A Parallel Gaussian–Bernoulli Restricted Boltzmann Machine for Mining Area Classification With Hyperspectral Imagery

Kun Tan, *Senior Member, IEEE*, Fuyu Wu, Qian Du, *Fellow, IEEE*, Peijun Du, *Senior Member, IEEE*, and Yu Chen

***Abstract*—In this paper, a novel feature extraction method is proposed for hyperspectral image classification using a Gaussian–Bernoulli restricted Boltzmann machine (GBRBM) in parallel. The proposed approach employs several GBRBMs with different hidden layers to extract deep features from hyperspectral images, which are nonlinear and local invariant. Based on the learned deep features, a logistic regression layer is trained for classification. The proposed approaches are carried out on two public hyperspectral datasets: Pavia University dataset and Salinas dataset, and a new dataset obtained by HySpex imaging spectrometer in the mining area in Xuzhou. The obtained results reveal that the proposed approach offers superior performance compared to traditional classifiers. The advantage of the proposed GBRBM is that it can extract deep features in an unsupervised way and reduce the prediction time by using GPU. In particular, the classification results of the mining area provide valuable suggestions to improve environmental protection.**

***Index Terms*—Deep learning, Gaussian–Bernoulli restricted Boltzmann machine (GBRBM), hyperspectral image classification.**

I. Introduction

HYPERSPECTRAL image classification has important remote sense applications. Many machine learning techniques have been applied, such as K-nearest neighbor [1], Bayes classifier [2], and support vector machine (SVM) [3], to achieve classification. However, the increase of spectral dimension leads to low efficiency of classification algorithms and large computational cost [4]. Therefore, it is important to reduce dimensionality before classification. Feature extraction is one of dimensionality reduction methods by transforming the original features into a lower dimensional feature space. The commonly used linear models include the independent component analysis [5], linear discriminant analysis [6], and their variants. In [7], Wen *et al*. presented a supervised linear manifold learning feature extraction method for hyperspectral image classification. In [8], Friedman *et al*. alternated the usual maximum likelihood estimates for the covariance matrices and proposed the regularized discriminant analysis, and quadratic discriminant analysis. In [9], Howland *et al*. modified the generalized singular value decomposition for the commonly used SVD algorithm to solve the small-sample-size problem. However, the linear models cannot deal with high-dimensional data effectively. In order to solve these problems, a variety of nonlinear models, such as locally linear embedding [10], Isomap [11], and Laplacian eigenmaps [12], were proposed. The nonlinear models can effectively exploit the intrinsic structure of data. In [13], Wu *et al*. introduced the kernel function into the PCA algorithm, which was applied to the near-infrared spectral data for feature extraction. In [14], Sun *et al*. proposed a fast and robust principal component analysis on the Laplacian graph for hyperspectral band selection, which regularized the Laplacian graph into a regular robust principal component analysis. In [15], Mika *et al*. presented a fisher discriminant analysis combined with a kernel function, but the disadvantage is that the model is complex and time-consuming. Sun *et al*. proposed a sparse and low-rank near-isometric linear embedding to extract features in hyperspectral imagery, which performed the best among comparative methods [16].

Recently, deep learning has made a major breakthrough in all aspects [17]. For instance, a deep neural network (DNN) can extract high-level features for efficient classification [18]. Currently, the DNN mainly includes deep belief networks (DBN) [19], deep Boltzmann machines [20], deep auto-encoder neural networks [21], and deep convolutional neural networks [22]. They also provide a new solution for feature extraction and classification of hyperspectral images. In [23], Zhang *et al*. gave a technical tutorial of deep learning for the remote sensing data analysis, which introduced the applications of deep learning in

Manuscript received February 9, 2018; revised June 17, 2018 and December 31, 2018; accepted January 5, 2019. Date of publication February 6, 2019; date of current version March 4, 2019. This work was supported in part by the Natural Science Foundation of China under Grants 41871337 and 41471356 and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions. *(Kun Tan and Fuyu Wu contributed equally to this work.) (Corresponding author: Kun Tan.)*

K. Tan is with the Key Laboratory for Land Environment and Disaster Monitoring of NASG, China University of Mining and Technology, Xuzhou 221116, China, and also with the Key Laboratory of Geographic Information, Ministry of Education, East China Normal University, Shanghai 200241, China (e-mail: tankuncu@gmail.com).

F. Wu and Y. Chen are with the Key Laboratory for Land Environment and Disaster Monitoring of NASG, China University of Mining and Technology, Xuzhou 221116, China (e-mail: wufuyu666@163.com; chenyurs@gmail.com).

Q. Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

P. J. Du is with the Key Laboratory for Satellite Mapping Technology and Applications of NASG, Nanjing University, Nanjing 210023, China (e-mail: dupjrs@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2019.2892975

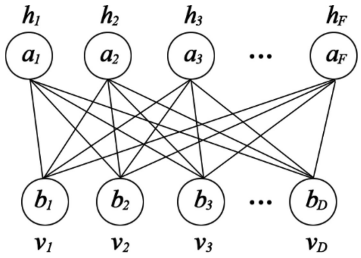

Fig. 1. Structure of RBM.

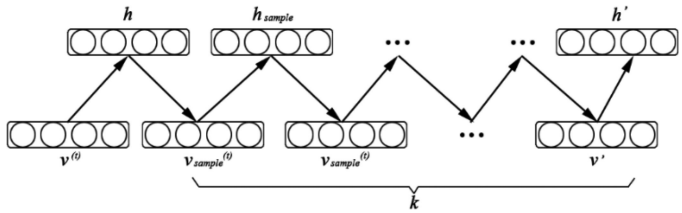

Fig. 2. Model of CD training procedure.

different perspectives of image preprocessing, such as pixel-based classification, target recognition, and scene understanding. In [24], a DBN-based classifier achieved a spectral–spatial classification of hyperspectral data. In [25], Lin *et al*. applied the deep auto-encoder neural networks for the classification of hyperspectral imagery. Meanwhile, many methods combining traditional classification strategies with deep learning were proposed. In [26], Ma *et al*. presented a novel semisupervised classification based on multi-decision labeling and deep feature learning. In [27], Wang *et al*. introduced a hybrid of PCA, guided filtering, and deep learning into hyperspectral image classification. Recently, multiple neural networks have been shown to be successful in computer vision. In [28], Farabet *et al*. introduced a discriminative framework using a multi-scale convolutional network operated on raw pixels to learn appropriate low-level and mid-level features for scene parsing. In [29], Längkvist *et al*. found that multiple CNNs in parallel with varying contest can achieve stable classification accuracy. In [30], Zhou *et al*. raised a group belief network based on unlabeled hyperspectral data, which was the first attempt to incorporate group knowledge of hyperspectral features for classification.

However, in the application of hyperspectral classification, it has been noticed that classification accuracy may be degraded with the increase of hidden layers. There are two main reasons for this phenomenon: first, the depth of DNN directly affects the number of parameters, and the deeper the DNN, the more the number of parameters. Thus, due to lack of training samples, the DNN cannot be trained effectively. Second, the classification of hyperspectral images is carried out pixel by pixel, which is different from natural images. The number of bands in hyperspectral images is about 200, which leads to the absence of DNNs for feature extraction. It has been argued that a single-hidden-layer network with well-tuned parameters may be more effective than DNN [31].

In this paper, we focus on single-layer Gaussian–Bernoulli restricted Boltzmann machines (GBRBM) and multiple GBRBMs in parallel to extract the features of hyperspectral data. The extracted features are then used by a logistic regression (LR) layer to address the classification problem. Compared with the contrast methods, the parallel GBRBM method can make full use of spectral information by combining different features. Moreover, this method has fast convergence due to fewer parameters. According to experiments, the parallel GBRBM can offer better classification accuracy with lower computational cost, which is more suitable for applications of hyperspectral datasets.

The rest of the paper is organized as follows. The theory of GBRBM and the structure of GBRBM in parallel are given in Section II. Datasets and experimental results are shown in Section III. An application of the model to a mining area is given in Section IV, and conclusions are drawn in Section V.

II. Proposed Method

A. Restricted Boltzmann Machine (RBM)

An RBM is a two-layer network, which contains a "visible" layer $v = [0, 1]^D$ and a "hidden" layer $h = [0, 1]^F$. The illustration of RBM is shown in Fig. 1.

A joint configuration (v, h) of the visible and hidden layers has an energy given by [32]

$$E(v, h; \theta) = -\sum_{i=1}^{D} b_i v_i - \sum_{j=1}^{F} a_j h_j - \sum_{i=1}^{D}\sum_{j=1}^{F} w_{ij} v_i h_j \quad (1)$$

where $\theta = [a_j, b_i, w_{ij}]$, w_{ij} represents the weight between the hidden and visible layers, and a_j and b_i are bias terms of the hidden and visible layers, respectively. The joint distribution over the layers is defined by

$$P(v, h; \theta) = \frac{1}{Z(\theta)} \exp\left(-E(v, h; \theta)\right) \quad (2)$$

where the "partition function" $Z(\theta)$ is given by summing over all possible pairs of hidden and visible vectors as

$$Z(\theta) = \sum \sum E(v, h; \theta). \quad (3)$$

The conditional distributions of the hidden layer *h* and visible layer *v* are given by

$$p(h_j = 1|v) = \sigma\left(a_j + \sum_i v_i W_{ij}\right) \quad (4)$$

$$p(v_i = 1|h) = \sigma\left(b_i + \sum_j h_j W_{ij}\right) \quad (5)$$

$$\sigma(x) = \frac{1}{1 + \exp(x)}. \quad (6)$$

In general, a method named contrastive divergence (CD) is used to train the RBM model. The model of CD is illustrated in Fig. 2.

During training, the parameters of RBM can be updated as

$$w_{ij} = w_{ij} + \varepsilon_w(E_{\text{data}}[v_i h_j] - E_{\text{model}}[v_i h_j]) \quad (7)$$

$$a_j = a_j + \varepsilon_a(E_{\text{data}}[h_j] - E_{\text{model}}[h_j]) \quad (8)$$

$$b_i = b_i + \varepsilon_b(E_{\text{data}}[v_i] - E_{\text{model}}[v_i]) \quad (9)$$

where ε is the learning rate, E_{data} is the expectation of visible layer in an active state, and E_{model} is the expectation of visible layer in any active states.

B. Gaussian–Bernoulli Restricted Boltzmann Machines

The conventional RBM requires the state of layers to follow binary distribution, which limits their application. To solve the problem, a popular approach is to replace the binary visible layer with Gaussian ones [33]. The modified model is named GBRBM whose structure is same as the RBM [34]. The energy function of GBRBM is given by

$$E(v,h) = \sum_{i=1}^{V} \frac{(v_i - b_i)}{2\sigma^2} - \sum_{j=1}^{H} c_j h_j - \sum_{i=1}^{V}\sum_{j=1}^{H} \frac{v_i}{\sigma_i} h_j w_{ij}. \quad (10)$$

Unlike the traditional RBM, the conditional distributions of the hidden layer *h* and visible layer *v* are

$$p(v|h) = \frac{e^{-E(v,h)}}{\int e^{-E(v,h)} du} \sim N\left(b_i + \sigma_i \sum_{j=1}^{H} h_j w_{ij}, \sigma^2\right) \quad (11)$$

$$p(h_j = 1|v) = \frac{1}{1 + e^{\left(\sum_{i=1}^{v} \frac{v_i}{\sigma_i} w_{ij} + c_j\right)}} = \sigma\left(\sum_{i=1}^{v} \frac{v_i}{\sigma_i} w_{ij} + c_j\right). \quad (12)$$

During model training, the parameters of GBRBM are updated as

$$w_{ij} = w_{ij} + \varepsilon_w \left(E_{\text{data}}\left[\frac{v_i h_j}{\sigma^2}\right] - E_{\text{model}}\left[\frac{v_i h_j}{\sigma^2}\right]\right) \quad (13)$$

$$c_j = c_j + \varepsilon_b(E_{\text{data}}[h_j] - E_{\text{mdoel}}[h_j]) \quad (14)$$

$$b_i = b_i + \varepsilon_b \left(E_{\text{data}}\left[\frac{v_i}{\sigma^2}\right] - E_{\text{model}}\left[\frac{v_i}{\sigma^2}\right]\right) \quad (15)$$

$$\sigma_i = \sigma_i + \varepsilon_\sigma \left(E_{\text{data}}\left(\frac{(v_i - b_i)^2}{\sigma^2} - \sum_{j=1}^{H} h_j \frac{w_{ij} v_i}{\sigma^2}\right) - E_{\text{model}}\left(\frac{(v_i - b_i)^2}{\sigma^2} - \sum_{j=1}^{H} h_j \frac{w_{ij} v_i}{\sigma^2}\right)\right). \quad (16)$$

The parameter σ has an influence on the gradient calculation of other parameters in the process of model optimization. In general, to avoid this problem, the standardization pretreatment of the input data should be conducted before feature extraction. In this research, the parameter σ is set to be a constant 1, and it does not participate in the optimization process to reduce model complexity.

After pre-processing, the GBRBM can extract features from spectral information in an unsupervised way. In this paper, the features are used as the input data for classification using LR. The entire structure is shown in Fig. 3.

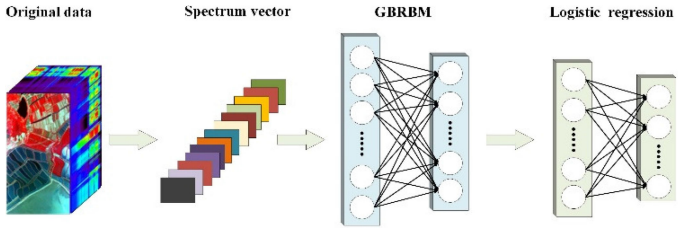

Fig. 3. Spectral classification using GBRBM-LR framework.

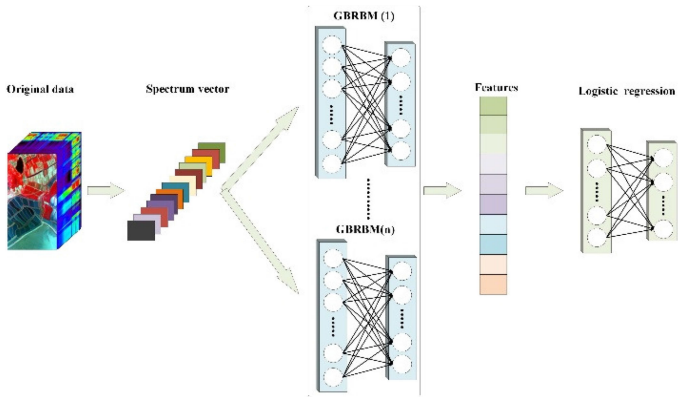

Fig. 4. Spectral classification using GBRBM in Parallel-LR framework.

C. Per-Pixel Classification Using Multiple GBRBM in Parallel

Recently, multiple feature learning has acquired better results in the field of computer vision. The features extracted by a single GBRBM model are always limited, and the combination of multiple features extracted by several GBRBM models is more powerful for classification. Thus, in this paper, a multiple feature learning based on GBRBM is proposed. The basic idea is to run several GBRBM models with varying hidden neurons in parallel, and then to combine the output from each model to the LR classifier. The whole structure is shown in Fig. 4. Each GBRBM model is trained and features are extracted separately in an unsupervised way. The features extracted from different models are combined together for the LR classification. Finally, the class labels are calculated in a supervised way by the LR. Another possibility of achieving multiscale feature learning is to use the same hidden neurons for GBRBM model, but increasing the depth of the GBRBM models. However, according to previous studies, in the application of hyperspectral classification, it has been noticed that classification accuracy may be degraded with the increase of hidden layers. Thus, we choose to increase the number of hidden neurons in this paper.

III. Experiments With Standard Data

In this section, we evaluate our method using standard hyperspectral data. Experiments were implemented based on the deep learning framework Theano and executed on a PC with an Intel single Core i5 CPU, NVIDIA GTX-1070 GPU

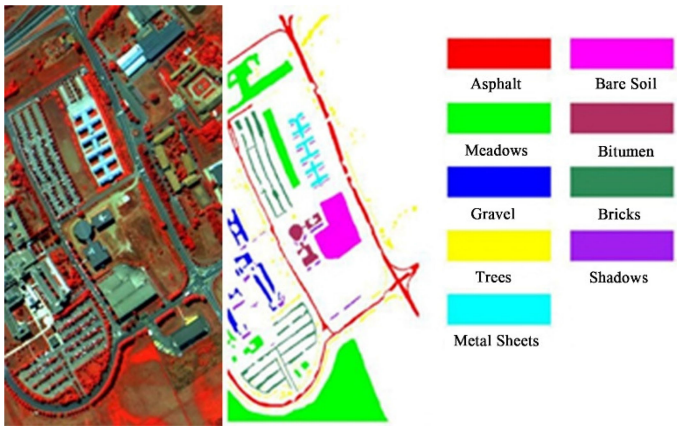

Fig. 5. Pavia University dataset.

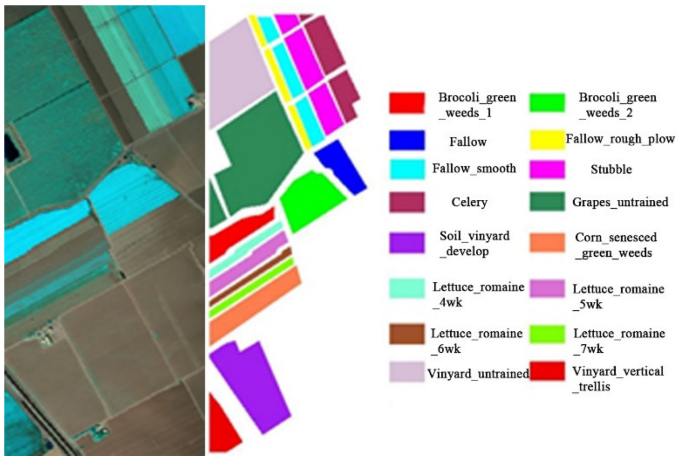

Fig. 6. Salinas dataset.

(8 GB video memory), and 16 GB RAM. The PC operating system is Ubuntu 16.04.

A. Dataset Description

In this paper, two standard hyperspectral datasets (i.e., Pavia University and Salinas dataset) are used to verify the effectiveness of the proposed model. The Pavia University dataset, as shown in Fig. 5, acquired by the ROSIS sensor during a flight campaign over Pavia. This dataset has 610 × 340 pixels and 103 spectral bands. There are nine classes in this dataset. The Salinas dataset with 16 classes, as shown in Fig. 6, was collected by the 224-band AVIRIS sensor over Salinas Valley, California, and is characterized by high spatial resolution.

To evaluate the classification accuracy, 10% of labeled samples were selected as training samples and rest for testing. The numbers of training and testing samples for each class are given in Tables I and II. As SVM and DBN have been widely used in the application of remote sensing classification, they were introduced for comparison in this study. Otherwise, multi-grained cascade-forest (Gcforest) was a new DNN that has produced excellent performance in some datasets, so it was also introduced for comparison. In the experiment, each classifier was run 20 times with different training samples to avoid the fortuity, and then overall accuracy (OA), average accuracy (AA), kappa statistic, training, and prediction time were reported.

TABLE I
LAND-COVER CLASSES AND NUMBER OF PIXELS IN THE PAVIA UNIVERSITY DATASET

Class code	Name	No. of samples	No. of training samples
1	Asphalt	6631	663
2	Meadows	18649	1864
3	Gravel	2099	209
4	Trees	3064	306
5	Metal Sheets	1345	134
6	Bare Soil	5029	502
7	Bitumen	1330	133
8	Bricks	3682	368
9	Shadows	947	94
Total		42776	4273

TABLE II
LAND-COVER CLASSES AND NUMBER OF PIXELS IN THE SALINAS DATASET

Class code	Name	No. of samples	No. of training samples
1	Brocoli_green_weeds_1	2009	200
2	Brocoli_green_weeds_2	3726	372
3	Fallow	1976	197
4	Fallow_rough_plow	1394	139
5	Fallow_smooth	2678	267
6	Stubble	3959	395
7	Celery	3579	357
8	Grapes_untrained	11271	1127
9	Soil_vinyard_develop	6203	620
10	Corn_senesced_green_weeds	3278	327
11	Lettuce_romaine_4wk	1068	106
12	Lettuce_romaine_5wk	1927	192
13	Lettuce_romaine_6wk	916	91
14	Lettuce_romaine_7wk	1070	107
15	Vinyard_untrained	7268	726
16	Vinyard_vertical_trellis	1807	180
Total		54129	5403

TABLE III
CLASSIFICATION WITH DIFFERENT NUMBER OF HIDDEN NEURONS FROM PAVIA UNIVERSITY DATASET

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
H=25	93.72	93.75	0.9164	63.03	0.06
H=50	93.85	93.93	0.9191	63.06	0.07
H=100	94.07	94.12	0.9204	66.00	0.08
H=200	94.21	94.35	0.9279	67.40	0.09
H=400	93.48	93.57	0.9131	69.20	0.12
H=600	93.50	94.10	0.9133	71.47	0.15

B. Spectral Classification Using GBRBM-LR Framework

In the GBRBM model, the number of hidden neurons is an important parameter, which affects the complexity of the model. Meanwhile, it also indicates the number of features to be extracted. In this section, the model with different numbers of hidden neurons are used. The results are shown in Tables III and IV. It can be seen that with the increase in hidden

TABLE IV
CLASSIFICATION WITH DIFFERENT NUMBER OF HIDDEN NEURONS FROM SALINAS DATASET

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
H=25	93.50	93.54	0.9279	82.17	0.08
H=50	93.68	93.83	0.9297	82.24	0.10
H=100	93.81	93.87	0.9310	93.66	0.14
H=200	93.29	93.42	0.9253	93.84	0.15
H=400	93.83	93.88	0.9228	92.80	0.18
H=600	93.79	93.81	0.9317	97.49	0.23

TABLE V
COMPARISON RESULTS OF CLASSIFICATION FROM PAVIA UNIVERSITY DATASET (10%)

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
GBRBM in Parallel	96.22	96.58	0.9491	238.93	0.29
SVM (Linear)	92.17	92.93	0.9128	0.41	5.34
SVM (RBF)	94.21	94.26	0.9420	0.53	10.33
DBN	92.39	91.38	0.8998	752.79	0.1292
Gcforest	90.67	90.76	0.8744	52.23	25.22

TABLE VI
COMPARISON RESULTS OF CLASSIFICATION FROM SALINAS UNIVERSITY DATASET (10%)

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
GBRBM in Parallel	95.94	95.98	0.9543	355.69	0.38
SVM (Linear)	91.65	92.93	0.9180	0.67	15.48
SVM (RBF)	94.27	94.24	0.9361	1.14	32.71
DBN	91.74	95.30	0.9083	1400.33	0.12
Gcforest	92.41	95.70	0.9154	60.21	32.15

neurons, classification accuracy is improved, but when the hidden neurons are increased to a certain level, classification accuracy becomes almost stable. According to these experiments, 100 hidden neurons can obtain satisfactory results.

Meanwhile, in order to verify the advantage of the GBRBM model, linear SVM, SVM with RBF kernel, DBN, and Gcforest are used for comparison. The number of hidden neurons is chosen to be 100 with the consideration of accuracy and efficiency. The accuracies over different classification methods are reported in Tables V and VI. Figs. 7 and 8 are classification maps of the model with different numbers of hidden neurons for the two datasets. The GBRBM model can extract the features of hyperspectral data effectively. However, the capability of a single GBRBM is still limited.

In this experiment, we use CUDA to make the GBRBM's training procedure faster. Due to complexity, the proposed method takes longer training time compared to other algorithms, as shown in Tables V and VI. Although the parallel GBRBM-LR framework has a longer training time, its shorter prediction time makes it suitable for practical applications.

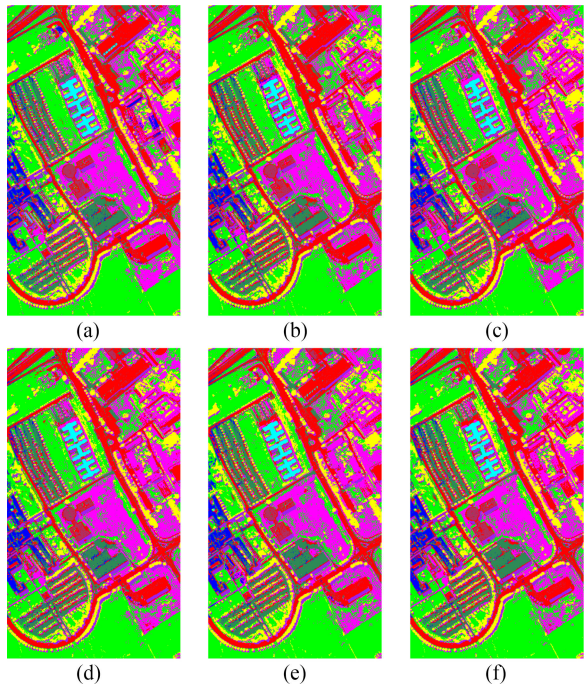

Fig. 7. Best classification results of the Pavia University dataset. (a) $H = 25$. (b) $H = 50$. (c) $H = 100$. (d) $H = 200$. (e) $H = 400$. (f) $H = 600$.

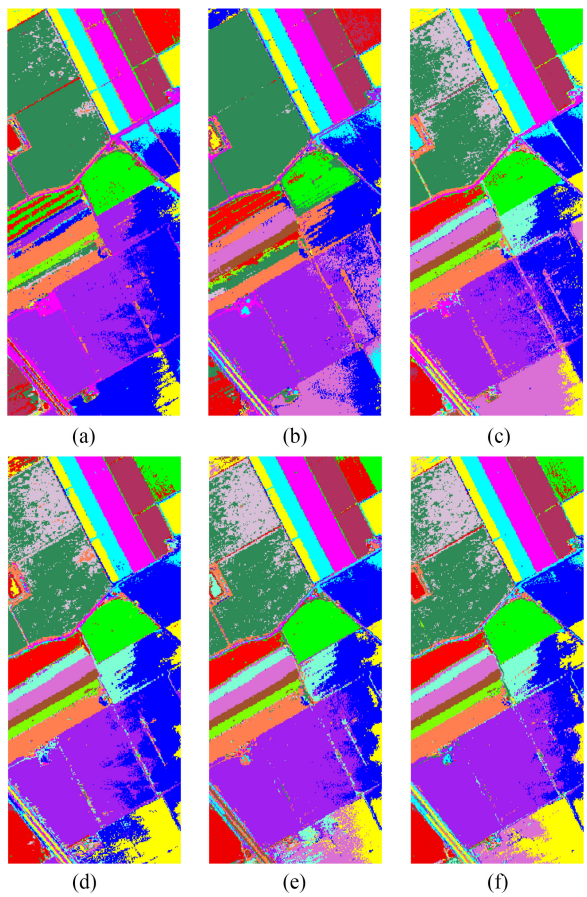

Fig. 8. Best classification results of the Salinas dataset. (a) $H = 25$. (b) $H = 50$. (c) $H = 100$. (d) $H = 200$. (e) $H = 400$. (f) $H = 600$.

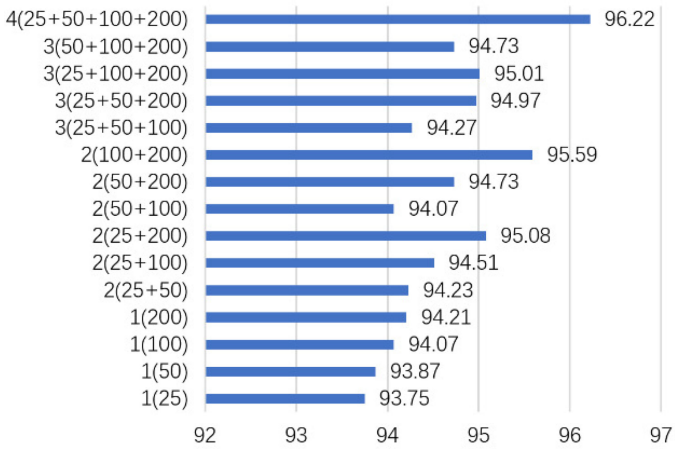

Fig. 9. Influence of the number of hidden neurons on OA in Pavia University dataset.

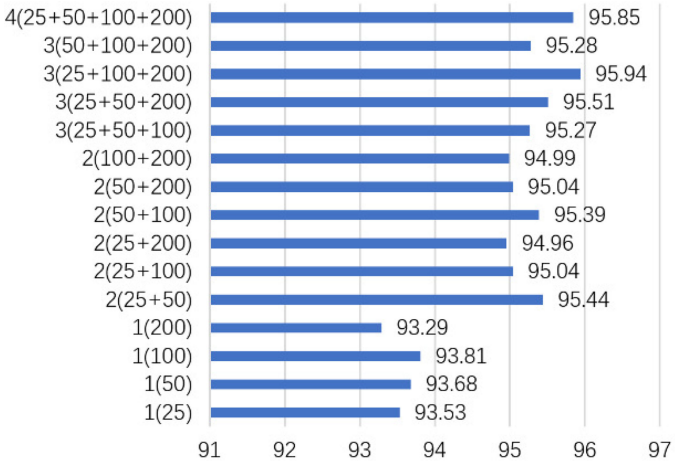

Fig. 10. Influence of the number of hidden neurons on OA in Salinas dataset.

C. Spectral Classification Using GBRBM in Parallel-LR Framework

Based on the experimental results in the previous section, after the number of hidden neurons reaches 200, classification accuracy will not further advance but the calculation time is significantly increased. Thus, considering the calculation time and accuracy, the hidden neurons are selected as 25, 50, 100, or 200 in the parallel-LR framework. Figs. 9 and 10 show the classification accuracy for all possible combinations of using 1, 2, 3, or all 4 of the GBRBMs in parallel. It can be seen from Fig. 9 that the classification accuracy of Pavia University dataset using $25 + 50 + 100 + 200$ in parallel is 96.22%, which is higher than the accuracy using a single layer with 100 hidden neurons. Fig. 10 shows that the classification accuracy of Salinas dataset using $25 + 50$ in parallel is 95.44%, which is higher than the accuracy using a single layer with 100 hidden neurons. Therefore, it can be concluded that classification accuracy using GBRBM in parallel is higher than a single layer with more hidden neurons.

Meanwhile, for the two datasets, the network with the highest accuracy is also different. The Salinas dataset can achieve the highest accuracy (95.94%) by using three layers ($25 + 100 + 200$) in parallel. The Pavia University dataset achieves the highest accuracy (96.22%) by using four layers ($25 + 50 + 100 + 200$) in parallel. The accuracy using GBRBM in parallel is better than the accuracy from other models. The comparative result is shown in Tables V and VI. Figs. 11 and 12 illustrate the classification maps. The experiment shows that the GBRBM in parallel-LR framework can classify effectively hyperspectral data, and the classification accuracy will be improved with the increase of hidden neurons. However, when the hidden neurons are increased to a certain level, classification accuracy becomes stable.

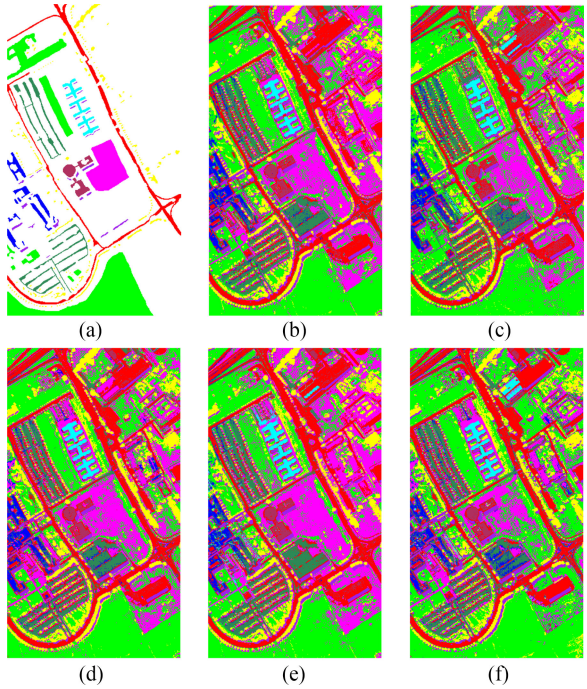

Fig. 11. Best classification results of the Pavia University dataset (10%). (a) Labeled map. (b) GBRBM in parallel ($25 + 50 + 100 + 200$). (c) SVM (linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

Moreover, to verify the stability of the model using less training samples, 5% of labeled samples were selected as training and rest for testing. The comparative result is shown in Tables VII and VIII. Figs. 13 and 14 illustrate the classification maps. According to the results, reducing the number of training samples will degrade the accuracy of the GBRBM in the parallel model, DBN, and Gcforest. And reducing the number of training samples does not affect the SVM model. The reason for this phenomenon may be that the models have a large number of parameters to be trained, but SVM does not have such a problem.

IV. Application to Mining Area

While promoting the development of national economy, the coal industry has also brought about a series of environmental problems, such as soil quality decline, land subsidence, ecosystem degradation, etc. With the advance of spatial information

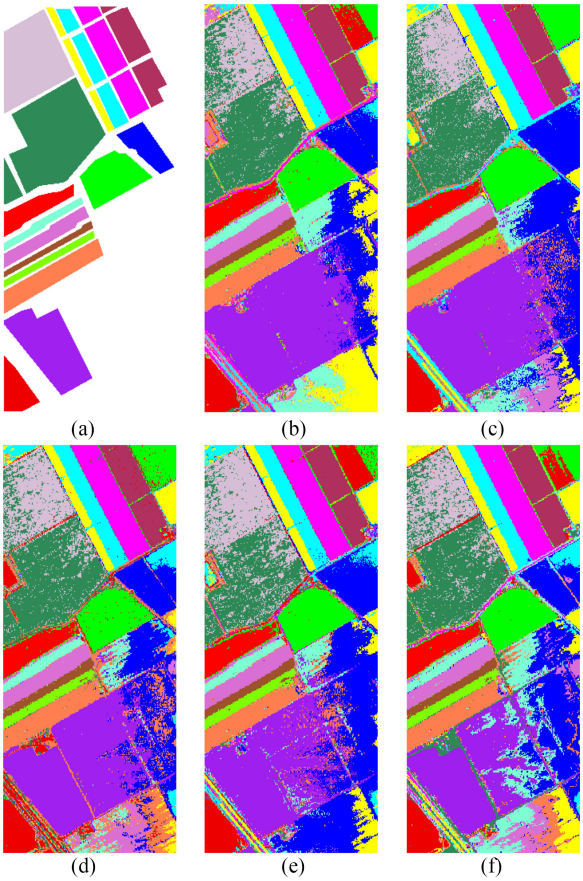

Fig. 12. Best classification results of the Salinas dataset (10%). (a) Labeled map. (b) GBRBM in parallel (25 + 100 + 200). (c) SVM (linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

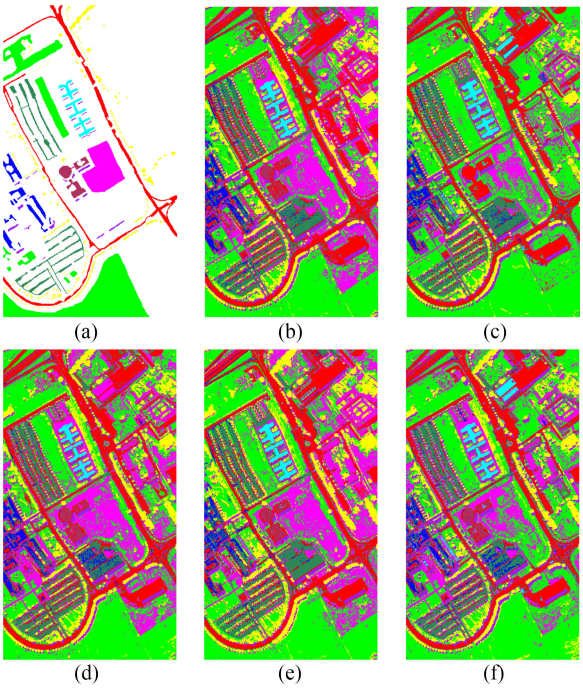

Fig. 13. Best classification results of the Pavia University dataset (5%). (a) Labeled map. (b) GBRBM in parallel (25 + 50 + 100 + 200). (c) SVM (linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

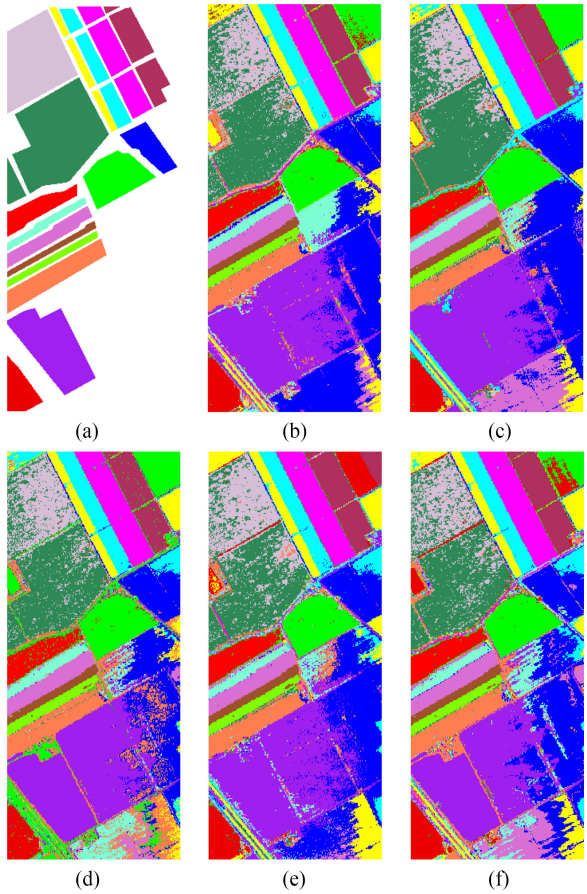

Fig. 14. Best classification results of the Salinas dataset (5%). (a) Labeled map. (b) GBRBM in parallel (25 + 100 + 200). (c) SVM (Linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

TABLE VII
COMPARISON RESULTS OF CLASSIFICATION FROM PAVIA UNIVERSITY DATASET (5%)

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
GBRBM in Parallel	92.20	90.93	0.8960	892.88	0.30
SVM (Linear)	86.82	79.53	0.8205	0.11	3.25
SVM (RBF)	93.13	92.67	0.9085	0.19	4.33
DBN	90.92	89.44	0.8741	407.15	0.10
Gcforest	89.12	88.96	0.8536	52.54	19.85

technology, remote sensing has been widely used in environmental monitoring and land-use/land-cover mapping.

Suggestions on environmental governance can be provided by the remote sensing technology.

The experiment was conducted in Xuzhou of Jiangsu Province. Xuzhou is an important coal producing area in China, which has confirmed reserves of more than 3.9 billion tons with an annual output of about 25 million tons. However, the coal mining area may lead to surface subsidence and soil quality declining, threatening the safety of residence and crop cultivation. Meanwhile, it can induce secondary geological disasters. Since the tailing reservoirs are stacking with tailings of high acid and

TABLE VIII
COMPARISON RESULTS OF CLASSIFICATION FROM SALINAS DATASET (5%)

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
GBRBM in Parallel	91.21	94.82	0.9020	962.32	0.43
SVM (Linear)	91.10	95.68	0.9004	0.20	9.29
SVM (RBF)	91.82	95.32	0.9088	0.61	13.32
DBN	90.72	94.21	0.8968	510.70	0.13
Gcforest	91.00	94.55	0.8997	58.25	21.37

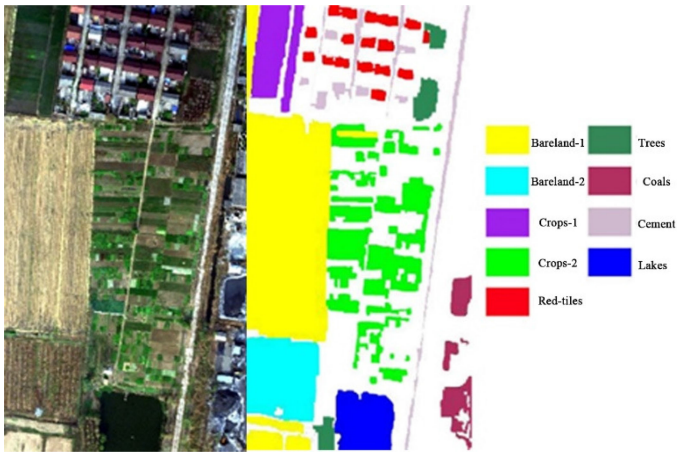

Fig. 15. Original image and ground truth of test area in the HySpex experiment.

alkali concentration, once it is destroyed by an extreme natural environment, it will cause flood plowing, damage roads, and other serious environmental pollution. The experimental area is located near the coal mining area, which has a security threat.

The experimental data were obtained by HySpex SWIR-384 and HySpex VNIR-1600 imaging spectrometer in Xuzhou on November 2014. The image obtained by HySpex VNIR-1600 with 0.19-m spatial resolution has 160 bands, which cover the spectral range of 415–992 nm, and the image obtained by HySpex SWIR-384 with 0.73-m spatial resolution has 288 bands, which cover the spectral range of 952–2508 nm. First, two images were preprocessed (including atmospheric correction and geometric correction). Second, the image obtained by HySpex VNIR-1600 was resampled to 0.73 m, and the two images were merged into one image after wiping OFF spectral overlap region and vapor absorption region. The new image has 436 bands and covers the spectral range of 415–2508 nm. Finally, an image with a size of 500 $\times$ 260 pixels was chosen as the experimental area. Based on field investigation, there are nine types of ground objects determined. The original image and ground truth image are shown in Fig. 15, and the number of labeled samples for each class is shown in Table IX. To evaluate the classification accuracy and verify the stability of the model, 10% and 5% of the labeled sample were selected as training samples separately, and rest for testing. According to the classification result by the parallel GBRBM, the best classification accuracy of the framework is using four layers (50 + 100 + 200 + 400). The accuracy

TABLE IX
NUMBER OF LABELED PIXELS IN EACH CLASS IN THE HYSPEX EXPERIMENT

Class code	Name	No. of samples
1	Bareland-1	26396
2	Lakes	4027
3	Coals	2783
4	Cement	5214
5	Crops-1	13184
6	Trees	2436
7	Bareland-2	6990
8	Crops-2	4777
9	Red-tiles	3070
Total		68877

TABLE X
COMPARISON RESULTS OF CLASSIFICATION IN THE HYSPEX EXPERIMENT (10%)

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
GBRBM in Parallel	97.84	97.92	0.9726	218.52	0.73
SVM (Linear)	96.73	96.84	0.9586	2.12	30.27
SVM (RBF)	97.08	97.12	0.9629	4.14	62.76
DBN	95.87	93.55	0.9473	807.57	0.16
Gcforest	94.38	93.67	0.9285	65.58	39.25

TABLE XI
COMPARISON RESULTS OF CLASSIFICATION IN THE HYSPEX EXPERIMENT (5%)

Model	OA(%)	AA(%)	Kappa statistic	Training time(s)	Prediction time(s)
GBRBM in Parallel	95.70	94.37	0.9455	693.33	0.57
SVM (Linear)	95.53	94.33	0.9434	0.64	19.25
SVM (RBF)	95.11	94.05	0.9377	1.04	30.59
DBN	95.53	94.39	0.9432	408.24	0.16
Gcforest	92.82	91.70	0.9085	63.48	35.47

comparison is shown in Tables X and XI, and classification maps are illustrated in Figs. 16 and 17. The results are similar to Section III-C; the GBRBM in parallel achieved the best accuracy and reducing the number of training samples also degraded the accuracy.

The classification maps show the land use in the experimental area, which can obtain the land use change by comparing the classification maps in different periods. From the current classification map, there are still large tracts of cultivated land around the tailing reservoirs, and the cultivated area has not been reduced by the existence of tailing. However, the potential risks still exist. The accumulation of tailings not only takes up a large of land resources but causes ecological pollution. It is mainly reflected in two aspects: one is dust pollution of surrounding land in the coal transportation, and the other is that heavy metals in tailing would pollute the cultivated land with groundwater. Therefore, it is helpful to find out the distribution of the tailing reservoirs for later recovery.

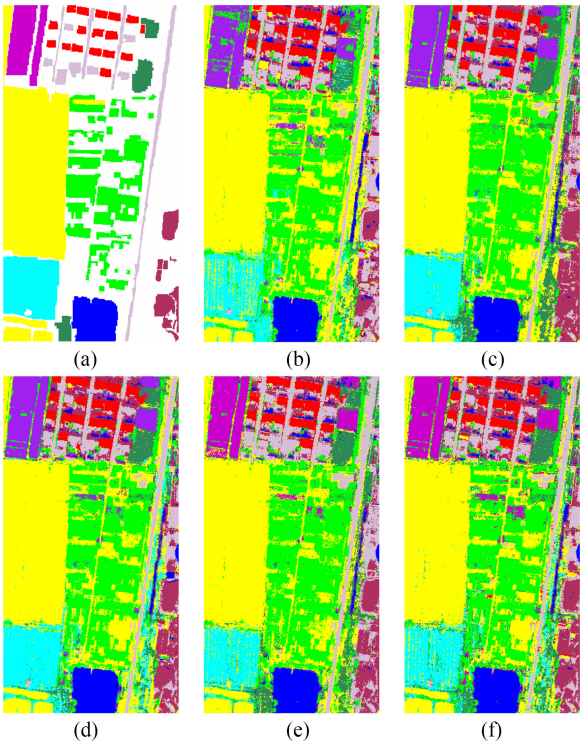

Fig. 16. Best classification results of the HySpex experiment (10%). (a) Labeled map. (b) GBRBM in parallel. (c) SVM (linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

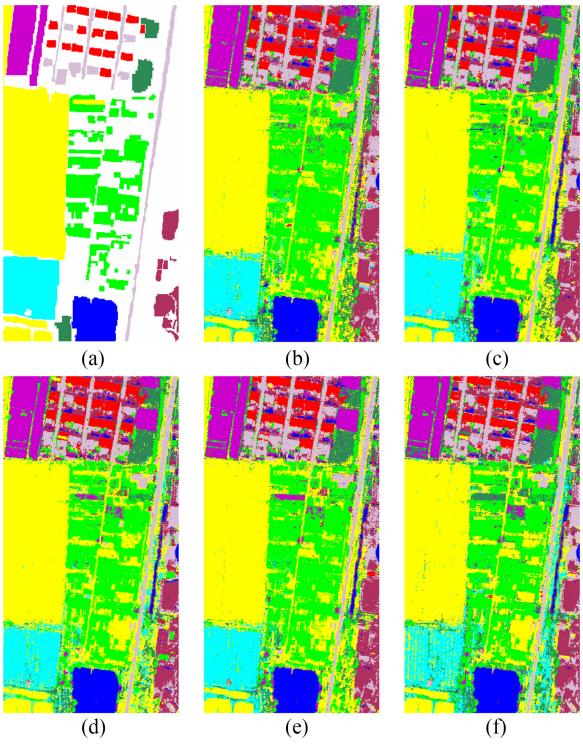

Fig. 17. Best classification results of the HySpex experiment (5%). (a) Labeled map. (b) GBRBM in parallel. (c) SVM (linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

V. Conclusion

In this paper, a novel classifier based on parallel GBRBM was proposed, which was tested on three hyperspectral datasets. The experiments demonstrate that the parallel GBRBM is an effective feature extraction method. Compared to traditional classifiers and several deep learning models, the accuracy by parallel GBRBM-LR can improve about 2% to 5%. Although the parallel GBRBM-LR framework has a longer training time, its shorter prediction time makes it more suitable for a wide range of hyperspectral data. There are two significant parameters that impact the model classification accuracy: the number of layers in parallel and the number of hidden neurons, which can only be determined by experiments. In the future study, to make full use of spatial and spectral information, feature fusion will be introduced into this framework.

Acknowledgment

The authors would like to thank Prof. D. Landgrebe and P. Gamba for providing the AVIRIS and ROSIS datasets used in the experiments. The authors would also like to thank the Jiangsu Innovation Teams, CUMT.

References

[1] M. Li, M. M. Crawford, X. Yang, and Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2832–2844, May 2015.

[2] A. Bhattacharya and D. Dunson, "Nonparametric Bayes classification and hypothesis testing on manifolds," *J. Multivariate Anal.*, vol. 111, no. 5, pp. 1–19, 2012.

[3] K. Tan, J. Zhang, D. Qian, and X. Wang, "GPU parallel implementation of support vector machines for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4647–4656, Oct. 2015.

[4] C. I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. New York, NY, USA: Wiley, 2007.

[5] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[6] Q. Du, "Modified Fisher's linear discriminant analysis for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 503–507, Oct. 2007.

[7] J. Wen, W. Yan, and L. Wei, "Supervised linear manifold learning feature extraction for hyperspectral image classification," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 3710–3713.

[8] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.

[9] P. Howland, J. Wang, and H. Park, "Solving the small sample size problem in face recognition using generalized discriminant analysis," *Pattern Recognit.*, vol. 39, no. 2, pp. 277–287, 2006.

[10] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[11] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[12] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances Neural Inform. Process. Syst.*, vol. 14, no. 6, pp. 585–591, 2001.

[13] W. Wu, D. L. Massart, and S. de Jong, "The kernel PCA algorithms for wide data. Part I: Theory and algorithms," *Chemometrics Intell. Lab. Syst.*, vol. 36, no. 2, pp. 165–172, 1997.

[14] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3185–3195, Jun. 2018.

[15] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," in *Proc. Neural Netw. Signal Process. IX, Proc. IEEE Signal Process. Soc. Workshop*, 1999, pp. 41–48.
[16] W. Sun, G. Yang, D. Bo, L. Zhang, and L. Zhang, "A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4032–4046, Jul. 2017.
[17] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
[18] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
[19] G. Hinton, "Deep belief nets," in Encyclopedia of Machine Learning, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, pp. 267–269, 2010.
[20] R. Salakhutdinov and G. Hinton, "An efficient learning procedure for deep Boltzmann machines," *Neural Comput.*, vol. 24, no. 8, pp. 1967–2006, 2012.
[21] S. Lange and M. Riedmiller, "Deep auto-encoder neural networks in reinforcement learning," *Neural Netw. (IJCNN) Int. Joint Conf.*, pp. 1–8, Jul. 2010.
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, pp. 84–90, 2017.
[23] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
[24] Y. Chen, X. Zing, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
[25] Z. Lin, Y. Chen, Z. Xing, and G. Wang, "Spectral-spatial classification of hyperspectral image using autoencoders," in *Proc. 9th Int. Conf. Inf., Commun. Signal Process.*, pp. 1–5, Dec. 2013.
[26] X. Ma, H. Wang, and J. Wang, "Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 120, pp. 99–107, 2016.
[27] L. Wang, J. Zhang, P. Liu, K.-K. R. Choo, and F. Huang, "Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification," *Soft Comput.*, vol. 21, no. 1, pp. 213–221, 2017.
[28] C. Farabet, C. Couprie, N. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
[29] M. Längkvist, A. Kiselev, M. Alirezaie, and A. Loutfi, "Classification and segmentation of satellite orthoimagery using convolutional neural networks," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 329.
[30] X. Zhou, S. Li, T. Fang, Q. Kai, S. Hu, and S. Liu, "Deep learning with grouped features for spatial spectral classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 97–101, Jan. 2017.
[31] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, pp. 215–223.
[32] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," Neural networks: Tricks of the trade. Berlin, Heidelberg: Springer, vol. 9, no. 1, 2012, pp. 599–619.
[33] K. H. Cho, A. Ilin, and T. Raiko, "Improved learning of Gaussian-Bernoulli restricted Boltzmann machines," in *Proc. 21th Int. Conf. Artif. Neural Netw.*, pp. 10–17, 2011.
[34] N. Wang, J. Melchior, and L. Wiskott, "An analysis of Gaussian-binary restricted Boltzmann machines for natural images," in *Proc. ESANN*, pp. 287–292, 2012.

Kun Tan (SM'16) received the B.S. degree in information and computer science from Hunan Normal University, Hunan, China, in 2004, and the Ph.D. degree in photogrammetric and remote sensing from the China University of Mining and Technology, Jiangsu, China, in 2010. From September 2008 to September 2009, he was working toward the Joint Ph.D. degree in remote sensing at Columbia University, New York, NY, USA.

From 2010 to 2018, he was with the Department of Surveying, Mapping, and Geoinformation, China University of Mining and Technology, Xuzhou, China. He is currently a Professor with East China Normal University, Shanghai, China. His research interests include hyperspectral image classification and detection, spectral unmixing, quantitative inversion of land surface parameters, and urban remote sensing.

Fuyu Wu was born in Shandong, China, in 1993. He received the B.S degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2015. He is currently working toward the Ph.D degree in photogrammetry and remote sensing at the China University of Mining and Technology, Xuzhou, China.

His research interests include the hyperspectral imagery processing and ecological monitoring.

Qian Du (S'98–M'00–SM'05–F'18) received the Ph.D. degree in electrical engineering from the University of Maryland—Baltimore County, Baltimore, MD, USA, in 2000.

She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include the hyperspectral remote sensing image analysis, pattern recognition, and machine learning.

Dr. Du was the Co-Chair for the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013, and the Chair for Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She is currently the Chief Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She is a fellow of SPIE–International Society for Optics and Photonics.

Peijun Du (M'07–SM'12) received the Ph.D. degree in photogrammetric and remote sensing from China University of Mining and Technology, Jiangsu, China, in 2001. He is currently a Professor of photogrammetry and remote sensing with the Department of Geographic Information Sciences, Nanjing University, Nanjing, China, and the Deputy Director of the Key Laboratory for Satellite Surveying Technology and Applications with the National Administration of Surveying and Geoinformation, China. His research interests include remote sensing image processing and pattern recognition; remote sensing applications; hyperspectral remote sensing information processing; multi-source geospatial information fusion and spatial data handling; integration and applications of geospatial information technologies; and environmental information science.

Prof. Du is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Yu Chen received the master's degree in photogrammetry and remote sensing from the China University of Mining and Technology (CUMT), Xuzhou, China, in 2012, and the Ph.D. degree in earth and planetary science from the University of Toulouse, Toulouse, France, in 2017.

In 2013, she was an Assistant Researcher with the Géoscience Environnement Toulouse Laboratory, French National Center for Scientific Research. She is currently a Lecturer with CUMT. Her current research focuses on SAR interferometry, with particular emphasis on its application for geophysical studies.