram (2001). "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines". *Journal of Machine Learning Research*. 2: 265–292.
25. Ho, Tin Kam (1995). "Random Decision Forests". Proceedings of the 3rd International Conference on *Document Analysis and Recognition*, Montreal, QC, 14–16 August 1995. pp. 278–282.
26. Ho TK (1998). "The Random Subspace Method for Constructing Decision Forests". IEEE Transactions on *Pattern Analysis and Machine Intelligence*. 20 (8): 832–844. DOI 10.1109/34.709601.

.

AUTHORS PROFILE

Kavita Bhosle completed BE and ME and started her Phd in Computer Science and Engineering. She is working as Associate Professor in Computer Science and Engineering department at Maharashtra Institute of Technology, Aurangabad, MH, India. Her research areas are artificial neural network, deep learning, Remote sensing.

Dr. Vijaya Musande PhD in Computer Science and Engineering, Department of Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MH, India. She is working as Professor in Computer Science and Engineering Department of MGM's Jawaharlal Nehru Engineering College, Aurangabad, MH, India. Her research areas are Remote sensing, machine learning, deep learning.