A Fast 3D CNN for Hyperspectral Image Classification

Muhammad Ahmad,

Abstract—**Hyperspectral imaging (HSI) has been extensively utilized for a number of real-world applications. HSI classification (HSIC) is a challenging task due to high inter-class similarity, high intra-class variability, overlapping, and nested regions. A 2D Convolutional Neural Network (CNN) is a viable approach whereby HSIC highly depends on both Spectral-Spatial information, therefore, 3D CNN can be an alternative but highly computational complex due to the data volume and spectral dimensions. Furthermore, these models do not extract quality feature maps and may underperform over the regions having similar textures. Therefore, this work proposed a 3D CNN model that utilizes both spatial and spectral feature maps to attain good performance. In order to achieve the said performance, the HSI cube is first divided into small overlapping 3D patches. Later these patches are processed to generate 3D feature maps using a 3D kernel function over multiple contiguous bands which persevere the spectral information as well. The class labels are decided based on the central pixel. Benchmark HSI datasets (i.e., Pavia University, Pavia Centre, Salinas, Indian Pines, Kennedy Space Centre and Botswana) are considered to validate the performance of our proposed method. The results are further compared with several state-of-the-art methods.**

Index Terms—**3D Convolutional Neural Network (CNN); Kernel Function; Classification; Hyperspectral Images (HSI);**

I. INTRODUCTION

HYPERSPECTRAL SENSOR collects the information (reflectance) in several hundreds of contiguous bands with a very high spectral resolution which enables us to classify the objects based on their spectral signatures. However, these images are in relatively low spatial resolution due to the sensor limitations, SNR, and complexity constraints which significantly affect the performance for several real-world applications [1]. The traditional classifiers, for instance, KNN [2], SVM [3], Maximum Likelihood [4], Logistic Regression [2] and Extreme Learning Machine (ELM) [5] are only works based on spectral information. These classifiers do not perform well due to spectral redundancy and high correlation among the spectral bands. Furthermore, these classifiers fail to preserve the important spatial variability of Hyperspectral data which also results in low performance.

The simplest way to improve the classification performance is to design a classifier that should incorporate both spectral and spatial information. Spatial information is considered as additional discriminative information related to the size, shape, and structure of the object which if provided correctly brings more competitive results. Spatial spectral classifiers can generally be classified into two groups. First category

explores the spatial and spectral information separately. The spatial information is extracted in advance using entropy [6], morphological operations [7, 8], low rant representation [9] and attribute profiles [10]. Later this information is combined with spectral information to perform pixel-level classification.

The second category fuses the spatial-spectral information to get the joint features [11], for instance, 3D wavelet, scattering wavelet and Gabor filter [12, 13] are generated at different frequencies and scales to extract the joint spatial-spectral features for classification. Hyperspectral images are in 3D cubes thus the former category can result in a large number of feature cubes comprising key information about local signal changes in the spectrum, space, spatial-spectral correlations which are vital for good performance. However, the classical feature extraction methods are based on shallow learning and handcrafted features which largely depend on domain knowledge. Deep models have been proposed to address the aforementioned issues i.e., automatically learn low to high-level features from raw data which have achieved tremendous success for Hyperspectral Image Classification (HSIC).

The last few years witnessed an intensive improvement in Convolutional Neural Network (CNN) for HSIC where the spatial features are tailored by a 2D CNN model [14, 15, 16]. However, these spatial features are usually extracted separately which to some extent void the reason to jointly exploit the spatial-spectral information for HISC. Therefore, in this paper, a novel 3D CNN for HSIC method is proposed. This work first divides the HSI cube into small overlapping 3D patches. These patches are processed to generate 3D feature maps using 3D kernel function over multiple contiguous bands to preserve the joint spatial and spectral information for the feature learning process which exploits important discrimination information for HSIC. As a preprocessing, incremental Principle Component Analysis (iPCA) is deployed to reduce the redundancy among the bands to process the few important wavelengths out of the entire HSI cube. Later the 3D CNN classifier is trained in an end-to-end fashion which involves fewer parameters than other 2D/3D CNN models. The comparative study is also carried out with several state-of-the-art 2D/3D CNN based HSIC methods proposed in the literature. Experimental/comparative results revealed that the proposed method outperforms the compared ones.

The rest of the paper is structured as follows; section II presents the proposed methodology. Section III describes the experimental Datasets, Results and discussion. Finally Section IV concludes the paper with possible future research directions.

M. Ahmad is with the Department of Computer Engineering, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, 64200, Pakistan.
E-mail: mahmad00@gmail.com

II. PROPOSED METHODOLOGY

Let us assume a Hyperspectral dataset can be expressed as $X = [x_1, x_2, x_3, ..., x_L]^T \in R^{L \times (N \times M)}$ consisting of $N \times M$ samples associated with C classes per band with total L bands, in which each sample is represented as (x_i, y_j), where y_j is the class label of x_i sample. In a nutshell i^{th} sample belongs to j^{th} class. Since the HSI pixels exhibit high inter-class similarity, high intra-class variability, overlapping, and nested regions which required intensive efforts to handle for any classification model. To overcome the aforesaid issues, incremental Principle Component Analysis (iPCA) is applied to the HSI cube to eliminate the redundant bands. iPCA reduces the number of bands (L to B, where $B \ll L$) while maintaining the spatial dimensions as shown in Figure 1.
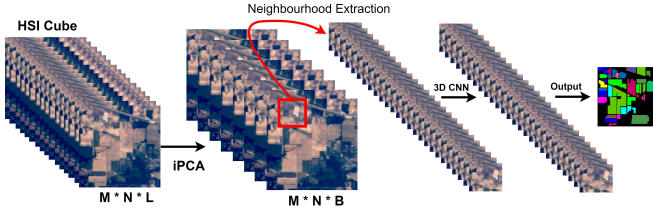

Fig. 1: Proposed 3D CNN Model for HSIC. 3D CNN Model details, i.e., the number of 3D Convectional and fully connected layers, can be found in Table I.

TABLE I: Layer based Summary of our Proposed 3D CNN Model architecture shown in Figure 2 with Window Size set as 11×11 for a sub-scene of Salinas Dataset.

Layer	Output Shape	# of Parameters
Input Layer	(11, 11, 20, 1)	0
Conv3D_1 (Conv3D)	(9, 9, 14, 8)	512
Conv3D_2 (Conv3D)	(7, 7, 10, 16)	5776
Conv3D_3 (Conv3D)	(5, 5, 8, 32)	13856
Conv3D_4 (Conv3D)	(3, 3, 6, 64)	55360
Flatten_1 (Flatten)	(3456)	0
Dense_1 (Dense)	(256)	884992
Dropout_1 (Dropout)	(256)	0
Dense_2 (Dense)	(128)	32896
Dropout_2 (Dropout)	(128)	0
Dense_3 (Dense)	(# of Classes)	774
Total Number of Trainable Parameters 994,166		

In ordered to pass the HSI cube to the model, it is divided into small overlapping 3D patches on which the ground labels are formed based on the central pixel as shown in Figure 2. The process creates neighboring patches $P \in R^{S \times S \times B}$ centered at the spatial location (α, β) covering $(S \times S)$ spatial window. The total of n patches given by $(M-S+1) \times (N-S+1)$. Thus, these patches cover the width from $\frac{\alpha-(S-1)}{2}$ to $\frac{\alpha+(S-1)}{2}$ and height from $\frac{\beta-(S-1)}{2}$ to $\frac{\beta+(S-1)}{2}$.

In 3D CNN, the input patches are convolved with 3D kernel function [17] which computes the sum of the dot product between input patch and kernel function. Later the convolved features are passed through the activation function which introduces the nonlinearity in the model. In this model, the activation values at spatial position (x, y, x) in the jth feature map of the i^{it} layer which is denoted as $v_{i,j}^{x,y,z}$ and generated as follows:
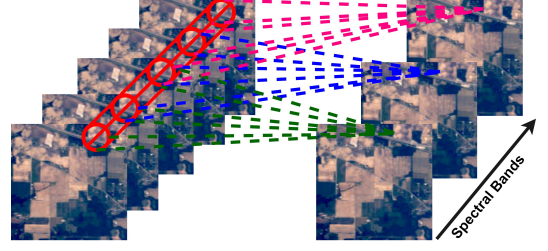

Fig. 2: 3D Convolution Operation

$$v_{i,j}^{x,y,z} = \phi\Big(\sum_{\tau=1}^{d_{l-1}} \sum_{\lambda=-\nu}^{\nu} \sum_{\rho=-\gamma}^{\gamma} \sum_{\phi=-\delta}^{\delta} w_{i,j,\tau}^{\nu,\rho,\lambda} \times$$
$$\times v_{(i-1),\tau}^{(x+\nu),(y+\rho),(z+\lambda)} + b_{i,j}\Big) \quad (1)$$

where ϕ is an activation function, $b_{i,j}$ is the bias, d_{l-1} be the number of feature maps at $(l-1)^{th}$ layer and $w_{I,j}$ be the depth of the kernel, $2\lambda + 1$, $2\delta + 1$ and $2\nu + 1$ be the width, height and depth of the kernel.

In a nutshell, the proposed 3D CNN convolutional kernels are as follows: $3D\_conv\_layer1 = 8 \times 3 \times 3 \times 7 \times 1$ where $K_1^1 = 3, K_2^1 = 3$ and $K_3^1 = 7$. $3D\_conv\_layer2 = 16 \times 3 \times 3 \times 5 \times 8$ where $K_1^2 = 3, K_2^2 = 3$ and $K_3^2 = 5$. $3D\_conv\_layer3 = 32 \times 3 \times 3 \times 3 \times 16$ where $K_1^3 = 3, K_2^3 = 3$ and $K_3^3 = 3$ and finally $3D\_conv\_layer4 = 64 \times 3 \times 3 \times 3 \times 16$ where $K_1^3 = 3, K_2^3 = 3$ and $K_3^3 = 3$. To increase the number of spatial-spectral feature maps, 4 3D convolutional layers are deployed before the flatten layer to make sure the model is able to discriminate the spatial information within different spectral bands without any loss. The further details regarding the proposed model can be found in Table I. The total number of parameters (i.e., tune-able weights) of our proposed 3D CNN model is 994, 166. The weights are initially randomized and optimized using Adam optimizer back-propagation with a soft-max loss function. The weights are updated using a mini-batch of size 256 with 50 epochs without batch normalization and augmentation.

III. EXPERIMENTAL DATASETS AND RESULTS

The Salinas dataset (SD) was acquired over Salinas Valley California using AVIRIS sensor. SD is of size $512 \times 217 \times 224$ with a 3.7 meter spatial resolution with 512×217 is spatial and 224 spectral dimensions. SD consists of vineyard fields, vegetables and bare soils. SD consist of 16 classes. A few water absorption bands $108 - 112, 154 - 167$ and 224 are removed before analysis.

Indian Pines Dataset (IPD) is obtained over northwestern Indiana's test site by Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) sensor. IPD is of size $145 \times 145 \times 224$ in the wavelength range $0.4 - 2.5 \times 10^{-6}$ meters where 145×145 is the spatial and 224 spectral dimensions. IPD consists of $1/3$ forest and $2/3$ agriculture area and other naturally evergreen vegetation. Some corps in the early stages of their growth is also present with approximately less than 5% of total coverage. Low-density housing, building and small

roads, Two dual-lane highway and a railway line are also a part of IPD. The IPD ground truth comprised of 16 classes which are not mutually exclusive. The water absorption bands have been removed before the experiments thus the remaining 200 bands are used in this experiment.

Pavia University Dataset (PUD) gathered over Pavia in northern Italy using a Reflective Optics System Imaging Spectrometer (ROSIS) optical sensor. PUD consists of 610×610 spatial and 103 spectral bands with a spatial resolution of 1.3 meters. PUD ground truth classes are 9. Further details about the experimental datasets can be found at [18]. The ground images of all the experimental datasets are shown in Figure 3.

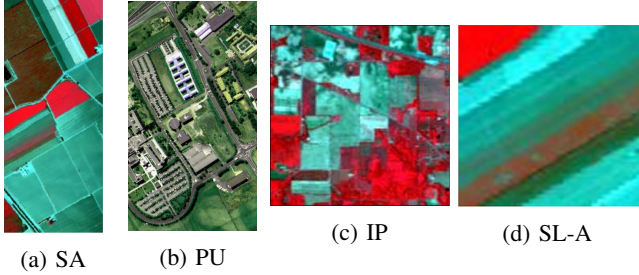(a) SA (b) PU (c) IP (d) SL-A

Fig. 3: Ground images of experimental datasets used in this work.

All the experiments were performed on an online platform known as Google Colab [19]. Google Colab is an online platform that requires a good speed of internet to run any environment. Google Colab provides an option to execute the codes on python 3 notebook with Graphical Processing Unit (GPU), 25 GB of Random Access Memory (RAM) and 358.27 GB of could storage for data computation. In all the experiments, the initial Test/Train set is divided into a 30/70% ratio on which Training samples (70% of the entire population) are further divided into 50/50% for the Training and Validation set.

To make the fair comparisons, the learning rate for all the experiments is set to 0.001, $relu$ as an activation function is used for all layers except last on which $softmax$ is used, the patch sizes are set as $11 \times 11 \times 20$, $13 \times 13 \times 20$, $15 \times 15 \times 20$, $17 \times 17 \times 20$, $19 \times 9 \times 20$, $21 \times 21 \times 20$ and $25 \times 25 \times 20$, respectively with 20 most informative bands selected by iPCA method. For evaluation purposes, Average Accuracy (AA), Overall Accuracy (OA) and Kappa (κ) coefficient have been computed form the confusion matrices. AA represents the average class-wise classification performance, OA is computed as the number of correctly classified examples out of the total test examples and finally, κ is known as a statistical metric that considered the mutual information regarding a strong agreement among classification and ground-truth maps. Along with OA, AA and κ metrics, several statistical tests are also being considered such as F1-Score, Precision and Recall.

The convergence loss and accuracy of our proposed 3D CNN model for a 50 number of epochs are shown in Figure 4. From these figures, one can conclude that the proposed model is converged almost around 325 echos. Whereas, the computational time of our proposed model is shown in Table II which also reveals a fast convergence and computational
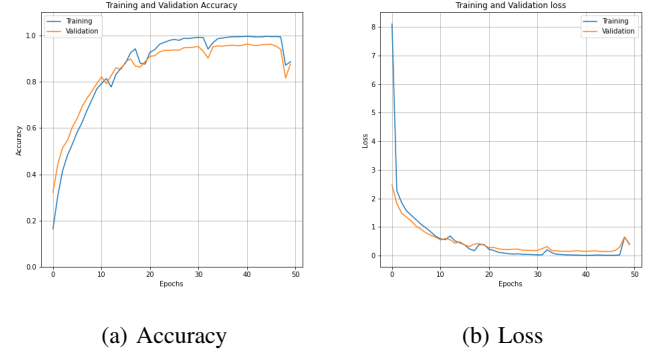

(a) Accuracy (b) Loss

Fig. 4: Accuracy and Loss for Training and Validation sets on Indian Pines Dataset with 11×11 window patch corresponds to the 50 number of Epochs.

efficiency of our proposed model. The computational time highly depends on the speed of the internet and available RAM.

TABLE II: Computational time in minutes for all the experimental datasets with several window sizes.

Dataset	Window Size						
	11×11	13×13	15×15	17×17	19×19	21×21	25×25
SL-A	0.22	0.23	0.56	0.28	0.98	0.37	0.45
SL	1.34	1.41	1.60	2.00	3.17	2.63	3.52
IP	0.33	0.33	0.61	0.78	0.62	0.58	0.76
PU	2.16	5.26	1.35	2.00	2.46	2.14	2.83

The accuracy analysis i.e., OA, AA, and κ based on the impact of spatial dimensions [1] processed by the proposed model is presented in Table III. While looking into the Table III, one can conclude that the window size of 11×11 is enough for Pavia University, Salinas and Salinas-A dataset whereas the window size of 13×13 and 25×25 both works almost the same.

Furthermore, the classification maps (geographical locations for each class) according to the different number of window sizes (spatial dimensions) are shown in Figures 6-8. In regards to comparison, the proposed model is compared with several state-of-the-art methods published in the recent few years. From experimental results listed in Table IV one can conclude that the proposed model has competitive results and to some extent better in regards to the other methods. The comparative methods includes Multi-scale-3D-CNN [20], 3D/2D-CNN [17, 21, 22, 23]. From experiments listed in Table IV shows

TABLE III: Impact of window size on our proposed model

Window	PU			IP			SA			SL-A		
	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ
11×11	99.94	99.89	99.92	88.65	83.52	87.11	99.80	99.91	99.78	100	100	100
13×13	99.81	99.65	99.75	95.38	94.14	94.72	99.93	99.94	99.93	100	100	100
15×15	99.85	99.62	99.80	93.69	93.09	92.79	99.99	99.99	99.99	100	100	100
17×17	99.05	98.49	98.75	91.80	91.74	90.62	99.95	99.97	99.95	99.93	99.93	99.92
19×19	99.93	99.78	99.91	93.13	93.42	92.15	98.04	94.02	97.81	100	100	100
21×21	99.78	99.43	99.72	94.34	91.31	93.52	99.99	99.99	99.99	100	100	100
25×25	98.79	97.67	98.39	97.75	96.17	97.44	99.96	99.93	99.95	100	100	100

the proposed method improves the results significantly then the state-of-the-art methods with even less number of training samples.

---

[1] The Confusion matrices (with per class accuracy) for each window size and every dataset is provided in the supplementary material.
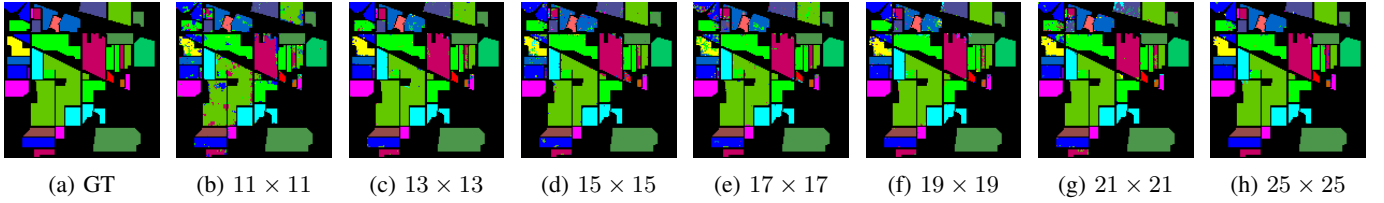
| (a) GT | (b) 11×11 | (c) 13×13 | (d) 15×15 | (e) 17×17 | (f) 19×19 | (g) 21×21 | (h) 25×25 |

Fig. 5: **Indian Pines Dataset** Ground Truths for each spatial dimensions processed through our proposed model.

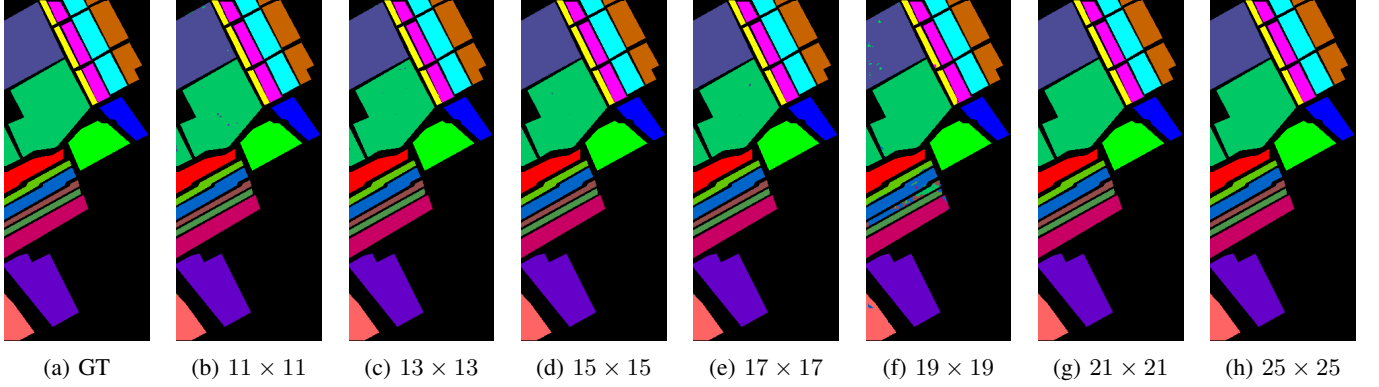| (a) GT | (b) 11×11 | (c) 13×13 | (d) 15×15 | (e) 17×17 | (f) 19×19 | (g) 21×21 | (h) 25×25 |

Fig. 6: **Salinas Dataset** Ground Truths for each spatial dimensions processed through our proposed model.

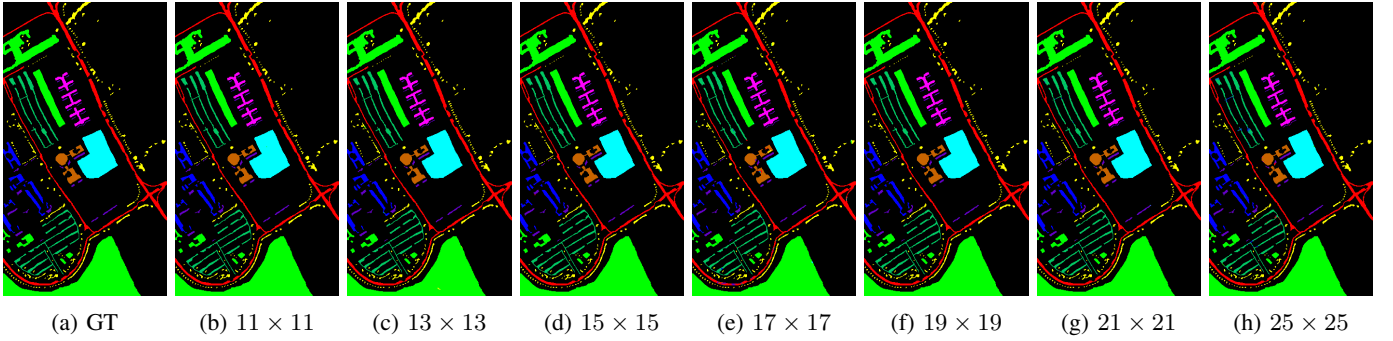| (a) GT | (b) 11×11 | (c) 13×13 | (d) 15×15 | (e) 17×17 | (f) 19×19 | (g) 21×21 | (h) 25×25 |

Fig. 7: **Pavia University Dataset** Ground Truths for each spatial dimensions processed through our proposed model.

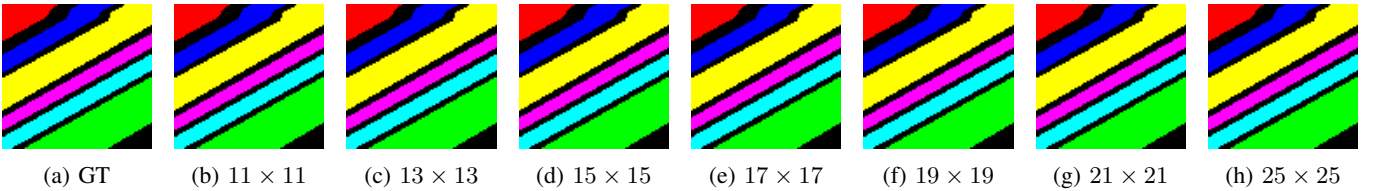| (a) GT | (b) 11×11 | (c) 13×13 | (d) 15×15 | (e) 17×17 | (f) 19×19 | (g) 21×21 | (h) 25×25 |

Fig. 8: **Salinas-A Dataset** Ground Truths for each spatial dimensions processed through our proposed model.

TABLE IV: Comparative evaluations with State-of-the-art methods while considering 11×11 Spatial dimensions and 10% of training samples.

dataset	Multi-scale-3D-CNN			3D-CNN			2D-CNN			Ours		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
PU	95.95	97.52	93.40	96.34	97.03	94.90	96.63	94.84	95.53	**98.40**	**97.89**	**97.89**
IP	81.39	75.22	81.20	82.62	76.51	79.25	80.27	68.32	75.26	**97.75**	**94.54**	**97.44**
SA	94.20	96.66	93.61	85.00	89.63	83.20	96.34	94.36	95.93	**98.06**	**98.80**	**97.85**

IV. CONCLUSION

Hyperspectral Image Classification (HSIC) is a challenging task due to high inter-class similarity and high intra-class variability. Therefore, this paper proposed a lightweight fast 3D CNN model which not only overcome the abovesaid challenges but also provide state of the art experimental results in a computationally efficient fashion on three benchmark Hyperspectral datasets. The inter-class similarity and high intra-class variability issues are being resolved using a spatial-spectral information using 3D convolutions. The experimental results reveal that the proposed method outperformed the state of the art methods, furthermore, the proposed model is less complex than the conventional 3D CNN models.

RUNNING CODE

The Running Demo/code can be found at ''''.

REFERENCES

[1] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sensing*, vol. 9, p. 1139, 11 2017.

[2] M. Ahmad, A. Khan, A. M. Khan, M. Mazzara, S. Distefano, A. Sohaib, and O. Nibouche, "Spatial prior fuzziness pool-based interactive classification of hyperspectral images," *Remote Sensing*, vol. 11, no. 9, May. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/9/1136

[3] Y. Wang, W. Yu, and Z. Fang, "Multiple kernel-based svm classification of hyperspectral images by combining spectral, spatial, and semantic information," *Remote Sensing*, vol. 12, no. 1, p. 120, 2020.

[4] A. Alcolea, M. E. Paoletti, J. M. Haut, J. Resano, and A. Plaza, "Inference in supervised spectral classifiers for on-board hyperspectral imaging: An overview," *Remote Sensing*, vol. 12, no. 3, p. 534, 2020.

[5] M. Ahmad, S. Shabbir, D. Oliva, M. Mazzara, and S. Distefano, "Spatial-prior generalized fuzziness extreme learning machine autoencoder-based active learning for hyperspectral image classification," *Optik-International Journal for Light and Electron Optics*, 2020.

[6] D. Tuia, M. Volpi, M. D. Mura, A. Rakotomamonjy, and R. Flamary, "Automatic feature learning for spatio-spectral image classification with sparse svm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 6062–6074, 2014.

[7] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2335–2353, 2015.

[8] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.

[9] S. Jia, X. Zhang, and Q. Li, "Spectral–spatial hyperspectral image classification using $\ell_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2473–2484, 2015.

[10] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 542–546, 2011.

[11] Y. Zhong, A. Ma, and L. Zhang, "An adaptive memetic fuzzy clustering algorithm with spatial information for remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1235–1248, 2014.

[12] L. Shen and S. Jia, "Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 5039–5046, 2011.

[13] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2276–2291, 2013.

[14] Y. Li and L. He, "An improved hybrid CNN for hyperspectral image classification," in *Eleventh International Conference on Graphics and Image Processing (ICGIP 2019)*, Z. Pan and X. Wang, Eds., vol. 11373, International Society for Optics and Photonics. SPIE, 2020, pp. 485 – 490. [Online]. Available: https://doi.org/10.1117/12.2557384

[15] B. Fang, Y. Bai, and Y. Li, "Combining spectral unmixing and 3d/2d dense networks with early-exiting strategy for hyperspectral image classification," *Remote Sensing*, vol. 12, p. 779, 02 2020.

[16] L. Huang and Y. Chen, "Dual-path siamese cnn for hyperspectral image classification with limited training samples," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.

[17] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, p. 67, 01 2017.

[18] *Hyperspectral Datasets Description*, 2020 (accessed 2020-01-12), http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[19] T. Carneiro, R. V. M. Da Nóbrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. Reboucas Filho, "Performance analysis of google colaboratory as a tool for accelerating deep learning applications," *IEEE Access*, vol. 6, pp. 61 677–61 685, 2018.

[20] M. He, B. Li, and H. Chen, "Multi-scale 3d deep convolutional neural network for hyperspectral image classification," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3904–3908.

[21] Y. Luo, J. Zou, C. Yao, T. Li, and G. Bai, "HSI-CNN: A novel convolution neural network for hyperspectral image," *CoRR*, vol. abs/1802.10478, 2018. [Online]. Available: http://arxiv.org/abs/1802.10478

[22] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben 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, 2018.

[23] B. Liu, X. Yu, P. Zhang, X. Tan, A. Yu, and Z. Xue, "A semi-supervised convolutional neural network for hyperspectral image classification," *Remote Sensing Letters*, vol. 8, no. 9, pp. 839–848, 2017. [Online]. Available: https://doi.org/10.1080/2150704X.2017.1331053