Rotation-Based Deep Forest for Hyperspectral Imagery Classification

Xianghai Cao, *Member, IEEE*, Li Wen, Yiming Ge, Jing Zhao, and Licheng Jiao, *Fellow, IEEE*

***Abstract*— In recent years, deep learning methods have been widely used for the classification of hyperspectral images (HSIs). However, the training of deep models is very time-consuming. In addition, the rare labeled samples of remote sensing images also limit the classification performance of deep models. In this letter, a simple deep learning model, a rotation-based deep forest (RBDF), is proposed for the classification of HSIs. Specifically, the output probability of each layer is used as the supplement feature of the next layer. The rotation forest is used to increase the discriminative power of spectral features and neighboring pixels are used to introduce spatial information. The RBDF consumes much less training time than traditional deep models. Experimental results based on three HSIs demonstrate that the proposed method achieves the state-of-the-art classification performance. In addition, the RBDF obtains satisfied classification results with very few training samples.**

***Index Terms*— Classification, deep forest, hyperspectral imagery, rotation forest (ROF).**

I. Introduction

With the rapid development of earth observation technology, the spectral resolution and spatial resolution of hyperspectral images (HSIs) have been significantly improved. Therefore, HSIs have been widely adopted in many fields [1], such as atmospheric science, geology minerals, military reconnaissance, and disguise identification. Hyperspectral remote sensing equipment can receive narrow and continuous spectrum bands with high spectral resolution. Multiband remote sensing images only contain several bands and HSIs often contain tens or even hundreds of bands. Although the abundant spectral information can facilitate the precise land cover types classification, the high dimensionality and high information redundancy also bring great challenges for HSIs processing [2]. Some traditional classifiers have been introduced for the classification of HSIs [3]. In [4], the support vector machine (SVM) is used to combine the spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. In [5], the random forest incorporates bagging of training samples and adaptive random subspace feature selection within a binary hierarchical classifier. With the rapid technique development, the spatial resolution of HSIs is also increased rapidly. The spatial information becomes more and more important for the classification of HSIs. The segmentation method is an effective way to introduce spatial information. Tarabalka *et al.* [6] combine the results of a pixelwise SVM classifier and the segmentation map obtained by partitional clustering using majority voting. In [7], two segmentation methods, fractional-order Darwinian particle swarm optimization and mean shift segmentation, are used for HSIs classification. The fast superpixel segmentation is also widely used for HSIs classification [8], [9]. The Markov random field (MRF) is also an effective algorithm to introduce spatial information. Yuan *et al.* [10] proposed a method that mainly focuses on multitask joint sparse representation and a stepwise MRF framework. In [11], the Gaussian mixture model and MRF are combined to classify the HSIs. In addition, the spectral–spatial feature extraction and spatial filtering are fast and effective spectral–spatial classification method [12], [13]. In [14], the extinction profile (EP) and multiple kernel learning are combined to improve the performance of EP. In [15], the spatial filtering and locality preserving discriminant analysis are combined for HSIs classification.

In recent years, deep learning methods have been successfully used for many tasks and obtain remarkable performance, such as scene segmentation, object detection and localization, face recognition, audio analysis, and so on. In very recent years, the deep learning methods are also introduced for hyperspectral imagery classification. The convolutional neural networks (CNNs) are the most commonly adopted deep models. In [16], deep CNNs are employed to classify HSIs directly in the spectral domain. Chen *et al.* [17] proposed a 3-D CNN-based feature extraction model with combined regularization to extract effective spectral–spatial features of hyperspectral imagery. Makantasis *et al.* [18] exploited a CNN to encode pixels' spectral and spatial information and a multilayer perceptron to conduct the classification task. Zhu *et al.* [19] introduced a deformable convolutional network in which the sampling locations can be adaptively adjusted according to spatial contexts. In the meantime, CNN is utilized to automatically find spatial-related features at high levels. Ghamisi *et al.* [20] proposed a novel framework for the fusion of hyperspectral and light detection and ranging-derived rasterized data using EPs and CNN. In addition to the CNN, the deep autoencoder networks and deep belief network (DBN) are also introduced for the classification of HSIs. In [21], the deep stacked autoencoders (SAEs) are used to

Manuscript received August 28, 2018; revised November 12, 2018; accepted December 25, 2018. Date of publication January 24, 2019; date of current version June 24, 2019. This work is supported by the Fundamental Research Funds for the Central Universities under Grant JB181708. *(Corresponding author: Xianghai Cao.)*

X. Cao, L. Wen, Y. Ge, and L. Jiao are with the School of Artificial Intelligence, Xidian University, Xi'an 710126, China (e-mail: xianghaicao@hotmail.com).

J. Zhao is with the Laboratories and Equipments Department, Xidian University, Xi'an 710126, China.

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

Digital Object Identifier 10.1109/LGRS.2019.2892117

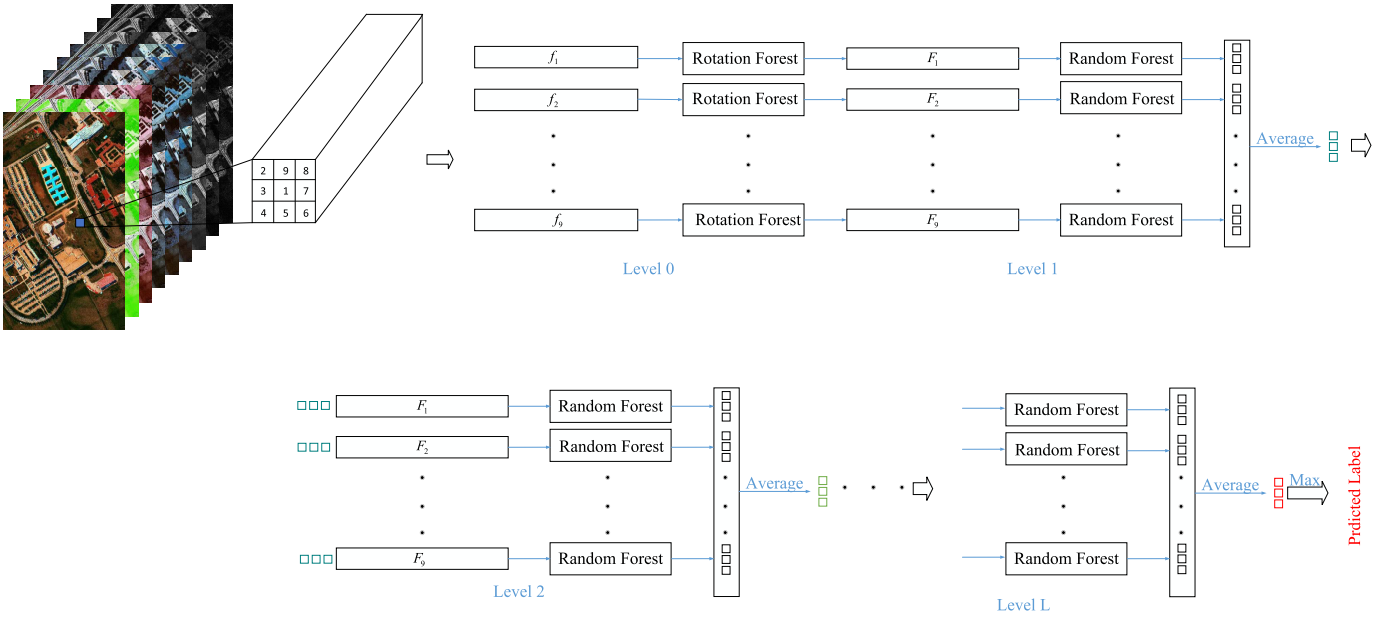

Fig. 1. Structure of RBDF. The 3×3 neighborhood size is adopted. There are nine ROFs in the first level of RBDF.

get useful high-level features. In [22], the segmented SAEs are proposed by confronting the original features into smaller data segments, which are separately processed by different smaller SAEs. In [23], new feature extraction and image classification framework are proposed for hyperspectral data analysis based on DBNs. Zhong *et al.* [24] developed a new diversified DBN through regularizing pretraining and fine-tuning procedures by a diversity-promoting prior over latent factors.

In order to obtain satisfied performance, traditional deep models often need a lot of training samples. However, the labeling task of HSIs is time-consuming and expensive. In addition, the training of traditional deep models will also consume much time. In this letter, inspired by the idea of deep forest [25], a new rotation-based deep forest model (RBDF) is proposed for the classification of HSIs, in which the rotation forest (ROF) [26] is used to improve the discriminative power of spectral feature, and neighboring pixels are used to introduce spatial information. The proposed RBDF can obtain satisfied classification performance with few training samples, besides, the training time is also much less than traditional deep models.

The rest of this letter is organized as follows. In Section II, the basic steps of ROF are first introduced, then the proposed RBDF model is presented. Experimental results and analyses are provided in Section III. Conclusions are provided in Section IV.

II. HSIs Classification With RBDF

In this section, we first briefly introduce the ROF. Then, the proposed RBDF is presented in detail.

A. Rotation Forests

Suppose $\{X, Y\} = \{x_i, y_i\}_{i=1}^{N}$ be the training samples and corresponding labels. F is the features with dimension D. First, the feature F is divided into K feature subsets with dimension M. Then, 75% of the training samples are randomly selected from each feature subset X_j $(j = 1, 2, \ldots, K)$ to form a new training set and a data transformation method (such as principle component analysis) is used to get the coefficients $a_1^j, a_2^j, \ldots, a_M^j$. A sparse rotation matrix R can be obtained as follows:

$$R = \begin{bmatrix} a_1^1, \ldots, a_M^1 & 0 & \ldots & 0 \\ 0 & a_1^2, \ldots, a_M^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_1^K, \ldots, a_M^K \end{bmatrix}.$$

The columns of R are rearranged to R^a that correspond to the order of original spectral features. Then, a new feature can be formed with XR^a. In fact, the above-mentioned steps can be repeated many times and the produced multiple new features can be fused to improve the classification performance of original features.

B. Rotation-Based Deep Forest

The deep forest model [25] consists of many layers and each layer consists of several random forests. This model is easy to use and fast to train. However, in the deep forest model, the output of each layer is only fed into the next layer, the output information is not fully exploited. In order to ensure the maximum information flow between layers in the network, for each layer, the output of all preceding layers is used as inputs, and its own output is used as inputs for all subsequent layers in the proposed RBDF model [27]. In addition, the multigrained scanning in the deep forest model is also not suited for the hyperspectral data processing. Therefore, ROF is adopted to improve the discriminative power of spectral features. The flowchart of the proposed RBDF is illustrated in Fig. 1.

As shown in Fig. 1, the RBDF model consists of three parts. Level 0 is the first part which is used to extract spectral features. In this part, the neighboring pixels of the

TABLE I
CLASSIFICATION ACCURACIES FOR THREE DATA SETS

Class	RF	ROF	ROF+MRF	EPF	3D-CNN	GFDN	RBDF
Pavia University	84.26	91.28	92.92	97.52	97.20	94.70	**99.42**
Salinas	88.10	92.06	94.14	95.72	95.31	94.08	**98.82**
Houston University	79.01	88.96	90.72	93.30	94.13	92.02	**97.29**

training sample are also used for training to introduce the spatial information. Take the nine neighborhoods for example, the spectral feature f_i of each training sample and its neighboring pixels are fed into an ROF. Each ROF will produce three rotation matrices and form three rotation features with the same dimension with the original spectral feature. Therefore, F_i includes three rotation feature vectors. The middle levels $2, 3, \ldots, L-1$ constitute the second part of the RBDF model. In this part, each rotation feature vector is fed into a random forest and the classification probability vector is produced with threefold cross validation to avoid the overfitting for the training samples. In fact, the classification probability vector contains the guidance information for the wrong classified samples which can be used together with the original input to adjust the classification results in the next levels. All the outputted probability vectors will be averaged to obtain a robust estimation. The averaged probability vector and the rotation feature are stacked together to form the input vector for the next level. It needs to note again, the averaged probability vector will be used for all subsequent levels. Level L is the last part of the RBDF model. After obtaining the averaged classification probability vector, the predicted label is obtained by finding the maximum probability value.

Different from the conventional end-to-end training strategy, the RBDF is trained level by level and does not need to compute the gradients. Therefore, the training speed of RBDF is much faster than traditional deep models. The random forest in the RBDF can also be replaced with any other classifiers. The random forest is chosen because it involves few parameters and robust to the parameters setting. The proposed RBDF is constructed based on the random forest. Similar to the random forest, the RBDF involves very few parameters. In addition, the RBDF does not need an end-to-end training to tune the parameters. Therefore, the RBDF is robust to the number of training samples.

III. EXPERIMENTS

A. Experimental Data Sets and Settings

In this letter, three widely used hyperspectral data sets are used to validate the proposed method: Pavia University, Salinas, and Houston University.

1) Pavia University Scene: The Pavia University scene was gathered by the reflective optics system imaging spectrometer during the Pavia flight in northern Italy. The HSI size is 610×340 with the spatial resolution of 1.3 m. In this experiment, 12 noisy channels were removed, and the remaining 103 channels are used for analysis. This hyperspectral data set has 42 776 labeled samples and 9 categories.

2) Salinas Scene: This scene was collected by the AVIRIS sensor in the Salinas Valley, California, with a spatial resolution of 3.7 m. The data have 512×217 pixels and 224 spectral channels. It mainly consists of vegetables, bare soil, and vineyards. As with the Indian pines scene, 20 noisy bands covering the region of water absorption were discarded and the remaining 204 bands were used in this letter.

3) Houston University Scene: The Houston University image size is 349×1905 with the spatial resolution of 2.5 m. It contains 144 spectral channels ranging from 380 to 1050 nm. Originally, there are 15 categories in Houston University image, and we remove the categories whose samples are less than 400, and now there are 14 categories for experimentation.

For the three hyperspectral data sets, 5% labeled samples of each category are randomly selected as the training samples, and the remaining 95% samples are used for test. For the proposed RBDF, M is set as 10 and principal component analysis (PCA) is used as the transformation in the first part. The neighborhood size is 9×9 and each random forest includes five trees.

B. HSIs Classification

The overall classification accuracy is used to evaluate the classification performance in the experiments. Six methods are introduced for performance comparison as follows.

1) The classic random forest classifier with spectral information only (RF) and the number of trees is 5.
2) The ROF classifier with PCA transformation [26]. The number of ensemble classifiers is set as 10, M is 10, and each RF includes five trees.
3) An improved rotation model with MRFs (ROF + MRF), the spatial neighbor's size τ is 12, and the regularization parameter μ is fixed to be 4.
4) A spatial filtering method that filters the classification probability maps to introduce the spatial information edge-preserving filtering (EPF). The local window size parameter δ_s is 3 and the intensity difference weight δ_r is 0.2.
5) A 3-D-CNN in which the 3-D data cube is used as the input of deep CNN [17] (3-D-CNN) and 9×9 neighborhood pixels are fed into the network.
6) A Gabor filtering-based deep model in which the Gabor features and autoencoder deep network (GFDN) are combined to classify the hyperspectral imagery [28]. The size of Gabor filters is 3×3 and the neighborhood size of input pixel is 9×9.

The classification results for Pavia University, Salinas, and Houston University are listed in Table I.

1) Classification Results of Pavia University Scene: Table I lists the classification accuracies for different methods. From which we can see that ROF obtains much higher accuracy than the RF; this result testifies that the rotation transformation does improve the discriminative power of the spectral features effectively. By introducing the spatial information with

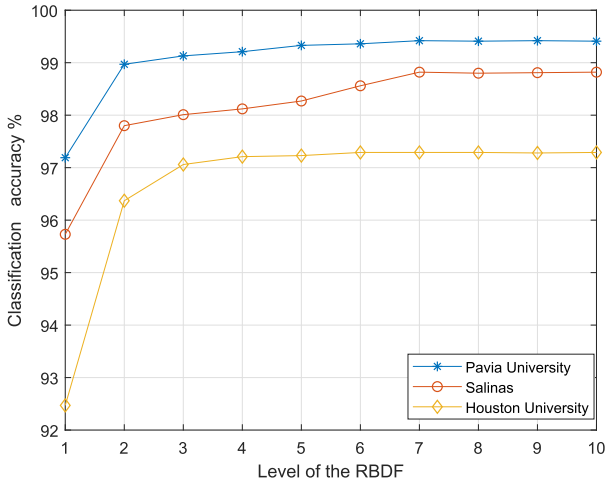

Fig. 2. Classification accuracy of RBDF with different levels.

MRF, the ROF + MRF method improves the classification further. The EPF method obtains higher accuracy than other methods except the proposed RBDF. The reason is that edge-preserving filtering can remove the pepper noise effectively and prevent the edge ambiguity between different land cover types. For two conventional deep models, 3-D-CNN obtains a higher classification than GFDN. Their overall classification accuracies are 97.20% and 94.70%, respectively. The proposed RBDF obtains the highest overall accuracy with 99.42%. This result indicates that RBDF is an effective deep model for hyperspectral imagery classification.

2) Classification Results of Salinas Scene: As shown in Table I, the classification accuracies of RF and ROF are 88.10% and 92.06%, respectively; this result testifies again that the ROF is an effective way to promote the classification performance of spectral features. With the MRF, the classification accuracy of ROF + MRF is promoted to 94.14%. The EPF still obtains a satisfied performance with 95.72%. Without enough training samples, the performance of 3-D-CNN and GFDN is also not satisfactory and the classification accuracies are 95.31% and 94.08%, respectively. With the extracted rotation features and densely connected structure, the proposed RBDF still obtains the highest accuracy with 98.82%.

3) Classification Results of Houston University Scene: For these data, the RF obtains a much lower accuracy than other methods. The ROF still improves the performance of RF effectively. The classification accuracies of two conventional deep models are 94.13% and 92.02%, which are far from satisfactory. The RBDF achieves the best performance with 97.29%.

C. Parameters Analysis

1) Analysis of Level L: Similar to the traditional deep models, the multilayer structure is the key for the performance improvement of the RBDF. The experimental results of Pavia University, Salinas, and Houston University with different levels are shown in Fig. 2. It can be observed that when the output of the first level is inputted into the second level, the classification accuracy is increased greatly. For the Pavia University, the accuracy is promoted from 97.2% to 99.0%. For the Salinas scene, the similar improvement is observed. The accuracy is promoted from 95.7% to 97.8%. A much higher improvement can be observed from Houston University, which is improved from 92.47% to 96.37%. With the increase in levels, the classification accuracy is also promoted steadily. However, when the level number exceeds 7, the classification accuracy reaches a stable value. For the Pavia University, this value is about 99.4%. For the Salinas scene, this value is about 98.8%. For the Houston University, the value is 97.29%. This result means that when there are too many levels in the RBDF, the output of the last several levels cannot provide useful information for classification anymore. Therefore, in other experiments, the level number is set as 7.

TABLE II
CLASSIFICATION ACCURACIES OF RBDF WITH DIFFERENT NEIGHBORHOOD SIZES

Data	1×1	3×3	5×5	7×7	9×9
Pavia University	92.74	97.16	98.40	99.14	99.42
Salinas	91.50	95.71	97.09	98.23	98.82
Houston University	88.43	93.78	95.13	96.68	97.29

2) Analysis of Neighborhood Size: For HSIs, the neighbor pixels often share the same labels. Therefore, in the proposed RBDF, the local spatial information is introduced by using the neighboring pixels. The neighborhood size is an important parameter which has a great influence on the classification accuracy. Table II lists the classification results of RBDF with different neighborhood sizes. As shown in Table II, with the size of the neighborhood increases, the classification accuracy will also increase. Take the Salinas scene for example, the accuracies for 3×3 and 9×9 are 95.71% and 98.82%, respectively. However, when the neighborhood size exceeds 9×9, the classification accuracy has only a very slight increase and the computation complexity is greatly increased. Therefore, the neighborhood size is set as 9×9 in other experiments. When the neighborhood size is 1×1, for the three data sets, the accuracies of RBDF are 92.74%, 91.50%, and 88.43%. For ROF, the accuracies are 91.28%, 92.06%, and 88.96%. However, it needs to note that the ROF includes ten rotation matrices and the RBDF only includes three rotation matrices. If the ROF also includes three rotation matrices, then its classification accuracies reduce to 90.89%, 90.56%, and 87.25%. This means that the RBDF can still improve the classification performance of ROF even when there is no spatial information involved.

3) Analysis of Number of Training Samples: Different from the traditional deep models, the RBDF can obtain satisfied performance with few training samples. The experimental results with different training sizes are listed in Table III. With 1% training samples, the RBDF obtains 97.73%, 96.90%, and 88.01% accuracies for three data sets. For 3-D-CNN, the accuracies are 94.78%, 85.13%, and 81.23%. For GFDN, the accuracies are 92.32%, 84.21%, and 78.56%.

TABLE III

CLASSIFICATION ACCURACIES OF RBDF WITH DIFFERENT TRAINING SIZES

Dataset	Methods	1%	2%	5%	10%
Pavia University	3D-CNN	94.78	96.03	97.20	97.86
	GFDN	92.32	94.01	94.70	95.14
	RBDF	97.73	98.91	99.42	99.52
Salinas	3D-CNN	85.13	89.45	95.31	96.89
	GFDN	84.21	88.01	94.08	95.27
	RBDF	96.90	97.57	98.82	99.09
Houston University	3D-CNN	81.23	84.17	94.13	95.65
	GFDN	78.56	80.23	92.02	94.23
	RBDF	88.01	90.23	97.29	98.60

TABLE IV

TRAINING TIME (SECONDS) OF DIFFERENT DEEP METHODS

Data	3D-CNN	GFDN	RBDF
Pavia University	9310	3165	693
Salinas	21340	8320	2211
Houston University	6510	1935	430

The training time of three deep models is listed in Table IV. Without the time-consuming backpropagation algorithm, the RBDF needs much less training time when compared with conventional deep learning methods.

IV. CONCLUSION

For conventional deep learning models, there are a lot of parameters that need to be tuned, so they require abundant training samples and much training time to get satisfied performance. In this letter, a new deep learning model, RBDF, is proposed. With the ROF, the spectral feature is enhanced effectively. With the densely connected structure, the output information is also fully exploited. The RBDF obtains competitive performance with few training samples. In future works, different classifiers will be introduced into RBDF to get better performance.

REFERENCES

[1] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.

[2] X. Cao, T. Xiong, and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 329–333, Mar. 2016.

[3] X. Cao, J. Han, S. Yang, D. Tao, and L. Jiao, "Band selection and evaluation with spatial information," *Int. J. Remote Sens.*, vol. 37, no. 19, pp. 4501–4520, 2016.

[4] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

[5] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[6] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[7] P. Ghamisi, M. S. Couceiro, M. Fauvel, and J. A. Benediktsson, "Integration of segmentation techniques for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 342–346, Jan. 2014.

[8] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

[9] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[10] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2016.

[11] W. Li, S. Prasad, and J. E. Fowler, "Hyperspectral image classification using Gaussian mixture models and Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 153–157, Jan. 2014.

[12] 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.

[13] X. Cao, B. Ji, Y. Ji, L. Wang, and L. Jiao, "Hyperspectral image classification based on filtering: A comparative study," *Proc. SPIE*, vol. 11, no. 3, p. 035007, 2017.

[14] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.

[15] M. Han, C. Zhang, and J. Wang, "Spectral-spatial classification of hyperspectral image based on locality preserving discriminant analysis," in *Proc. Int. Symp. Neural Netw.*, 2016, pp. 21–29.

[16] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, Jan. 2015, Art. no. 258619.

[17] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[18] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 4959–4962.

[19] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018.

[20] P. Ghamisi, B. Höfle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[21] 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.

[22] J. Zabalza *et al.*, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1–10, Apr. 2016.

[23] 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.

[24] P. Zhong, Z. Gong, S. Li, and C.-B. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3516–3530, Jun. 2017.

[25] Z.-H. Zhou and J. Feng. (2017). "Deep forest: Towards an alternative to deep neural networks." [Online]. Available: https://arxiv.org/abs/1702.08835

[26] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 239–243, Jan. 2014.

[27] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.

[28] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166–1178, Apr. 2018.