Robust discriminative broad learning system for hyperspectral image classification[*]

ZHAO Liguo[1], HAN Zhe[1]**, and LUO Yong[2]

1. *School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang 471023, China*

2. *Guizhou Cloud Big Data Industry Development Co., Ltd., Guiyang 550001, China*

With the advantages of simple structure and fast training speed, broad learning system (BLS) has attracted attention in hyperspectral images (HSIs). However, BLS cannot make good use of the discriminative information contained in HSI, which limits the classification performance of BLS. In this paper, we propose a robust discriminative broad learning system (RDBLS). For the HSI classification, RDBLS introduces the total scatter matrix to construct a new loss function to participate in the training of BLS, and at the same time minimizes the feature distance within a class and maximizes the feature distance between classes, so as to improve the discriminative ability of BLS features. RDBLS inherits the advantages of the BLS, and to a certain extent, it solves the problem of insufficient learning in the limited HSI samples. The classification results of RDBLS are verified on three HSI datasets and are superior to other comparison methods.

The spatial resolution of hyperspectral images (HSIs) has been greatly improved, which enhances the expression ability of hyperspectral data for ground objects. Using the characteristics of hyperspectral data map integration and wideband coverage can greatly improve the ability to distinguish and identify the types of ground objects. Then, hyperspectral remote sensing technology has been used in precision agriculture and geology[1,2].

In the past decade, machine learning methods have been widely used in HSI classification[3], such as the K-nearest neighbor method[4], support vector machine (SVM) decision tree[5]. Other methods such as sparse representation[6] for hyperspectral separation have also been widely used. However, since the high dimensionality and complexity of HSI, it is usually time-consuming to find optimal parameters for classification through machine learning algorithms.

Recently, deep learning can obtain the high-level features of data through the training set, instead of selecting the classification features through the feature engineering like the traditional method, so that the classification model can better represent the features of the data and improve the classification accuracy. Deep learning has made many applications in the field of image recognition. CHEN et al[7] designed a method for regularized deep feature extraction and virtual sample augmentation

based on convolutional neural networks (CNNs) to achieve better model generalization. HE et al[8] established a three-dimensional (3D) CNN, which has become an effective selection of high-order spatial-spectral features of HSI. QIN et al[9] extended the graph convolution neural network (GCNN) to the second-order version.

The above HSI classification methods have achieved good classification performance. However, the structure of a deep network involves a large number of super parameters, which makes deep learning very time-consuming in the training process. Therefore, CHEN et al[10] proposed a simple fast incremental learning neural network, broad learning system (BLS). BLS first learns the sparse mapping features from the original input through the feature nodes and then obtains the enhanced features through the nonlinear expansion of the enhanced nodes. The two feature expressions are connected in parallel as the final total input to the output layer for classification and recognition. In addition, CHEN et al[11] also proposed a series of incremental learning algorithms to further improve the BLS model without the whole retraining process from scratch. KONG et al[12] proposed a semi-supervised BLS for HSI classification. MA et al[13] combined a CNN with the BLS, extracted deep-seated features of HSIs through CNN, and then used the BLS for classification. ZHAO et al[14] input the fused spectral-spatial features

into the BLS.

Researchers have proposed various improved methods based on the BLS. However, BLS and its improved methods do not fully consider the geometry structure between data samples and the discrimination information contained in data. This paper proposes a robust discriminative broad learning system (RDBLS), which considers the discriminative information and geometric structure of data samples, constructs the discriminative regularization framework, and then introduces the discriminative regularization framework into BLS.

As shown in Fig.1, the hidden layer of BLS includes two parts, feature node and enhancement node.

$$\boldsymbol{Z}_i = \phi_i(\boldsymbol{X}\boldsymbol{W}_{ei} + \boldsymbol{\beta}_{ei}), \ i = 1,2,\cdots,n, \qquad \boldsymbol{H}_j = \xi_j(\boldsymbol{Z}^n\boldsymbol{W}_{hj} + \boldsymbol{\beta}_{hj}), \ j = 1,2,\cdots,m,$$
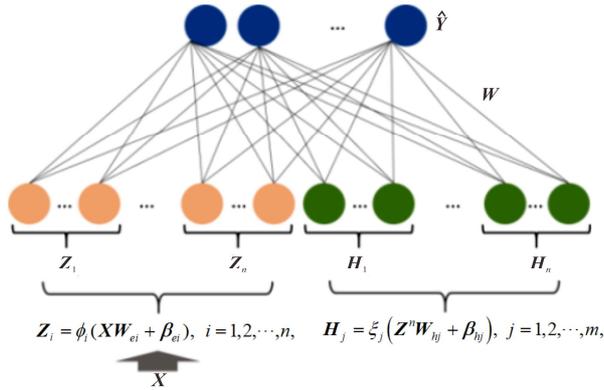
Fig.1 Structure of BLS

Using $\boldsymbol{Z}_i$ to represent the *i*-th group of feature nodes containing *q* neurons, there is

$$\boldsymbol{Z}_i = \phi_i(\boldsymbol{X}\boldsymbol{W}_{ei} + \boldsymbol{\beta}_{ei}), \ i = 1,2,\cdots,n, \qquad (1)$$

where ϕ_i is a linear function, and $\boldsymbol{W}_{ei}$ and $\boldsymbol{\beta}_{ei}$ are input weight matrix and hidden layer node offset matrix, respectively. $\boldsymbol{W}_{ei}$ and $\boldsymbol{\beta}_{ei}$ are fine-tuned by sparse self-encoder. The total feature matrix is

$$\boldsymbol{Z}^n = [\boldsymbol{Z}_1, \boldsymbol{Z}_2 \cdots, \boldsymbol{Z}_n]. \qquad (2)$$

If $\boldsymbol{H}_i$ is used to represent the *r*-th group of enhancement nodes containing *r* neurons, there is

$$\boldsymbol{H}_j = \xi_j(\boldsymbol{Z}^n\boldsymbol{W}_{hj} + \boldsymbol{\beta}_{hj}), \ j = 1,2,\cdots,m, \qquad (3)$$

where ξ_i is the nonlinear activation function, and $\boldsymbol{W}_{hi}$ and $\boldsymbol{\beta}_{hi}$ are input weight matrix and hidden layer node offset matrix, respectively.

The total enhanced matrix is

$$\boldsymbol{H}^m = [\boldsymbol{H}_1, \boldsymbol{H}_2 \cdots, \boldsymbol{H}_m] \in R^{N \times mr}. \qquad (4)$$

Let $\boldsymbol{A} = [\boldsymbol{Z}^n \mid \boldsymbol{H}^m]$, and the output of BLS is

$$\hat{\boldsymbol{Y}} = \boldsymbol{A}\boldsymbol{W}, \qquad (5)$$

where $\boldsymbol{W}$ is the weight from the hidden layer to the output layer. Since $\boldsymbol{W}_{ei}$, $\boldsymbol{\beta}_{ei}$, $\boldsymbol{W}_{hi}$ and $\boldsymbol{\beta}_{hi}$ are randomly generated, the weight learned by the network is only $\boldsymbol{W}$. The function of BLS is

$$\operatorname*{argmin}_{\boldsymbol{W}} J_{\mathrm{BLS}} = \left\| \boldsymbol{Y} - \hat{\boldsymbol{Y}} \right\|_2^2 + \frac{1}{2}\lambda\|\boldsymbol{W}\|_2^2, \qquad (6)$$

where $\left\| \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{W} \right\|_2^2$ is to minimize the training error, $\|\boldsymbol{W}\|_2^2$ is used to prevent the model from overfitting, and λ is the regularization coefficient. We can solve Eq.(6) to get

$$\boldsymbol{W} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \lambda\boldsymbol{I} \right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y}, \qquad (7)$$

where $\boldsymbol{I}$ is the identity matrix.

Fig.2 shows the framework of the proposed method. For N different data samples, $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots \boldsymbol{x}_i, \cdots, \boldsymbol{x}_N\}$. The total scatter matrix is $\boldsymbol{S}_t$ which can be represented as

$$\boldsymbol{S}_t = \sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{u})(\boldsymbol{x}_i - \boldsymbol{u})^{\mathrm{T}}, \qquad (8)$$

where $\boldsymbol{u} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i$ is the total mean of samples, which reflects the distribution features and discriminative information of input data sample space.

For the robustness of RDBLS, we use cosine metric to represent $\boldsymbol{S}_t$, and we can compute the cosine value of $\phi_i = \langle(\boldsymbol{x}_i - \boldsymbol{u}), \boldsymbol{x}_i - \boldsymbol{u}\rangle$ as follows

$$\tilde{\boldsymbol{S}}_t = \sum_{i=1}^{N}\cos^2\phi_i = \sum_{i=1}^{N}\frac{\left[(\boldsymbol{x}_i - \boldsymbol{u})\cdot(\boldsymbol{x}_i - \boldsymbol{u})^{\mathrm{T}}\right]^2}{\|\boldsymbol{c}\|^2} = \\ \sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{u})_e(\boldsymbol{x}_i - \boldsymbol{u})_e^{\mathrm{T}}. \qquad (9)$$
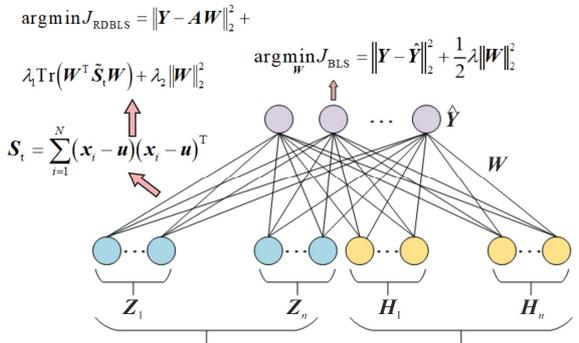
Fig.2 Framework of the proposed method

We introduce the total scatter matrix $\boldsymbol{S}_t$ into BLS. In the BLS method, $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{W}$. We replace $\boldsymbol{x}_i$ with $\boldsymbol{A}_i$ in the total scatter matrix, $\boldsymbol{u} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i$. The function of RDBLS can be represented as

$$\operatorname*{argmin} J_{\mathrm{RDBLS}} = \left\| \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{W} \right\|_2^2 + \\ \lambda_1\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\tilde{\boldsymbol{S}}_t\boldsymbol{W}) + \lambda_2\|\boldsymbol{W}\|_2^2, \qquad (10)$$

where $\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\tilde{\boldsymbol{S}}_t\boldsymbol{W})$ is the discriminative regularization framework, and λ_1 and λ_2 are regularization coefficients. Differentiate Eq.(10) $\partial F_{\mathrm{RDBLS}}/\partial\boldsymbol{W} = 0$ as follows

$$\frac{\partial F_{\mathrm{RDBLS}}}{\partial\boldsymbol{W}} = \frac{\partial\left(\left\| \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{W} \right\|^2 + \lambda_1(\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\tilde{\boldsymbol{S}}_t\boldsymbol{W}) + \lambda_2\|\boldsymbol{W}\|_2^2\right)}{\partial\boldsymbol{W}} =$$

$$\left(2A^{\mathrm{T}}AW - 2A^{\mathrm{T}}Y\right) + 2\lambda_1 A^{\mathrm{T}}\tilde{S}_{\mathrm{t}} A^{\mathrm{T}}W + 2\lambda_2 W. \qquad (11)$$

Order $\dfrac{\partial F_{\mathrm{RDBLS}}}{\partial W} = 0$, and then we get

$$\left(A^{\mathrm{T}}AW - A^{\mathrm{T}}Y\right) + \lambda_1 A^{\mathrm{T}}\tilde{S}_{\mathrm{t}} A^{\mathrm{T}}W + \lambda_2 W = 0. \qquad (12)$$

The output weight matrix can be calculated according to Eq.(12) as

$$W = \left[A^{\mathrm{T}}A + \lambda_1 \tilde{S}_{\mathrm{t}} + \lambda_2 I\right]^{-1} A^{\mathrm{T}}Y. \qquad (13)$$

In this paper, three HSI datasets (The four HSI benchmarks are available from http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral _ Remote _ Sensing _ Scenes.) are used to verify RDBLS. The evaluation metrics include overall accuracy (*OA*), average accuracy (*AA*), and Kappa coefficient.

The first experiment is on the Indian Pines dataset. The image contains 200 bands, and the image size is 145×145 and contains 16 feature categories. The second experiment is on the Pavia University dataset. It contains 9 figure categories and 103 spectral bands, and the image size is 610×340. The third experiment is on the Salinas Scene dataset. The image contains 204 bands and 16 feature categories, and the image size is 512×217.

For the above three datasets, 30, 20, and 10 are selected as training samples and the rest as test samples. To avoid random bias, each method repeated 10 experiments and calculated the average results. RDBLS is compared with BLS[10], random multigraphs ensemble learning (RMGE)[14], extreme learning machine with composite kernels (ELMCK)[15], and Gabor filtering based deep network (GFDN)[16]. In experiments, we use edge-preserving filtering to extract the spatial-spectral joint features. For the RDBLS, 10-fold cross-validation and grid search method are used for parameter optimization. The regularization parameters λ_1 and λ_2 of the RDBLS are chosen from $\{2^{-10}, 2^{-4},...., 2^{-1}, 2^0\}$. Tabs.1—3 respectively give the comparison results of the classification accuracy of the three HSI datasets under different classification methods.

The comparison results of RDBLS with RMGE, ELMCK, GFDN, and BLS are shown in Tab.1. From the Indian Pines dataset, RDBLS is 1.43% higher than the RMGE in *OA*, 6.42% higher than ELMCK, 3.08% higher than GFDN, and 2.12% higher than BLS. For the metric *AA*, RDBLS is 0.58% higher than the RMGE, 3.11% higher than ELMCK, 1.35% higher than GFDN, and 1.18% higher than BLS. GFDN as a deep learning method, it is 3.34% higher than ELMCK in *OA*. For the metric *OA*, BLS is 4.3% higher than ELMCK, and 0.96% higher than GFDN.

From Tab.2, RDBLS achieves a better classification on the Pavia University dataset. The RDBLS is 4.48% higher than the RMGE in *OA*, 8.72% higher than ELMCK,

6.14% higher than GFDN, and 5.47% higher than BLS. For the metric *AA*, RDBLS is 5.1% higher than the RMGE, 7.52% higher than ELMCK, 3.08% higher than GFDN, and 2.85% higher than BLS. GFDN is 2.58% higher than ELMCK in *OA*. For the metric *OA*, BLS is 3.75% higher than the RMGE, 2.25% higher than ELMCK, and 0.67% higher than GFDN.

From Tab.3, RDBLS achieves the best performance in *OA*, *AA*, and Kappa coefficient, which are 98.69%, 98.82%, and 98.54%, respectively. For the metric *OA*, RDBLS is 1.05% higher than the RMGE, 7.28% higher than ELMCK, 9.06% higher than GFDN, and 2.74% higher than BLS. For the metric Kappa coefficient, RDBLS is 0.93% higher than the RMGE, 8.1% higher than ELMCK, 10.07% higher than GFDN, and 3.04% higher than BLS. For the Kappa coefficient, BLS is 5.06% higher than ELMCK, and 7.03% higher than GFDN.

Figs.3—5 show the classification results of RDBLS on three HSI datasets. We can see that the classification graph of this algorithm is smoother.

Tab.1 Experimental results of different algorithms on the Indian Pines dataset (%)

No.	ELMCK	GFDN	RMGE	BLS	RDBLS
1	100.00	95.65	100.00	95.65	**100.00**
2	90.41	88.98	85.64	**98.35**	96.92
3	89.88	98.63	95.56	89.75	**91.13**
4	100.00	100.00	100.00	100.00	**100.00**
5	84.77	91.17	95.24	89.62	**96.03**
6	98.71	99.71	99.86	**100.00**	99.71
7	100.00	100.00	100.00	100.00	**100.00**
8	100.00	100.00	100.00	100.00	**100.00**
9	100.00	100.00	100.00	100.00	**100.00**
10	87.69	87.15	86.42	**92.46**	91.08
11	82.27	89.20	97.15	88.45	**96.58**
12	87.57	97.51	98.82	**98.76**	98.22
13	100.00	99.43	99.51	100.00	**100.00**
14	97.33	99.19	100.00	99.92	**99.92**
15	99.16	99.72	100.00	97.19	**99.44**
16	100.00	100.00	100.00	98.41	98.41
OA	90.37	93.71	95.36	94.67	**96.79**
AA	94.86	96.65	97.39	96.79	**97.97**
Kappa	89.06	92.86	94.73	93.92	**96.32**

(a)   (b)
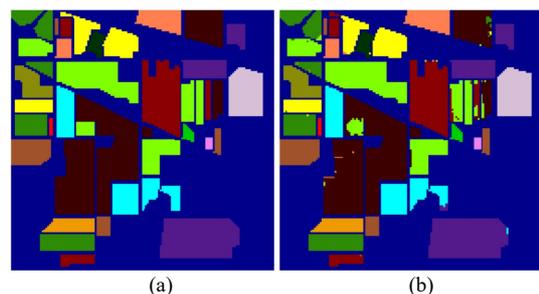
Fig.3 Classification results on the Indian Pines dataset: (a) Ground-truth image; (b) RDBLS

Tab.2 Experimental results of different algorithms on the Pavia University dataset (%)

No.	ELMCK	GFDN	RMGE	BLS	RDBLS
1	78.93	86.01	89.67	90.64	**93.59**
2	90.89	90.00	98.95	88.41	**98.26**
3	91.29	94.71	94.24	96.20	**97.11**
4	89.52	94.88	68.70	85.28	**93.00**
5	**100.00**	96.60	99.93	96.30	99.70
6	94.87	86.42	78.45	99.68	**97.56**
7	99.16	92.60	95.86	94.73	**100.00**
8	70.18	**96.31**	94.08	93.36	95.66
9	82.63	**99.89**	99.26	94.93	90.29
OA	88.00	90.58	92.24	91.25	**96.72**
AA	88.61	93.05	91.03	93.28	**96.13**
Kappa	84.32	87.92	89.53	88.62	**95.67**

Tab.3 Experimental results of different algorithms on the Salinas dataset (%)

No.	ELMCK	GFDN	RMGE	BLS	RDBLS
1	100.00	99.40	100.00	100.00	**100.00**
2	99.97	97.98	99.97	99.19	**100.00**
3	100.00	95.37	100.00	100.00	**100.00**
4	97.04	97.54	98.13	**100.00**	99.21
5	**98.50**	90.44	99.40	98.69	97.83
6	**100.00**	96.38	99.55	99.85	99.85
7	99.55	95.60	99.94	99.86	**100.00**
8	79.98	62.82	94.79	91.78	**96.55**
9	100.00	98.50	100.00	100.00	**100.00**
10	93.88	98.07	98.44	70.93	**98.75**
11	99.91	99.34	99.63	97.83	**100.00**
12	91.24	91.29	96.83	95.15	91.60
13	99.01	**99.56**	99.45	98.57	97.68
14	99.81	98.96	94.30	99.25	**99.91**
15	73.93	96.11	95.46	98.65	**99.72**
16	99.33	99.22	100.00	100.00	**100.00**
OA	91.41	89.63	97.64	95.95	**98.69**
AA	95.76	94.79	98.49	96.86	**98.82**
Kappa	90.44	88.47	97.61	95.50	**98.54**

In summary, RDBLS considers the geometry between data samples and discriminative information contained in the data, and it can effectively solve the problem of insufficient learning of BLS and enhance the discrimination ability of BLS.
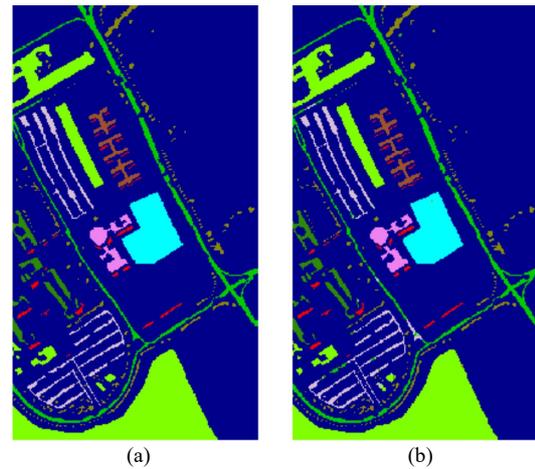
Fig.4 Classification results on the Pavia University dataset: (a) Ground-truth image; (b) RDBLS
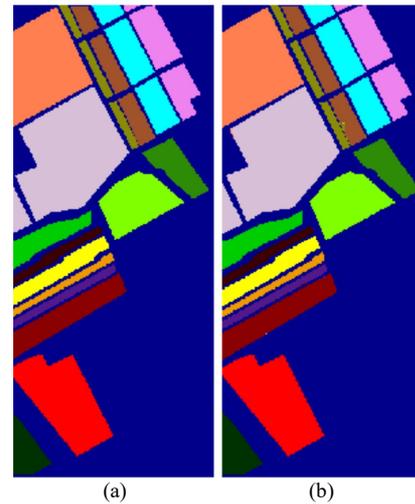
Fig.5 Classification results on the Salinas dataset: (a) Ground-truth image; (b) RDBLS

In this paper, we propose an HSI classification method based on RDBLS. RDBLS takes the connection and difference information of HSI data into account. Therefore, the concept of total scatter is introduced into BLS, which reflects the discriminative information and geometry of input samples. By minimizing the total scatter, the projection direction of the BLS is optimized. In the future work, we will try to use the manifold structure information of data to enhance the discrimination performance of BLS.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

[1]   LEE M A, YAO H, et al. Determining the effects of storage on cotton and soybean leaf samples for hyperspectral

analysis[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2014, 7(6)：2562-2570.

[2]   MAHESH S, JAYAS D S, PALIWAL J, et al. Hyperspectral imaging to classify and monitor quality of agricultural materials[J]. Journal of stored products research, 2015, 61：17-26.

[3]   LIU R, NING X, CAI W, et al. Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification[J]. Mobile information systems, 2021, 2021：1-15.

[4]   MELGANI F, BRUZZONE L. Classification of hyperspectral remote sensing images with support vector machines[J]. IEEE transactions on geoscience and remote sensing, 2004, 42(8)：1778-1790.

[5]   WANG M, GAO K, WANG L, et al. A novel hyperspectral classification method based on C5.0 decision tree of multiple combined classifiers[C]//2012 Fourth International Conference on Computational and Information Sciences, August 17-19, 2012, Chongqing, China. New York：IEEE, 2012：373-376.

[6]   MOOKAMBIGA A, GOMATHI V. Kernel eigenmaps based multiscale sparse model for hyperspectral image classification[J]. Signal processing：image communication, 2021, 99：116416.

[7]   CHEN Y, JIANG H, LI C, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks[J]. IEEE transactions on geoscience and remote sensing, 2016, 54(10)：6232-6251.

[8]   HE M, LI B, CHEN H. Multi-scale 3D deep convolutional neural network for hyperspectral image classification[C]//2017 IEEE International Conference on Image Processing (ICIP), September 17-20, 2017, Beijing, China. New York：IEEE, 2017：3904-3908.

[9]   QIN A, SHANG Z, TIAN J, et al. Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification[J]. IEEE geoscience and remote sensing letters, 2018, 16(2)：241-245.

[10]   CHEN C L, LIU Z. Broad learning system：an effective and efficient incremental learning system without the need for deep architecture[J]. IEEE transactions on neural networks and learning systems, 2017, 29(1)：10-24.

[11]   KONG Y, WANG X, CHENG Y, et al. Hyperspectral imagery classification based on semi-supervised broad learning system[J]. Remote sensing, 2018, 10(5)：685.

[12]   MA Y, LIU Z, CHEN C L P. Multiscale random convolution broad learning system for hyperspectral image classification[J]. IEEE geoscience and remote sensing letters, 2021, 19：1-5.

[13]   ZHAO G, WANG X, KONG Y, et al. Spectral-spatial joint classification of hyperspectral image based on broad learning system[J]. Remote sensing, 2021, 13(4)：583.

[14]   MIAO Y, CHEN M, YUAN Y, et al. Hyperspectral imagery classification via random multigraphs ensemble learning[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2021, 15：641-653.

[15]   ZHOU Y, PENG J, CHEN C L P. Extreme learning machine with composite kernels for hyperspectral image classification[J]. IEEE journal of selected topics in applied earth observations and remote sensing, 2014, 8(6)：2351-2360.

[16]   LI C, LI S, KANG X, et al. Gabor filtering based deep network for hyperspectral image classification[C]// IEEE Geoscience & Remote Sensing Symposium, July 23-28, 2017, Fort Worth, TX, USA. New York：IEEE, 2017：17414009.