IADIS International Journal on Computer Science and Information Systems
Vol. 16, No. 2, pp. 29-44
ISSN: 1646-3692

SPECTRAL-SPATIAL CLASSIFICATION OF HYPERSPECTRAL DATA USING 3D-2D CONVOLUTIONAL NEURAL NETWORK AND INCEPTION NETWORK

Douglas Omwenga Nyabuga and Guohua Liu
School of Computer Science and Technology, Donghua University
Shanghai City 201620, China

ABSTRACT

Hyperspectral imaging (HSI) classification has recently become a field of interest in the remote sensing (RS) community. However, such data contain multidimensional dynamic features that make it difficult for precise identification. Also, it covers structurally nonlinear affinity within the gathered spectral bands and the related materials. To systematically facilitate the HSI categorization, we propose a spectral-spatial classification of HSI data using a 3D-2D convolutional neural network and inception network to extract and learn the in-depth spectral-spatial feature vectors. We first applied the principal component analysis (PCA) on the entire HSI image to reduce the original space dimensionality. Second, the exploitation of the spatial hyperspectral input features contiguous information by 2-D CNN. Besides, we used 3-D CNN without relying on any preprocessing to extract deep spectral-spatial fused features efficiently. The learned spectral-spatial characteristics are concatenated and fed to the inception network layer for joint spectral-spatial learning. Furthermore, we learned and achieved the correct classification with a softmax regression classifier. Finally, we evaluated our model performance on different training set sizes of two hyperspectral remote sensing data sets (HSRSI), namely Botswana (BT) and Kennedy Space Center (KSC), and compared the experimental results with deep learning-based and state-of-the-art (SOTA) classification methods. The experiment results show that our model provides competitive classification results with state-of-the-art techniques, demonstrating the considerable potential for HSRSI classification.

KEYWORDS

Convolution Neural Network, Remote Sensing, Hyperspectral, Spectral-Spatial, Inception

1. INTRODUCTION

Hyperspectral remote sensing image (HSRSI) holds very vital information about the land objects, i.e., the spatial and spectral information (Tang et al., 2020), which contributes toward efficient and accurate image processing. Also, HSRSI classification (Tong et al., 2014) is vital for its interpretation and processing. Hyperspectral sensors acquire hundreds of information from consecutive land material segments, with a supply of rich spectral information and improved capability to discriminate unevenly distributed land-use materials (Bing, 2011). The HSRSIs contains hundreds of adjacent channels with rich spectral-spatial signatures, making it possible to discriminate earth objects. Thus, it has contributed to its wide use in crop analysis, urban administration, and environmental management. Some traditional HSRSI classification approaches have been used to classify and distinguish these images, but nearly all are founded both on the handcraft features (Yanshan Li et al., 2019; Wang et al., 2019) and traditional classifiers (Melgani & Bruzzone, 2004).

Because neighbor intensities are closely correlated at such a high spatial resolution, a spectral signature involves a wealth of redundant data. Often, problems such as the curse of dimensionality and class imbalance occur in the HSRSI data set. Dimensionality reduction (DR) is a fundamental pre-processing step to complement an HSRSI data set classification process, which has received much attention lately. Several studies, for example (Datta et al., 2018; Koumoutsou et al., 2021; Rodarmel & Shan, 2002), deployed the well-known technique, i.e., the PCA, to fix the problems mentioned above.

In past years, scholars mined spatial features from raw HSRSI data cube by cropping the spectral resolution through the deployment of some DR techniques, for instance, principal component analysis (PCA), and reducing spatial regions of $k \times k$ pixel-based regions. The authors such as (Haut et al., 2019) trained a CNN2-D with a single principal component (PC), at the same time (C. Li et al., 2015) applied three PCs in the training of the CNN2-D in addition to post-processing of the unearthed spatial information with sparse coding (SC) (Song et al., 2014) to produce more typical spatial representations of sparse dictionaries for categorization. For instance, the simple process involves providing an approach having 3-D contiguous domains of $k \times k \times n_{channels}$ where $n_{channels}$ can represent a definite count of PCs. Hereof, some techniques make an initial DR crop the redundancy and correlation of spectral channels. Motivated by the stated advantage in our proposed model, we applied PCA to surmount the problems caused by the high dimensionality covered in the HSIRS images.

In recent years, deep learning (DL) methods have experienced tremendous advancement for the study of HSRSI data. The extracted features through the convolutional neural network (CNN) possess more excellent semantic representation and stronger robustness than handcraft features. CNN has also been deployed to the categorization of HSRSI. For example, (Hu et al., 2015) presented a 1-D-CNN-based method for the extraction of in-depth feature vectors of HSRSI in spectral measurement. (Ying Li et al., 2017) employed 3-D CNNs to extract the complete fusion of spectral-spatial features from the original 3D cube image data concurrently. However, due to the CNN structure's simplistic stacking and the lack of an effective feature aggregation method to deepen network information transmission, it could not effectively deduce deep features. The ability of CNN to express features improves as its depth grows. As a result, the residual network was used to deepen the network using a feature aggregation method. (Zhong et al., 2018) proposed using a 3-D CNN to deploy a supervised spatial-spectral residual Network (SSRN) and generate spatial and spectral residual modules to derive spatial-spectral properties constantly. This approach, however, only takes as input a single-scale neighborhood block. In terms of overall classification accuracy, single-scale features perform poorly.

Moreover, to classify the HSI images (Zhang et al., 2019), presented a method known as a multi-scale dense Network (MSDN). The authors applied various dimension information in the proposed framework design and fused dimensional features throughout the framework. In the process of obtaining the necessary feature vectors, the framework implemented the discriminating and imbalanced of two dimensions. The framework deployed the rebuilding of extraction of abstract features and multi-scale fusion for the classifying of HSI data. It had better performance concerning the representation of HSI images. Also, it improved the speed of the training of the framework and its accuracy, which specifically enhanced the convergence speed. Even though there was an added feature network layer on the dimension, which condensed the deep framework in feature extraction, which is attributable to the threshold response of complexity, the precision improvement was not any longer robust after the complexity got to a particular level. (Roy et al., 2019) combined a CNN3-D with a CNN2-D, where they applied CNN3-D basically to extract spatial-spectral features and subsequently distinguished by the CNN2-D. However, the proposed method faces the problem of overfitting. Hence, it is an open challenge to achieve higher interpretation accuracies when processing such increased spectral-spatial measurement imagery.

To this end, we, therefore, propose a spectral-spatial classification of hyperspectral data using 3D-2D convolutional neural network and inception network, which contributes to the correct classification of two publicly available HSRSI data sets, i.e., BT and KSC. The inclusion of the 3D-inception network immediately after the 3D convolutional filters for discriminative spatial feature learning is the key difference between our proposed model and spatial-spectral approaches. Furthermore, the following are the key contributions of our proposed model:

1) First, we adopted the principal component analysis (PCA) for dimensional reduction of spectral channels that are very much correlated while preserving the desirable information

2) We fused the ordered spectral-spatial related features extracted for CNN2-D and CNN3-D with our model to exploit spatial features. To train the CNN2-D and CNN3-D pixels, we convolved the input data from the CNN-2D and CNN-3D with 2D kernels and 3D kernels, respectively.

3) We incorporated the inception network layer in our model to decrease the complexity of the model compared to training the layers separately.

4) We further introduced a softmax regression classifier for the correct classification of HSRSI images.

The study demonstrates that this model can improve the classification accuracy of HSRSI and offer new scientific ideas and references for related research.

The remaining part of the paper takes the following structure; the related work is reviewed in part 2, our proposed methodology is discussed in part 3. The experimental setup, dataset, compared SOTA methods, and parameter setting takes part 4; in part 5, we discuss the experimental results. Finally, we give a conclusion of the study in part 6.

2. RELATED WORK

Lately, HSRSI classification technology based on DL has become a hot research area (Zhu et al., 2017). When likened with artificially constructed features, the DL methods spontaneously extract deep feature vectors of small details to those of essential facts and translates imageries into feature vectors that are simple to identify and classify. For instance, the deep brief network (DBN) (Chen et al., 2015) method apply unsupervised approaches to mine extensive feature vectors in a layer-based training mode. In achieving the objective, flattening the 1-D training sample set must be performed to fit the required standard size as the model inputs, which often suffer a loss of the spatial details of the raw image. The CNN2D was introduced by (Zhao & Du, 2016) to extract spatial contents from the representation of the reduced dimensions via PCA and mines spectral information deploying their proposed method and lastly, fused with spatial/spectral information to enhance the HSI classification precision. Nevertheless, the 2-D CNN method by (Makantasis et al., 2015) mines spatial/spectral feature vectors solely but needs convoluted pre-processing.

Spatial-spectral learning can be accomplished by 2D-CNN frameworks presenting spectral-spatial handcrafted information. For example, (He et al., 2019) trained the 2D-CNN method under the concept of covariance matrices (CM), which encodes the spatial-spectral features of various measurements, i.e., regions of twenty principal components (PCs), attaining multi-scale covariance maps. (Yue et al., 2015) designed a 2D-CNN framework to study the spectral-spatial contents by blending the spectral contextual as maps of three distinct features and combining them to the spatial patches (downsized three PCs through PCA). Additionally, a manifold of approaches combines the 2D-CNN with other methods to extract the spatial-spectral features solely. Although the aforementioned methods were used in the image classification task, they have limitations in their application due to only 2D-CNN approaches.

The discriminating spatial dimension of HSRSIs facilitates in supplying a diversity of features of low-levels indicating detailed spatial information. On the other hand, the spectral contents offer essential and distinct details to exhibit the characteristics of land-use materials (Du et al., 2016). Making excellent use of rich spectral-spatial contents, therefore, facilitates the improvement of the HSRSI classification accuracy. (Ying Li et al., 2017) proposed 3-D CNN's approach to obtaining a complete blending of spectral-spatial features concurrently from the raw 3-D data cube. However, their proposed method could not efficiently extract extensive information attributable to the stacked CNN layer and its simplicity. Besides, the limitation of an active feature combination approach increases the depth of the network, which enables it to transmit information. The CNN potentiality feature expression grows as the layers increase the depth. As a result, the current DL techniques have enhanced the classification of HSRSI images, for instance, stacking 2D-CNN and 3D-CNN with a dense layer network.

Nevertheless, the best learning process can have a steady and length improvement, but discriminate and meaningful features mostly get lost or even dissipate in the course of depth transfer. Besides, for the 1D CNN and 2D CNN methods, there is a frequent implementation of the 3-D-CNN method in classifying spatial-spectral feature vectors. The 3-D filters with a size of $f^{(k)} \times f^{(k)} \times f^{(k)} \times z^{(k)}$ can automatically perform the extraction of complex spatial-spectral features, where the extracting f represents an output feature dimensions.

The process of analyzing and classifying HSI is complicated and involves diverse representations; these CNN3D frameworks have specific limitations. Therefore, it is important to engage both CNN2-D and CNN3-D architectures to study HSRSI images from the perspective of structural features. In addition, we fused the inception network layer in our model to decrease the complexity of the model compared to training CNN-2D and CNN-3D separately. The inception network layer has not been utilized in the study of HSRSI images for the previous works.

3. METHOD

In this part, we describe our proposed model, which we illustrate in Figure 1. We subjected the deep extracted 3D spectral-spatial features to the principal component analysis (PCA) technique for dimensionality reduction and removal of redundant features. Next, our model jointly employed the spectral/spatial HSRSI information. The hyperspectral tensor, which takes three-dimensional cubes, i.e., $S \times S \times L$, where $S \times S$ is the width and height of the spatial dimensions and L is the spectral dimension represents our HSRSI input image, which expressed as

$$I \in \mathbf{R}^{W \times H \times C} \tag{1}$$

where I denotes the raw HSRS input image, the W and H is the width and height of the spatial features and C denotes the spectral channels in the image.

3.1 Dimensionality Reduction

PCA is an unsupervised linear transformation method that embeds data into a reduced linear subspace. It aims to discover a linear transformation T that augments the covariance matrix $Cov(Y)$ mathematically. The p primary eigenvectors, or principal components (PCs) of the covariance matrix with mean $(\mu = 0)$ data, produce this linear transformation. It augments $Cov(Y)$ relative to T conditionally that $T = 1$. In brief, PCA seeks to identify the regions of apex variance in high-dimensionality data and transform these pixels into a new subclass with comparable or lesser dimensionalities. The orthogonal axes (PCs) of the new subclass can be selected as the directions with the largest variance.

Due to the redundancy and the intra/inter-class similarities, we adopted the PCA as a classic dimensionality reduction technique which we directly applied to the extracted feature vectors of the HSRSI input image, and it takes the form of equation (2).

$$P \in \mathbf{R}^{W \times H \times T} \tag{2}$$

The batch sizes are set as 25×25×13 and 25×25×15 for BT and KSC data sets, respectively, with 13 and
15 most informative bands selected by PCA strategy.

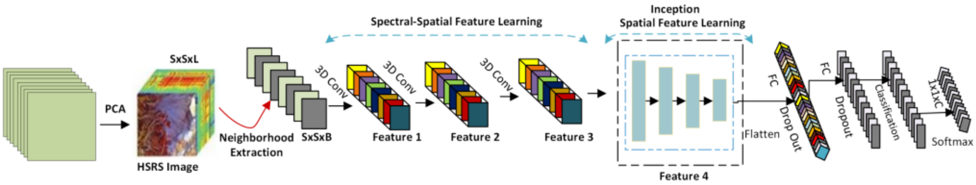

Figure 1. The 2D-3D convolutional neural network and inception network proposed architecture

To learn the spatial-spectral features, we convoluted a 3-D kernel with the 3-D-data. In our proposed architecture for HSRSI data, through employing the 3-D kernel over a manifold of neighborhood channels in the input layer, the feature maps of the convolution layer are produced; this apprehends the spectral features. The expression in the 3-D convolution, the activation value at the spatial domain $\left(x,y,z\right)$ in the j^{th} feature map of the i^{th} layer represented as $v_{(i,j)}^{(x,y,z)}$ we define it as

$$v_{i,j}^{x,y,z} = \phi\left(b_{i,j} + \sum_{r-1}^{d_{l-1}} \sum_{\lambda=-\eta}^{\eta} \sum_{\rho=-\gamma}^{\gamma} \sum_{\sigma=-\delta}^{\delta} w_{i,j,\tau}^{\sigma,\rho,\lambda} \times v_{i-1,\tau}^{x+\sigma,y+\rho,z+\lambda} \right) \tag{3}$$

where ϕ is the activation function, $b_{i,j}$ is the bias parameter for the j^{th} feature map of the i^{th} layer, d_{l-1} is the number of feature map in $l-1^{th}$ layer, $2\gamma+1$ is the width of kernel, $2\delta+1$ is the height of kernel, $2\eta+1$ is the depth of kernel along spectral resolution, and $w_{i,j}$ is the value of weight parameter for the j^{th} feature map of the i^{th} layer.

We trained the bias b and the kernel weight w parameters using supervised strategies with the aid of a gradient descent optimization technique. The 3-D-CNN kernel finally derives the feature vector depictions of spectral-spatial contents concurrently from HSRSI data, but at the cost of increased computational complexity. To further learn more abstract level spatial representation, we apply CNN2-D on top of the CNN3-D.

By convolving the input image from the CNN3-D with the 2D kernels, we learned the spatial feature. We achieved the convolutions by calculating the sum of the dot product between input data and kernel filter of the spatial resolutions, which only present all the feature maps of earlier network layers resulting in 2-D discriminative feature maps. The kernel filter is stride over the input data to cover the full spatial region. The convolved features are passed through the activation function to insert the non-linearity in the model. In 2-D convolution, the activation value at spatial position (x,y) in the j^{th} feature map of the i^{th} layer is denoted as $v_{(i,j)}^{(x,y)}$ and can be generated using equation (4).

$$v_{i,j}^{x,y,z} = \phi\left(b_{i,j} + \sum_{r-1}^{d_{l-1}} \sum_{\rho=-\gamma}^{\gamma} \sum_{\sigma=-\delta}^{\delta} w_{i,j}^{\sigma,\rho} \times v_{i-1,\tau}^{x+\sigma,y+\rho} \right) \tag{4}$$

where ϕ denotes the activation function, $b_{i,j}$ denotes the bias parameter for the j^{th} feature map of the i^{th} layer. The d_{l-1} represents the value of feature map in the $l-1^{th}$ network layer, and $w_{i,j}$ represents the depth of kernel filter for the j^{th} feature map of the i^{th} network layer. $2\gamma + 1$ represents the width of kernel filter, $2\delta + 1$ is the height of kernel filter, and $w_{i,j}$ represents the value of weight parameter for the j^{th} feature map of the i^{th} network layer.

To avoid the overfitting problem and fast-track the model's convergence speed and enhance the accuracy, we selected a rectified linear unit (ReLU) (Maas et al., 2013) as the nonlinear activation function. Hence, defined as

$$f(x) = \max(0, x) \tag{5}$$

We randomly initialized all the model's weights and trained them through Back-propagation (BP) algorithm with the gradient drop algorithm, i.e., the adaptive moment estimation (Adam) optimizer, by applying the softmax loss. We trained our proposed framework for 100 epochs without data augmentation and batch normalization (BN). The softmax regression layer, which is used to predict the conditional probability distribution of each class for correct classification of the data set, is expressed as

$$P(y = \frac{i}{x} W, b) = \frac{e^{(W_{ix} + b_i)}}{\sum_{i=1}^{k} e^{(W_{jx} + b_j)}} \tag{6}$$

where W and b denotes the weights and bias of the softmax regression layer classifier, respectively.

4. EXPERIMENTS

4.1 Dataset

In evaluating the performance of our proposed model, this study analyzes two classic HSRSI data set, namely the Botswana (BT) and Kennedy Space Center (KSC) data set (*Hyperspectral Remote Sensing Scenes - Grupo de Inteligencia Computacional (GIC)*, n.d.). We divided all sampled data into two set sizes, i.e., the train/test size sets. We adopted a ratio of 1:9 and 3:7, i.e., 10/30% for training set size and 90/70% for testing set size. During the validation process, we selected the network with the highest classification accuracy and retained the corresponding weight parameters—this attributed to the early stopping generalization technique that we applied in our model. Finally, we employed the test set to evaluate the classification capability of the trained model. We quantitatively carried a random of ten trains and determined the average and standard deviation of the multiple sets of overall accuracy (OA), mean accuracy (AA), and kappa coefficient (k), which we picked as overall accuracy. The classification accuracy of the best performance for the validation set was thus, saved during the training process.

Moreover, we compared our proposed model with the SOTA methods, namely the 3D-CNN (Ying Li et al., 2017), SSRN (Zhong et al., 2018), HybridSN (Roy et al., 2019), and MSDN (Zhang et al., 2019), respectively, to show the performance of our model on the two HSRSI data sets. Table 3 and Table 4 report per class accuracy classification result of the methods with their respective OA, k, and AA. The classification maps of our model are as shown in Figure 2 and Figure 3.

4.1.1 The Botswana (BT) Data Set

NASA in the year 2001-2004, through the EO-1 satellite, gathered BS data in sequence structure. The BS data set has 242 bands, covering a wavelength/spectrum range of 400-2500 nm and 30 m—spatial resolution over a 7.7 km strip. The personnel discarded the lousy inter-detector miscalibration effects, detectors, and recurrent abnormalities from the data set after preprocessing the BT data. They also removed noisy, disordered, and imperfect bands that cover water absorption features (10-55, 82-97, 102-119, 134-164, & 187-220) and thus remained with 145 bands. The data set consists of 14 classes considered for our experiments. The total number of classification samples is 3248.

4.1.2 The Kennedy Space Center (KSC) Data Set

The dataset was captured by NASA scientists using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument in the year 1996. AVIRIS captures data with 400-2500 nm center wavelengths in—224 bands of 10nm width and 18 m-spatial resolution. Discarding the low SNR bands and water absorption from the dataset, the KSC personnel used 176 bands and 13 classes for their study. We used all 13 classes for classification, representing diverse land cover types in nature settings. The selection of our training data solely depends on the land cover. The total number of classification samples is 5211.

4.2 Parameter Setting

The proposed model via the gradient of the back-propagation (BP) loss function updated the parameters of the 3-D convolution in the architecture. We adopted Adam optimizer as the gradient descent algorithm in our model to optimize the loss function. For both KSC and BT data set, the learning rate is very crucial. First, learning rates (lrs) regulate each training iteration's learning step. Inappropriate lr settings, in particular, will result in delayed convergence or divergence. As a result, we applied the grid search strategy and executed each experiment for 100 epochs to determine the best lr for each data set from (0.01, 0.003, 0.001, 0.1). We simulated our experiments on a MacBook Pro laptop computer with Intel Core i5, 2.3 GHz, the Intel Iris Plus Graphics 640 1536 MB, and the memory of 8 GB 2133 MHz LPDDR3 as the hardware, and all experiments executed on GPU-Google Colab ltd platform.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

5.1 The Effectiveness of Spatialized Inputs

We evaluated our suggested model using input cubes of various spatial sizes to assess the impact of the spatialized input. Because our model learns distinguishing spatial aspects of input data, the results in Table 1 and Table 2 illustrates that the proposed model performs well for various spatial sizes if these sizes are equal to 25 × 25. The classification accuracy in both data sets improves as the spatial size of the input cubes increases. This shows how essential spatial size is for the model's efficiency. We set the spatial size of input HSI data to 25 × 25 to make a fair comparison between various classification SOTA methods because large input sizes contribute to higher classification accuracy.

Table 1. The classification accuracy (OA%) of different spatialized input sizes for BT data set

Class.No	11×11×13	13×13×13	15×15×13	17×17×13	25×25×13
1	100	100	99.47	100	100
2	100	100	100	100	100
3	100	100	100	100	100
4	100	100	100	100	100
5	99.47	97.87	100	100	99.94
6	98.54	100	98.40	98.40	100
7	100	100	100	100	100
8	100	100	100	100	100
9	100	100	100	100	100
10	100	100	100	100	100
11	100	100	100	100	98.40
12	100	100	100	100	100
13	100	100	100	100	100
14	100	100	100	100	100
OA (%)	99.85	99.86	99.82	99.87	99.87
Kappa × 100	99.83	99.85	99.81	99.86	99.89
AA	99.85	99.88	99.85	99.87	99.86

Table 2. The classification accuracy (OA%) of different spatialized input sizes for KSC data set

Class.No	11×11×15	13×13×15	15×15×15	17×17×15	25×25×15
1	97.75	97.37	98.69	97.75	97.73
2	75.29	86.47	95.29	98.24	98.89
3	75.42	84.92	97.77	93.30	100
4	44.32	59.66	87.50	87.50	98.98
5	79.65	83.19	94.69	95.58	100
6	32.50	55.63	91.88	84.38	100
7	91.89	72.97	95.95	83.78	100
8	86.76	97.02	98.34	98.34	100
9	85.16	48.90	84.62	73.63	100
10	86.22	93.64	97.53	100	100
11	99.32	100	100	100	100
12	98.86	97.16	98.30	93.18	100
13	100	100	100	100	99.62
OA (%)	81.06	87.17	96.24	94.08	99.48
Kappa × 100	85.54	85.67	95.82	93.40	99.48
AA (%)	81.01	82.84	95.43	92.74	99.42

5.2 BT Data Set Performance Results

Further, we used the BT data set to demonstrate the performance of the 3D-CNN, SSRN, MSDN, HybridSN, and our proposed model. It comprises 14 classes with a total of 3248 samples. We applied a dropout rate of 55% to prevent the method's overfitting due to the limitation of labeled samplings, which produced a high performance. Table 3 lists the accuracy classification performance. On individual classes, we can see that our proposed model significantly improved on the classification of classes such as Floodplain Grasses

1 = 99.94%, Reeds 1 = 98.40%, Riparian =100%, and Water =100% compared to the baseline methods. Our model's overall accuracy (OA) outperformed the other SOTA methods, i.e., 3D-CNN, SSRN, MSDN, and HybridSN. The apparent predicted reference in Figure 2 illustrates the BT data set.

Table 3. Per class performance of our proposed method comparing to other SOTA methods, i.e., 3D-CNN, SSRN, MSDN, and HybridSN methods on 30% training set size of BT data set

Class.No	Class Label	Train/Test Samples	3D-CNN	SSRN	MSDN	HybridSN	Proposed
C1	Acacia grasslands	92/92	95.29	100	100	100	100
C2	Acacia shrublands	74/74	99.80	100	100	100	100
C3	Acacia woodlands	94/94	99.57	100	99.45	100	100
C4	Exposed soils	29/67	97.58	99.84	100	96.97	100
C5	Floodplain Grasses 1	75/176	99.04	99.29	99.45	99.86	**99.94**
C6	Firescar 2	78/181	97.94	100	100	100	100
C7	Floodplain Grasses 2	65/151	93.68	100	100	100	100
C8	Hippo grass	30/71	98.61	99.39	100	100	100
C9	Island interior	61/142	99.62	100	100	100	100
C10	Mixed mopane	80/188	99.84	100	100	100	100
C11	Reeds 1	81/188	94.98	95.72	96.76	94.68	**98.40**
C12	Riparian	81/188	87.31	99.16	92.89	99.47	**100**
C13	Short mopane	54/127	94.99	99.69	100	100	100
C14	Water	81/189	95.34	98.76	97.35	98.90	**100**

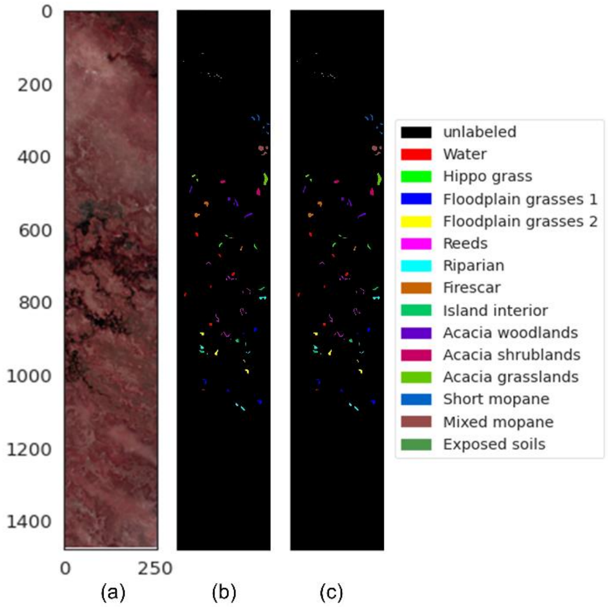

Figure 2. The Botswana (BT) hyperspectral remote sensing data set, a) the false-color composite, b) the ground truth label and c) the predicted reference

5.3 KSC Data Set Performance Results

The KSC dataset had a smaller number of classes (i.e., 13 classes) but comparatively broad feature sampling resolutions. From Table 4, we can observe that the proposed model's overall accuracy performance excels compared to 3D-CNN, SSRN, HybridSN, and the MSDN models. We can also see that our model was superior in individual classes with higher accuracy, such as Cabbage-palm/oak-hammock (98.89%), Spartina-marsh (98.98%), Willow-swamp (100%), Hardwood-swamp (100%), and Oak/broadleaf-hammock (100%). Figure 3 represents the false-color composite, ground-truth label, and the predicted reference of our proposed model, respectively.

Table 4. Per class performance of our proposed method comparing to other SOTA methods, i.e., 3D-CNN, SSRN, MSDN, and HybridSN methods on 30% training set size of KSC data set

Class. No	Class label	Train/Test Samples	3D-CNN	SSRN	MSDN	HybridSN	Proposed
C1	Cabbage-palm/oak-hammock	76/176	97.81	99.84	100	95.46	97.73
C2	Cabbage-palm-hammock	77/179	87.21	98.68	98.41	97.21	**98.89**
C3	Slash-pine	48/113	95.24	94.76	99.10	100	100
C4	Spartina-marsh	156/364	63.87	95.05	96.67	98.35	**98.98**
C5	Water	278/649	80.69	87.50	100	100	100
C6	Willow-swamp	73/170	87.92	97.34	99.89	98.82	**100**
C7	Cattail-marsh	121/283	92.63	100	100	96.11	100
C8	Graminoid-marsh	129/302	97.68	100	100	99.34	100

C9	Hardwood-swamp	32/74	96.79	99.97	99.75	94.6	**100**
C10	Mud-flats	151/352	94.23	99.99	99.71	95.17	**100**
C11	Oak/broadleaf-hammock	69/160	98.41	99.96	96.99	95.63	**100**
C12	Salt-marsh	126/293	95.36	98.76	100	100	100
C13	Scurb	228/533	96.29	100	100	99.44	99.62

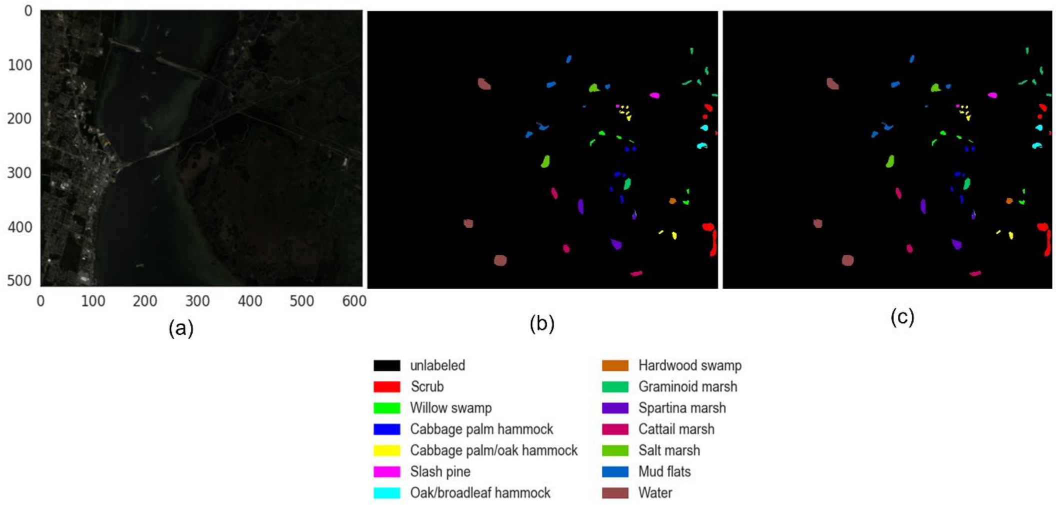

Figure 3. The Kennedy Space Center (KSC) hyperspectral remote sensing data set, a) the false-color composite, b) the ground truth label and c) the predicted reference

Figures 4 and 5 depict our model's accuracy and loss convergence graphs on 100 epochs for the two data sets we analyzed. With the accuracy convergence of our model in Figure 4, we can see that the model converged quickly at approximately 55 epochs of the BT dataset. Also, we can see that the accuracy convergence in Figure 5, i.e., the KSC dataset, our model converged faster at approximately 60 epochs. This is attributable to the introduction of the early-stopping regularization technique in addition to the inception network layer within our framework.

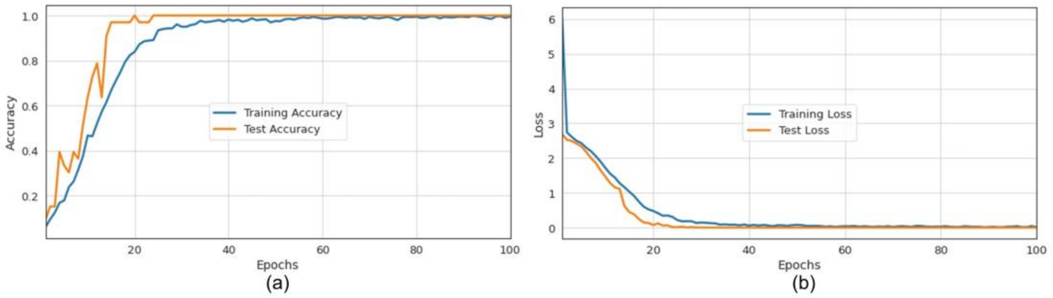

Figure 4. The convergence graphs for model accuracy and loss for 100 epochs on BT hyperspectral remote sensing data set

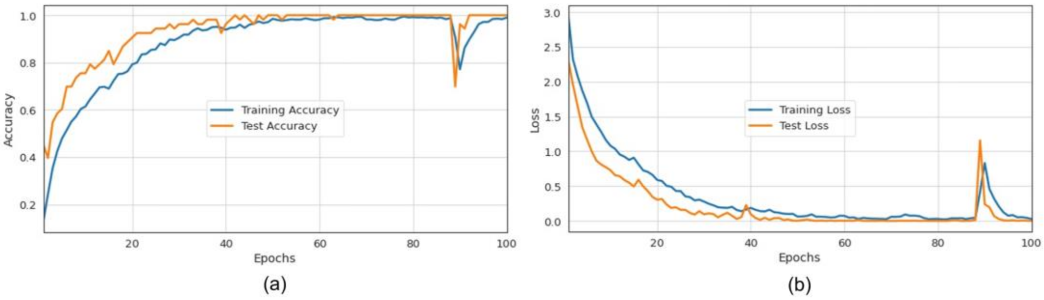

Figure 5. The convergence graphs for model accuracy and loss for 100 epochs on KSC hyperspectral remote sensing data set

On 10% and 30% training sets, we had the best classification performance in most classes, with the best classification results in terms of OA, AA, and κ × 100 (see Table 5 and Table 6). This is due to the inclusion of an inception network layer in our model, which allowed for more efficient computation by stacking 3D and 2D features for learning.

Table 5. The performance results for BT data set on 10% and 30% training set size

Train set	Metrics	3D-CNN	SSRN	MSDN	HybridSN	Proposed
10%	**OA (%)**	97.81	98.40	98.86	99.35	99.59
	AA (%)	97.72	98.65	98.80	99.27	99.50
	κ × 100	97.62	98.26	98.84	99.30	99.56
30%	**OA (%)**	95.56	98.01	98.77	99.43	99.87
	AA (%)	96.54	98.57	98.96	99.36	99.89
	κ × 100	95.06	97.89	98.66	99.38	99.86

Table 6. The performance results for KSC data set on 10% and 30% training set size

Train set	Metrics	3D-CNN	SSRN	MSDN	HybridSN	Proposed
10%	**OA (%)**	87.08	89.61	88.90	88.70	91.71
	AA (%)	85.09	89.33	87.56	86.49	90.32
	κ × 100	86.74	89.56	88.47	87.40	90.76
30%	**OA (%)**	93.63	97.87	99.45	98.22	99.48
	AA (%)	91.09	97.14	99.42	97.71	99.48
	κ × 100	92.92	97.74	99.39	98.02	99.42

The overall performance of both the BT and KSC data sets is depicted in Figure 6 (i.e., 10% training set). From the Figure, we can see that even when the training data is only 10%, our model maintains a high classification accuracy compared to other SOTA approaches. Further, in Figure 7 on 30%, it can also be seen that our proposed model’s performance is excellent compared to the other SOTA methods.

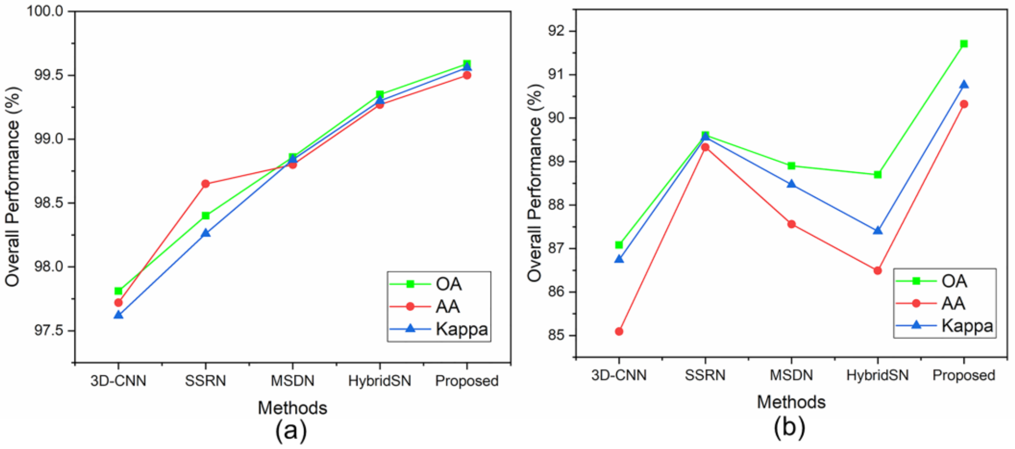

Figure 6. The overall performance on 10% of our proposed vs SOTA methods. a) The overall performance of the BT data set and b) The overall performance of the KSC data set

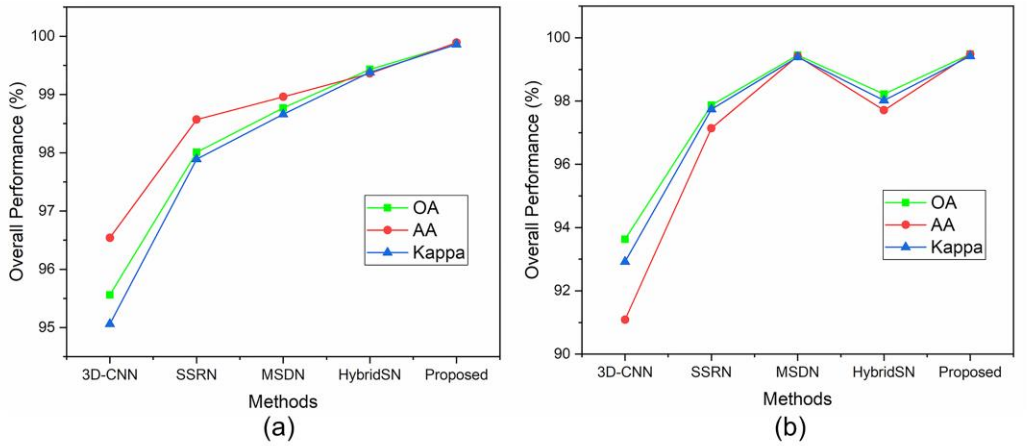

Figure 7. The overall performance on 30% of our proposed vs SOTA methods. a) The overall performance of the BT data set and b) The overall performance of the KSC data set

Table 7 lists the computation time in seconds for our proposed model with other SOTA methods. Regarding the computational complexity of our proposed model to the baseline methods, ours performed significantly good attributed to the use of inception network in our model. Furthermore, it contributed to decreasing the model's complexity compared to training CNN-2D and CNN-3D separately.

Table 7. The comparison of computation time (s) results

	BS Data Set		KSC Data Set	
Methods	Training time (s)	Testing time (s)	Training time (s)	Testing time (s)
MSDN	5595.2	25.49	11111.42	51.4
SSRN	4209.1	23.82	10764.3	48.8
HybridSN	4213.8	22.36	10521.4	46.5
Ours	4020.23	18.73	908.69	38.7

6. CONCLUSION

This work presented a spectral-spatial classification of hyperspectral data using a 3D-2D convolutional neural network and inception network. The proposed framework, which includes PCA and successive 3D-2D spectral-spatial convolutional layers with an inception network layer, has addressed the decreasing accuracy issue. The results of the experiments showed that our proposed model consistently produced the best classification accuracy for both types of HSRSI data sets, namely Botswana (BT) and Kennedy Space Center (KSC), demonstrating its significant superiority over the other SOTA approaches. It is important to note that this model has produced reliable classification results with both small and large numbers of unequal training data. Furthermore, the inception network layer reduced processing complexity and increased classification accuracy. Finally, our proposed model may easily be applied to different remote-sensing applications because of its consistent structural design and deep feature learning capabilities.

ACKNOWLEDGEMENT

This work was supported in part by the Development of Shanghai Industrial Internet (Grant No. 2019-GYHLW-01004).

REFERENCES

Bing, Z. (2011). Intelligent remote sensing satellite system. *Journal of Remote Sensing*.

Chen, Y., Zhao, X., & Jia, X. (2015). Spectral-Spatial Classification of Hyperspectral Data Based on Deep Belief Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(6), 2381–2392.

Datta, A., Ghosh, S., & Ghosh, A. (2018). *PCA, Kernel PCA and Dimensionality Reduction in Hyperspectral Images*. Advances in Principal Component Analysis.

Du, P., Xia, J., Xue, Z., Tan, K., Su, H., & Bao, R. (2016). Review of hyperspectral remote sensing image classification. *Yaogan Xuebao/J. Remote Sensing*, *20*(2), 236–256.

Haut, J. M., Paoletti, M. E., Plaza, J., Plaza, A., & Li, J. (2019). Hyperspectral Image Classification Using Random Occlusion Data Augmentation. *IEEE Geoscience and Remote Sensing Letters*, *16*(11), 1751–1755.

He, N., Paoletti, M. E., Haut, J. M., Fang, L., Li, S., Plaza, A., & Plaza, J. (2019). Feature extraction with multiscale covariance maps for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(2), 755–769.

Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, *2015*.

Hyperspectral Remote Sensing Scenes - Grupo de Inteligencia Computacional (GIC). (n.d.). Retrieved January 6, 2020, from http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

Koumoutsou, D., Charou, E., Siolas, G., & Stamou, G. (2021). *Class-Wise Principal Component Analysis for hyperspectral image feature extraction*.

Li, C., Chen, C., Carlson, D., & Carin, L. (2015). Preconditioned Stochastic Gradient Langevin Dynamics for Deep Neural Networks. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, 1788–1794.

Li, Yanshan, Li, Q., Liu, Y., & Xie, W. (2019). A spatial-spectral SIFT for hyperspectral image matching and classification. *Pattern Recognition Letters*, *127*, 18–26.

Li, Ying, Zhang, H., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing*, *9*(1).

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *In ICML Workshop on Deep Learning for Audio, Speech and Language Processing.*

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

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(8), 1778–1790.

Rodarmel, C., & Shan, J. (2002). Principal Component Analysis for Hyperspectral Image Classification. In *Information Systems* (Vol. 62, Issue 2).

Roy, S. K., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. (2019). HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters*, 1–5.

Song, B., Li, J., Dalla Mura, M., Li, P., Plaza, A., Bioucas-Dias, J. M., Benediktsson, J. A., & Chanussot, J. (2014). Remotely sensed image classification using sparse representations of morphological attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(8), 5122–5136.

Tang, H., Li, Y., Han, X., Xie, W., & Huang, Q. (2020). A Spatial-Spectral Prototypical Network for Hyperspectral Remote Sensing Image. *IEEE Geoscience and Remote Sensing Letters*, *17*(1).

Tong, Q., Xue, Y., & Zhang, L. (2014). Progress in hyperspectral remote sensing science and technology in China over the past three decades. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(1), 70–91.

Wang, Q., He, X., & Li, X. (2019). Locality and structure regularized low rank representation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(2), 911–923.

Yue, J., Zhao, W., Mao, S., & Liu, H. (2015). Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, *6*(6), 468–477.

Zhang, C., Li, G., & Du, S. (2019). Multi-Scale Dense Networks for Hyperspectral Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9201–9222.

Zhao, W., & Du, S. (2016). Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4544–4554.

Zhong, Z., Li, J., Luo, Z., & Chapman, M. (2018). Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(2), 847–858.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36.