Hyperspectral Image Classification Using Spectral-Spatial LSTMs

Feng Zhou *, Renlong Hang, Qingshan Liu(✉), and Xiaotong Yuan

Jiangsu Key Laboratory of Big Data Analysis Technology,
School of Information and Control,
Nanjing University of Information Science and Technology, Nanjing 210044, China
qsliu@nuist.edu.cn

Abstract. In this paper, we propose a hyperspectral image (HSI) classification method using spectral-spatial long short term memory (LSTM) networks. Specifically, for each pixel, we feed its spectral values in different channels into Spectral LSTM one by one to learn the spectral feature. Meanwhile, we firstly use principle component analysis (PCA) to extract the first principle component from a HSI, and then select local image patches centered at each pixel from it. After that, we feed the row vectors of each image patch into Spatial LSTM one by one to learn the spatial feature for the center pixel. In the classification stage, the spectral and spatial features of each pixel are fed into softmax classifiers respectively to derive two different results, and a decision fusion strategy is further used to obtain a joint spectral-spatial results. Experiments are conducted on two widely used HSIs, and the results show that our method can achieve higher performance than other state-of-the-art methods.

Keywords: Deep learning · Long short term memory · Decision fusion · Hyperspectral image classification

1 Introduction

With the development of hyperspectral sensors, it is convenient to acquire images with high spectral and spatial resolutions simultaneously. Hyperspectral data is becoming a valuable tool to monitor the Earth's surface. The classification of hyperspectral image (HSI) has become one of the most important tasks for many applications including both commercial and military domains.

Many methods have been proposed to deal with HSI classification. Traditional methods, such as k-nearest-neighbors and logistic regression, often use the high-dimensional spectral information as features, thus suffering from the issue of "curse of dimensionality". To address this issue, dimensionality reduction methods are widely used. These methods include principal component analysis (PCA) [14, 21] and linear discriminant analysis (LDA) [8, 1]. In [20], a promising method called support vector machine (SVM) was successfully applied to HSI

* He is currently working toward the Master degree in the School of Information and Control, Nanjing University of Information Science and Technology.

classification. It exhibits low sensitivity to the data with high dimensionality and small sample size. In most cases, SVM-based classifiers can obtain superior performance as compared to other methods. However, SVM is still a shallow architecture. As discussed in [7], these shallow architectures have shown effectiveness in solving many simple or well-constrained problems, but their limited representational power is insufficient in complex cases.

In the past few years, with the advances of the computing power of computers and the availability of large-scale datasets, deep learning techniques [19] have gained great success in a variety of machine learning tasks. Among these techniques, CNN [6, 18] has been recognized as a state-of-the-art feature extraction method for various computer vision tasks [16, 22] owing to its local connections and weight sharing properties. Besides, recurrent neural network (RNN) and its variants have been widely used in sequential data modeling such as speech recognition [9, 10] and machine translation [5, 23].

Recently, deep learning has been introduced into the remote sensing community especially for HSI classification [26]. For example, in [3], a stacked autoencoder model was proposed to extract high-level features in an unsupervised manner. Inspired from it, Tao *et al.* proposed an improved autoencoder model by adding a regularization term into the energy function [24]. In [4], deep belief network (DBN) was applied to extract features and classification results were obtained by logistic regression classifier. For these models, inputs are high-dimensional vectors. Therefore, to learn the spatial feature from HSIs, an alternative method is flattening a local image patch into a vector and then feeding it into them. However, this method may destroy the two-dimensional structure of images, leading to the loss of spatial information. To address this issue, a two dimensional CNN model was proposed in [27]. Due to the use of the first principal component of HSIs as input, two dimensional CNN may lose the spectral information. To simultaneously learn the spectral and spatial features, three-dimensional CNN considers the local cube as inputs [2].

Since hyperspectral data are densely sampled from the entire spectrum, they are expected to have dependencies between different spectral bands. First, it is easy to observe that for any material, the adjacent spectral bands tend to have very similar values, which implies that adjacent spectral bands are highly dependent on each other. In addition, some materials also demonstrate long-term dependency between non-adjacent spectral bands [25]. In this paper, we regard each hyperspectral pixel as a data sequence and use long short term memory (LSTM) [13] to model the dependency in the spectral domain. Similar to spectral channels, pixels of the image also depend on each other in the spatial domain. Thus, we can also use LSTM to extract spatial features. The extracted spectral and spatial features for each pixel are then fed into softmax classifiers. The classification results can be combined to derive a joint spectral-spatial result.

The rest of this paper is structured as follows. In the following section, we will present the proposed method in detail. The experiments are reported in Section 3, followed by the conclusion in Section 4.

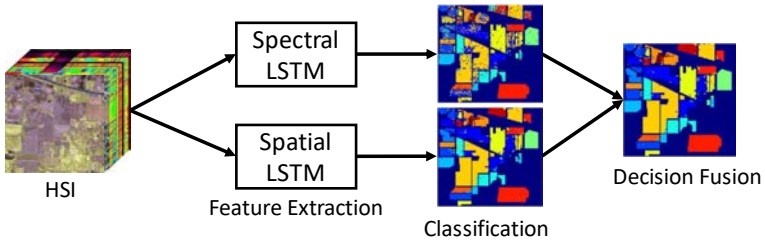

Fig. 1. Flowchart of the proposed SSLSTMs.

2 Methodology

The flowchart of the proposed spectral-spatial LSTMs (SSLSTMs) is shown in Fig. 1. From this figure, we can observe that SSLSTMs consist of two important components: Spectral LSTM (SeLSTM) and Spatial LSTM (SaLSTM). For each pixel in a given HSI, we feed its spectral values into the SeLSTM to learn the spectral feature and then derive a classification result. Similarly, for the local patch of each pixel, we feed it into a SaLSTM to extract the spatial feature and then obtain a classification result. To fuse the spectral-spatial results, we finally combine these two classification results in a weighted sum manner. In the following subsections, we will introduce these processes in detail.

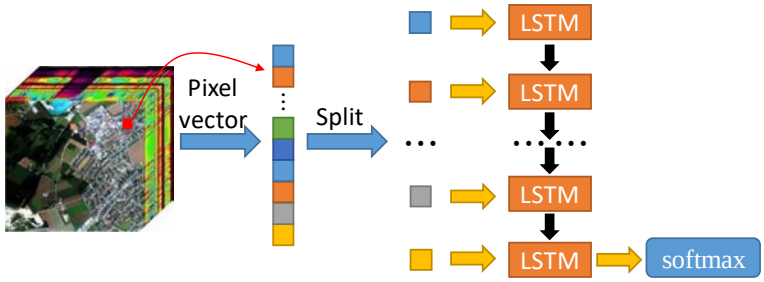

Fig. 2. Flowchart of SeLSTM.

2.1 Spectral LSTM

Hundreds of spectral bands in HSIs provide different spectral characteristics of the object in the same location. Due to the complex situation of lighting, rotations of the sensor, different atmospheric scattering conditions and so on, spectra have complex variations. Therefore, we need to extract robust and invariant features for classification. It is believed that deep architectures can potentially lead to progressively more abstract features at higher layers, and more abstract features are generally invariant to most local changes of the input. In this paper, we

consider the spectral values in different channels as an input sequence and use LSTM discussed above to extract spectral features for HSI classification. Fig. 2 shows the flowchart of the proposed classification scheme with spectral features. First, we choose the pixel vector $x_i \in \mathbf{R}^{1\times K}$ where K indicates the number of spectral bands from a given HSI. Second, we transform the vector to a K-length sequence $\{x_i^1, \cdots, x_i^k, \cdots, x_i^K\}$ where $x_i^k \in \mathbf{R}^{1\times 1}$ indicates the pixel value of k-th spectral band. Then, the sequence is fed into LSTM one by one and the last output is fed to softmax classifier. We set the loss function to cross entropy and optimize it by Adam algorithm [15]. Finally, we can obtain the probability value $P_{spe}(y = j|x_i), j \in \{1, 2, \cdots, C\}$ where C indicates the number of classes.

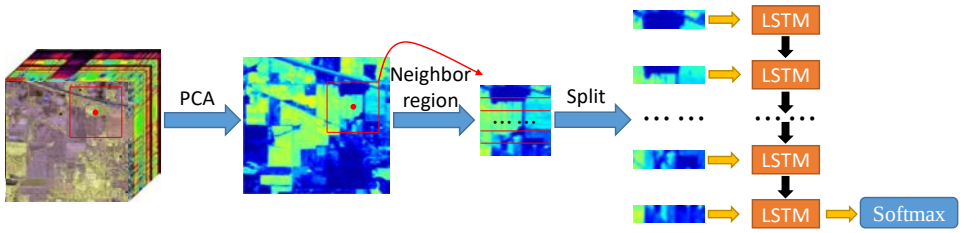

Fig. 3. Flowchart of SaLSTM.

2.2 Spatial LSTM

To extract the spatial feature of a specific pixel, we take a neighborhood region of it into consideration. Due to the hundreds of channels along the spectral dimension, it always has tens of thousands of dimensions. A large neighborhood region will result in too large input dimension for the classifier, containing too large amount of redundancy [3]. Motivated by the works in [3] and [27], we firstly use PCA to extract the first principle component. Second, for a given pixel x_i, we choose a neighborhood $\mathbf{X}_i \in \mathbf{R}^{S\times S}$ centered at it. After that, we transform the rows in this neighborhood to a S-length sequence $\{X_i^1, \cdots, X_i^l, \cdots, X_i^S\}$ where X_i^l indicates the l-th row of $\mathbf{X}_i$. Finally, we feed the sequence into LSTM to extract the spatial feature of x_i. Similar to spectral features-based classification, we use the last output of LSTM as an input to the softmax layer and achieve the probability value $P_{spa}(y = j|x_i), j \in \{1, 2, \cdots, C\}$. The configurations of loss function and optimization algorithm in SaLSTM is the same as those of SeLSTM. The overall flowchart of the proposed spatial features-based classification method is demonstrated in Fig. 3.

2.3 Joint spectral-spatial classification

The above two subsections introduce the classification methods based on spectral and spatial features respectively. With the development of imaging spectroscopy

technologies, current sensors can acquire HSIs with very high spatial resolutions. Therefore, the pixels in a small spatial neighborhood belong to the same class with a high probability. For a large homogeneous region, the pixels may have different spectral responses. If we only use the spectral features, the pixels will be classified into different subregions. On the contrary, for multiple neighboring regions, if we only use the spatial information, these regions will be classified as the same one. Thus, for accurate classifications, it is essential to take into account the spatial and spectral information simultaneously [12]. Based on the posterior probabilities $P_{spe}(y = j|x_i)$ and $P_{spa}(y = j|x_i)$, an intuitive method to combine the spectral and spatial feature is to fuse these two results in a weighted sum manner, which can be formulated as $P(y = j|x_i) = w_{spe}P_{spe}(y = j|x_i) + w_{spa}P_{spa}(y = j|x_i)$, where w_{spe} and w_{spa} are fusion weights that satisfy $w_{spe} + w_{spa} = 1$. For simplicity, we use uniform weights in our implementation, i.e., $w_{spe} = w_{spa} = \frac{1}{2}$.

3 Experimental Results

3.1 Datasets

We test the proposed method on two famous HSI datasets, which are widely used to evaluate classification algorithms.

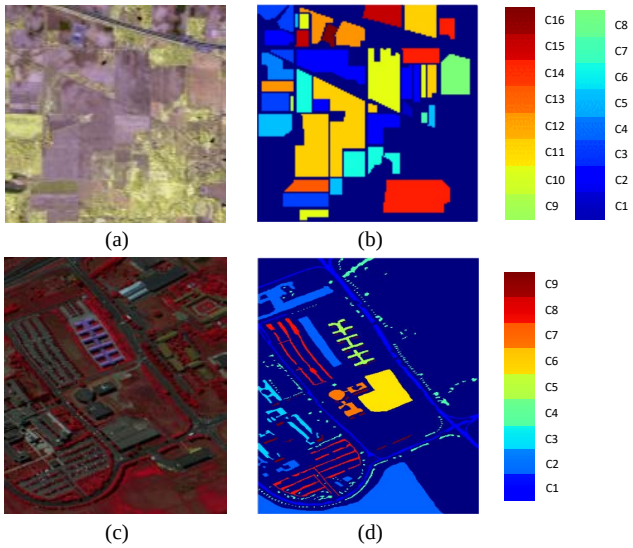

Fig. 4. False-color composite images and ground-truth maps of (a)-(b) IP, (c)-(d) PUS.

Indian Pines Scene (IP): The first dataset was acquired by the AVIRIS sensor over the Indian Pine test site in northwestern Indiana, USA, on June 12, 1992 and it contains 224 spectral bands. We utilize 200 bands after removing four bands containing zero values and 20 noisy bands affected by water absorption. The spatial size of the image is 145×145 pixels, and the spatial resolution is 20 m. The false-colour composite image and the ground-truth map are shown in Fig. 4(a)-(b). The available number of samples is 10249 ranging from 20 to 2455 in each class, which is reported in Table 1.

Pavia University Scene (PUS): The second dataset was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy, on July 8, 2002. The original image was recorded with 115 spectral channels ranging from 0.43 μm to 0.86 μm. After removing noisy bands, 103 bands are used. The image size is 610×340 pixels with a spatial resolution of 1.3 m. A three band false-colour composite image and the ground-truth map are shown in Fig. 4(c)-(d). In the ground-truth map, there are nine classes of land covers with more than 1000 labeled pixels for each class shown in Table 1.

Table 1. Number of pixels for training/testing and the total number of pixels for each class in IP and PUS ground truth maps.

	IP			PUS		
Class	Total	Training	Testing	Total	Training	Testing
1	46	5	41	6631	548	6083
2	1428	143	1285	18649	540	18109
3	830	83	747	2099	392	1707
4	237	24	213	3064	524	2540
5	483	48	435	1345	265	1080
6	730	73	657	5029	532	4497
7	28	3	25	1330	375	955
8	478	48	430	3682	514	3168
9	20	2	18	947	231	716
10	972	97	875			
11	2455	246	2209			
12	593	59	534			
13	205	21	184			
14	1265	127	1138			
15	386	39	347			
16	93	9	84			

3.2 Experimental Setup

To demonstrate the effectiveness of the proposed LSTM-based classification method, we quantitatively and qualitatively evaluate the performance of SeL-

STM, SaLSTM and SSLSTMs. Besides, we compare them with several state-of-the-art methods, including PCA, LDA, NWFE [17], RLDE [28], MDA [11] and CNN [2]. We also directly use the original pixels as a benchmark. For LDA, the within-class scatter matrix $\mathbf{S}_W$ is replaced by $\mathbf{S}_W + \varepsilon\mathbf{I}$, where $\varepsilon = 10^{-3}$, to alleviate the singular problem. The optimal reduced dimensions for PCA, LDA, NWFE and RLDE are chosen from $[2, 30]$. For MDA, the optimal window size is selected from a given set $\{3, 5, 7, 9, 11\}$. For CNN, the number of layers and the size of filters are the same as the network in [2]. For LSTM, we only use one hidden layer, and the number of optimal hidden nodes are selected from a given set $\{16, 32, 64, 128, 256\}$.

For IP dateset, we randomly select 10% pixels from each class as the training set, and the remaining pixels as the testing set. For PUS dataset, we randomly choose 3921 pixels as the training set and the rest of pixels as the testing set [11]. The detailed numbers of training and testing samples are listed in Table 1. In order to reduce the effects of random selection, all the algorithms are repeated five times and the average results are reported. The classification performance is evaluated by overall accuracy (OA), average accuracy (AA), per-class accuracy, and Kappa coefficient κ. OA defines the ratio between the number of correctly classified pixels to the total number of pixels in the testing set, AA refers to the average of accuracies in all classes, and κ is the percentage of agreement corrected by the number of agreements that would be expected purely by chance.

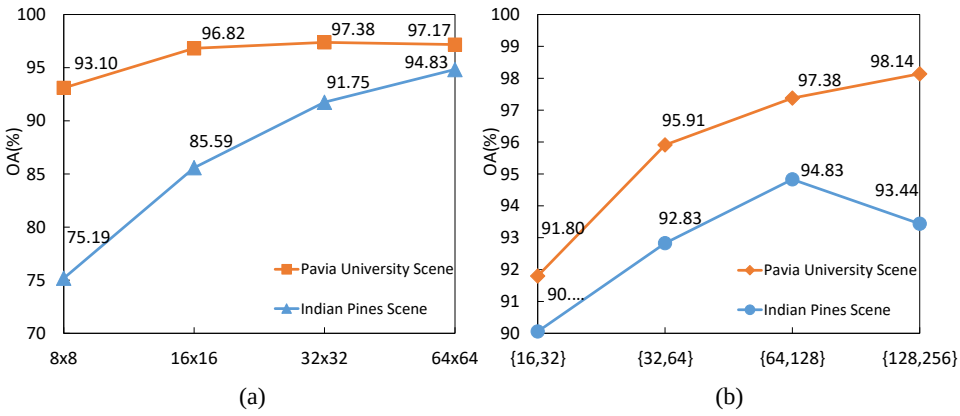

Fig. 5. (a) OAs of the SSLSTMs with different size of neighborhood regions. (b) OAs of SeLSTM and SaLSTM with different numbers of hidden nodes.

3.3 Parameter Selection

There are two important parameters in the proposed classification framework, including the size of neighborhood regions and the number of hidden nodes.

Firstly, we fix the number of hidden nodes and select the optimal region size from a given set $\{8 \times 8, 16 \times 16, 32 \times 32, 64 \times 64\}$. Fig. 5(a) demonstrates OAs of the SSLSTMs method on two datasets. From this figure, we can observe that as the region size increases, OA will firstly increase and then decrease on PUS dataset. Therefore, the optimal size is chosen as 32×32. For IP dataset, OA will increase as the size increases. However, larger sizes significantly increase the computation time. Thus, we set the optimal size as 64×64 for IP dataset.

Secondly, we fix the region size and search for the optimal number of hidden nodes for SeLSTM and SaLSTM from four different combinations $\{16, 32\}$, $\{32, 64\}$, $\{64, 128\}$ and $\{128, 256\}$. As shown in Fig. 5(b), when the number of hidden nodes for SeLSTM and SaLSTM are set to 64 and 128 respectively, the SSLSTMs method achieves the highest OA on IP dataset. Similarly, we can see that SSLSTMs obtains the highest OA on PUS dataset when the number of hidden nodes for SeLSTM and SaLSTM are set to 128 and 256 respectively.

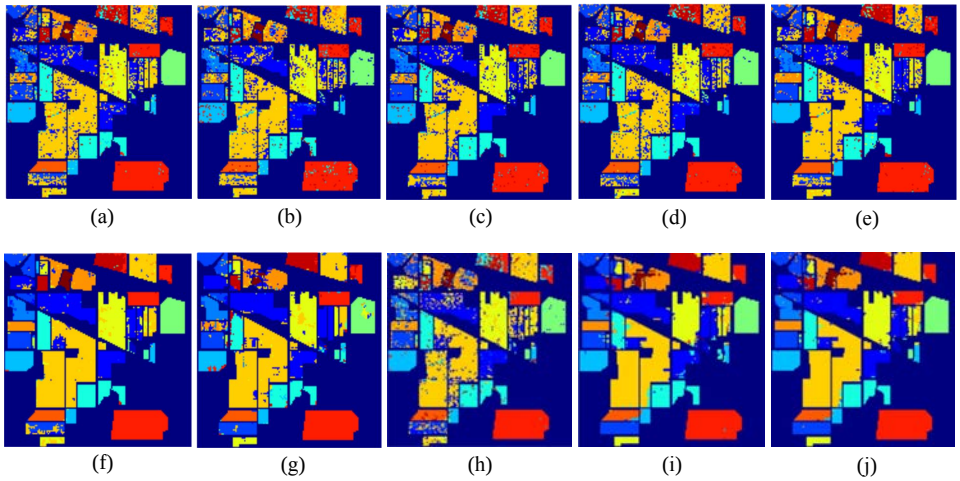

Fig. 6. Classification maps on the IP dataset. (a) Original. (b) PCA. (c) LDA. (d) NWFE. (e) RLDE. (f) MDA. (g) CNN. (h) SeLSTM. (i) SaLSTM. (j) SSLSTMs.

3.4 Performance Comparison

Table 2 reports the quantitative results acquired by ten methods on IP dataset. From these results, we can observe that PCA achieves the lowest OA among ten methods, mainly because PCA directly extracts spectral features for classification without considering spatial features. Although LDA and NWFE are still spectral-based methods, they achieve better results than PCA due to the use of label information in training samples. Besides, MDA achieves better performance than the other LDA-related methods which consider spectral information

Table 2. OA, AA, per-class accuracy (%), and κ performed by ten methods on IP dataset using 10% pixels from each class as the training set.

Label	Original	PCA	LDA	NWFE	RLDE	MDA	CNN	SeLSTM	SaLSTM	SSLSTMs
OA	77.44	72.58	76.67	78.47	80.97	92.31	90.14	72.22	91.72	95.00
AA	74.94	70.19	72.88	76.08	80.94	89.54	85.66	61.72	83.51	91.69
κ	74.32	68.58	73.27	75.34	78.25	91.21	88.73	68.24	90.56	94.29
C1	56.96	59.57	63.04	62.17	64.78	73.17	71.22	25.85	85.85	88.78
C2	79.75	68.75	72.04	76.27	78.39	93.48	90.10	66.60	89.56	93.76
C3	66.60	53.95	57.54	59.64	68.10	84.02	91.03	54.83	91.43	92.42
C4	59.24	55.19	46.58	59.83	70.80	83.57	85.73	43.94	90.61	86.38
C5	90.31	83.85	91.76	88.49	92.17	96.69	83.36	83.45	88.60	89.79
C6	95.78	91.23	94.41	96.19	94.90	99.15	91.99	87.76	90.81	97.41
C7	80.00	82.86	72.14	82.14	85.71	93.60	85.60	23.20	51.20	84.80
C8	97.41	93.97	98.74	99.04	99.12	99.91	97.35	95.40	99.02	99.91
C9	35.00	34.00	26.00	44.00	73.00	63.33	54.45	30.00	38.89	74.44
C10	66.32	64.18	60.91	69.18	69.73	82.15	75.38	71.29	88.64	95.95
C11	70.77	74.96	76.45	77.78	79.38	92.76	94.36	75.08	94.62	96.93
C12	64.42	41.72	67.45	64.05	72.28	91.35	78.73	54.49	86.10	89.18
C13	95.41	93.46	96.00	97.56	97.56	99.13	95.98	91.85	90.11	98.48
C14	92.66	89.45	93.79	93.49	92.36	98.22	96.80	90.37	98.10	98.08
C15	60.88	47.77	65.54	58.50	67.10	87.84	96.54	30.49	88.59	92.85
C16	87.53	88.17	83.66	89.03	89.68	94.29	81.90	62.86	64.05	87.86

only, because it can extract spatial and spectral features simultaneously. This indicates the importance of spatial features for HSI classification. So, as spatial based methods, CNN and SaLSTM perform better than other spectral-based methods. However, they only use the first principal component of all spectral bands, leading to the loss of spectral information. Therefore, the performance obtained by CNN or SaLSTM is inferior to that by MDA. Nevertheless, if we combine the spectral information and spatial information together, SSLSTMs can significantly improve the performance as compared to SeLSTM and SaLSTM. Additionally, as a kind of neural network, SSLSTMs is able to capture the non-linear distribution of hyperspectral data, while the linear method MDA may fail. Therefore, SSLSTMs obtains better results than MDA. Fig. 6 demonstrates classification maps achieved by different methods on the IP dataset. It can be

Table 3. OA, AA, per-class accuracy (%), and κ performed by ten methods on PUS dataset using 3921 pixels as the training set.

Label	Original	PCA	LDA	NWFE	RLDE	MDA	CNN	SeLSTM	SaLSTM	SSLSTMs
OA	89.12	88.63	84.08	88.73	88.82	96.95	96.55	93.20	94.98	98.48
AA	90.50	90.18	87.23	90.38	90.45	96.86	97.19	93.13	94.86	98.51
κ	85.81	85.18	79.59	85.31	85.43	95.93	95.30	90.43	92.84	97.56
C1	87.25	87.07	82.91	86.86	87.20	96.69	96.72	91.33	92.20	96.83
C2	89.10	88.38	80.68	88.50	88.40	97.76	96.31	94.58	95.86	98.74
C3	81.99	81.96	69.21	82.20	81.69	90.69	97.15	83.93	92.42	96.57
C4	95.65	95.14	95.99	95.27	95.79	98.44	96.16	97.78	91.59	98.43
C5	99.76	99.76	99.90	99.81	99.87	100.00	99.81	99.46	98.70	99.94
C6	88.78	88.06	89.53	88.16	88.67	96.26	94.87	91.73	96.91	99.43
C7	85.92	85.32	81.11	86.57	86.06	97.95	97.44	90.76	98.74	99.31
C8	86.14	86.06	85.81	86.13	86.42	93.98	98.23	88.78	94.79	97.98
C9	99.92	99.92	99.92	99.89	99.94	100.00	98.04	99.83	92.54	99.39

observed that SSLSTMs obtains a more homogeneous map than other methods.

Similar conclusions can be observed from the PUS dataset in Table 3 and Fig. 7. Again, MDA, CNN, and LSTM-based methods achieve better performance than other methods. Specifically, OA, AA and κ obtained by CNN are almost the same as MDA, and SSLSTMs obtains better performance than CNN and MDA. It is worth noting that the improvement of OA, AA and κ from MDA or CNN to SSLSTMs is not remarkable as those on IP dataset, because CNN and MDA have already obtained a high performance and a further improvement is very difficult.

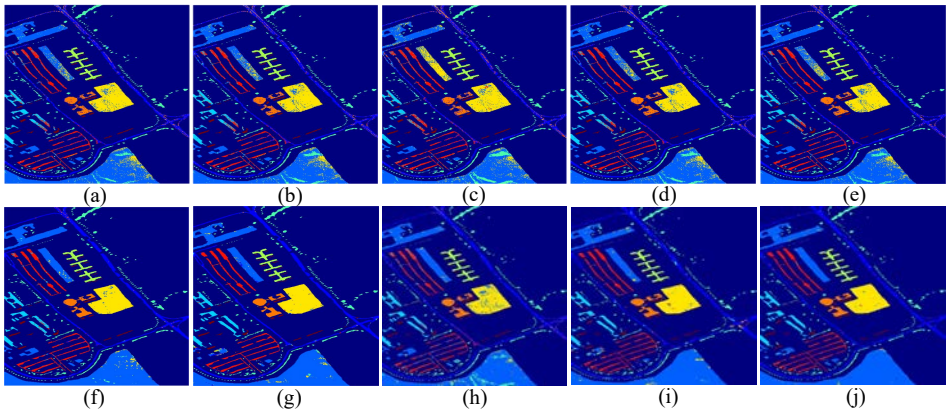

Fig. 7. Classification maps on the PUS dataset. (a) Original. (b) PCA. (c) LDA. (d) NWFE. (e) RLDE. (f) MDA. (g) CNN. (h) SeLSTM. (i) SaLSTM. (j) SSLSTMs.

4 Conclusion

In this paper, we have proposed a HSI classification method based on a LSTM network. Both the spectral feature extraction and the spatial feature extraction issues were considered as sequence learning problems, and LSTM was naturally applied to address them. Specifically, for a given pixel in HSIs, its spectral values in different channels were fed into LSTM one by one to learn spectral features. For the spatial feature extraction, a local image patch centered at the pixel was firstly selected from the first principal component of HSIs, and then the rows of the patch were fed into LSTM one by one. By conducting experiments on two HSIs collected by different instruments (AVIRIS and ROSIS), we compared the proposed method with state-of-the-art methods including CNN. The experimental results indicate that using spectral and spatial information simultaneously improves the classification performance and results in more homogeneous regions in classification maps compared to only using spectral information. We also

evaluated the influences of different parameters in the network, including the patch size and the number of hidden nodes.

References

1. Bandos, T.V., Bruzzone, L., Camps-Valls, G.: Classification of hyperspectral images with regularized linear discriminant analysis. IEEE Transactions on Geoscience and Remote Sensing 47(3), 862–873 (2009)
2. Chen, Y., Jiang, H., Li, C., Jia, X.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing 54(10), 1–20 (2016)
3. Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y.: Deep learning-based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6), 2094–2107 (2014)
4. Chen, Y., Zhao, X., Jia, X.: Spectral-spatial classification of hyperspectral data based on deep belief network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(6), 2381–2392 (2015)
5. Cho, K., Merrienboer, B.V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. Computer Science (2014)
6. Cun, Y.L., Boser, B., Denker, J.S., Howard, R.E., Habbard, W., Jackel, L.D., Henderson, D.: Handwritten digit recognition with a back-propagation network. In: Advances in Neural Information Processing Systems. pp. 396–404 (1990)
7. Deng, L.: A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing 3, e2 (2014)
8. Friedman, J.H.: Regularized discriminant analysis. Journal of the American Statistical Association 84(405), 165–175 (1989)
9. Graves, A., Jaitly, N.: Towards end-to-end speech recognition with recurrent neural networks. In: International Conference on Machine Learning. pp. 1764–1772 (2014)
10. Graves, A., Mohamed, A.R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 6645–6649 (2013)
11. Hang, R., Liu, Q., Song, H., Sun, Y.: Matrix-based discriminant subspace ensemble for hyperspectral image spatial-spectral feature fusion. IEEE Transactions on Geoscience and Remote Sensing 54(2), 783–794 (2016)
12. Hang, R., Liu, Q., Sun, Y., Yuan, X., Pei, H., Plaza, J., Plaza, A.: Robust matrix discriminative analysis for feature extraction from hyperspectral images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10(5), 2002–2011 (2017)
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Springer Berlin Heidelberg (1997)
14. Jolliffe, I.: Principal component analysis. Wiley Online Library (2002)
15. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. Computer Science (2015)
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
17. Kuo, B.C., Landgrebe, D.A.: Nonparametric weighted feature extraction for classification. IEEE Transactions on Geoscience and Remote Sensing 42(5), 1096–1105 (2004)

18. Lecun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541–551 (1989)
19. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015)
20. Melgani, F., Bruzzone, L.: Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing 42(8), 1778–1790 (2004)
21. Palsson, F., Sveinsson, J.R., Ulfarsson, M.O., Benediktsson, J.A.: Model-based fusion of multi- and hyperspectral images using pca and wavelets. IEEE Transactions on Geoscience and Remote Sensing 53(5), 2652–2663 (2015)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computer Science (2014)
23. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Advances in Neural Information Processing Systems 4, 3104–3112 (2014)
24. Tao, C., Pan, H., Li, Y., Zou, Z.: Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. IEEE Geoscience and Remote Sensing Letters 12(12), 2438–2442 (2015)
25. Wu, H., Prasad, S.: Convolutional recurrent neural networks for hyperspectral data classification. Remote Sensing 9(3), 298 (2017)
26. Zhang, L., Zhang, L., Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and Remote Sensing Magazine 4(2), 22–40 (2016)
27. Zhao, W., Du, S.: Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing 54(8), 4544–4554 (2016)
28. Zhou, Y., Peng, J., Chen, C.L.P.: Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 53(2), 1082–1095 (2015)