Hyperspectral Image Classification Based on Nonlinear Spectral–Spatial Network

Bin Pan, Zhenwei Shi, Ning Zhang, and Shaobiao Xie

***Abstract*—Recently, for the task of hyperspectral image classification, deep-learning-based methods have revealed promising performance. However, the complex network structure and the time-consuming training process have restricted their applications. In this letter, we construct a much simpler network, i.e., the nonlinear spectral–spatial network (NSSNet), for hyperspectral image classification. NSSNet is developed from the basic structure of a principal component analysis network. Nonlinear information is included in NSSNet, to generate a more discriminative feature expression. Moreover, spectral and spatial features are combined to further improve the classification accuracy. Experimental results indicate that our method achieves better performance than state-of-the-art deep-learning-based methods.**

***Index Terms*—Deep learning, hyperspectral image classification, nonlinear spectral–spatial network (NSSNet).**

I. Introduction

HYPERSPECTRAL remote sensing is an advanced Earth observation tool, which is developed based on spectroscopy technique. Hyperspectral imaging sensors can provide data with spatial and spectral information simultaneously. Hundreds of contiguous and narrow spectral band values are recorded as a data cube, with the spectral resolution up to nanometer level. A common application of the hyperspectral data is materials classification for each pixel. In the last decade, plenty of hyperspectral image classification methods have been developed. Researchers usually design reliable classifiers [1]–[3] or further utilize spatial–spectral information [4]–[7].

Recently, deep-learning-based algorithms have been introduced to hyperspectral image classification and presented promising performance. The idea of deep learning is to extract higher level features which represent more abstract semantics of the original data. To put it simply, deep learning may be considered as a nonlinear feature expression process, whereas the network structure could be regarded as the mapping relationship. Autoencoder (AE) is a popular deep-architecture-based model. In [8], Chen *et al.* proposed a deep learning framework to deal with hyperspectral image classification for the first time, where AE is used to learn deep features of hyperspectral data in an unsupervised manner. Then, in [9], the framework was improved by deep belief network (DBN). Some variations of AE were studied in the literature, such as stacked denoising AE [10], [11] and convolutional AE [12]. Another kind of deep learning method, convolutional neural networks (CNNs) were also reported in recent research on hyperspectral image classification [13]–[15]. However, deep learning methods usually suffer from a complex network structure and a time-consuming training process. Many tricks have to be used in the process of designing the basic structure of the network, and the experimental results are difficult to reproduce.

In this letter, we propose a nonlinear spectral–spatial network (NSSNet) for hyperspectral image classification, which is developed from a principal component analysis (PCA) network (PCANet) [16]. PCANet is a much simpler network while it is already on par with state-of-the-art deep-learning-based methods for many image classification tasks [16]. Motivated by this, we attempt to propose a new network for hyperspectral image classification, while achieving better performance than deep learning methods. However, the spectral data cannot be directly utilized by PCANet. Moreover, spatial information, which provides additional discriminant information that may lead to more accurate classification results, is not considered by PCANet. In this letter, we propose a novel joint spectral–spatial network utilizing the spectral and spatial information simultaneously. Furthermore, considering that nonlinear information may generate a more discriminative feature expression for each pixel, we improve the structure of PCANet by adding some nonlinear elements. Our work can be regarded as a simplified deep learning method, but better performances are observed in the experiments.

The remainder of this letter is organized as follows. In Section II, we present detailed description of the proposed NSSNet-based hyperspectral image classification method. Section III reports the experimental comparison with state-of-the-art deep-learning-based methods. Conclusions are drawn in Section IV.

II. NSSNet-Based Hyperspectral Image Classification

The NSSNet-based hyperspectral image classification method includes two parallel modules: spectral and spatial features extraction. In each module, a different imaging strategy

Manuscript received June 30, 2016; revised August 26, 2016; accepted September 9, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61671037 and Grant 61273245; by the Beijing Natural Science Foundation under Grant 4152031; by the funding project of the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under Grant BUAA-VR-16ZZ-03; and by the Shanghai Association for Science and Technology under Grant SAST2016090. *(Corresponding author: Zhenwei Shi.)*

B. Pan and Z. Shi are with the Image Processing Center, State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics and the Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China (e-mail: panbin@buaa.edu.cn; shizhenwei@buaa.edu.cn).

N. Zhang is with Shanghai Aerospace Electronic Technology Institute, Shanghai 201109, China (e-mail: dzs_zhangning@163.com).

S. Xie is with Shanghai Academy of Spaceflight Technology, Shanghai 201109, China (e-mail: boyish9747@outlook.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/LGRS.2016.2608963

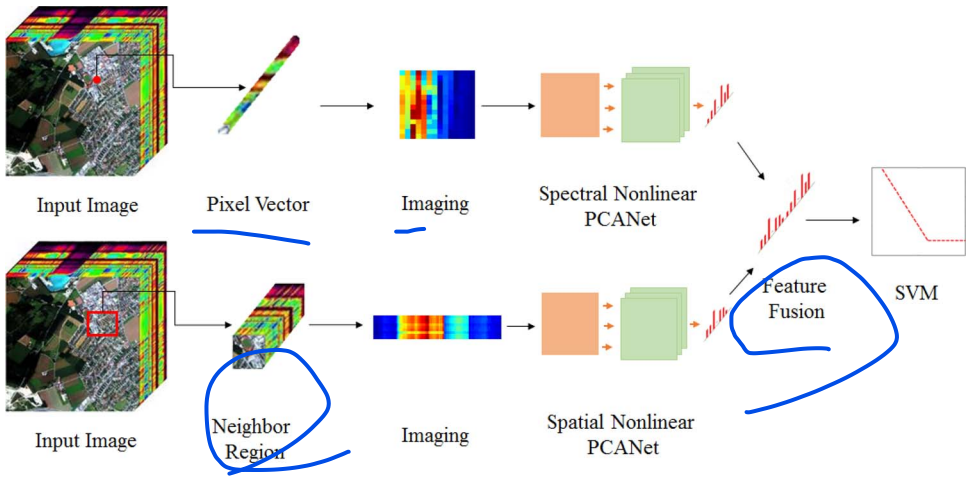

Fig. 1. Flowchart of the NSSNet method.

is adopted to generate 2-D samples, and an improved PCANet with nonlinear information is followed for feature extraction. Then, we fuse the spectral and spatial features, and support vector machine (SVM) is used to obtain the final results. The flowchart of our NSSNet-based method is shown in Fig. 1. Here, we first give a brief introduction to PCANet, and then we discuss how nonlinear information is used to improve the performance of our network. Finally, spectral information and spatial information are combined to construct the proposed NSSNet.

A. Brief Introduction to PCANet

PCANet is first proposed to handle the task of classification for a single image. The works about PCANet try to achieve two goals: The first is designing a simple deep learning network. The second is providing a baseline for different tasks where more advanced and sophisticated architectures can be utilized [16]. Compared with CNN, the most important change in PCANet is that the original data-adopting convolution filter banks between layers are replaced by basic PCA filters, and thus, the complex supervised optimizing process is avoided. Furthermore, binary quantization (hashing), histogram features, and linear SVM are also adopted to simplify traditional deep learning methods such as CNN. Generally, PCANet contains three processing components: PCA filters, binary hashing, and histogram features. More details are shown in [16].

PCA filters are used to construct a cascaded network. Suppose $\mathbf{I}_i$ is a given input image of size $m \times m$. PCANet first take a $k_1 \times k_2$ patch around each pixel and collect all the vectorized patches, the mean values of which have been removed, to form a matrix $\mathbf{X_i} \in \mathbb{R}^{(k_1 k_2)\times n}$, where n is the number of patches extracted from $\mathbf{I}_i$. Then, it constructs the same matrix for each input image and combines all the matrices, i.e.,

$$\mathbf{X} = [\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_N}] \in \mathbb{R}^{(k_1 k_2)\times(Nn)} \tag{1}$$

where N is the number of input images. PCA minimizes the reconstruction error within a family of orthonormal filters, i.e.,

$$\min_{V \in \mathbb{R}^{(k_1 k_2)\times L_1}} \|\mathbf{X} - \mathbf{V}\mathbf{V}^{\mathbf{T}}\mathbf{X}\|_F^2, \quad \text{s.t. } \mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{E}_{L_1} \tag{2}$$

where $\mathbf{E}$ is an identity matrix with size $L_1 \times L_1$, $\mathbf{V}$ are the principal eigenvectors of $\mathbf{X}\mathbf{X}^{\mathbf{T}}$, and L_1 denotes the number of principal eigenvectors. In PCANet, L_1 also corresponds to the number of filters in the first layer. Therefore, the filters can be expressed as

$$\mathbf{W}_l^1 \doteq \text{mat}_{(k_1 k_2)}\big(q_l(\mathbf{X}\mathbf{X}^{\mathbf{T}})\big) \in \mathbb{R}^{k_1 \times k_2}, \qquad l = 1, 2, \ldots, L_1 \tag{3}$$

where $\text{mat}_{(k_1 k_2)}(\cdot)$ is a function that reshapes a vector to a matrix, and $q_l(\mathbf{X}\mathbf{X}^{\mathbf{T}})$ denotes the lth principal eigenvector of $\mathbf{X}\mathbf{X}^{\mathbf{T}}$. Then, the first layer of PCANet can be obtained by

$$\mathbf{I}_i^l \doteq \mathbf{I}_i * \mathbf{W}_l^1, \qquad i = 1, 2, \ldots, N. \tag{4}$$

Conducting the same process as in the first layer for all the obtained $\mathbf{I}_i^l$, the second layer of PCANet is obtained. Assuming that the number of filters used in the second layer is L_2, PCANet would output $L_1 L_2$ images. One can build a multiple-layer PCA network, by repeating the aforementioned process, to extract higher level features. In [16], the authors suggested that two layers are enough for most tasks.

After establishing the structure of the network, a binary quantization process is followed. Subsequently, each of the L_1 images is separated into many local blocks. The histogram of each block (with 2^{L_2} bins) is computed, and all the histograms are concatenated into one vector. This vector is the feature expression for image $\mathbf{I}_i$. Finally, a linear SVM is used to determine the classification results.

B. Nonlinear Information

Original PCANet is completely a linear network, where the construction of convolution kernels is simple PCA filters. By contrast, in typical (convolutional) deep-learning-based methods such as CNN, the convolution kernels are obtained by a complex nonlinear optimization process. In NSSNet, we adopt two nonlinear strategies, namely, kernel method and nonlinear mapping, to improve the performance of PCANet. Note that, although nonlinear information is added, NSSNet is still much simpler than traditional deep-learning-based methods.

The kernel method is an effective approach for transforming input data to a higher dimensional space. If the selected kernel meets the Mercer theorem, then the obtained space is actually a reproducing kernel Hilbert space. In NSSNet, we utilize the kernel PCA (KPCA) to replace the original PCA filter in PCANet. KPCA is a nonlinear extension of the classical PCA. Compared with PCA, KPCA is able to capture the nonlinear information among input data, while requiring nothing about the data distribution in the original space. Based on KPCA, we modify the construction of filters by

$$\mathbf{KW}_l^j \doteq \text{mat}_{(k_1 k_2)}(\psi_l(\boldsymbol{K}(\mathbf{x}_a, \mathbf{x}_b))) \in \mathbb{R}^{k_1 \times k_2}, \quad l = 1, 2, \ldots, L_j \tag{5}$$

where $\boldsymbol{K}(\mathbf{x}_a, \mathbf{x}_b)$ is a kernel function, $\mathbf{x}_a, \mathbf{x}_b \in \mathbf{X}$ are column vectors, j is the jth layer, and $\psi_l(\cdot)$ denote the lth principal dimension-reduced data. Instead of using principal eigenvectors, in NSSNet, we construct the filter banks by the dimension-reduced data directly because KPCA cannot give the explicit eigenvectors for the input data. Equation (5) indicates that the inner product operation is simply replaced by a kernel function. Consequently, the computation cost does not increase a lot. There are many popular kernel functions that can be used, such as the linear kernel, the radial basis function (RBF) kernel, and the polynomial kernel. In NSSNet, the most popular kernel, i.e., RBF, is adopted, which is defined by

$$\boldsymbol{K}(\mathbf{x}_a, \mathbf{x}_b) = \exp\left(-\frac{\|\mathbf{x}_a - \mathbf{x}_b\|^2}{2\sigma^2}\right) \tag{6}$$

where σ is a parameter that controls the width of the radial basis function.

To further enhance the nonlinear structure, we introduce nonlinear mapping for NSSNet. In traditional deep-learning-based methods such as CNN, the hidden neurons usually include a *sigmoid* function for nonlinear mapping. This strategy not only can avoid the multilayer networks equaling to a single layer but also can extract more nonlinear information from the original data. However, in PCANet, there is no nonlinear mapping between layers. In NSSNet, we improve the original PCANet by adding a sigmoid function to the convolved results, which can be mathematically expressed by

$$\mathbf{I}_{nl}^i = \frac{1}{1 + \exp\left(-\mathbf{I}_l^i\right)}. \tag{7}$$

Note that the nonlinear mapping is conducted on $\mathbf{I}_l^i$.

Overall, both the kernel method and the sigmoid function are used in NSSNet, to extract nonlinear information in hyperspectral data.

C. Spectral–Spatial Classification Network

The inputs for PCANet are many single images, and the goal of PCANet is determining the label of each image. However, in hyperspectral image classification, we aim at giving a label to each pixel which is represented by an observed 1-D spectrum. Furthermore, it has been proved by many researchers that spatial information can significantly improve the classification accuracy, whereas the original PCANet does not consider the spatial relationship. In NSSNet, we propose the imaging strategy to handle the problem of hyperspectral image classification. Spectral and spatial features are extracted by different imaging methods, and a feature fusion process is followed to obtain the final feature expression for each pixel.

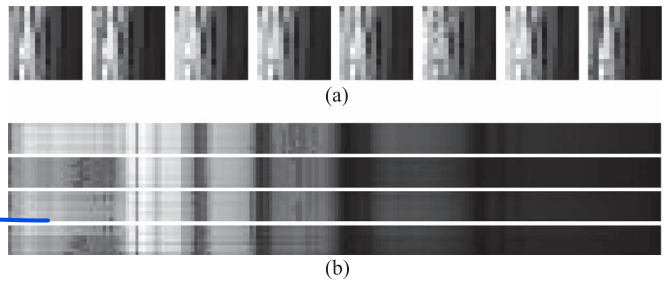

Fig. 2. Samples of imaged data. (a) Spectral images. (b) Spatial images.

1) Spectral Feature Extraction: Instead of using the spectra vectors directly, we first conduct data imaging. Data imaging refers to transforming a 1-D vector to a 2-D image. Let $\mathbf{y}_i \in \mathbb{R}^p$ denote an observed spectrum vector of a pixel, then the imaging form of $\mathbf{y}$ can be expressed by

$$\mathbf{y}_i \xrightarrow{\text{reshape}} \mathbf{Y}_i^{\text{Spec}} \in \mathbb{R}^{k \times k}. \tag{8}$$

Here, we set the width and height of the imaged data the same for simplifying. In the imaging process, the reflectance value in each band of a pixel vector is considered as the "texture" in an image. Compared with real-world images, the generated imaging data can be regarded as texture-stable images which are immune to illumination, scales, and rotation. This means that the problem of hyperspectral image classification is transformed to image classification. Subsequently, the improved nonlinear PCANet is used to extract the feature extraction for each pixel. Some samples of imaged spectra for the *Indian Pines* data set are shown in Fig. 2(a).

2) Spatial Feature Extraction: Researchers have verified that spatial information can significantly improve the accuracy of deep-learning-based hyperspectral image classification [8], [9]. In NSSNet, we construct an image for each pixel by utilizing the spectra of the pixel and its neighbors. Different from the imaging process for spectral information, for a pixel spectrum $\mathbf{y}_i$, here, we connect $\mathbf{y}_i$ and its neighbors side by side, which can be denoted by

$$[\mathbf{y}_{i-s}, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_{i+s}] \xrightarrow{\text{reshape}} \mathbf{Y}_i^{\text{Spat}} \in \mathbb{R}^{s^2 \times p} \tag{9}$$

where $s \times s$ is the window size for a pixel. Generally, 3×3 is appropriate to describe the spatial information. We adopt this imaging strategy mainly because the spatial correlation between a pixel and its neighbors is an important concern. Some samples of spatial imaged data for the Indian Pines data set are displayed in Fig. 2(b). Note that this is only an available imaging method, other similar approaches also work.

After extracting the spectral and spatial features, we directly connect them to compose the final feature expression for each pixel. Finally, an SVM classifier is used to determine the labels, which is also adopted in the original PCANet [16].

没考虑spatial feature

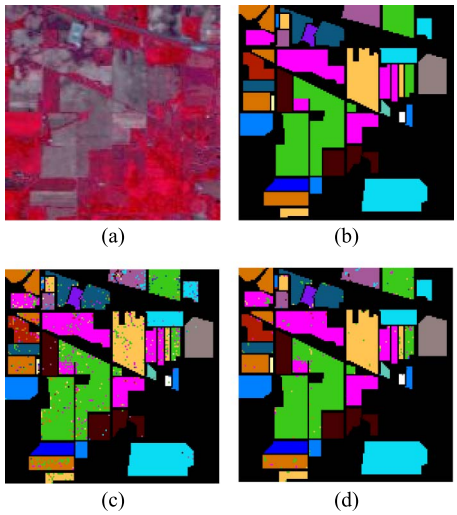

Fig. 3. Classification results for the Indian Pines data set. (a) False-color composition image. (b) Ground truth. (c) Results of PCANet. (d) Results of NSSNet.

III. Experimental Results

A. Data Sets and Experimental Setup

Two popular real-world hyperspectral images are used in the experiments, namely, the Indian Pines and Pavia University scenes.[1]

- The Indian Pines image was collected by the Airborne Visible/Infrared Imaging Spectrometer in Northwestern Indiana, with a size of 145 × 145 pixels. After removing the water absorption bands, 200 spectral bands remain in this image. There are a total of 10 249 labeled pixels in the ground truth, and they are classified into 16 classes [see Fig. 3(a) and (b)].
- The Pavia University image was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-3) sensor over the city of Pavia, Italy. The size of this image is 610 × 340, and 103 bands are used for classification after removing the noise bands. A total of 42 776 labeled pixels is available in the ground truth, and they are classified into nine different classes [see Fig. 4(a) and (b)].

The labeled samples are randomly divided into training set and test set with a ratio of 1:1, as adopted in [9], for a fair comparison. In addition, the default network parameters setting of PCANet is adopted by NSSNet: two layers, patch size of 7 × 7, eight filters in each layer, size of block histograms of 7 × 7, and overlap ratio of 0.8. As discussed in [16], different parameters only have little influence on the final results. Our unreported works also indicate the same phenomenon. Therefore, the same setup as in PCANet is used to construct the basic structure of NSSNet. Note that our spectral- and spatial-based networks share the same parameters.

[1]Available online at http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

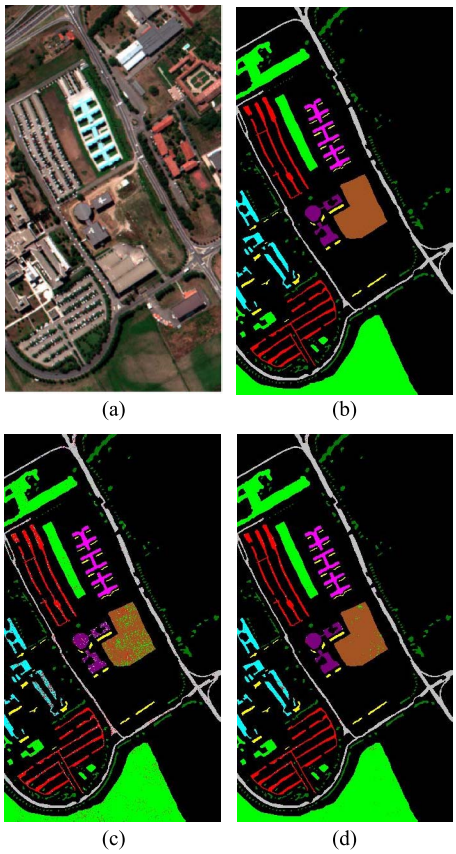

Fig. 4. Classification results for the Pavia University data set. (a) False-color composition image. (b) Ground truth. (c) Results of PCANet. (d) Results of NSSNet.

B. Compared Methods

Since we aim at designing a simple deep learning method for hyperspectral image classification, we compare our NSSNet with two state-of-the-art deep-learning-based methods, namely, Stacked Autoencoder-Logistic Regression (SAE-LR) [8] and Deep Belief Network-Logistic Regression (DBN-LR) [9]. Both the spectral information and the spatial information are used in SAE-LR and DBN-LR. Furthermore, to demonstrate the effectiveness of the nonlinear and spectral–spatial joint information used in NSSNet, we also compare our method with the original PCANet [16]. However, as discussed earlier, the original PCANet is used for single-image classification, whereas it cannot be directly used for hyperspectral image classification. Here, we adopt the same imaging process as NSSNet so that PCANet is able to conduct. In addition, a traditional method, i.e., the representative multiple-kernel learning (RMKL) [3], is also compared in the experiments.

C. Results and Discussion

The classification results for the two data sets by the original PCANet and the proposed NSSNet are depicted in Figs. 3(c) and (d) and 4(c) and (d). We can see that the performance of NSSNet is better than that of PCANet, which may indicate that the nonlinear and spectral–spatial joint information has improved the feature expression capability of the network. In Table I, the quantified evaluation results about accuracies and

TABLE I
CLASSIFICATION RESULTS AND TRAINING TIME FOR THE TWO DATA SETS

	Indian Pines				Pavia University			
	OA(%)	AA(%)	κ	Time(m)	OA(%)	AA(%)	κ	Time(m)
PCANet	86.58	85.16	0.8471	**12.4**	93.20	91.01	0.8997	**28.6**
RMKL	95.61	94.20	0.9499	-	96.06	94.48	0.9443	-
SAE-LR	92.58	90.38	0.9152	359.6	98.69	98.17	0.9829	1788.4
DBN-LR	95.95	95.45	0.9539	-	99.05	98.48	0.9875	-
NSSNet	**96.08**	**96.40**	**0.9547**	20.3	**99.50**	**99.03**	**0.9910**	66.1

running time are displayed. We note that all the compared methods perform well in hyperspectral image classification, whereas our method achieves the best results in both data sets. NSSNet takes more time than PCANet in the training process, but the overall accuracies (OA) of NSSNet surpass those of PCANet by about 9.5% in the Indian Pines image and 6.2% in the Pavia University image, which verifies the effectiveness of our improvements. Similar results are also observed in average accuracies (AA) and Kappa coefficients (κ). The accuracies of RMKL are lower than those of the proposed method, particularly in the Pavia University data set. The reason may be that the Pavia University data set can provide more training samples, which would improve the performance of the proposed method significantly. Compared with SAE-LR and DBN-LR, NSSNet outperforms them slightly. However, as discussed in [8] and [9], the training process of deep-learning-based methods is quite time consuming and usually requires special and expensive equipment (such as graphics processing units). For example, the training time of SAE-LR for the Pavia University image is up to 1788.4 min,[2] whereas the same process of NSSNet takes only 66.1 min. NSSNet performs similarly, and even better, when compared with some state-of-the-art deep-learning-based methods, whereas it has a simpler network structure and a faster processing speed.

Furthermore, we also perform a paired t-test between the NSSNet and DBN-LR methods to validate whether the observed improvements in OA are statistically significant (at a confidence level of 95%) [9]. We assume that the mean OA of NSSNet ($\overline{a_1}$) is larger than that of DBN-LR ($\overline{a_2}$). We accept the hypothesis only if

$$\frac{(\overline{a_1}-\overline{a_2})\sqrt{n_1+n_2-2}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)(n_1s_1^2+n_2s_2^2)}} > t_{1-\alpha}[n_1+n_2-2] \quad (10)$$

where s_1 and s_2 are the standard deviations of the two results, and n_1 and n_2 are the numbers of realizations of experiments reported, which are set as 20 in this letter. For the Indian Pines and Pavia University data sets, the standard deviations of our method are 0.1675% and 0.0853%, respectively. Therefore, the increases in OA are statistically significant (at the level of 95%) for the two test data sets.

Overall, we may infer from the experiments that NSSNet is a promising and efficient method for the task of hyperspectral image classification.

[2]TITAN X, i7-6700 K, 32-G random access memory, Theano.

IV. CONCLUSION

In this letter, we have introduced a novel hyperspectral image classification framework, i.e., NSSNet. NSSNet could be regarded as a simplified deep-learning-based method, where the original hyperspectral data were transformed to a new feature space by nonlinear mapping. Furthermore, in addition to spectral features, the spatial information was also involved in NSSNet, which has significantly improved the final classification results. Experiments indicated that NSSNet outperforms some state-of-the-art deep-learning-based methods. Future work will be devoted to reducing the number of training samples and further improving the accuracies.

ACKNOWLEDGMENT

The authors would like to thank Prof. Y. Chen for providing the codes for SAE-LR and Prof. Y. Gu for providing the codes for RMKL.

REFERENCES

[1] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE 27th AIPR Workshop, Adv. Comput.-Assisted Recognit.*, 1999, pp. 221–232.
[2] M. Chi, Q. Kun, J. A. Benediktsson, and R. Feng, "Ensemble classification algorithm for hyperspectral remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 762–766, Oct. 2009.
[3] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
[4] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[5] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
[6] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
[7] J. Li *et al.*, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.
[8] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[9] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
[10] X. Ma, J. Geng, and H. Wang, "Hyperspectral image classification via contextual deep learning," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, pp. 1–12, 2015.
[11] Y. Liu, G. Cao, Q. Sun, and M. Siegel, "Hyperspectral classification via deep networks and superpixel segmentation," *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3459–3482, 2015.
[12] W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo, "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery," *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3368–3379, 2015.
[13] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE IGARSS*, 2015, pp. 4959–4962.
[14] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sens.*, vol. 2015, 2015, Art. no. 258619.
[15] H. Liang and Q. Li, "Hyperspectral imagery classification using sparse representations of convolutional neural network features," *Remote Sens.*, vol. 8, no. 2, p. 99, 2016.
[16] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.