HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification

Swalpa Kumar Roy, *Student Member, IEEE,* Gopal Krishna, Shiv Ram Dubey, *Member, IEEE,* and Bidyut B. Chaudhuri, *Life Fellow, IEEE*

arXiv:1902.06701v3 [cs.CV] 3 Jul 2019

Abstract—Hyperspectral image (HSI) classification is widely used for the analysis of remotely sensed images. Hyperspectral imagery includes varying bands of images. Convolutional Neural Network (CNN) is one of the most frequently used deep learning based methods for visual data processing. The use of CNN for HSI classification is also visible in recent works. These approaches are mostly based on 2D CNN. Whereas, the HSI classification performance is highly dependent on both spatial and spectral information. Very few methods have utilized the 3D CNN because of increased computational complexity. This letter proposes a Hybrid Spectral Convolutional Neural Network (HybridSN) for HSI classification. Basically, the HybridSN is a spectral-spatial 3D-CNN followed by spatial 2D-CNN. The 3D-CNN facilitates the joint spatial-spectral feature representation from a stack of spectral bands. The 2D-CNN on top of the 3D-CNN further learns more abstract level spatial representation. Moreover, the use of hybrid CNNs reduces the complexity of the model compared to 3D-CNN alone. To test the performance of this hybrid approach, very rigorous HSI classification experiments are performed over Indian Pines, Pavia University and Salinas Scene remote sensing datasets. The results are compared with the state-of-the-art hand-crafted as well as end-to-end deep learning based methods. A very satisfactory performance is obtained using the proposed HybridSN for HSI classification. The source code can be found at https://github.com/gokriznastic/HybridSN.

Index Terms—Deep Learning, Convolutional Neural Networks, Spectral-Spatial, 3D-CNN, 2D-CNN, Remote Sensing, Hyperspectral Image Classification, HybridSN.

I. INTRODUCTION

THE research in hyperspectral image analysis is important due to its potential applications in real life [1]. Hyperspectral imaging results in multiple bands of images which makes the analysis challenging due to increased volume of data. The spectral, as well as the spatial correlation between different bands convey useful information regarding the scene of interest. Recently, Camps-Valls et al. have surveyed the advances in hyperspectral image (HSI) classification [2]. The HSI classification is tackled in two ways, one with hand-designed feature extraction technique and another with learning based feature extraction technique.

Several HSI classification approaches have been developed using the hand-designed feature description [3], [4]. Yang and Qian have proposed a joint collaborative representation by using the locally adaptive dictionary [3]. It reduces the adverse impact of useless pixels and improves the HSI classification performance. Fang et al. have utilized the local covariance matrix to encode the relationship between different spectral bands [4]. They used these matrices for HSI training and classification using Support Vector Machine (SVM). A composite kernel is used to combine spatial and spectral information for HSI classification [5]. Li et al. have applied the learning over the combination of multiple features for the classification of hyperspectral scenes [6]. Some other hand crafted approaches are Joint Sparse Model and Discontinuity Preserving Relaxation [7], Boltzmann Entropy-Based Band Selection [8], Sparse Self-Representation [9], Fusing Correlation Coefficient and Sparse Representation [10], Multiscale Superpixels and Guided Filter [11], and etc.

Recently, the Convolutional Neural Network (CNN) has become very popular due to drastic performance gain over the hand-designed features [12]. The CNN has shown very promising performance in many applications where visual information processing is required, such as image classification [13], [14], object detection [15], semantic segmentation [16], colon cancer classification [17], depth estimation [18], face anti-spoofing [19], etc. In recent years, a huge progress is also made in deep learning for hyperspectral image analysis. A dual-path network (DPN) by combining the residual network and dense convolutional network is proposed for the HSI classification [20]. Yu et al. have proposed a greedy layer-wise approach for unsupervised training to represent the remote sensing images [21]. Li et al. introduced a pixel-block pair (PBP) based data augmentation technique to generalize the deep learning for HSI classification [22]. Song et al. have proposed deep feature fusion network [23] and Cheng et al. have used the off-the-shelf CNN models for HSI classification [24]. Basically, they extracted the hierarchical deep spatial features and used with SVM for training and classification. Recently, the low power consuming hardwares for deep learning based HSI classification is also explored [25]. Chen et al. have used the deep feature extraction of 3D-CNN for HSI classification [26]. Zhong et al. have proposed the spectral-spatial residual network (SSRN) [27]. The residual blocks in SSRN use the identity mapping to connect every other 3-D convolutional layer. Mou et al. have investigated the residual conv-deconv network, an unsupervised model, for HSI classification [28]. Recently, Paoletti et al. have proposed the Deep Pyramidal Residual Networks (DPRN) specially for the HSI data [29]. Very recently, Paoletti et al. have also proposed spectral-spatial

S.K. Roy and G. Krishna are with Computer Science and Engineering Department at Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal-735102, India (email: swalpa@cse.jgec.ac.in; gk1948@cse.jgec.ac.in).

S.R. Dubey is with Computer Vision Group, Indian Institute of Information Technology, Sri City, Chittoor, Andhra Pradesh-517646, India (e-mail: srdubey@iiits.in).

B.B. Chaudhuri is with Computer Vision and Pattern Recognition Unit at Indian Statistical Institute, Kolkata-700108, India (email: bbc@isical.ac.in).
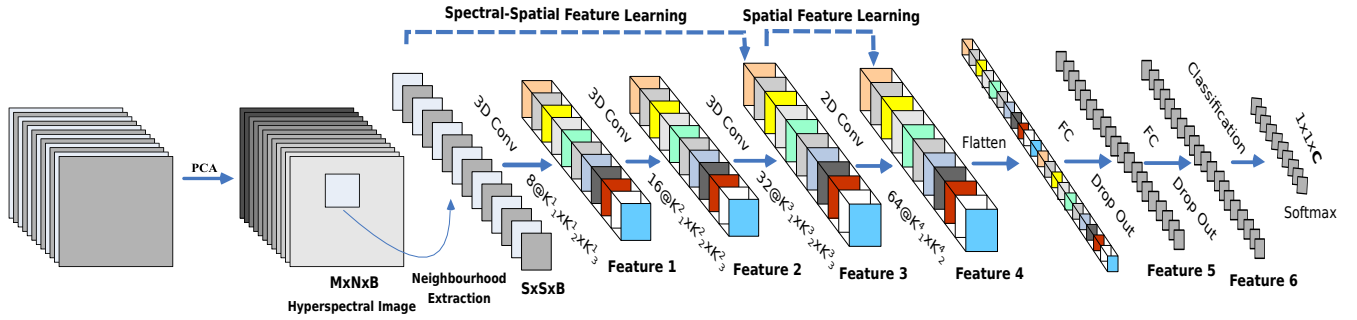
Fig. 1: Proposed HybridSpectralNet (HybridSN) Model which integrates 3D and 2D convolutions for hyperspectral image (HSI) classification.

capsule networks to learn the hyperspectral features [30], whereas Fang et al. introduced deep hashing neural networks for hyperspectral image feature extraction [31].

It is evident from the literature that using just 2D-CNN or 3D-CNN had a few shortcomings such as missing channel relationship information or very complex model, respectively. It also prevented these methods from achieving a better accuracy on hyperspectral images. The main reason is due to the fact that hyperspectral images are volumetric data and have a spectral dimension as well. The 2D-CNN alone isn't able to extract good discriminating feature maps from the spectral dimensions. Similarly, a deep 3D-CNN is more computationally complex and alone seems to perform worse for classes having similar textures over many spectral bands. This is the motivation for us to propose a hybrid-CNN model which overcomes these shortcomings of the previous models. The 3D-CNN and 2D-CNN layers are assembled for the proposed model in such a way that they utilise both the spectral as well as spatial feature maps to their full extent to achieve maximum possible accuracy.

This letter proposes the HybridSN in Section 2; presents the experiments and analysis in Section 3; and highlights the concluding remarks in section 4.

II. PROPOSED HYBRIDSN MODEL

Let the spectral-spatial hyperspectral data cube be denoted by $\mathbf{I} \in \mathcal{R}^{M \times N \times D}$, where $\mathbf{I}$ is the original input, M is the width, N is the height, and D is the number of spectral bands/depth. Every HSI pixel in $\mathbf{I}$ contains D spectral measures and forms a one-hot label vector $Y = (y_1, y_2, \ldots y_C) \in \mathcal{R}^{1 \times 1 \times C}$, where C represents the land-cover categories. However, the hyperspectral pixels exhibit the mixed land-cover classes, introducing the high intra-class variability and inter-class similarity into $\mathbf{I}$. It is of great challenge for any model to tackle this problem. To remove the spectral redundancy first the traditional principal component analysis (PCA) is applied over the original HSI data ($\mathbf{I}$) along spectral bands. The PCA reduces the number of spectral bands from D to B while maintaining the same spatial dimensions (i.e., width M and height N). We have reduced only spectral bands such that it preserves the spatial information which is very important for recognising any object. We represent the PCA reduced data cube by $\mathbf{X} \in \mathcal{R}^{M \times N \times B}$, where $\mathbf{X}$ is the modified input after PCA, M is the width, N is the height, and B is the number of spectral bands after PCA.

TABLE I: The layer wise summary of the proposed $HybridSN$ architecture with window size 25×25. The last layer is based on the Indian Pines dataset.

Layer (type)	Output Shape	# Parameter
input_1 (InputLayer)	(25, 25, 30, 1)	0
conv3d_1 (Conv3D)	(23, 23, 24, 8)	512
conv3d_2 (Conv3D)	(21, 21, 20, 16)	5776
conv3d_3 (Conv3D)	(19, 19, 18, 32)	13856
reshape_1 (Reshape)	(19, 19, 576)	0
conv2d_1 (Conv2D)	(17, 17, 64)	331840
flatten_1 (Flatten)	(18496)	0
dense_1 (Dense)	(256)	4735232
dropout_1 (Dropout)	(256)	0
dense_2 (Dense)	(128)	32896
dropout_2 (Dropout)	(128)	0
dense_3 (Dense)	(16)	2064

Total Trainable Parameters: $5, 122, 176$

In order to utilize the image classification techniques, the HSI data cube is divided into small overlapping 3D-patches, the truth labels of which are decided by the label of the centred pixel. We have created the 3D neighboring patches $P \in \mathcal{R}^{S \times S \times B}$ from $\mathbf{X}$, centered at the spatial location (α, β), covering the $S \times S$ window or spatial extent and all B spectral bands. The total number of generated 3D-patches (n) from X is given by $(M - S + 1) \times (N - S + 1)$. Thus, the 3D-patch at location (α, β), denoted by $P_{\alpha, \beta}$, covers the width from $\alpha - (S - 1)/2$ to $\alpha + (S - 1)/2$, height from $\beta - (S - 1)/2$ to $\beta + (S - 1)/2$ and all B spectral bands of PCA reduced data cube X.

In 2D-CNN, the input data are convolved with 2D kernels. The convolution happens by computing the sum of the dot product between input data and kernel. The kernel is strided over the input data to cover full spatial dimension. The convolved features are passed through the activation function to introduce the non-linearity in the model. In 2D convolution, the activation value at spatial position (x, y) in the jth feature map of the ith layer, denoted as $v_{i,j}^{x,y}$, is generated using the following equation,

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

where ϕ is the activation function, $b_{i,j}$ is the bias parameter for the jth feature map of the ith layer, d_{l-1} is the number of feature map in $(l-1)^{th}$ layer and the depth of kernel $w_{i,j}$ for the jth feature map of the ith layer, $2\gamma + 1$ is the width

(1) Overall Accuracy
(2) Kappa ,Kappa [-1,1] O
(3) average accuracy(AA),AA recall O

TABLE II: The classification accuracies (in percentages) on Indian Pines, University of Pavia, and Salinas Scene datasets using proposed and state-of-the-art methods.

Methods	Indian Pines Dataset			University of Pavia Dataset			Salinas Scene Dataset		
	OA	Kappa	AA	OA	Kappa	AA	OA	Kappa	AA
SVM	85.30 ± 2.8	83.10 ± 3.2	79.03 ± 2.7	94.34 ± 0.2	92.50 ± 0.7	92.98 ± 0.4	92.95 ± 0.3	92.11 ± 0.2	94.60 ± 2.3
2D-CNN	89.48 ± 0.2	87.96 ± 0.5	86.14 ± 0.8	97.86 ± 0.2	97.16 ± 0.5	96.55 ± 0.0	97.38 ± 0.0	97.08 ± 0.1	98.84 ± 0.1
3D-CNN	91.10 ± 0.4	89.98 ± 0.5	91.58 ± 0.2	96.53 ± 0.1	95.51 ± 0.2	97.57 ± 1.3	93.96 ± 0.2	93.32 ± 0.5	97.01 ± 0.6
M3D-CNN	95.32 ± 0.1	94.70 ± 0.2	96.41 ± 0.7	95.76 ± 0.2	94.50 ± 0.2	95.08 ± 1.2	94.79 ± 0.3	94.20 ± 0.2	96.25 ± 0.6
SSRN	99.19 ± 0.3	99.07 ± 0.3	98.93 ± 0.6	99.90 ± 0.0	99.87 ± 0.0	99.91 ± 0.0	99.98 ± 0.1	99.97 ± 0.1	99.97 ± 0.0
HybridSN	99.75 ± 0.1	99.71 ± 0.1	99.63 ± 0.2	99.98 ± 0.0	99.98 ± 0.0	99.97 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0

of kernel, $2\delta + 1$ is the height of kernel, and $w_{i,j}$ is the value of weight parameter for the jth feature map of the ith layer.

The 3D convolution [32] is done by convolving a 3D kernel with the 3D-data. In the proposed model for HSI data, the feature maps of convolution layer are generated using the 3D kernel over multiple contiguous bands in the input layer; this captures the spectral information. In 3D convolution, the activation value at spatial position (x, y, z) in the jth feature map of the ith layer, denoted as $v_{i,j}^{x,y,z}$, is generated as follows,

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

(2)

where $2\eta + 1$ is the depth of kernel along spectral dimension and other parameters are the same as in (Eqn. 1).

The parameters of CNN, such as the bias b and the kernel weight w, are usually trained using ~~supervised approaches~~ [12] with the help of ~~a gradient descent~~ optimization technique. In conventional 2D CNNs, the convolutions are applied over the spatial dimensions only, covering all the feature maps of the previous layer, to compute the 2D discriminative feature maps. ~~Whereas~~, for the HSI classification problem, it is desirable to capture the spectral information, encoded in multiple bands along with the spatial information. The 2D-CNNs are not able to handle the spectral information. ~~On the other hand~~, the 3D-CNN kernel can extract the spectral and spatial feature representation simultaneously from HSI data, but at the cost of increased computational complexity. ~~In order to~~ take the advantages of the automatic feature learning capability of both 2D and 3D CNN, we propose a hybrid feature learning framework called $HybridSN$ for HSI classification. The flow diagram of the proposed $HybridSN$ network is shown in Fig. 1. It comprises of three 3D convolutions (Eqn. 2), one 2D convolution (Eqn. 1) and three fully connected layers.

In $HybridSN$ framework, the dimensions of 3D convolution kernels are $8 \times 3 \times 3 \times 7 \times 1$ (i.e., $K_1^1 = 3$, $K_2^1 = 3$, and $K_3^1 = 7$ in Fig. 1), $16 \times 3 \times 3 \times 5 \times 8$ (i.e., $K_1^2 = 3$, $K_2^2 = 3$, and $K_3^2 = 5$ in Fig. 1) and $32 \times 3 \times 3 \times 3 \times 16$ (i.e., $K_1^3 = 3$, $K_2^3 = 3$, and $K_3^3 = 3$ in Fig. 1) in the subsequent 1st, 2nd and 3rd convolution layers, respectively, where $16 \times 3 \times 3 \times 5 \times 8$ means 16 3D-kernels of dimension $3 \times 3 \times 5$ (i.e., two spatial and one spectral dimension) for all 8 3D input feature maps. Whereas, the dimension of 2D convolution kernel is $64 \times 3 \times 3 \times 576$ (i.e., $K_1^4 = 3$ and $K_2^4 = 3$ in Fig. 1), where 64 is the number of 2D-kernels, 3×3 represents the spatial dimension of 2D-kernel, and 576 is the number of 2D input feature maps. To increase the number of spectral-spatial feature maps simultaneously, 3D convolutions are applied thrice and can

preserve the spectral information of input HSI data in the output volume. The 2D convolution is applied once before the $flatten$ layer by keeping in mind that it strongly discriminates the spatial information within the different spectral bands without substantial loss of spectral information, which is very important for HSI data. A detailed summary of the proposed model in terms of the layer types, output map dimensions and number of parameters is given in Table I. It can be seen that the highest number of parameters are present in the 1st dense layer. The number of node in the last dense layer is 16, which is same as the number of classes in Indian Pines dataset. ~~Thus~~, the total number of parameters in the proposed model depends on the number of classes in a dataset. The total number of trainable weight parameters in $HybridSN$ is $5,122,176$ for Indian Pines dataset. All weights are randomly initialised and trained using back-propagation algorithm with the $Adam$ optimiser by using the $softmax$ loss. We use mini-batches of size 256 and train the network for 100 epochs with no batch normalization and data augmentation.

III. EXPERIMENTS AND DISCUSSION

A. Dataset Description and Training Details

We have used three publicly available hyperspectral image datasets[1], namely Indian Pines, University of Pavia and Salinas Scene. The **Indian Pines (IP) dataset** has images with 145×145 spatial dimension and 224 spectral bands in the wavelength range of 400 to 2500 nm, out of which 24 spectral bands covering the region of water absorption have been discarded. The ground truth available is designated into 16 classes of vegetation. The **University of Pavia (UP) dataset** consists of 610×340 spatial dimension pixels with 103 spectral bands in the wavelength range of 430 to 860 nm. The ground truth is divided into 9 urban land-cover classes. The **Salinas Scene (SA) dataset** contains the images with 512×217 spatial dimension and 224 spectral bands in the wavelength range of 360 to 2500 nm. The 20 water absorbing spectral bands have been discarded. In total 16 classes are present in this dataset.

All experiments are conducted on an Acer Predator-Helios laptop with the GTX 1060 Graphical Processing Unit (GPU) and 16 GB of RAM. We have chosen the optimal learning rate of 0.001, based on the classification outcomes. In order to make the fair comparison, we have extracted the same spatial dimension in 3D-patches of input volume for different datasets, such as $25 \times 25 \times 30$ for IP and $25 \times 25 \times 15$ for UP and SA, respectively.

[1]www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

(a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)
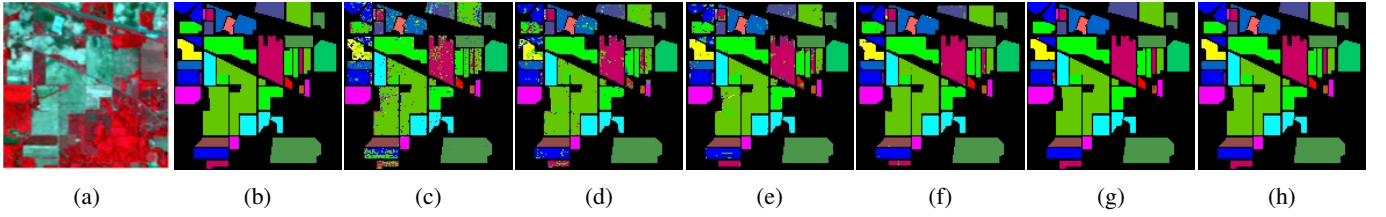
Fig. 2: The classification map for Indian Pines, (a) False color image (b) Ground truth (c)-(h) Predicted classification maps for SVM, 2D-CNN, 3D-CNN, M3D-CNN, SSRN, and proposed HybridSN, respectively.
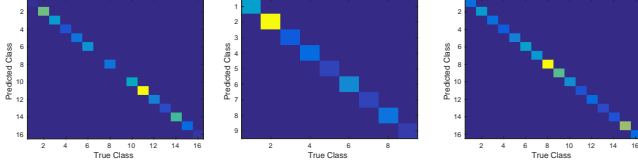

Fig. 3: The confusion matrix using proposed method over Indian Pines, University of Pavia, and Salinas Scene datasets in 1st, 2nd, and 3rd matrix, respectively.
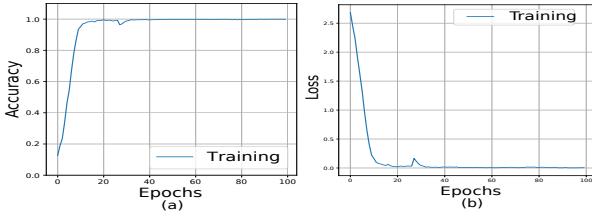

Fig. 4: The accuracy and loss convergence vs epochs over Indian Pines dataset.

B. Classification Results

In this letter, we have used the Overall Accuracy (OA), Average Accuracy (AA) and Kappa Coefficient (Kappa) evaluation measures to judge the HSI classification performance. Here, OA represents the number of correctly classified samples out of the total test samples; AA represents the average of class-wise classification accuracies; and Kappa is a metric of statistical measurement which provides mutual information regarding a strong agreement between the ground truth map and classification map. The results of the proposed $HybridSN$ model are compared with the most widely used supervised methods, such as SVM [33], 2D-CNN [34], 3D-CNN [35], M3D-CNN [36], and SSRN [27]. The 30% and 70% of the data are randomly divided into training and testing groups, respectively[2]. We have used the publicly available code[3] of the compared methods to compute the results.

Table II shows the results in terms of the OA, AA, and Kappa coefficient for different methods[4]. It can be observed from Table II that the $HybridSN$ outperforms all the compared methods over each dataset while maintaining the minimum standard deviation. The proposed $HybridSN$ is based on the hierarchical representation of spectral-spatial 3D CNN followed by a spatial 2D CNN, which are complementary to each other. It is also observed from these results that the performance of 3D-CNN is poor than 2D-CNN over Salinas Scene dataset. To the best of the our

---

[2]More details of dataset are provided in the supplementary material

[3]https://github.com/eecn/Hyperspectral-Classification

[4]The class-wise accuracy is provided in the supplementary material.

TABLE III: The training time in minutes (m) and test time in seconds (s) over IP, UP, and SA datasets using 2D-CNN, 3D-CNN and $HybridSN$ architectures.

Data	2D CNN		3D CNN		HybridSN	
	Train(m)	Test(s)	Train(m)	Test(s)	Train(m)	Test(s)
IP	1.9	1.1	15.2	4.3	14.1	4.8
UP	1.8	1.3	58.0	10.6	20.3	6.6
SA	2.2	2.0	74	15.2	25.5	9.0

TABLE IV: The impact of spatial window size over the performance of $HybridSN$.

Window	IP(%)	UP(%)	SA(%)	Window	IP(%)	UP(%)	SA(%)
19×19	99.74	99.98	99.99	23×23	99.31	99.96	99.71
21×21	99.73	99.90	99.69	25×25	99.75	99.98	100

TABLE V: The classification accuracies (in percentages) using proposed and state-of-the-art methods on less amount of training data, i.e., 10% only.

Methods	Indian Pines			Univ. of Pavia			Salinas Scene		
	OA	Kappa	AA	OA	Kappa	AA	OA	Kappa	AA
2D-CNN	80.27	78.26	68.32	96.63	95.53	94.84	96.34	95.93	94.36
3D-CNN	82.62	79.25	76.51	96.34	94.90	97.03	85.00	83.20	89.63
M3D-CNN	81.39	81.20	75.22	95.95	93.40	97.52	94.20	93.61	96.66
SSRN	98.45	98.23	86.19	99.62	99.50	99.49	99.64	99.60	99.76
HybridSN	98.39	98.16	98.01	99.72	99.64	99.20	99.98	99.98	99.98

knowledge, this is probably due to the presence of two classes in the Salinas dataset (namely Grapes-untrained and Vinyard-untrained) which have very similar textures over most spectral bands. Hence, due to the increased redundancy among the spectral bands, the 2D-CNN outperforms the 3D-CNN over Salinas Scene dataset. Moreover, the performance of SSRN and $HybridSN$ is always far better than M3D-CNN. It is evident that 3D or 2D convolution alone is not able to represent the highly discriminative feature compared to hybrid 3D and 2D convolutions.

The classification map for an example hyperspectral image is illustrated in Fig. 2 using SVM, 2D-CNN, 3D-CNN, M3D-CNN, SSRN and $HybridSN$ methods. The quality of classification map of SSRN and $HybridSN$ is far better than other methods. Among SSRN and HybridSN, the maps generated by $HybridSN$ in small segment are better than SSRN. Fig. 3 shows the confusion matrix for the HSI classification performance of the proposed $HybridSN$ over IP, UP and SA datasets, respectively. The accuracy and loss convergence for 100 epochs of training and validation sets are portrayed in Fig. 4 for the proposed method. It can be seen that the convergence is achieved in approximately 50 epochs which points out the fast convergence of our method. The computational efficiency of $HybridSN$ model appears in term of training and testing times in Table III. The proposed model is more

efficient than 3D-CNN. The impact of spatial dimension over the performance of $HybridSN$ model is reported in Table IV. It has been found that the used 25×25 spatial dimension is most suitable for the proposed method. We have also computed the results with an even less training data, i.e., only 10% of total samples and have summarized the results in Table V. It is observed from this experiment that the performance of each model decreases slightly, whereas the proposed method is still able to outperform other methods in almost all cases.

IV. CONCLUSION

This letter has introduced a hybrid 3D and 2D model for hyperspectral image classification. The proposed $HybridSN$ model basically combines the complementary information of spatio-spectral and spectral in the form of 3D and 2D convolutions, respectively. The experiments over three benchmark datasets compared with recent state-of-the-art methods confirm the superiority of the proposed method. The proposed model is computationally efficient than the 3D-CNN model. It also shows the superior performance for small training data.

REFERENCES

[1] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer Science & Business Media, 2003, vol. 1.

[2] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 45–54, 2014.

[3] J. Yang and J. Qian, "Hyperspectral image classification via multiscale joint collaborative representation with locally adaptive dictionary," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 112–116, 2018.

[4] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial–spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3534–3546, 2018.

[5] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, 2006.

[6] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1592–1606, 2015.

[7] Q. Gao, S. Lim, and X. Jia, "Hyperspectral image classification using joint sparse model and discontinuity preserving relaxation," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 78–82, 2018.

[8] P. Gao, J. Wang, H. Zhang, and Z. Li, "Boltzmann entropy-based unsupervised band selection for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[9] P. Hu, X. Liu, Y. Cai, and Z. Cai, "Band selection of hyperspectral images using multiobjective optimization-based sparse self-representation," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[10] B. Tu, X. Zhang, X. Kang, G. Zhang, J. Wang, and J. Wu, "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 340–344, 2018.

[11] T. Dundar and T. Ince, "Sparse representation-based hyperspectral image classification using multiscale superpixels and guided filter," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[13] 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*, 2016, pp. 770–778.

[14] S. K. Roy, S. Manna, S. R. Dubey, and B. B. Chaudhuri, "Lisht: Non-parametric linearly scaled hyperbolic tangent activation function for neural networks," *arXiv preprint arXiv:1901.05894*, 2019.

[15] 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*, 2015, pp. 91–99.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2980–2988.

[17] S. S. Basha, S. Ghosh, K. K. Babu, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Rccnet: An efficient convolutional neural network for histological routine colon cancer nuclei classification," in *15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2018, pp. 1222–1227.

[18] V. K. Repala and S. R. Dubey, "Dual cnn models for unsupervised monocular depth estimation," *arXiv preprint arXiv:1804.06324*, 2018.

[19] C. Nagpal and S. R. Dubey, "A performance evaluation of convolutional neural networks for face anti spoofing," *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2019.

[20] X. Kang, B. Zhuo, and P. Duan, "Dual-path network-based hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[21] Y. Yu, Z. Gong, C. Wang, and P. Zhong, "An unsupervised convolutional feature fusion network for deep representation of remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 23–27, 2018.

[22] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep cnn," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[23] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3173–3184, 2018.

[24] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, no. 99, pp. 1–11, 2018.

[25] J. M. Haut, S. Bernabé, M. E. Paoletti, R. Fernandez-Beltran, A. Plaza, and J. Plaza, "Low-high-power consumption architectures for deep-learning models applied to hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2018.

[26] 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.

[27] 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, 2018.

[28] L. Mou, P. Ghamisi, and X. X. Zhu, "Unsupervised spectral–spatial feature learning via deep residual conv–deconv network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 391–406, 2018.

[29] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, no. 99, pp. 1–15, 2018.

[30] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza, J. Li, and F. Pla, "Capsule networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

[31] L. Fang, Z. Liu, and W. Song, "Deep hashing neural networks for hyperspectral image feature extraction," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2019.

[32] 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, 2013.

[33] 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.

[34] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 4959–4962.

[35] 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.

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