



HAL
open science

A Multi-level Mixed Perception Network for Hyperspectral Image Classification

Huai Wang, Qinghua He, Miaomiao Liang

► **To cite this version:**

Huai Wang, Qinghua He, Miaomiao Liang. A Multi-level Mixed Perception Network for Hyperspectral Image Classification. 5th International Conference on Intelligence Science (ICIS), Oct 2022, Xi'an, China. pp.284-293, 10.1007/978-3-031-14903-0_30 . hal-04666441

HAL Id: hal-04666441

<https://hal.science/hal-04666441v1>

Submitted on 1 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

A Multi-level Mixed Perception Network for Hyperspectral Image Classification

Huai Wang, Qinghua He, and Miaomiao Liang[✉]

School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China
6920190632@mail.jxust.edu.cn, hqh@mail.jxust.edu.cn, liangmiaom@jxust.edu.cn

Abstract. Objects in hyperspectral images (HSI) exist many subtle information differences, thus multi-level spectral-spatial perception will be beneficial to discriminative feature learning for HSI. We propose a multi-level mixed perception network (MMPN) for HSI classification, which is composed of three perceptrons: compact global and partition spectral perceptron (CSeP), pixel-wise spectral-partition perceptron (PSeP), and local spatial perceptron (LSaP). Specifically, we partition the object-centered block from HSI into non-overlapping spectral patches equidistantly. CSeP is designed on the squeezed feature to model spectral dependencies from overall and intra patches, respectively. The outputs are embedded together into the original patches for spectral information calibration. Then, PSeP is followed to avoid subtle spectra confusion, and LSaP is concurrently followed for multiscale spatial feature extraction. The learned features from each patch are used for label prediction respectively, and finally soft voting the classification result. Experimental results across two HSI datasets indicate that MMPN achieves expect performance in object classification when compared with the state-of-the-art methods.

Keywords: hyperspectral image classification, spectral partition, multilayer perceptron, feature fusion

1 Introduction

Hyperspectral images (HSI) contain abundant spectral and spatial information of ground objects and have aroused great concern in various fields, including environmental monitoring, urban planning, and mineral exploration [16, 1, 20], etc. HSI provides more abundant information for recognition of the target with a slight difference. However, it also presents new challenges for building models with a strong ability in capturing detailed information, which is subtler and easily overlooked in feature representation.

Deep learning (DL) has achieved great success in feature learning, such as multilayer perceptrons (MLP) and convolutional neural networks (CNN). Stacked autoencoders (SAEs) [15], deep belief networks (DBNs) [22], recursive autoencoders (RAEs) [21], and other modified MLP forms are introduced to extract spectral-spatial information. Subsequently, CNN gradually becomes a standard with the benefits of local feature learning by weight sharing, of which the parameter steep decline compared to MLP, especially face to deep models [7, 13, 10]. However, inefficient long-range interaction gives rise to limited performance in scale adaptive contextual sensing. Some recent works design self-attention mechanisms to enhance long-range dependence [18, 19, 4]. Such as vision transformer (ViT) [4] divides an image into equal patches, and passes them through a series of multi-head attention layers for the transformer encoder. The attention block could excellently weigh the important region in the entire image while getting local perception on image patches.

Recently, researchers try to clarify why a transformer works so well. They proposed a series of pure MLP structured networks to demonstrate that MLP with no transformer performs the same well on ImageNet. Ding et al. [3] proposed a re-parameters multi-layer perceptron (RepMLP), where feature extraction is achieved in three levels: global perception, local perception, and partition perception. In addition, they introduced a

parameter reconstruction approach to insert the convolutional parameters into the FC layer for more efficient inference. Luke et al. [9] replace the attention layer with a feed-forward layer and attempt to reveal the reason for transformer with good performance. Hugo et al. [3] designed a residual multi-layer perceptron (ResMLP), where affine transformation is used to achieve a similar effect as layer normalization and translation invariance. Furthermore, ResMLP uses two residual modules for feature extraction, one for linear interaction between patches, and the other for local feature learning in each patch. Ilya et al. [17] proposed a "mixing" multi-layer perceptron (MLP-Mixer), which uses two "mixing" strategies: channel-mixing that can be regarded as convolution with a kernel of size $N \times N$ and token-mixing do as a convolution with a kernel of size 1×1 .

As mentioned above, all the methods divide the input image into patches in the spatial dimension, and then use MLP for feature extraction and object classification. HSI classification is a pixel-level object recognition task and usually takes a patch surrounded center pixel as the input of a network for feature learning. The neighborhood pixels usually act as assistant information for the central object identification and usually with a much small size, while the hundreds of spectrums in HSI provide a lot more cues for the physical properties of the object. Many subtle differences exist in the spectral response, while most recent deep models pay less attention to its local and meanwhile global perception. Therefore, we aim to build a multi-level spectral-spatial perception network for discriminative feature learning and classification of HSI.

Inspired by RepMLP [3], we design a lightweight network with MLP and CNN operation for HSI classification in this work. The input HSI is divided into equal blocks in spectral dimension, and then a multi-level mixed perception network is built with fully connected layers on squeezed features for information interaction in global and partition spectrum, with 3D convolution layers on each partition for pixel-wise spectral partition perception, and with multi-scale 2D convolution layers for multi-level spatial contextual extraction. Compared with other state-of-the-art methods, our proposed MMPN achieves good classification performance on two real HSI data sets, with fewer parameters and faster inference speed.

The remainder of this paper is organized as follows. Section 2 introduces the proposed method in detail. Section 3 shows the experimental results and analysis. Section 4 gives the conclusion of this paper.

2 Methodology

The backbone of our proposed MMPN is mainly divided into four parts: compact global and partition spectral perceptron (CSeP-g and CSeP-p), pixel-wise spectral-partition perceptron (PSeP), local spatial perceptron (LSaP), and the final classification module.

As shown in Fig.1(a), block $\mathbf{M} \in \mathbb{R}^{C \times H \times W}$ that spitted from the original HSI is divided into $\frac{C}{b}$ patches in the spectral dimension, where C , H , W denote respectively the channel number, the height and width of the block, and b is the band number of each partition patch. Then, the input is in size of $(\frac{C}{b}, b, H, W)$ and is defined as $\mathbf{M}^{(\text{in})}$. In Fig.1(c), global average pooling (GAP) is used to squeeze the spatial information, and the output is of size $(\frac{C}{b}, b, 1, 1)$.

$$\mathbf{M}^{(\text{GAP1})} = \text{GAP}(\mathbf{M}^{(\text{in})}, (\frac{C}{b}, b, 1, 1)), \quad (1)$$

Then, CSeP-g and the CSeP-p modules are set to achieve global and intra-patches dependencies. For CSeP-g, we reshape (RS) the $\mathbf{M}^{(\text{GAP1})}$ as the size of $(1, C)$, and feed them into FC layers to learn global spectral dependencies, which can be denoted as,

$$\begin{aligned} \mathbf{M}_{\text{in}}^{(\text{CSeP-g})} &= \text{RS}(\mathbf{M}^{(\text{GAP1})}, (1, C)), \\ \mathbf{M}_{\text{out}}^{(\text{CSeP-g})} &= \text{MLP1}(\mathbf{M}_{\text{in}}^{(\text{CSeP-g})}, \mathbf{F}_1). \end{aligned} \quad (2)$$

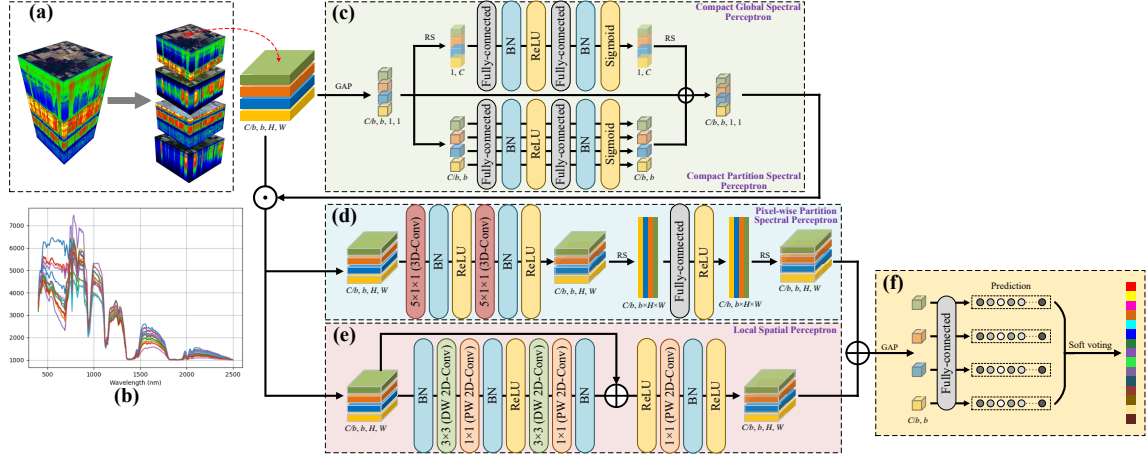


Fig. 1. The framework of MMPN for HSI classification. (a) Input from the original HSI is partitioned into non-overlapping patches in spectral dimension; (b) Spectral curves from some different ground objects; (c) CSeP module; (d) Pixel-wise spectral-partition perceptron; (e) Local spatial perceptron; (f) Classification with multiple group predictions.

where the MLP1 contains two FC layers, two batch normalization layers, one ReLU activation function, and one Sigmoid function, as illustrated in Fig. 1. \mathbf{F}_1 is the corresponding weight matrix. The CSeP-p module is built the same as CSeP-g but performs on each partition, which means,

$$\mathbf{M}_{\text{out},i}^{(\text{CSeP-p})} = \text{MLP1} \left(\mathbf{M}_i^{(\text{GAP1})}, \mathbf{F}_{2,i} \right), \quad i = 1, 2, \dots, \frac{C}{b}. \quad (3)$$

Finally, we sum up the two outputs together to calibrate the spectral bands, the compact weighting matrix \mathbf{W} and the output of CSeP module is defined as,

$$\begin{aligned} \mathbf{W} &= \text{RS} \left(\mathbf{M}_{\text{out}}^{(\text{CSeP-g})}, \left(\frac{C}{b}, b, 1, 1 \right) \right) + \mathbf{M}_{\text{out}}^{(\text{CSeP-p})}, \\ \mathbf{M}_{\text{out}}^{(\text{CSeP})} &= \mathbf{M}^{(\text{in})} \odot \mathbf{W}. \end{aligned} \quad (4)$$

In PSeP module, we focus on intra-patches dependencies, which means each band partition is considered as an independent sample during feature learning. Fig. 1 (d) gives the sketch of PSeP module that follows behind $\mathbf{M}_{\text{out}}^{(\text{CSeP})}$. Here, we design two sub-modules for pixel-wise spectral feature learning. One is built by two 3D convolutional layers with respectively one filter kernel, which is to overcome the defect of spectral confusion caused by spatial squeeze and thus worse performance in key spectral perception. The other sub-module is built by one FC layer when the inputs are reshaped to size of $(\frac{C}{b}, b \times H \times W)$, a design for overall partition spectral perception from all the neighborhood. It should be noted that all partitions share the same filter weights to avoid parameter increases. the operation can be written as follows,

$$\begin{aligned} \mathbf{M}_{\text{out},i}^{(\text{Conv3D})} &= \text{Conv3D} \left(\mathbf{M}_{\text{out},i}^{(\text{CSeP})}, \mathbf{W}_1 \right), \\ \mathbf{M}_{\text{out},i}^{(\text{MLP2})} &= \text{MLP2} \left(\text{RS} \left(\mathbf{M}_{\text{out},i}^{(\text{Conv3D})}, \left(\frac{C}{b}, b \times H \times W \right) \right), \mathbf{F}_3 \right). \end{aligned} \quad (5)$$

where \mathbf{W}_1 is the 3D filter parameter of size $5 \times 1 \times 1$, and \mathbf{F}_3 is the weights of FC layer with size of (bHW, bHW) . Here, MLP2 contains only one FC layer and one activation Layer by ReLU function. Finally, we shape $\mathbf{M}_{\text{out},i}^{(\text{MLP2})}$

back to $(\frac{C}{b}, b, H, W)$ and obtain the output $\mathbf{M}_{\text{out},i}^{(\text{PSeP})}$ of PSeP module for further feature fusion,

$$\mathbf{M}_{\text{out},i}^{(\text{PSeP})} = \text{RS}(\mathbf{M}_{\text{out},i}^{(\text{MLP2})}, (\frac{C}{b}, b, H, W)). \quad (6)$$

Many studies have shown that local spatial perception has great benefits in effective HSI classification[6, 5]. The pure FC layer is suitable for learning large-range dependencies and location information but is poor in capturing local spatial features. In Fig.1(e), we employ depth-wise separable convolution to extract local spatial features in LSaP module which is parallel with PSeP part. This part can be defined as,

$$\begin{aligned} \mathbf{M}_{\text{out},i}^{(\text{DSCConv})} &= \text{DSCConv}(\mathbf{M}_{\text{out},i}^{(\text{CSeP})}, \mathbf{W}_2) + \mathbf{M}_{\text{out},i}^{(\text{CSeP})}, \\ \mathbf{M}_{\text{out},i}^{(\text{LSaP})} &= \text{PConv}(\mathbf{M}_{\text{out},i}^{(\text{DSCConv})}, \mathbf{W}_3), \end{aligned} \quad (7)$$

where $\text{DSCConv}(\cdot, \cdot)$ contains two depthwise separable convolution layers with kernel of size 3×3 in depth-wise layers and $\text{PConv}(\cdot, \cdot)$ contains one point-wise convolution layer for further feature recombination. Besides, skip connection is used here for smooth flow of information.

Finally, features from PSeP and LSaP modules are fused together for object classification. Here, we perform class prediction separately on each patch and obtain the final results by soft voting. As shown in Fig.1(f), each classifier consists of a mean statistics description of each feature map by a GAP layer and linear class prediction by an FC layer. The classification module can be represented as,

$$\begin{aligned} \mathbf{M}_i^{(\text{GAP2})} &= \text{GAP}(\mathbf{M}_{\text{out},i}^{(\text{PSeP})} + \mathbf{M}_{\text{out},i}^{(\text{LSaP})}, (1, b, 1, 1)), \\ \mathbf{L} &= \frac{b}{C} \sum_{i=1}^{C/b} \text{FC}(\mathbf{M}_i^{(\text{GAP2})}, \mathbf{F}_{4,i}), \end{aligned} \quad (8)$$

where $\mathbf{F}_{4,i} \in b \times L$ and L is the class number. \mathbf{L} is the probability that a sample belongs to each class.

3 Experiments Results and Analysis

We evaluate our proposed MMPN model from parameter analysis, ablation study, and comparison with some state-of-the-art methods on two real HSI data sets. Indian Pines (IN) dataset contains 145×145 pixels with a spatial resolution of 20 meters per pixel, and 200 spectral bands after some noisy bands are removed. There are 16 classes of ground objects in the ground truth. The Pavia University (UP) dataset contains 610×340 pixels and 103 bands after some noisy bands are removed. 9 classes of urban area ground objects are labeled in the ground truth. Figs. 3 and 4 shows the false-color images and the ground truth of the two datasets.

All experiments in this work are implemented on the platform with Intel Xeon W-2133 CPU and NVIDIA GeForce RTX 2080 Ti GPU. The software environment is Python 3.8.3, PyTorch 1.7.0, and CUDA 11.0. We randomly select 10% and 3% samples from the labeled IN and UP dataset for model training, and the rest for testing. SGD optimizer is used to update the network parameters, the batch size is set respectively to 16 and 32, the learning rate is set to 0.01, weight decay is set to 0.0001, and momentum is set to 0.9. Overall accuracy (OA) is used to measure classification performance.

3.1 Parameters Analysis and Ablation Study

In MMPN, the band number b in each partition is a critical parameter that affects the classification performance of HSI. Thus, we verify the OAs as the b ranging from 10 to 200 for the IN dataset, and from 10 to 100 for

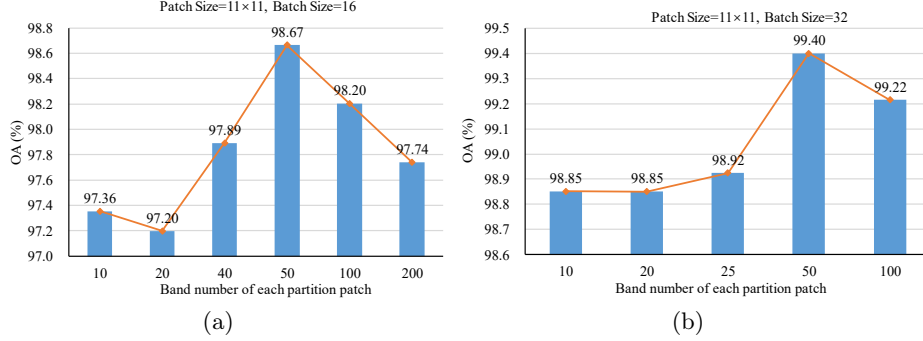


Fig. 2. Classification OAs (%) on the (a) IN and (b) UP data sets as the band number b of each partition ranging from 10 to all.

Table 1. Ablation study of each component in MMPN. \checkmark denotes the corresponding module present in MMPN.

CSeP-g	Components			OA	IN			UP		
	CSeP-p	PSeP	LSaP		AA	Kappa	OA	AA	Kappa	
	\checkmark	\checkmark	\checkmark	98.33	96.72	98.10	99.32	98.83	99.10	
\checkmark		\checkmark	\checkmark	98.28	96.74	98.04	99.17	98.68	98.90	
		\checkmark	\checkmark	97.14	95.19	96.75	99.18	98.58	98.92	
\checkmark	\checkmark		\checkmark	98.65	97.22	98.47	99.35	98.78	99.11	
\checkmark	\checkmark	\checkmark		95.70	92.71	95.10	97.05	96.09	96.13	
\checkmark	\checkmark	\checkmark	\checkmark	98.67	97.30	98.48	99.40	98.88	99.21	

the UP dataset when the last 3 spectra are removed for equal division. Here, ‘10’ means that we divide the IN data into 20 patches from spectral dimension and ‘200’ means there is no partition in feature learning, the same situation for the UP data. As the results in Fig. 2, we can observe that it does not look well to partition too coarse patches, or too finer. The best partition is $d = 50$ for both two experimental datasets.

Besides, three main components are designed for discriminative feature learning, which are CSeP for compact spectral calibration, PSeP for pixel-wise partition perception, and LSaP for local spatial filtering. In this part, we discuss the importance of those components in MMPN through several sets of ablation studies. As the results reported in Table 1, the classification accuracy decreases more or less when any of the components are absent in MMPN. Specifically, CSeP achieves a 1-2% accuracy increase in the IN dataset, but the advantage is not obvious in the UP dataset. This indicates that CSeP contributes more to distinguishing the ground objects with a highly similar spectrum (such as IN dataset) but is not beneficial to the recognition of objects with a large spectral difference. LSaP is crucial for both of the datasets, boosting the accuracy increase by 2 to 3 points. PSeP is a less elegant solution for accuracy increase but will be beneficial for finer boundary location.

3.2 Comparison with State-of-the-Art Methods

We further compare our proposed MMPN model with seven deep learning-based methods, including one RNN-based model (RNN [11]), two spectral-spatial convolution networks (SSRN [23] and S2FEF [2]), and four attention mechanism related models (S3EResBoF [12], RSSAN [24], SSAN [14], LMFN [8]). The parameter settings of the above deep models are set according to the original paper. We exhibit the best classification

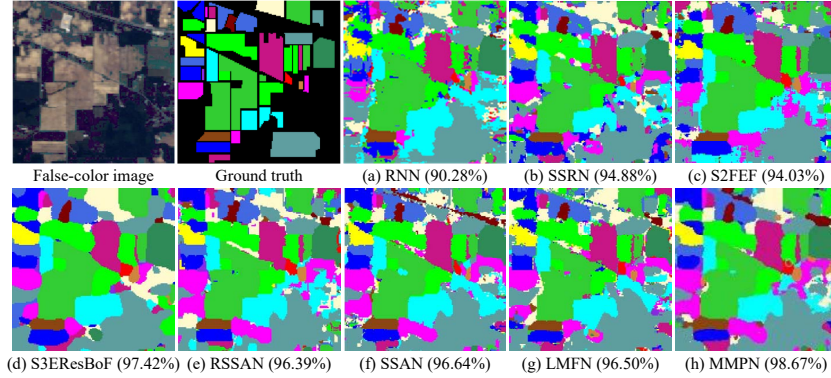


Fig. 3. Comparison of the classification maps with the state-of-the-art methods on IN data set.

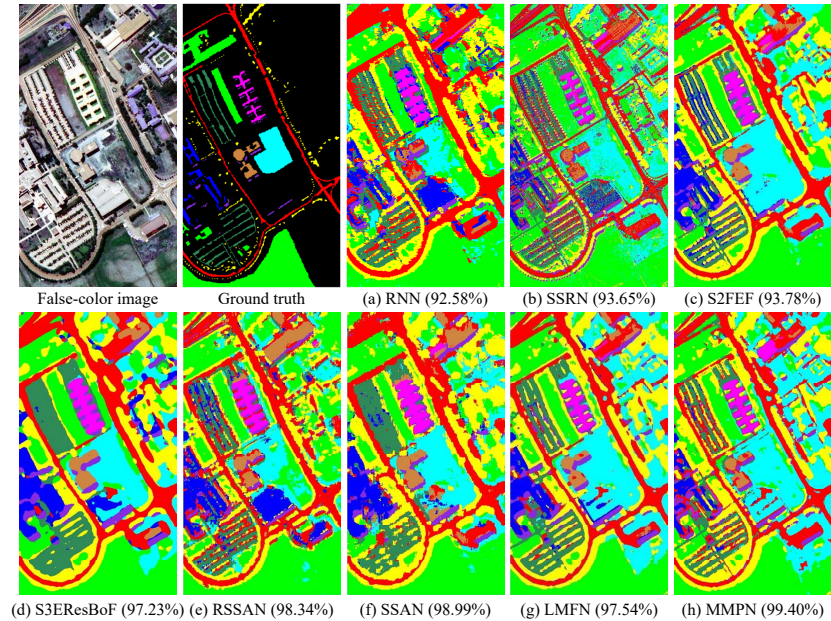


Fig. 4. Comparison of the classification maps with the state-of-the-art methods on UP data set.

maps and report the OAs from all the comparing methods in Fig. 3 and Fig. 4, respectively. From the results, we can conclude that our method is superior in classification accuracy, meanwhile giving more clear and clean boundary positioning.

To demonstrate the stability of our model to the number of training samples, we further show the classification accuracies when 1% to 15% labeled samples in each class are randomly sampled for model training, and all the results are reported as the mean of ten runs. The results in Fig. 5 show that our method achieves the best classification accuracy, except for a slightly lower OA on IN dataset with 1% samples per class for model training.

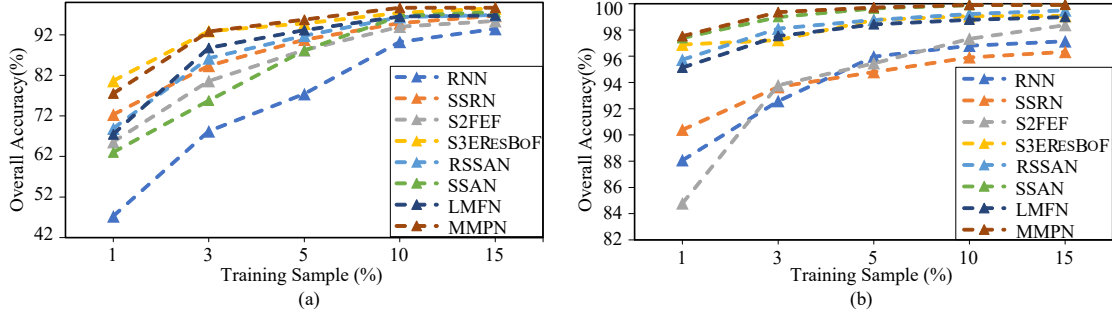


Fig. 5. Classification performance OA (%) on the IN (a) and UP (b) data sets with the number of training samples per class ranging from 1% to 15%.

4 Conclusion

In this paper, we introduce a multi-level mixed perception network for HSI classification. With the characteristics of the fully connected network and convolution network, MMPN divides the HSI into patches from the spectral dimension and builds three perceptrons: compact global and partition spectral perceptron, pixel-wise spectral-partition perceptron, and local spatial perceptron, to extract multi-scale spectral-spatial information from multiple perspectives. Besides, soft voting from each patch is designed at the end of the network to achieve the final class prediction. This multi-view feature extraction re-examines the importance of spectral information in HSI classification. Experimental results on two hyperspectral data sets show that MMPN presents some advantages in classification accuracy, including robustness and generalization.

In the future, we will pay close attention to automatic spectral partition and self-supervised learning of the potential rules or manifold.

Acknowledgements This research was funded by the National Natural Science Foundation of China (Nos. 61901198, 62066018); the Natural Science Basic Research Plan in Shaanxi Province of China (No. 2022JQ-704); and the Program of Qingjiang Excellent Young Talents, Jiangxi University of Science and Technology (No. JXUSTQJYX2020019).

References

1. Bioucas-Dias, J.M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J.: Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine* **1**(2), 6–36 (2013). <https://doi.org/10.1109/MGRS.2013.2244672>
2. Chen, L., Wei, Z., Xu, Y.: A lightweight spectral–spatial feature extraction and fusion network for hyperspectral image classification. *Remote Sensing* **12**(9), 1395 (2020)
3. Ding, X., Zhang, X., Han, J., Ding, G.: Repmlp: Re-parameterizing convolutions into fully-connected layers for image recognition (2021)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale (2021)

5. Ghamisi, P., Maggiori, E., Li, S., Souza, R., Tarablaka, Y., Moser, G., De Giorgi, A., Fang, L., Chen, Y., Chi, M., Serpico, S.B., Benediktsson, J.A.: New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, markov random fields, segmentation, sparse representation, and deep learning. *IEEE Geoscience and Remote Sensing Magazine* **6**(3), 10–43 (2018). <https://doi.org/10.1109/MGRS.2018.2854840>
6. He, L., Li, J., Liu, C., Li, S.: Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines. *IEEE Transactions on Geoscience and Remote Sensing* **56**(3), 1579–1597 (2018). <https://doi.org/10.1109/TGRS.2017.2765364>
7. Lee, H., Kwon, H.: Going deeper with contextual cnn for hyperspectral image classification. *IEEE Transactions on Image Processing* **26**(10), 4843–4855 (2017). <https://doi.org/10.1109/TIP.2017.2725580>
8. Liang, M., Wang, H., Yu, X., Meng, Z., Yi, J., Jiao, L.: Lightweight multilevel feature fusion network for hyperspectral image classification. *Remote Sensing* **14**(1), 79 (2021)
9. Melas-Kyriazi, L.: Do you even need attention? a stack of feed-forward layers does surprisingly well on imagenet (2021)
10. Meng, Z., Li, L., Jiao, L., Feng, Z., Tang, X., Liang, M.: Fully dense multiscale fusion network for hyperspectral image classification. *Remote Sensing* **11**(22), 2718 (2019)
11. Mou, L., Ghamisi, P., Zhu, X.X.: Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **55**(7), 3639–3655 (2017)
12. Roy, S.K., Chatterjee, S., Bhattacharyya, S., Chaudhuri, B.B., Platoš, J.: Lightweight spectral–spatial squeeze-and-excitation residual bag-of-features learning for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing* **58**(8), 5277–5290 (2020)
13. Song, W., Li, S., Fang, L., Lu, T.: Hyperspectral image classification with deep feature fusion network. *IEEE Transactions on Geoscience and Remote Sensing* **56**(6), 3173–3184 (2018). <https://doi.org/10.1109/TGRS.2018.2794326>
14. Sun, H., Zheng, X., Lu, X., Wu, S.: Spectral–spatial attention network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **58**(5), 3232–3245 (2020)
15. 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). <https://doi.org/10.1109/LGRS.2015.2482520>
16. Tian, A., Fu, C., Yau, H.T., Su, X.Y., Xiong, H.: A new methodology of soil salinization degree classification by probability neural network model based on centroid of fractional lorenz chaos self-synchronization error dynamics. *IEEE Transactions on Geoscience and Remote Sensing* **58**(2), 799–810 (2020). <https://doi.org/10.1109/TGRS.2019.2940592>
17. Tolstikhin, I., Houlisby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., Dosovitskiy, A.: Mlp-mixer: An all-mlp architecture for vision (2021)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Advances in neural information processing systems*. pp. 5998–6008 (2017)
19. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018)
20. Yokoya, N., Chan, J.C.W., Segl, K.: Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated enmap and sentinel-2 images. *Remote Sensing* **8**(3) (2016). <https://doi.org/10.3390/rs8030172>, <https://www.mdpi.com/2072-4292/8/3/172>
21. Zhang, X., Liang, Y., Li, C., Huyan, N., Jiao, L., Zhou, H.: Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* **14**(11), 1928–1932 (2017). <https://doi.org/10.1109/LGRS.2017.2737823>
22. Zhong, P., Gong, Z., Li, S., Schönlieb, C.B.: Learning to diversify deep belief networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **55**(6), 3516–3530 (2017)
23. Zhong, Z., Li, J., Luo, Z., Chapman, M.: Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing* **56**(2), 847–858 (2018)
24. Zhu, M., Jiao, L., Liu, F., Yang, S., Wang, J.: Residual spectral–spatial attention network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **59**(1), 449–462 (2020)