TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 06/11 pp183–194
DOI: 10.26599/TST.2018.9010043
Volume 24, Number 2, April 2019

Multiple Deep-Belief-Network-Based Spectral-Spatial Classification of Hyperspectral Images

Atif Mughees* and Linmi Tao

Abstract: A deep-learning-based feature extraction has recently been proposed for HyperSpectral Images (HSI) classification. A Deep Belief Network (DBN), as part of deep learning, has been used in HSI classification for deep and abstract feature extraction. However, DBN has to simultaneously deal with hundreds of features from the HSI hyper-cube, which results into complexity and leads to limited feature abstraction and performance in the presence of limited training data. Moreover, a dimensional-reduction-based solution to this issue results in the loss of valuable spectral information, thereby affecting classification performance. To address the issue, this paper presents a Spectral-Adaptive Segmented DBN (SAS-DBN) for spectral-spatial HSI classification that exploits the deep abstract features by segmenting the original spectral bands into small sets/groups of related spectral bands and processing each group separately by using local DBNs. Furthermore, spatial features are also incorporated by first applying hyper-segmentation on the HSI. These results improved data abstraction with reduced complexity and enhanced the performance of HSI classification. Local application of DBN-based feature extraction to each group of bands reduces the computational complexity and results in better feature extraction improving classification accuracy. In general, exploiting spectral features effectively through a segmented-DBN process and spatial features through hyper-segmentation and integration of spectral and spatial features for HSI classification has a major effect on the performance of HSI classification. Experimental evaluation of the proposed technique on well-known HSI standard data sets with different contexts and resolutions establishes the efficacy of the proposed techniques, wherein the results are comparable to several recently proposed HSI classification techniques.

Key words: hyperspectral image classification; segmentation; deep belief network; support vector machine

1 Introduction

Advancements in optical technology have enabled HyperSpectral Image (HSI) acquisition sensors to acquire reflectance in hundreds of different wavelengths from the Earths' surface, ranging from visible to infrared spectra, in the form of hundreds of continuous, narrow spectral bands. Each band in HSI is a two-dimensional (2-D) spatial map with varying spectral radiance. As a result, each pixel in HSI is a vector of measurement of the spectral characteristics of a particular spatial position in the image[1]. Such a wide range of spectral information is extremely useful in many applications, such as mineral detection[2], precision farming[3], urban planning, environmental monitoring and management[4], and target detection and surveillance[5]. However, such abundance of information in HSI comes with more challenges, cost of complexity and undesirable statistical and

• Atif Mughees and Linmi Tao are with Pervasive Computing Division, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: maoz14@mails.tsinghua.edu.cn; linmi@mail.tsinghua.edu.cn.

∗ To whom correspondence should be addressed.

Manuscript received: 2017-08-08; revised: 2017-11-23; accepted: 2017-12-11

geometrical properties[6], which require advanced techniques[7]. High dimensionality in the spectral domain[8] coupled with the curse of dimensionality[9] and a limited number of labeled samples makes HSI classification a challenging problem[10].

High dimensionality of hyperspectral data is a major problem. Handling hundreds of features at the same time lead to complexity and error, and deprives Deep Belief Network (DBN) from extraction of more abstract and deep features. As for each pixel in HSI, more than 200 spectral values exist and therefore an equal number of input nodes are implemented at the input layer to handle these spectral values. The internal parameters of the learning and training phases, as a result of a large number of spectral values, are also enormous and complex. Moreover, an increase in the spectral feature space requires the increase in sampled training data at the same rate. However, the availability of sampled data is a major bottleneck in HSI. The imbalance between the high-dimensional data and limited number of training samples leads to Hughes phenomena[9].

In general, HSI classification performance depends on the efficient use of high-dimensional data in the presence of a limited amount of sampled data as each band in hyperspectral cube presents valuable information on each class in HSI. Two main dimensional reduction approaches are commonly used to handle high dimensionality, Feature Selection (FS) and Feature Extraction (FE). FS also known as band selection, involves the selection of lesser number bands of the hypercube to represent the entire image cube[11, 12]. In contrast, the FE approach achieves dimensional reduction by projecting the original information to a lower-dimensional feature space, through principal component analysis[13], independent component analysis[14], and maximum noise fraction[15]. FE transforms the data into the feature space and seeks a subset of effective and appropriate features in the feature space[16, 17]. However, the major disadvantage of dimensional reduction techniques is the loss of important, valuable information as each spectral band among hundreds of bands contains a significant information concerning the material. Hence, dimensionality-reduction technique results in compromising the classification performance of HSI.

Recently, Deep Learning (DL) architectures have demonstrated their capability in effective feature extraction of HSI data. These architectures have produced encouraging performance in image, speech, and language classification[18–21]. The deep-learning architecture is the latest advancement in neural networks with multiple processing layers capable of extracting more abstract and deep invariant features that results in improved classification performance. Some DL models have recently been implemented on HSI. A deep-learning based Stacked Auto-Encoder (SAE) is proposed in Ref. [8], and DBN was proposed in Ref. [22]. These methods extract deep features. In recent years, there is a growing interest in integrating spatial features along with the spectral features, as it has recently been demonstrated that combining spectral and spatial features significantly improves the classification results[23–27]. In Ref. [28], the integration of spatial information through morphological profiles and original spectral information is performed. Loopy belief propagation was used to exploit the spectral and spatial features in Ref. [29]. Spatial-spectral kernel sparse representation is exploited for HSI classification in Ref. [30]. Incorporation of spatial information before the classifier[27, 31], in the classifier, and after the classification through majority voting[26, 32] is also analyzed. Many classifiers have been investigated for HSI classification. Traditional HSI classification methods are mainly spectral feature-based methods such as logistic regression[33] Bayesian classifier, random forests, neural networks, K-Nearest-Neighbor classifier (K-NN), Conditional Random Fields (CRFs), sparse coding, and Support Vector Machine (SVM). The SVM performes well due to its ability to handle high-dimensional data. However, these techniques also pose some limitations. A full and complete connection between the different hidden layers makes the deep-learning based HSI feature extraction very complex as HSI comprises hundreds of spectral bands and hidden units are required to evaluate the input and estimate the parameters of all the available spectral bands that are concurrently in the same activation function. This complex connection leads to a lack of proper abstraction. Moreover, the full connection of layers also demands more training data to train the parameters, which is a major limitation in HSI as limited training data is available for HSI. Thus, SAE and DBN are unable to extract the spatial information efficiently as fixed-sized window is used to extract the spatial information, which may contain multiple classes or subsets of the same class.

Here, we propose a Spectral-Adaptive Segmented

DBN (SAS-DBN) that performs spectral segmentation to efficiently reduce and divide the complexity, and fully exploits the available spectral and spatial information for better feature extraction in the presence of a limited number of available sampled data. This paper introduces the segmented-DBN technique, where local DBNs are applied to each segment of the spectral channels. It consists of two main steps: First, adaptive boundary adjustment-based segmentation is performed to exploit spatial information. Second, spectral segmentation is performed, where similar spectral contiguous bands are grouped together and DBN is applied seperatley to each spatial-spectral segmented group of bands. Locally applying DBN-based feature extraction to each group of bands reduces the computational complexity and simultaneoulsy results in better features, improving the classification accuracy. The main contributions of this study are listed as follows:

- Spectral information in each channel is fully utilized. Spectrally similar contiguous bands are grouped together, and DBN is applied to these segments of the spectrum for efficient feature extraction that divides and reduces the complexity and allows the local extraction of features in the presence of limited training data, resulting in improved classification accuracy.
- Spatial information is fully exploited by replacing the fixed-sized spatial window with hyper-segmented structures that adaptively change the size and shape according to the actual structure in the HSI.

The main focus of this work is the efficient utilization of all the valuable information provided in the spectral channels without compromising the computational complexity for HSI classification and without reducing the dimension of the hyperspectral data. The organization of the rest of the paper is as follows. A brief overview of DBN and proposed SAS-DBN is described in Section 2. Section 3 describes the experimental analysis on real and standard HSIs while concluding remarks are drawn in Section 4.

2 Spectral-Adaptive Segmented DBN for HSI Classification

This paper proposes exploiting spectral features through adaptive spectral segmentation-based DBN to enhance the DBN process, where highly correlated bands are grouped together and fed into a local DBN to extract more abstract features. A general flowchart of the proposed framework is shown in Fig. 1. Feature extraction process should consider the following spectral-spatial factors:

- There is a high probability that HSI pixels with identical spectral signatures contribute to the same class label.
- Neighboring pixels in spatial domain that are highly relevant in spectral signatures also

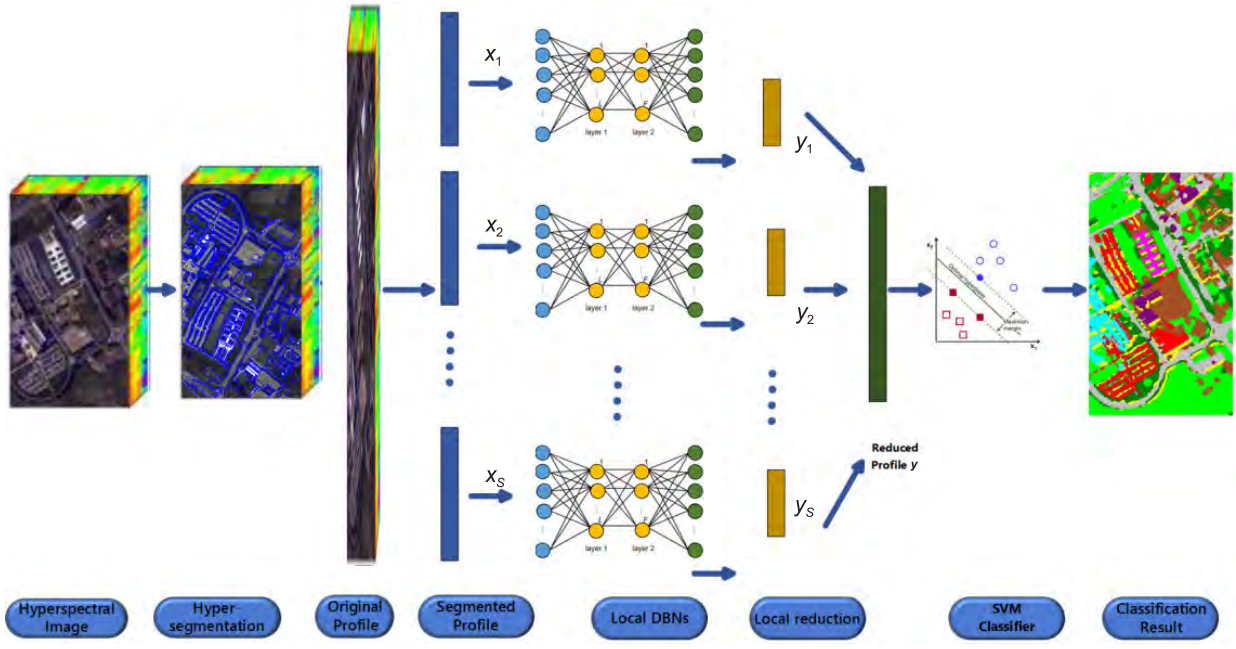


Fig. 1 Framework of the SAS-DBN process.

contribute to the same class label.

It is evident from the factors mentioned that pixels with high spectral and spatial correlations share the same class with a high probability. To effectively employ the mentioned factors, we improved the DBN to exploit spectral-spatial features by replacing the conventional, fixed-sized spatial window with an adaptive structural window, and by dividing similar spectral channels for higher HSI classification. The framework of the proposed technique is shown in Fig. 1. In the first phase, the spatially adaptive segmentation approach is applied[34] to segment the image into spatially similar regions. In the second phase, HSI spectral channels are grouped into different spectrally similar segments, and then in the third and final phases, the DBN is applied to each spectrally segmented group of channels with spatially similar regions.

2.1 Hyper-segmentation-based spatial feature extraction

As discussed above, considering the spatial information can result in improved classification performance. According to some researchers[8], two facts must be considered for effective HSI classification: (1) It is highly possible that pixels with similar spectral features share the same class and (2) It is highly possible that spatially neighboring pixels with similar spectral signatures share the same class. To consider these, effective segmentation plays an important role as segmentation leading to a grouping of spatially similar pixels.

In Ref. [35], the authors exploited the spatial information for HSI classification by using a static window size. However, the use of the static window size limits the exploitation of the spatial information, which degrades performance, as the window may contain more than one class or it may contain a sub-part of one class as shown in Fig. 2. In hyper-segmentation, neighboring pixels with similar spatial characteristics are grouped together. Adaptive boundary adjustment-based approach[34] is utilized to effectively segment the HSI into spatially similar regions. Hence, spatial consistency is ensured using the hyper-segmentation approach[34] that segments the HSI using local spatial regularity, where the size and shape of a structure is adaptively adjusted based on the actual boundaries. The main steps for hyper-segmentation are as follows:

(1) Divide the HSI into m initial hexagonal segments.

(2) Iterate over the following steps until no pixel goes into a different segment.

- Calculate the major class of each initial segments.
- Calculate the gradient of current boundary pixels of each segment.
- Calculate the straightness factor of each segment boundary.
- Based on the energy function, calculate the measure of similarity, between centroid of each class and its edge.
- Evolve the boundary based on the similarity measure.

(3) For each channel group, assign a weight to each boundary pixel according to certain criteria to award pixels that reflect an actual boundary.

The detailed procedure of the hyper-segmentation is illustrated in Fig. 3.

(a) Fixed size window (b) Adaptive boundary

Fig. 2 Spatial region selection by (a) fixed-square window and (b) adaptive boundary adjustment-based segmentation.

2.2 Spectral-spatial feature extraction by segmented DBN

2.2.1 DBN

A deep belief net is a formation of generative neural network-based learning modules, each of which is a Restricted Boltzmann Machine (RBM) that consists of an input layer to receive data and a hidden layer that learns to distinguish features that capture higher-order correlations in the input data, as shown in Fig. 4. The two RBM layers are connected by a matrix of symmetrically weighted connections, $\boldsymbol{w}$, with no visible-visible or hidden-hidden connections. This restriction makes the hidden units conditionally independent. A combined formulation of the energy for hidden unit $\boldsymbol{u}$ and input units $\boldsymbol{x}$ is given by[22]

$$E(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta}) = -\sum_{j=1}^{n}\frac{(x_j-b_j)^2}{2\sigma^2} - \sum_{i=1}^{m} a_i u_i - \sum_{j=1}^{n}\sum_{i=1}^{m} w_{ji}\frac{x_j}{\sigma_i}u_i \quad (1)$$

where $\boldsymbol{\theta} = (a_i, b_i, w_{ji})$, w_{ji} is the weight between the

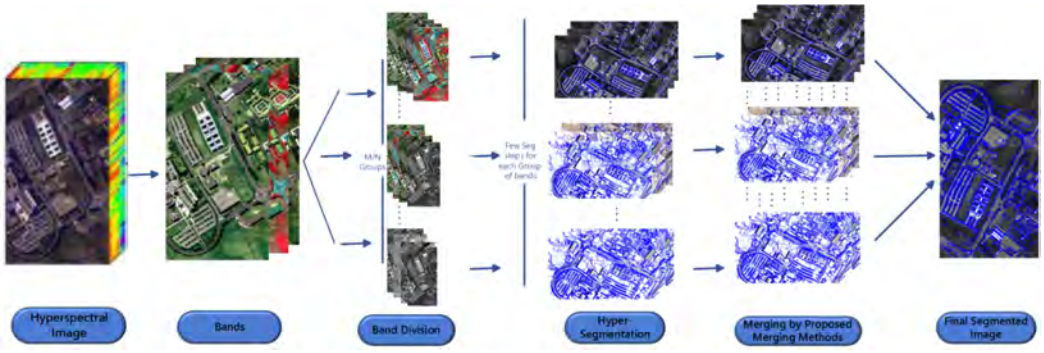


Fig. 3 Framework of segmentation[34].

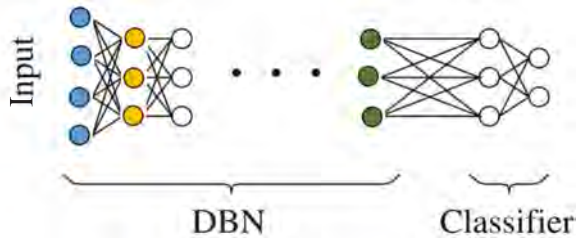


Fig. 4 Framework of the DBN based pixel-wise classification.

visible and hidden units, and b_j and a_i are bias terms of visible and hidden units, respectively. The joint distribution is given by

$$P(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta}) = \frac{a}{N(\boldsymbol{\theta})}\exp(-E(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta})) \tag{2}$$

where $N(\boldsymbol{\theta})$ is the normalizing constant. The network assigns a probabilistic value to each input vector through $E(\boldsymbol{x},\boldsymbol{u},\boldsymbol{\theta})$. The conditional distribution is given by

$$P(u_i|\boldsymbol{x};\boldsymbol{\theta}) = g\left(\sum_{j=1}^{n} w_{ji}x_j + a_i\right) \tag{3}$$

$$P(x_i|\boldsymbol{u};\boldsymbol{\theta}) = N\left(\sum_{i=1}^{m} w_{ji}u_i\sigma_j^2 + b_j\right) \tag{4}$$

where σ is the standard deviation of a Gaussian visible unit, and $N(\cdot)$ is the Gaussian distribution. The weights are learned by Constrastive Divergence (CD)[36] and are updated by

$$\Delta w_{ij} = \psi(x_i u_{j\,data} - x_i u_{j\,reconstruction}) \tag{5}$$

where ψ is the learning rate, and x_i and u_j are the input and the hidden units, respectively.

The strength of the RBM is in the restoration of the original information. In the reconstruction phase, the information in the hidden units is employed. The learned parameters are considered efficient if the framework can regain the original data. It employs that hidden units have preserved enough information of the original data. A single hidden layer is not enough to extract features in the given HSI data. Therefore, after training one RBM layer, learned features can be used as an input to the second RBM layer. In this way, the RBM layers are stacked together and trained in a greedy way to form the DBN. The entire process can be summarized as follows:

(1) The dataset is used to pre-train a single layer of DBN. CD[36] is used to train the RBM.

(2) The output of the first layer is used as input to the second layer, which is trained as a second RBM.

(3) Repeat Steps 1 and 2 for the desired number of layers.

(4) Fine tune all learned parameters with the available labeled training samples.

(5) In this paper, we used SVM as classifier to classify the features obtained as output from the last DBN layer.

In summary, we first utilize the hyper-segmented spatial region and the spectral information as an input. A DBN is then applied to learn deep and abstract features from the inputs through multi-layer DBN. Finally, SVM is utilized to classify and label the pixels based on the learned features as shown in Fig. 5. For each channel group, spectral and spatial information is fed into the network as shown in Fig. 5.

2.2.2 Segmented DBN

In traditional application of DBN in HSI, all spectral bands are treated equally and concurrently. This enhances complexity because the hidden units in the first layer are bombarded with the original feature dimension, which consists of a hundred of channels. Moreover, the correlation among different spectral regions of the HSI is not considered. Therefore, DBN can be employed in parts, in different segments of the spectral channels. This concept of applying a particular algorithm into different segments is

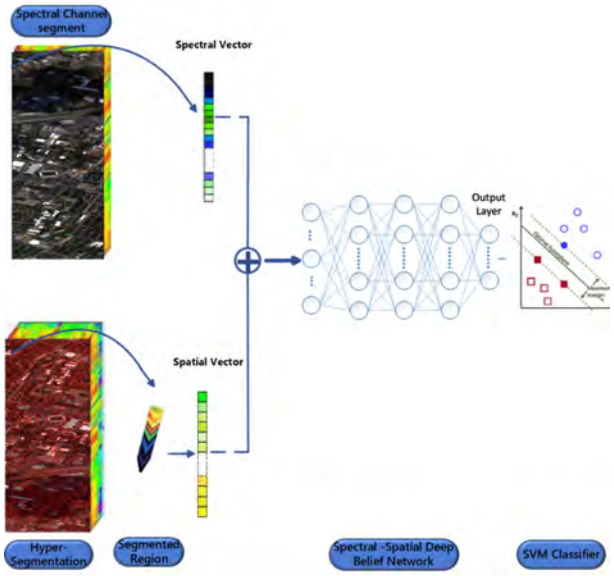


Fig. 5 Framework of spectral-spatial classification.

introduced and successfully used in other FE methods, such as segmented PCT[37] and other comparable modifications[15].

A detailed description of the proposed segmented DBN is presented in Fig. 1, where the original number of spectral channels $\boldsymbol{x}$ is segmented into S different regions x_s, $s \in [1, S]$ and then DBN is independently applied to each segmented region of spectral channels.

As the local DBNs deal with a small subset of the original spectral channels, they utilize a reduced number of hidden units. Hence SAS-DBN employs a number of DBNs that are however simpler than the ones utilized in the traditional applications. Moreover, more abstract and deep features are extracted in an easier way. Finally, the reduced features from the local regions y_s are integrated ($\sum_{s=1}^{S} F_s = F$) to form a reduced feature vector.

To segment HSI data channels, a correlation matrix among the spectral bands is exploited. The correlation matrix can be effectively employed to compute the correlation among different channels. High correlation segments[15] from correlation matrix can be utilized to define the SAS-DBN segments. The correlation matrix is formulated based on a covariance matrix. Mathematically, the covariance matrix can be defined as

$$\mathbf{Cov} = E\{(\boldsymbol{q} - E(\boldsymbol{q}))(\boldsymbol{q} - E(\boldsymbol{q}))^{\mathrm{T}}\} \tag{6}$$

where $E\{\cdot\}$ is the expectation operator, $\boldsymbol{q}$ is a random variable. Based on this, each element in the correlation matrix can be defined as

$$\mathbf{Corr}(r, c) = \frac{\mathbf{Cov}(r, c)}{\sqrt{\mathbf{Cov}(r, r)\mathbf{Cov}(c, c)}} \tag{7}$$

where $\mathbf{Cov}(r, r)$ and $\mathbf{Cov}(c, c)$ represents the covariance of the r-th and c-th channels, respectively. Therefore, $\mathbf{Corr}(r, c)$ represents the correlation among the r-th and c-th spectral channels in the HSI dataset. Hence, the complete correlation matrix offers a correlation among the spectral channel in the HSI dataset. Hence correlation information can be effectively employed to define the regions.

3 Experimental Results and Performance Comparisons

To validate and evaluate the performance of the proposed technique, several experiments were conducted on real-world urban and natural-scene hyperspectral image datasets. The detailed description of each dataset is given in the following subsections.

3.1 ROSIS urban scene: Pavia University dataset

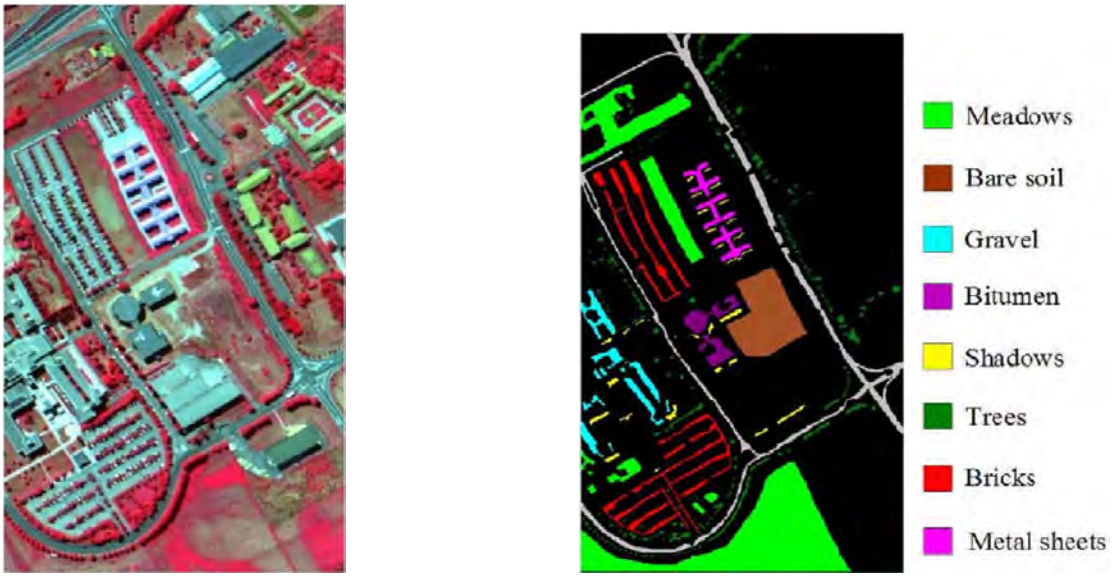
The Pavia University Scene was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the urban area of Pavia University, Italy. The Pavia scene comprises a spatial resolution of 610×340 pixels and a spectral resolution of 115 channels. The spatial resolution of the scene is 1.3 m/pixel and the spectral range is $0.43 - 0.86$ μm. A total of 12 noisy bands were removed due to water absorption with 103 remaining bands. Nine standard classes are utilized for Pavia scene classification. The false-color composite and reference ground truth are shown in Fig. 6.

3.2 AVIRIS scene: Houston image

The Houston database was acquired over the University of Houston and neighboring urban region. The database consists of 144 spectral channels with a spectral resolution of 380 to 1050 nm and a spatial area of 349×1905 pixels. It also consists of 15 different ground cover classes as shown in the false-color composite and ground truth in Fig. 7.

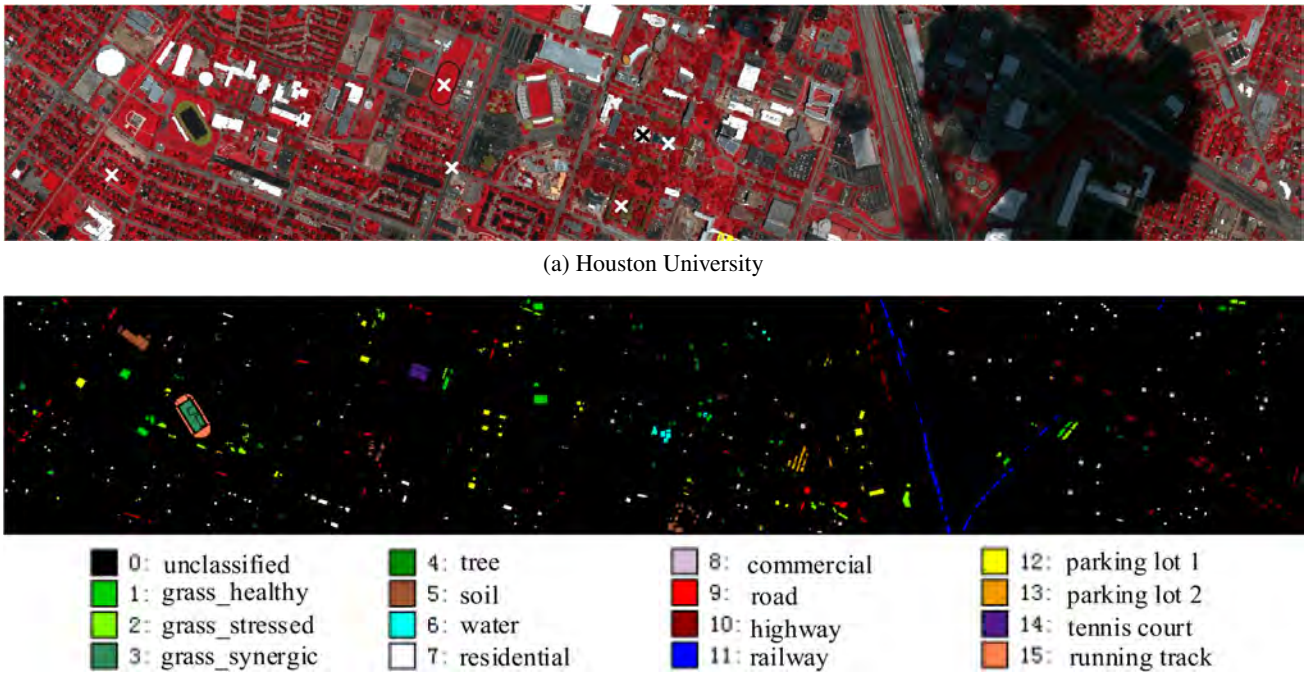
3.3 Experimental setup

The experiments were conducted on a 4.0-GHz processor with NVIDIA GeForce GTX 970 on a Windows 7 operating system. Theano was used for code implementation. A conventional DBN can be employed in different ways. The conventional DBN delivers the best performance with $2 - 6$ layers and $20 - 60$ hidden units as suggested in Ref. [22]. The hidden units of the deepest layer are equal in number to the features



(a) Pavia University false color (b) Pavia Univeriy ground truth and classes

Fig. 6 Hyperspectral image datasets of Pavia University.



(a) Houston University

(b) Ground truth and classes

Fig. 7 Hyperspectral image datasets of Houston University.

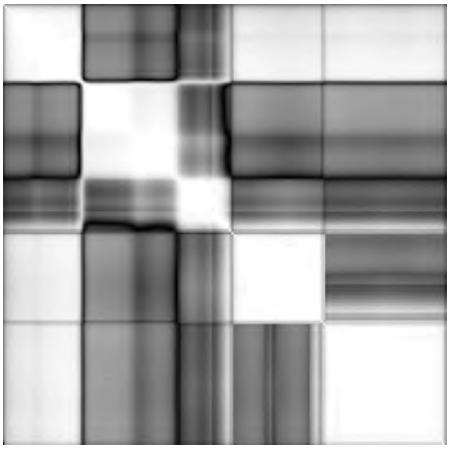
that are required. From these results, we employed a two-layer DBN architecture with 40 hidden units in each. In total, *three* layers from the Pavia University and Houston University datasets were used with one input and two hidden layers, with a configuration of $x - 40 \cdots - F$, where x corresponds to the number of channels as an input to the DBN, and F is the number of the corresponding required features that are also required as output. A greater number of hidden layers or hidden units do not have a significant effect on the classification performance, as specified in Ref. [22].

The proposed method first segments the spectral data into different groups so that the DBN can be applied to each group separately. The correlation among spectral bands can be effectively exploited using a Correlation Matrix (CM)[15]. The CM is closely

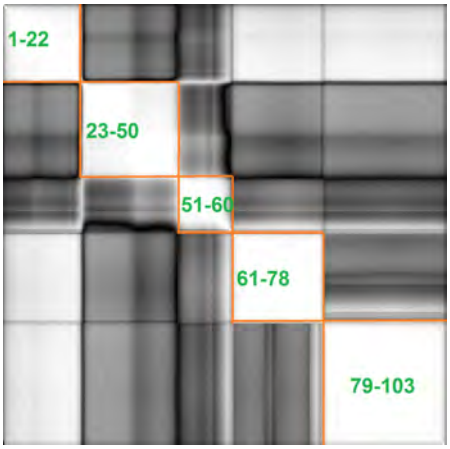
associated with a covariance matrix. A particular value from CM represents the correlation between the corresponding channels. For instance, CM(i, j) represents the corresponding relation between the i-th and j-th bands of a particular HSI. The entire CM delivers the association between every pair of bands in the hyperspectral channels, which can be efficiently utilized to define the spectral segmented groups, as can be observed in Fig. 8. As recommended in Ref. [15], the segmented group of channels can be derived from the main correlation groups obtained as a result of CM. Consequently, Table 1 presents the proposed groups of spectral channels. The total number of features is also divided (5, 10, 15, 20) evenly among the groups.

3.4 Spectral-spatial HSI classification

Experimental results of the proposed approach were compared with those of well-known existing techniques, the SVM[38], the recently developed Recurrent Neural Network (RNN)[39], the DBN with Logistic Regression (DBN-LR)[22], and the newly developed deep CNN[40]. In the case of the DBN-LR, only spectral data is considered an input.



(a) Correlation matrix (white=1, black=0)



(b) Selected regions

Fig. 8 Correlation matrix and corresponding band grouping for the Pavia University dataset.

Table 1 Segmented-DBN configuration for Pavia University and Houston University datasets.

Dataset	Segment No.	Grouped channels	Hidden-layer units	Reduced features
Pavia University	Segment S_1	$1-22$	$L_1=13$ $L_2=F$	1 3 3 6
	Segment S_2	$23-60$	$L_1=13$ $L_2=F$	2 3 6 7
	Segment S_3	$61-103$	$L_1=14$ $L_2=F$	2 4 6 7
Houston University	Segment S_1	$1-76$	$L_1=13$ $L_2=F$	2 4 7 6
	Segment S_2	$77-110$	$L_1=14$ $L_2=F$	1 3 7 6
	Segment S_3	$111-144$	$L_1=13$ $L_2=F$	2 3 6 3

The following standard evaluation criteria are used to access classification performance:

(1) *Overall Accuracy (OA)*: OA represents the number of HSI pixels correctly classified and divided by the total number of test samples taken.

(2) *Average Accuracy (AA)*: AA calculates the mean of the classification accuracies of all classes.

(3) *Kappa Coefficient (k)*: Kappa coefficient determines the agreement between the final classified map and the actual ground truth map. It is generally classified as a more accurate measurement as it takes into consideration the agreement occurring by chance[41].

The Houston University dataset is considered a challenging dataset due to the presence of small spatial regions. Only 10% of samples randomly chosen from each class was used as training samples and 50 features are used in each dataset for classification. Tables 2 and 3 shows the class-level accuracies of the Houston University and Pavia University datasets respectively.

Tables 2 and 3 further show that the SAS-DBN delievered better performance in OA, AA, and k among all the four well-known existing techniques, although the DBN-LR yielded better performance in classes 1 and 2. This is due to the availability of a very small number of samples for those classes. The proposed technique performed particularly well in classes with

Table 2 Classification accuracy (%) of each class for the Houston University dataset obtained by the SVM[38], RNN[39], CNN[40], DBN-LR[22], and proposed SAS-DBN.

Class	Training	Test	SVM	RNN	CNN	DBN-LR	SAS-DBN
1	125	1126	97.47	82.53	81.20	99.20	98.80
2	125	1129	98.32	83.36	83.55	99.60	99.0
3	70	627	99.37	100.0	99.41	100.0	100.0
4	124	1120	98.01	90.53	91.57	99.60	99.60
5	124	1118	96.01	97.82	94.79	99.60	99.60
6	33	292	99.83	93.01	95.10	97.2	98.81
7	127	1141	91.23	75.37	63.53	97.0	98.11
8	124	1120	86.23	42.36	42.64	97.8	98.0
9	125	1127	86.99	77.62	58.17	94.0	95.11
10	123	1104	91.42	57.63	41.80	97.4	97.75
11	124	1111	91.67	77.42	75.71	97.3	97.95
12	123	1110	87.05	69.74	84.15	95.2	96.05
13	47	422	78.16	66.32	40.00	88.0	90.55
14	43	385	97.42	100.0	98.79	100.0	100.0
15	66	594	99.49	95.98	97.89	100.0	100.0
Overall Accuracy			93.06	89.85	85.42	97.70	98.35
Average Accuracy			93.25	80.65	76.55	97.50	98.06
Kappa Coefficient			0.925	0.7606	0.7200	0.975	0.9805

small spatial regions. The over-all classification results for both the datasets demonstrate that the proposed SAS-DBN approach led to better performance than existing state-of-the-art approaches.

The visual results of the classification maps for all labeled samples of various existing classification methods for Pavia University and Houston University are presented in Figs. 9 and 10, respectively. From the subsequent visual results, it is evident that incorporation of spatial information along with extracted features through the proposed technique improved the results. Some existing techniques resulted in scattered and noisy labeled points, where the proposed method

Table 3 Classification accuracy (%) of each class for the Pavia University dataset obtained by the SVM[38], RNN[39], CNN[40], DBN-LR[22], and proposed SAS-DBN.

Class	Training	Test	SVM	RNN	CNN	DBN-LR	SAS-DBN
1	597	6034	97.50	84.45	87.34	87.37	89.11
2	1681	16971	97.70	85.24	94.63	92.10	93.55
3	189	1910	78.53	54.31	86.47	85.57	87.50
4	276	2788	89.29	95.17	96.29	95.11	97.35
5	121	1224	98.77	99.93	99.65	99.74	99.19
6	453	4576	83.04	80.99	93.23	91.94	93.85
7	120	1210	64.58	88.35	93.19	92.21	93.55
8	331	3351	86.90	88.62	86.42	87.02	88.05
9	85	862	99.92	99.89	100.0	100.0	100.0
Overall Accuracy			92.04	88.85	92.56	91.18	93.15
Average Accuracy			88.47	86.33	93.02	92.34	93.06
Kappa Coefficient			0.903	0.8048	0.9006	0.8828	0.9105

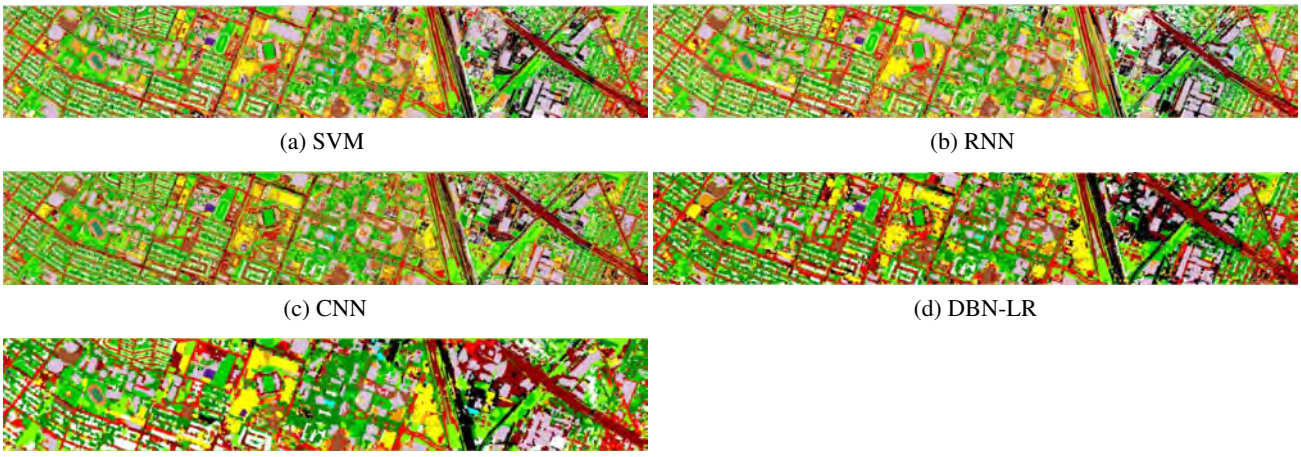
overcame this deficiency.

4 Conclusion

In this paper, an HSI classification SAS-DBN approach based on Deep Belief Networks and hyper-segmentation was proposed by considering spectral and spatial information. We proposed a deep-learning-based SAS-DBN architecture by analyzing a DBN and solving relevant concerns by spectral and spatial segmentation. Despite its learning capability, tackling hundreds of features at the same time leads to complexity and affects the performance of the DBN. We, thus, proposed a two-step classification approach combining the use of segmented spatial data and spectral segmentation, dividing the original spectral domain into different correlated bands, and applying DBN separately to each segment. Hence reducing the complexity of the learning process and extracting local features make it simpler for DBN to effectively extract

(a) SVM (b) RNN (c) CNN (d) DBN-LR (e) SAS-DBN

Fig. 9 Classification maps of various techniques for Pavia University.



(a) SVM (b) RNN

(c) CNN (d) DBN-LR

(e) SAS-DBN

Fig. 10 Classification maps of various techniques for the Houston University.

the spectral-spatial features. The proposed SAS-DBN delivers better HSI classification performance.

Experiments indicated that the two-step classification approach effectively combines spectral and spatial features to capture the generalities and specific details of datasets. Moreover, the multi-scale modules enhance the level of classification without excessively increasing the number of parameters.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (No. 61672017) and the National High-Tech Research and Development Program of China (No. 2012AA011602).

References

[1] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection, *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, 2014.

[2] F. Van Der Meer, Analysis of spectral absorption features in hyperspectral imagery, *Int. J. Appl. Earth Obs. Geoinf.*, vol. 5, no. 1, pp. 55–68, 2004.

[3] F. M. Lacar, M. M. Lewis, and I. T. Grierson, Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia, in *Proc. 2001 Int. Geoscience and Remote Sensing Symp. on Scanning the Present and Resolving the Future*, Sydney, Australia, 2001, pp. 2875–2877.

[4] T. J. Malthus and P. J. Mumby, Remote sensing of the coastal zone: An overview and priorities for future research, *Int. J. Remote Sens.*, vol. 24, no. 13, pp. 2805–2815, 2003.

[5] P. W. T. Yuen and M. Richardson, An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition, *Imaging Sci. J.*, vol. 58, no. 5, pp. 241–253, 2010.

[6] O. Kuybeda, D. Malah, and M. Barzohar, Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors, *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, 2007.

[7] X. P. Jia, B. C. Kuo, and M. M. Crawford, Feature mining for hyperspectral image classification, *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, 2013.

[8] Y. S. Chen, Z. H. Lin, X. Zhao, G. Wang, and Y. F. Gu, Deep learning-based classification of hyperspectral data, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, 2014.

[9] G. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[10] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, 2013.

[11] W. Di, L. Zhang, D. Zhang, and Q. Pan, Studies on hyperspectral face recognition in visible spectrum with feature band selection, *IEEE Trans. Syst., Man, Cybern.-Part A: Syst. Humans*, vol. 40, no. 6, pp. 1354–1361, 2010.

[12] A. Mughees, X. Q. Chen, R. C. Du, and L. M. Tao, AB^{3}C: Adaptive boundary-based band-categorization of hyperspectral images, *J. Appl. Remote Sens.*, vol. 10, no. 4, p. 046009, 2016.

[13] K. Y. Yeung and W. L. Ruzzo, Principal component analysis for clustering gene expression data, *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.

[14] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar, A method for making group inferences from functional MRI data using independent component analysis, *Human Brain Mapp.*, vol. 14, no. 3, pp. 140–151, 2001.

[15] J. Zabalza, J. C. Ren, M. Q. Yang, Y. Zhang, J. Wang, S. Marshall, and J. W. Han, Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing, *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 112–122, 2014.

[16] L. O. Jimenez-Rodriguez, E. Arzuaga-Cruz, and M. Velez-Reyes, Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data, *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 469–483, 2007.

[17] S. B. Serpico and G. Moser, Extraction of spectral channels from hyperspectral images for classification purposes, *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484–495, 2007.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proc. 25th Int. Conf. Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[19] Z. L. Yuan, Y. Q. Lu, and Y. B. Xue, Droiddetector: Android malware characterization and detection using deep learning, *Tsinghua Sci. Technol.*, vol. 21, no. 1, pp. 114–123, 2016.

[20] Y. Gao and Q. H. Dai, Efficient view-based 3-D object retrieval via hypergraph learning, *Tsinghua Sci. Technol.*, vol. 19, no. 3, pp. 250–256, 2014.

[21] X. P. Chen and H. Huang, Immune feed-forward neural network for fault detection, *Tsinghua Sci. Technol.*, vol. 16, no. 3, pp. 272–277, 2011.

[22] Y. S. Chen, X. Zhao, and X. P. Jia, Spectral–spatial classification of hyperspectral data based on deep belief network, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, 2015.

[23] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, Advances in spectral-spatial classification of hyperspectral images, *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, 2013.

[24] W. Z. Zhao and S. H. Du, Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach, *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, 2016.

[25] W. Z. Zhao and S. H. Du, Learning multiscale and deep representations for classifying remotely sensed imagery, *ISPRS J. Photogramm. Remote Sens.*, vol. 113, pp. 155–165, 2016.

[26] A. Mughees and L. M. Tao, Efficient deep auto-encoder learning for the classification of hyperspectral images, in *Proc. 2016 Int. Conf. Virtual Reality and Visualization*, Hangzhou, China, 2016, pp. 44–51.

[27] A. Mughees, A. Ali, and L. M. Tao, Hyperspectral image classification via shape-adaptive deep learning, in *Proc. 2017 IEEE Int. Conf. Image Processing*, Beijing, China, 2017.

[28] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles, *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, 2008.

[29] J. Li, J. M. Bioucas-Dias, and A. Plaza, Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning, *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, 2013.

[30] J. J. Liu, Z. B. Wu, Z. H. Wei, L. Xiao, and L. Sun, Spatial-spectral kernel sparse representation for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 6, no. 6, pp. 2462–2471, 2013.

[31] A. Mughees and L. M. Tao, Spectral-spatial hyperspectral image classification via boundary-adaptive deep learning, in *Proc. 2017 Int. Conf. Digital Image Computing: Techniques and Applications*, Sydney, Australia, 2017.

[32] A. Mughees and L. M. Tao, Hyperspectral image classification based on deep auto-encoder and hidden markov random field, in *Proc. 13th Int. Conf. Natural Computation, Fuzzy Systems and Knowledge Discovery*, Guilin, China, 2017.

[33] G. M. Foody and A. Mathur, A relative evaluation of multiclass image classification by support vector machines, *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, 2004.

[34] A. Mughees, X. Q. Chen, and L. M. Tao, Unsupervised hyperspectral image segmentation: Merging spectral and spatial information in boundary adjustment, in *Proc. 55th Annual Conf. the Society of Instrument and Control Engineers of Japan*, Tsukuba, Japan, 2016, pp. 1466–1471.

[35] A Ambikapathi, T. H. Chan, C. H. Lin, and C. Y. Chi, Convex geometry based outlier-insensitive estimation of number of endmembers in hyperspectral images, *Signal*, vol. 1, pp. 1–20, 2012.

[36] G. E. Hinton, Training products of experts by minimizing contrastive divergence, *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[37] X. P. Jia and J. A. Richards, Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification, *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, 1999.

[38] F. Melgani and L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.

[39] L. C. Mou, P. Ghamisi, and X. X. Zhu, Deep recurrent neural networks for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, 2017.

[40] W. Hu, Y. Y. Huang, L. Wei, F. Zhang, and H. C. Li, Deep convolutional neural networks for hyperspectral image classification, *J. Sensors*, vol. 2015, p. 258619, 2015.

[41] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Norwood, MA, USA: Artech House, 2015.

Atif Mughees received the MS degree from the National University of Science and Technology, Islamabad, Pakistan in 2010. Currently, he is working toward the PhD degree in the Key Laboratory of Pervasive Computing, Department of Computer Science and Technology, Tsinghua University, China. His research interests include image processing, remote sensing applications, and machine learning with a special focus on spectral and spatial techniques using deep learning for hyperspectral image classification.

Linmi Tao received the MS degree in cognitive science from Chinese Academy of Sciences, Beijing, China in 1991, and the PhD degree in computer science from Tsinghua University, China, in 2001. Currently, he is an associate professor with the Department of Computer Science and Technology, Tsinghua University. He has studied and worked with the International Institute for Advanced Scientific Studies and the University of Verona, Italy, and Tsinghua University on computational visual perception, 3-D visual information processing, and computer vision.