Deep Feature Aggregation Network for Hyperspectral Remote Sensing Image Classification

Chunju Zhang, Guandong Li, Runmin Lei, Shihong Du, Xueying Zhang, Hui Zheng, and Zhaofu Wu

***Abstract*—Hyperspectral remote sensing images (HSIs) are rich in spectral–spatial information. The deep learning models can help to automatically extract and discover this rich information from HSIs for classifying HSIs. However, the sampling of the models and the design of the hyperparameters depend on the number of samples and the size of each sample's input space. In the case of limited samples, the description dimension of features is also limited and overfitting to other remote sensing image datasets is evident. This study proposes a novel deep feature aggregation network for HSI classification based on a 3-D convolutional neural network from the perspective of feature aggregation patterns. By introducing the residual learning and dense connectivity strategies, we established a deep feature residual network (DFRN) and a deep feature dense network (DFDN) to exploit the low-, middle-, and high-level features in HSIs. For the Indian Pines and Kennedy Space Center datasets, the DFRN model was determined to be more accurate. On the Pavia University dataset, both the DFDN and DFRN have basically the same accuracy, but the DFDN has faster convergence speed and more stable performance on the validation set than the DFRN. Therefore, when faced with different HSI data, the corresponding aggregation method can be chosen more flexibly according to the requirements on number of training samples and the convergence speed. This is beneficial in the HSI classification.**

***Index Terms*—Dense connectivity, feature fusion, hyperspectral image classification, residual learning, 3-D convolutional neural network (3D-CNN).**

I. Introduction

HYPERSPECTRAL remote sensing images (HSIs) are rich in spectral–spatial information, which has important applications in land use, resource investigation, natural disasters, global environment, interstellar exploration, etc. Improving the classification accuracy of HSIs has become a hot topic in the field of remote sensing [1]–[4]. Existing studies often use a support vector machine (SVM) [5], neural network [6], multinomial logistic regression [7], and other methods to construct a pixel-based classifier to resolve the HSI classification. In general, although these methods utilize the spectral information, they do not consider spatial information, and the classification maps often contain noises. Recently, several spectral–spatial feature-based classification frameworks have been proposed to consider spatial information in pixel-based classifiers. For example, Benediktsson *et al.* [8] utilize multiple morphological operations to construct the spectral–spatial features of HSIs. Multiple kernel learning based on spectral–spatial information (e.g., composite kernel [9] and morphological kernel [10]) was designed to improve the SVM classifier. Su *et al.* [11] proposed the joint collaborative representation classification with correlation matrix for HSIs, which could keep the local structural information in HSIs. To further improve HSIs classification accuracy, Su *et al.* [12] integrated ensemble learning and tangent space collaborative representation classification. However, due to the high-dimensional characteristic of the signal information redundancy and several uncertainties, the HSI data structure is highly nonlinear to some classification models rooted in statistical pattern recognition, making it difficult to classify the original hyperspectral data directly [1], [4]. When the number of training samples involved in supervised learning is limited, Hughes phenomenon in which the classification accuracy decreases with an increase in the feature dimension is present [13].

Deep learning methods, such as the stacked autoencoder (SAE) [14] and deep brief network [15], can automatically extract abstract features from the bottom to the high level of semantics and convert images into more easily recognizable features. They propose the use of a 2D-CNN to extract spatial features and use of the principal component analysis to reduce the dimension of HSIs, and finally, use spectral–spatial features to improve the classification accuracy [16]–[18]. However, these methods extract spatial and spectral features separately and require complex preprocessing. To make full use of spectral–spatial information in HSIs, Chen *et al.* [19] use a 3-D convolutional neural network (3D-CNN) to directly extract features from the original image, taking 3-D cube data as 3D-CNN input in a small space size, showing a good classification performance. Based on the 3D-CNN, a spectral–spatial residuals network (SSRNs) is proposed in [20], in which the spatial and spectral residual modules are designed to extract the spatial and spectral

Manuscript received April 26, 2020; revised July 14, 2020 and August 20, 2020; accepted August 22, 2020. Date of publication September 1, 2020; date of current version September 17, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41401451 and Grant 4167145, and the Open Fund of Key Laboratory of the Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education under Grant GTYR201907. *(Corresponding author: Shihong Du.)*

Chunju Zhang is with the School of Civil Engineering and the Intelligent Interconnected Systems Laboratory of Anhui Province, Hefei University of Technology, Hefei 230009, China (e-mail: zcjtwz@sina.com).

Guandong Li, Runmin Lei, and Zhaofu Wu are with the School of Civil Engineering, Hefei University of Technology, Hefei 230009, China (e-mail: leeguandon@gmail.com; 893259405@qq.com; wzfhh@163.com).

Shihong Du is with the Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China (e-mail: dshgis@hotmail.com).

Xueying Zhang is with the Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing 210023, China (e-mail: zhangsnowy@163.com).

Hui Zheng is with the Key Laboratory of Geospatial Technology for the Middle and Yellow River Region (Henan University), Ministry of Education, Kaifeng 475004, China (e-mail: zhenghui842@163.com).

Digital Object Identifier 10.1109/JSTARS.2020.3020733

information, respectively. However, due to the relatively shallow network used in the HSI classification, including only a few convolution layers [19], [21], deep features cannot be extracted effectively. In addition, considering that the 3D-CNN network is composed of multiple layers, strong complementary relational information between the different layers has not been fully utilized in previous work. Song *et al.* [22] introduced residual learning, designing the feature fusion between multiple layers of information, to extract more discriminative features. However, its main drawback is that the optimal feature fusion mechanism depends on a hand-crafted setting with abundant experiments [2], and the convergence speed is too slow. Moreover, in the face of the diversity of hyperspectral data, the frequent adjustment of the network structure is a time-consuming and laborious measure and the performance of migration to other HSI datasets is not as expected; thus, the accurate classification of geographic objects cannot be achieved.

This study proposes a novel deep feature aggregation network (DFAN) for HSI classification based on a 3D-CNN from the perspective of feature aggregation patterns. We introduced residual learning [23] and dense connectivity [24], respectively, where residual learning aggregates features by summation and dense connectivity aggregates them by concatenation. Both residual learning and dense connectivity have increased the depth of the network and enhanced the flow of information. In addition, considering that different layers can extract features at different levels and provide complementary information, a fusion mechanism is required to utilize the features at the multiple layers. Aiming at resolving this issue, we proposed two kinds of DFANs, the deep feature residual network (DFRN), and the deep feature dense network (DFDN). When faced with different HSI data, the corresponding aggregation method can be designed more flexibly. This study provides a direction for HSI classification.

II. Aggregation View

In this article, we proposed aggregation functions, including summation and concatenation operations. In our structure, a series of units are defined to make nonlinear transformations of feature information, including convolution layers, batch normalization (BN) [25], Relu [26], and pooling layers. The blocks in both the DFRN and the DFDN are based on continuous BN, Relu, and convolution layers (composite function). They obtain the output from the aggregation function and continue to transfer to the next aggregation function. In the final feature fusion, the DFRN uses summation to fuse the features of the three blocks, whereas the DFDN uses concatenation to aggregate the features of the three blocks.

A. Connections in ResNet and DenseNet

The skip connection in ResNet ensures that the gradient can be continuously transmitted through each residual block, which consists of nonlinear units and a shortcut identity mapping. The dense connection mode is a core connection in DenseNet. Compared with ResNet, the dense connection not only connects the next layer but also directly implements the cross-layer connection. The gradient obtained from each layer is the gradient addition of the previous layers for both DenseNet and ResNet.

We unravel the view [24], [27] and carefully examine the connection manner in DenseNet and ResNet for HSI classification. In our design, the blocks in both the DFRN and the DFDN are set to 3; that is, each block contains multiple composite functions with each function consisting of BN–Relu–Conv–BN–Relu–Conv. This means that there are three residual connections in the residual block and three dense connections in the dense block. The number and distribution of the end-to-end paths and the nonlinear units in the dense connection structure are same with that in the residual connection structure. It is worth noting that any path in the dense connection structure cannot continuously skip connections. When a feature map is concatenated with others after going through a skip connection, it must be immediately forwarded into the following basic nonlinear unit instead of taking another skip connection [28]. In the residual network, the feature maps from the previous layer are summarized by the input composite function and, then, transferred to the next composite function after fusion. In the first skip connection of each block, we set up a $1 \times 1 \times 1$ convolution to ensure that the input and output are consistent. The two connection modes are shown in Fig. 1.

B. ResNet and DenseNet Under Aggregation View

Inspired by deeper aggregation, we used the residual and dense structures in the HSI classification network. In the DFRN, we use $\oplus$ to represent the summation operation as

$$x_{l+1} = y_1 \oplus y_{l-1} \oplus \ldots \oplus y_1. \tag{1}$$

In the DFDN, we use concatenation $\otimes$ instead of $\oplus$, as

$$x_{l+1} = y_1 \otimes y_{l-1} \otimes \ldots \otimes y_1. \tag{2}$$

These two equations describe similarities in the aggregation of features in the DFRN and the DFDN. The output of each basic unit is the result of the nonlinear transformation of the feature aggregation of the previous unit. Both DenseNet and ResNet aggregate features from the previous basic units differently. The ResNet aggregates features by summation, whereas the DenseNet aggregates features by concatenation.

DenseNet and ResNet train the networks with more than 100 layers due to feature aggregation. In our HSIs, both the DFRN and the DFDN are relatively deep networks; thus, feature aggregation can create a large number of shortcut connections [27], [29]. It helps to not only enhance the ability of learning deep features, but also alleviate the gradient disappearance and explosion problems. From the forward propagation neural network perspective, feature aggregation enables features to be extracted at each layer without the influence of intermediate nonlinear transformation units.

The form of aggregation is also diversified, including the summation of ResNet and the concatenation of DenseNet. The summation aggregation ensures that each layer feature is not directly read by subsequent layers. On the other hand, the loss of information during transmission increases with the increase of the number of network layers. The concatenation aggregation method can effectively pass each layer of information to all its sublayers; as a result, we believe this model has the potential to learn the best combination of input feature maps.

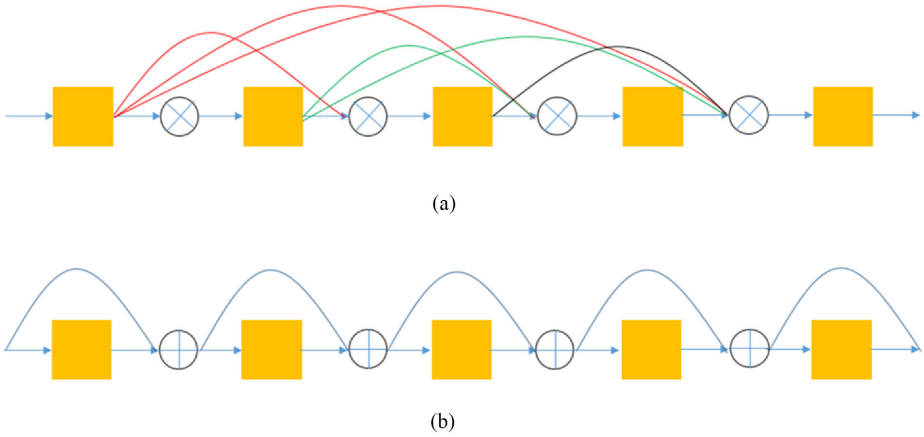

Fig. 1. Two connection modes. (a) Dense connectivity. (b) Residual connection.

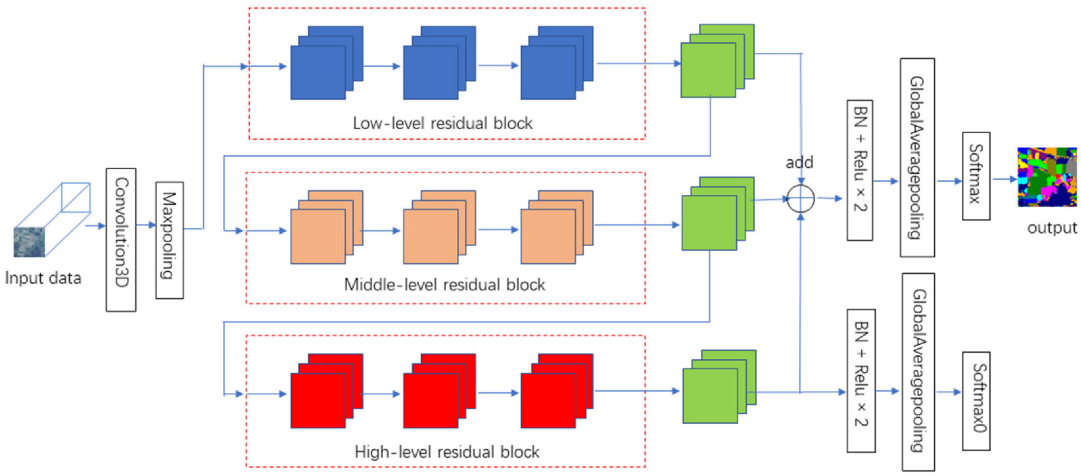

Fig. 2. Network structure of the DFRN.

III. DFAN for HSI Classification

The spatial features of HSIs are relatively sparse, and there is a large amount of redundant information in the spectrum. Shallow-layer networks with several convolutional layers cannot extract deep features effectively, leading to the poor generalization. The performance of migration to other HSIs is not as expected and accurate classification of geographic objects cannot be achieved. However, without the corresponding aggregation methods, only stacking the 3D-CNN structure often results in several problems such as the risk of overfitting [30] the model. Therefore, from the perspective of aggregation view, this study designs two deep 3D-CNN models based on different aggregation methods for considering the strong complementary information between the traditional block of 3D-CNNs. In addition, we also built a deep network to extract more discriminating features of HSIs and adopted a fusion mechanism to make full use of the network features. Three consecutive residual and dense blocks are considered with no transition layer among them for avoiding feature information loss. The features of the three layers are not only transferred to the next block, but are also fused at the end of the block.

A. Deep Feature Residual Network

The network structure of the DFRN is shown in Fig. 2. The input 3-D data cube of HSIs, followed by a $3 \times 3 \times 3$ 3-D convolution (64 filters), BN and Relu, is connected to the residual block. All the structures include three parts with 16, 32, and 64 filters, respectively. The three blocks excel in extracting low-, middle-, and high-level HSI features.

The proposed network consists of three residual blocks with each containing multiple composite functions (see Fig. 3). The composite functions are essentially the same for the residual and dense connections, including BN–Relu–$3 \times 3 \times 3$Conv–BN–Relu-$3 \times 3 \times 3$Conv [31]. BN makes a normalization operation and the Relu layer generates a nonlinear operation and increases the complexity of the neural network. To avoid gradient dispersion and explosion, BN normalizes the input of the upper layer, which prevents the feature from being either too

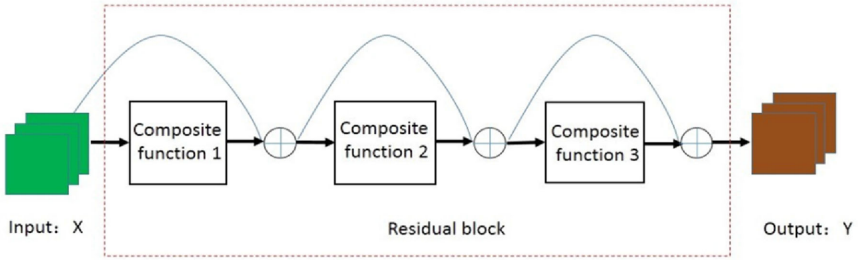

Fig. 3. Residual block.

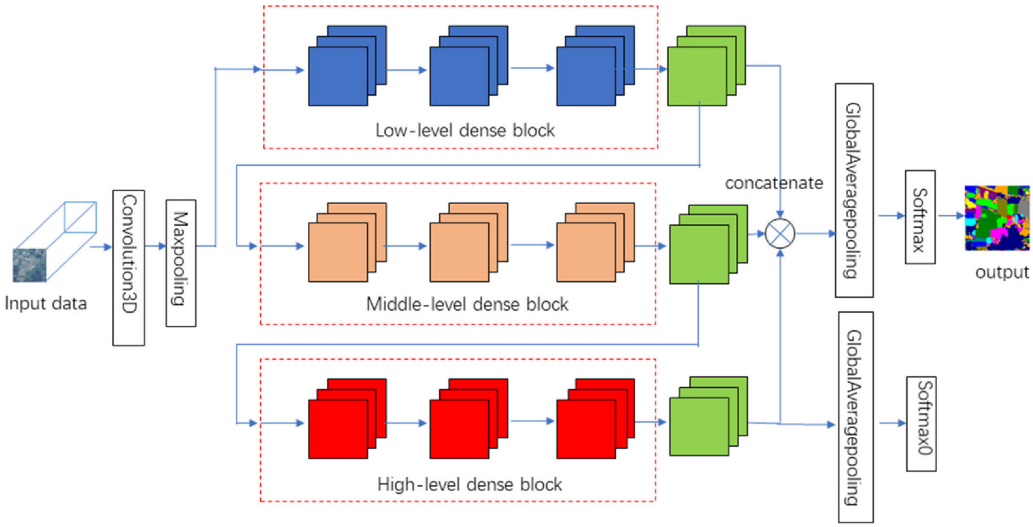

Fig. 4. Network structure of the DFDN.

large or too small in the input field of the activation function. The skip connection between the composite function is used for information transfer.

Considering that the numbers of feature maps produced by different blocks are not matched very well, we introduce the feature fusion mechanism to explore the complementary information implied in each block. The numbers of feature channels are 16, 32, and 64 in the low-, middle-, and high-level residual blocks, respectively. We used a $1 \times 1 \times 1$ 3-D convolution to match the feature maps of the first and second blocks by turning the number of feature maps for aggregation from 16 and 32 to 64, respectively. After feature aggregation, two sets of BN + Relu are used to accelerate the network regularization. Finally, a GlobalAveragePooling3D is used to send information to the classifier. We use the softmax classifier and RMSprop optimizer [32] to optimize the loss function.

B. Deep Feature Dense Network

The network structure of the DFDN is shown in Fig. 4. The input 3-D data cube of HSIs, followed by a $3 \times 3 \times 3$ 3-D convolution (64 filters) and max-pooling, is connected to the dense block. All the structures include three parts, in which three blocks are used to extract low-, middle-, and high-level HSI features.

The DFDN consists of three dense blocks with each block consisting of multiple composite functions (see Fig. 5). Although the composite function in the dense structure also contains two 3-D convolutions, the first convolution is used as the bottleneck layer ($1 \times 1 \times 1$ 3-D convolution with 128 filters) to reduce the parameters of the model, whereas the second $3 \times 3 \times 3$ convolution (32 filters) is used to extract the feature map. Dense connectivity is used in the three dense blocks and concatenation is used for aggregation between each function.

Like the DFRN, we also introduced a feature fusion mechanism into the DFDN to make use of complementary information after each block for deep feature extraction. In the DFDN, concatenation aggregation is used and it is not necessary to consider the matching of the number of feature channels. After feature aggregation, a GlobalAveragePooling3D is used to send information to the classifier. We use the softmax classifier and RMSprop optimizer to optimize the loss function.

Inspired that deep features are more discriminative, we add an auxiliary classifier after the high-level block for the DFDN and the DFRN to help training the model, which could enhance the importance of deep features in the final fusion features. The total loss of the DFDN and the DFRN is a sum of the loss of the final classifier and the auxiliary classifier, where α is the coefficient of loss of auxiliary classifier, as

$$\text{Loss}_{\text{total}} = \text{Loss} + \alpha * 0.5\, \text{Loss}_{\text{auxiliary}}. \tag{3}$$

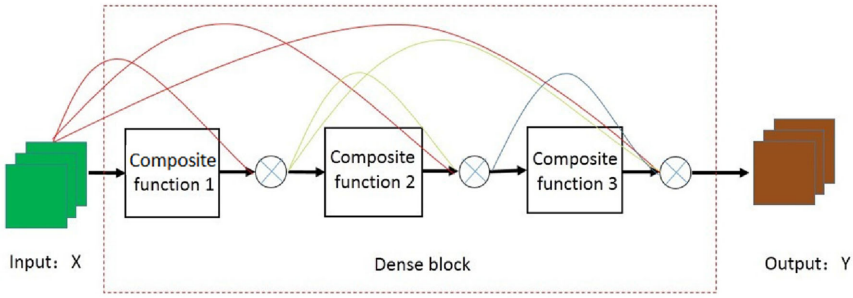

Fig. 5. Dense block.

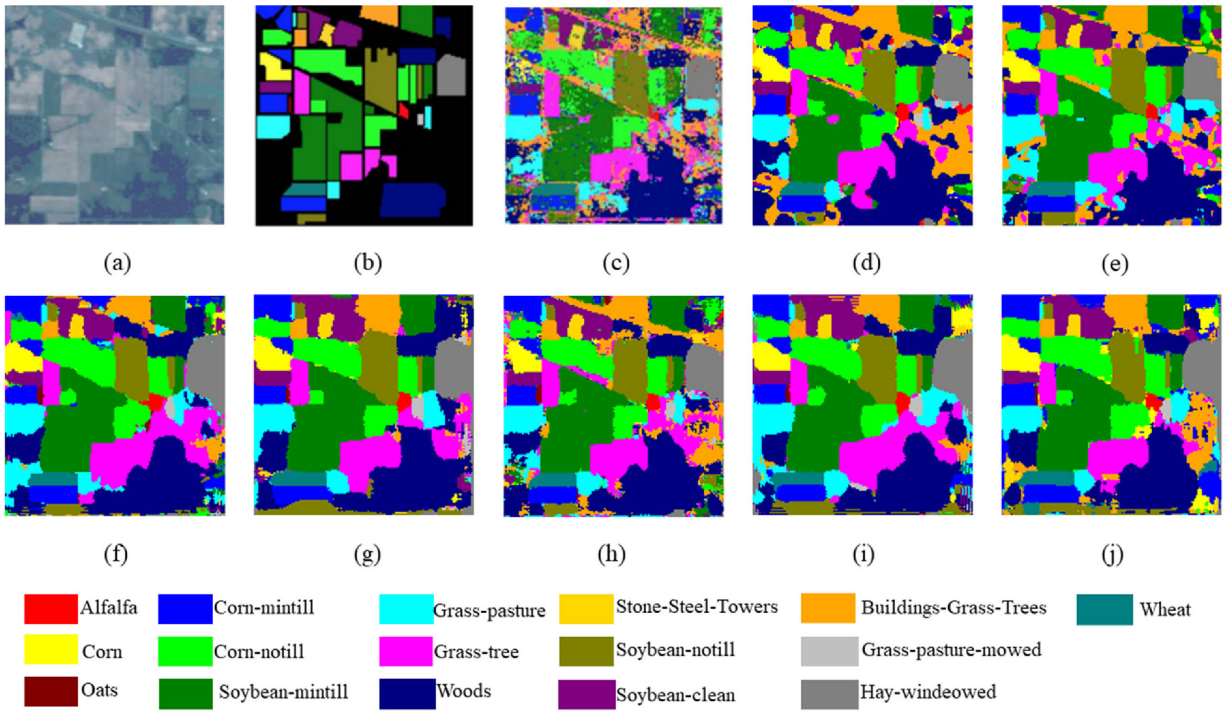

Fig. 6. Classification results of the Indian Pines dataset. (a) False color image. (b) Ground-truth labels. Results of (c) SAE, (d) 3D-CNN, (e) SSRN, (f) 3D-DenseNet, (g) 3D-ResNet, (h) MSDN, (i) DFDN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7), and (j) DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7).

IV. Experiments and Results

To evaluate the performance of the DFAN model, this article introduced three representative HSI datasets, namely, the Indian Pines, the Pavia University, and Kennedy Space Center datasets. All sampled data were divided into three groups, namely the training set, validation set, and test set [31]. During the test, we integrated the best-retained models and calculated the mean and standard deviation of the multiple sets of overall accuracy (OA), mean accuracy (AA), and kappa coefficient (k) [33]. The input data for the HSI datasets were normalized to unit variance. For all the datasets, we trained 200 epochs and the batch size was 16. We conducted experiments for the DFDN and the DFRN in each dataset, hoping to select the optimal learning rate for the two models from {0.01, 0.03, 0.05, 0.001, 0.003, 0.005, 0.0001, 0.0003, 0.0005}. According to the results, the learning rate was set to 0.0003 for the DFDN and 0.0005 for the DFRN. For the coefficient of loss of auxiliary classifier, α is set to 0.5 after experiments.

The experimental hardware platform was a desktop computer with the CPU of Intel i5-8500k, the GPU of GTX1080Ti, and the memory of 16 GB. We discussed the influences of the ratios of the training dataset, the neighboring pixel block size, and the number of composite functions and selected the optimal parameter set for the Indian Pines, Pavia University, and Kennedy Space Center datasets, respectively.

A. Experimental Datasets

The Indian Pines dataset was obtained by the AVIRIS spectral imager in northwestern Indiana in 1992. This dataset contains 145×145 pixels with 16 classes. The Pavia University dataset was collected by the ROSIS sensor over the Pavia region of northern Italy in 2001, and the data size was 610×340 with

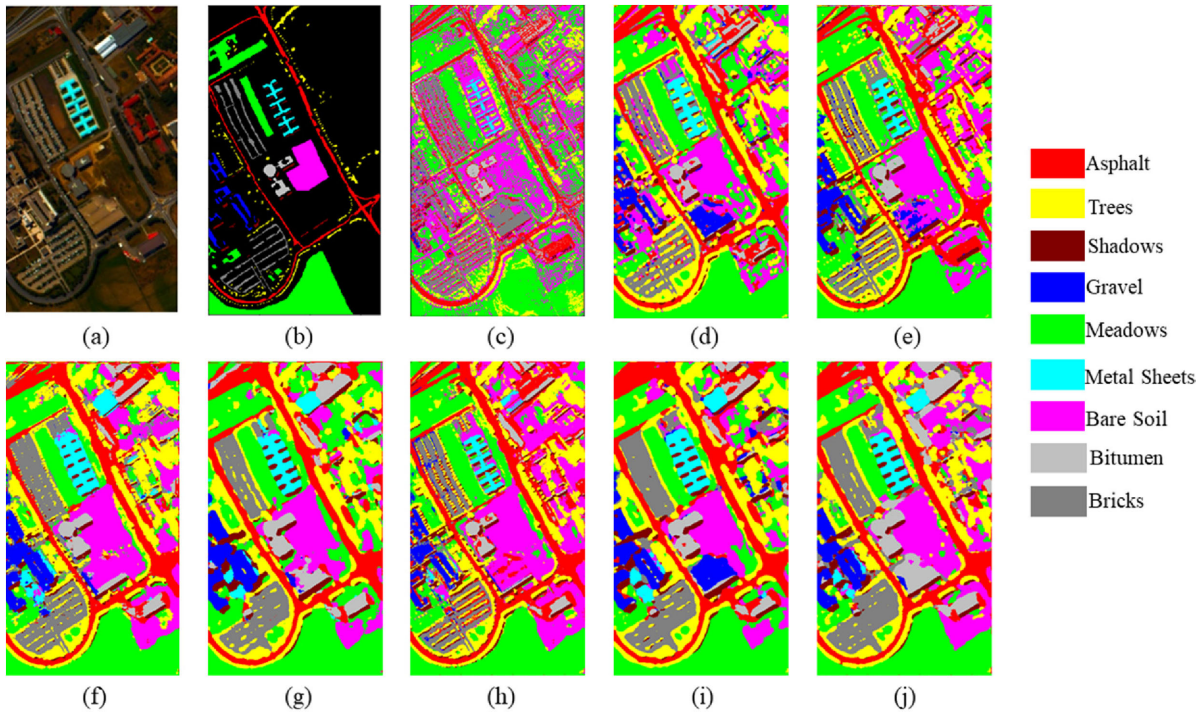

Fig. 7. Classification results of the Pavia University dataset. (a) False-color image. (b) Ground-truth labels. Results of (c) SAE, (d) 3D-CNN, (e) SSRN, (f) 3D-DenseNet, (g) 3D-ResNet, (h) MSDN, (i) DFDN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8), and (j) DFRN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8).

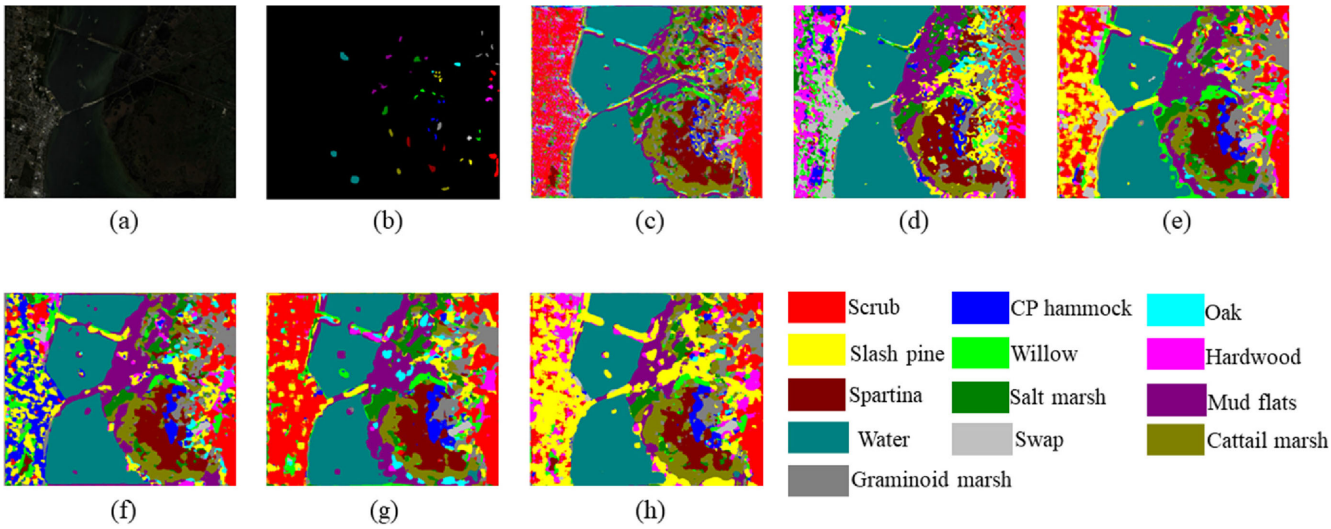

Fig. 8. Classification results of the Kennedy Space Center dataset. (a) False color image. (b) Ground-truth labels. Results of (c) MSDN, (d) FDSSC, (e) DFDN ($M \times N = 17 \times 17$; $c = 3$; ratio = 1:1:8), (f) DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 1:1:8), (g) DFDN ($M \times N = 17 \times 17$; $c = 3$; ratio = 2:1:7), and (h) DFRN($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7).

nine kinds of ground cover. The Kennedy Space Center dataset was acquired by the AVIRIS instrument in Florida in 1996. It contains 512 × 614 pixels with 176 bands and 13 categories.

B. Ratios of Training Datasets in the DFAN

The DFAN is relatively sensitive to the sample size and training set. Therefore, the model performances under different ratios of training, validation, and test sets are discussed (Table I). The numbers of composite functions in the DFRN and the DFDN models are 3 and the neighboring pixel block sizes are 11. Finally, the DFDN achieves the highest accuracy with a ratio of 5:1:4 of training, validation, and test sets, whereas the DFRN chose a ratio of training, validation, and test sets is 4:1:5. For the dense connection structure, due to the continuous reuse feature, feature utilization is high and can achieve high accuracy.

TABLE I
TRAINING TIME, TEST TIME, AND OA UNDER DIFFERENT TRAINING DATASET RATIOS ON THE INDIAN PINES DATASET FOR DFDN AND DFRN

Ratios	DFDN			DFRN		
	Training time (s)	Test time (s)	OA (%)	Training time (s)	Test time (s)	OA (%)
1:1:8	2326.39	25.88	94.40	3365.26	33.95	94.24
2:1:7	4050.70	22.56	98.51	5859.62	30.14	98.84
3:1:6	5777.66	19.41	99.43	8351.96	25.84	99.43
4:1:5	7504.27	16.36	99.57	10851.72	21.64	99.74
5:1:4	9227.60	13.62	99.80	13330.61	17.78	99.58

TABLE II
TRAINING TIME, TEST TIME, AND OA FOR DIFFERENT NEIGHBORING PIXEL BLOCK SIZES ON THE INDIAN PINES DATASET FOR DFDN AND DFRN

Neighboring pixel block size（M=N）	DFDN			DFRN		
	Training time (s)	Test time (s)	OA (%)	Training time (s)	Test time (s)	OA (%)
9	2834.48	15.26	97.54	3833.39	19.11	96.83
11	4050.70	22.56	98.51	5859.62	30.14	98.84
13	5573.12	31.77	98.74	8401.61	43.96	98.73
15	7429.04	42.95	99.09	11445.32	58.42	99.29
17	9512.79	57.08	99.30	14950.47	81.80	99.25
19	11911.28	74.28	99.05	18760.19	105.53	99.41

TABLE III
DFDN PARAMETERS, TRAINING, TEST TIME, AND PRECISION COMPARISON TABLE OF MIXED FUNCTION CHANGES ON THE INDIAN PINES DATASET

Number of composite functions（c）	DFDN				DFRN			
	Params	Training time (s)	Test time (s)	OA (%)	Params	Training time (s)	Test time (s)	OA(%)
3	1242128	9512.79	57.08	99.30	816736	14950.47	81.80	99.25
4	1730576	13374.15	76.47	98.75	1108160	19441.60	103.31	99.46
5	2257040	17752.21	98.26	99.23	1399584	23875.10	125.55	99.33
6	2821520	/	/	/	1691008	28321.18	147.65	99.20

C. Neighboring Pixel Block Sizes in the DFAN

Comprehensively considering the training time and OA with limited small-size training samples, the ratio of training, validation, and test sets was 2:1:7 for the DFDN and the DFRN. The number of composite functions is 3 for the two models. From Table II, in the DFDN, the neighboring pixel blocks ranging from $M \times N = 9 \times 9$ to $M \times N = 19 \times 19$ ($M \times N$ refers to the spatial size of the sample) present the results fluctuating in OA. From the analysis point-of-view, the larger the neighboring pixel block, the larger the receptive field of the model, the more local information, thus the accuracy of the model increases with the increase of the neighboring pixel block. However, the DFDN does not behave this way, the highest OA was obtained while the neighboring pixel block is $M \times N = 17 \times 17$. In the DFRN, the OA increases with an increase in neighboring pixel blocks, but the trend of accuracy increases fluctuates slightly. The neighboring pixel block of $M \times N = 19 \times 19$ may be a good choice.

D. Number of Composite Functions in the DFAN

The number of composite functions is the most intuitive control factor for the model depth. In the deep aggregation network, the depth of the network model can be deepened by the choice of aggregation mode. The neighboring pixel block is $M \times N = 17 \times 17$, training set ratio is 2:1:7 for the Indian Pines dataset and the depth variation of the DFDN is shown in Table III. The DFDN model achieved the highest accuracy when the composite function was 3. Due to hardware limitations, higher memory cannot be provided, thus the accuracy of the hybrid function is not discussed. When the depth increases, the probability of high accuracy will decrease slightly due to the risk of overfitting. As shown in Table III, the neighboring pixel

TABLE IV
PERFORMANCE OF DFDN AND DFRN ON THE INDIAN PINES DATASET WITH PREPROCESSING

Pretreatment method	DFDN			DFRN		
	Training time (s)	Test time (s)	OA(%)	Training time (s)	Test time (s)	OA (%)
No	11911.28	74.28	99.05	19049.94	107.22	99.40
Gussian	11901.63	74.51	99.62	19065.19	106.59	99.72
Median	11903.89	74.50	99.12	19092.48	106.48	99.05

TABLE V
PERFORMANCE OF DFDN ON THE INDIAN PINES DATASET WITH DIFFERENT GAUSSIAN FILTER SLIDING WINDOW SIZES

Sliding window size	DFDN			DFRN		
	Training time (s)	Test time (s)	OA(%)	Training time (s)	Test time (s)	OA (%)
3	11896.41	74.35	99.36	19131.88	106.77	99.60
5	11901.63	74.51	99.62	19065.19	106.59	99.72
7	11950.85	73.75	99.47	19288.45	106.37	99.30

blocks for the DFRN on the Indian Pines dataset is $M \times N = 17 \times 17$, with a training set ratio of 2:1:7. The DFRN achieves the highest accuracy when the composite function is 4. However, the accuracy gain from depth decreases, so that a significant threshold effect is formed.

E. Discussion on the Accuracy of the Training and Validation Sets

Since the curve variation is large on the training and the validation sets, the model attempts to preprocess the data using different filter methods, which reduces the data noise. Table IV shows the performance of the DFDN and DFRN on the Indian Pines dataset. For the DFDN and DFRN, the neighboring pixel block is $M \times N = 19 \times 19$, the number of composite functions is 3, and the training set ratio is 2:1:7. The accuracies under Gaussian and median filters were compared. It can be seen that the classification accuracy has been improved after processing the data with a Gaussian filter.

For the DFDN and DFRN model, a Gaussian filter is selected for preprocessing and the influence of the filter sliding window size on the accuracy of the model is further discussed. Table V shows the performance of the DFDN and DFRN when Gaussian filters are used. It was found that a Gaussian filter with a sliding window of 5 has been selected on the training and validation curve, which will improve the OA and have a certain inhibitory effect on the fluctuation.

F. Experimental Results

We evaluated the performance of the DFAN on the Indian Pines, Pavia University, and Kennedy Space Center datasets. Using the DFDN and DFRN for the Indian Pines dataset, the neighboring pixel block size was 19, the number of composite functions was 3, and the training set ratio was 2:1:7. The Pavia University dataset was with the neighboring pixel block size of 19, the number of composite functions of 5, and the training set ratio of 1:1:8 for the DFDN and DFRN. Using the DFDN and DFRN for the Kennedy Space Center dataset, the neighboring pixel block size was 17 and 19, respectively, the number of composite functions was 3, and the training set ratio was 2:1:7 and 1:1:8, respectively. To evaluate the performance, we compared the models in this study with the SAE [14], 3D-CNN [19], SSRN [20], 3D-DenseNet [31], 3D-ResNet [34], and MSDN [33] models (see Tables VI and VII, Figs. 6 and 7). For SAE, the training set ratio was 6:2:2. For 3D-CNN, SSRN, 3D-DenseNet, 3D-ResNet, MSDN, DFDN, and DFRN methods, the training set ratio was 2:1:7 for the Indian Pines dataset, and 1:1:8 for the Pavia University dataset. To further explore the performance of the proposed models, we compared the DFDN ($M \times N = 17 \times 17$, $c = 3$; ratio = 1:1:8) and the DFRN ($M \times N = 19 \times 19$, $c = 3$; ratio = 1:1:8) with the MSDN and FDSSC [35] on Kennedy Space Center dataset (see Table VIII and Fig. 8). MSDN is proposed to make full use of different scale information in the network structure and combined scale information throughout the network, which integrated feature aggregation and dense connection. FDSSC is a refinement of SSRN, which could converge faster and achieve better classification accuracy with limited small-size training samples.

Compared with other outstanding HSI classification algorithms, the DFAN achieved the highest accuracy on the Indian Pines dataset. The highest OA was 99.62% for the DFDN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7) and 99.72% for the DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7). With the Pavia University datasets, the highest OA was 99.90% for the DFDN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8) and 99.91% for the DFRN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8). For the Kennedy Space Center dataset, the DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7) and DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 1:1:8) achieved the highest accuracy. Meanwhile, the OA was 99.69% for the DFDN ($M \times N = 17 \times 17$; $c = 3$; ratio = 1:1:8) and 99.78% for the DFDN ($M \times N = 17 \times 17$; $c = 3$; ratio = 2:1:7) that were higher than MSDN and FDSSC. For the ratio of 2:1:7 and 1:1:8, the DFRN ($M \times N = 19 \times 19$; $c = 3$) had a higher accuracy than DFDN ($M \times N = 17 \times 17$; $c = 3$).

TABLE VI
CLASSIFICATION ACCURACIES (%) OF DIFFERENT METHODS FOR THE INDIAN PINES DATASETS

No	SAE	3D-CNN	SSRN	3D-DenseNet	3D-ResNet	MSDN	DFDN(M×N=19×19, c=3)	DFRN(M×N=19×19, c=3)
1	81.82	94.44	100	97.22	100	100	100	100
2	82.16	96.10	99.02	99.31	100	99.21	98.63	99.61
3	77.54	97.11	99.65	99.31	99.64	98.46	100	99.83
4	68.11	95.86	97.06	97.66	95.98	97.66	98.82	97.66
5	94.36	93.72	98.29	100	100	99.71	100	100
6	94.45	98.46	99.81	100	99.42	99.61	100	100
7	94.70	100	100	99.70	100	100	100	100
8	94.36	100	100	100	100	99.70	100	100
9	82.56	100	0	100	100	91.67	100	100
10	81.28	97.91	100	99.65	100	100	99.71	99.56
11	84.47	99.57	98.95	97.89	99.18	99.53	99.94	99.82
12	83.77	93.99	100	98.57	94.93	99.28	98.57	99.28
13	96.42	100	100	98.57	100	99.28	100	100
14	92.27	98.75	99.66	100	98.87	99.89	100	100
15	80.63	90.13	100	100	99.28	99.63	100	100
16	81.82	90.67	97.18	95.65	91.89	97.18	97.06	97.59
OA	85.47±0.58	97.34±0.82	99.34±0.78	99.50±0.74	99.08±0.70	99.41±0.39	99.62±0.23	99.72±0.01
AA	86.31±1.14	96.67±0.71	93.10±0.60	99.05±0.68	98.70±0.74	98.80±0.35	99.55±0.28	99.55±0.21
K	83.42±0.66	96.97±0.93	99.25±0.89	99.40±0.87	98.95±0.79	99.33±0.40	99.57±0.35	99.68±0.57

TABLE VII
COMPARISON OF CLASSIFICATION ACCURACY OF DIFFERENT METHODS ON PAVIA UNIVERSITY DATASETS

No	SAE	3D-CNN	SSRN	3D-DenseNet	3D-ResNet	MSDN	DFDN(M×N=19×19, c=5)	DFRN(M×N=19×19, c=5)
1	87.24	97.93	99.98	99.79	99.62	99.28	99.89	99.87
2	89.93	99.95	99.87	99.99	100	99.90	100	99.99
3	86.48	92.50	99.29	99.58	99.70	98.58	100	100
4	99.95	92.19	100	99.97	99.46	99.96	99.54	99.83
5	95.78	100	100	99.75	98.72	99.91	100	100
6	97.69	98.94	99.98	100	100	99.93	100	100
7	95.44	98.30	99.91	99.81	99.63	98.57	99.91	99.44
8	84.40	98.92	97.73	99.63	99.43	98.25	99.66	99.56
9	100	96.49	100	99.69	100	100	99.34	100
OA	90.58±0.18	98.33±0.41	99.69±0.17	99.88±0.02	99.79±0.03	99.57±0.04	99.90±0.02	99.91±0.04
AA	92.99±0.39	97.25±0.31	99.64±0.17	99.80±0.02	99.62±0.05	99.38±0.04	99.81±0.02	99.85±0.02
K	87.21±0.25	97.79±0.21	99.59±0.22	99.85±0.03	99.72±0.04	99.43±0.05	99.87±0.04	99.88±0.04

The experiment outputs the loss and accuracy changes of the DFDN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7) and the DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7) on the Indian Pines dataset (see Fig. 9), and the DFDN ($M \times N = 17 \times 17$; $c = 3$; ratio = 1:1:8) and the DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 1:1:8) on the Kennedy Space Center dataset (see Fig. 10), as well as the DFDN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8) and the DFRN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8) on the Pavia University dataset during training and validation (see Fig. 11). In the Indian Pines and Kennedy Space Center datasets, there was a large fluctuation in the loss of the validation set, but the accuracy still achieved good results and the training loss and accuracy maintained suitable trends. For the Pavia University dataset, the volatility on the validation set was significantly weakened and essentially stable in the interval [0, 1]. The validation accuracy of the DFRN is slightly different from that of the DFDN. The DFDN variation on the 0–25 round validation set shows lateral fluctuation but it

TABLE VIII
COMPARISON OF CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON KENNEDY SPACE CENTER DATASETS

No	MSDN	FDSSC	DFDN($M\times N$=17×17; c=3; ratio=1:1:8)	DFRN($M\times N$=19×19; c=3; ratio=1:1:8)	DFDN($M\times N$=17×17; c=3; ratio=2:1:7)	DFRN($M\times N$=19×19; c=3; ratio=2:1:7)
1	93.04	100	100	100	100	100
2	96.08	100	100	99.48	97.63	100
3	91.59	98.19	97.95	100	100	100
4	89.10	100	98.02	100	98.29	100
5	93.85	87.34	99.20	98.45	100	100
6	97.60	100	97.93	100	98.77	100
7	98.84	100	100	100	100	100
8	94.75	98.10	100	100	100	100
9	93.33	100	100	100	100	100
10	95.32	100	100	100	100	100
11	100	99.73	100	100	100	100
12	95.60	100	100	100	99.72	100
13	99.32	100	100	100	100	100
OA	95.53	99.28	99.69	99.93	99.78	100
AA	95.26	98.72	99.47	99.84	99.66	100
K	95.02	99.19	99.65	99.92	99.76	100

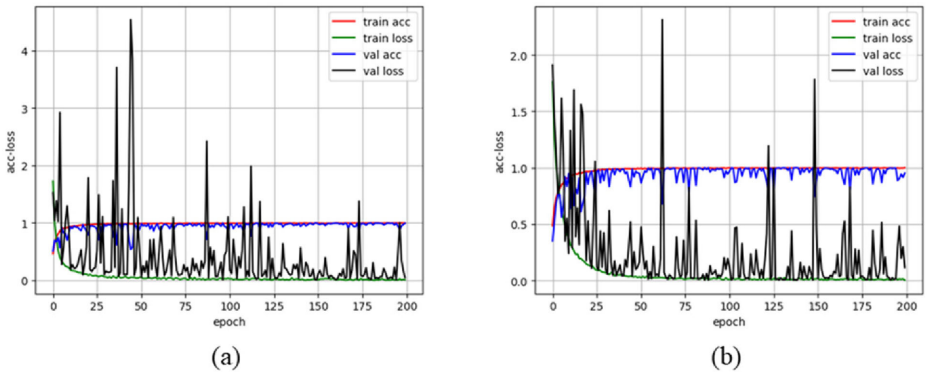

Fig. 9. Loss and accuracy changes in training and validation in the Indian Pines dataset of (a) the DFDN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7) and (b) the DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 2:1:7).

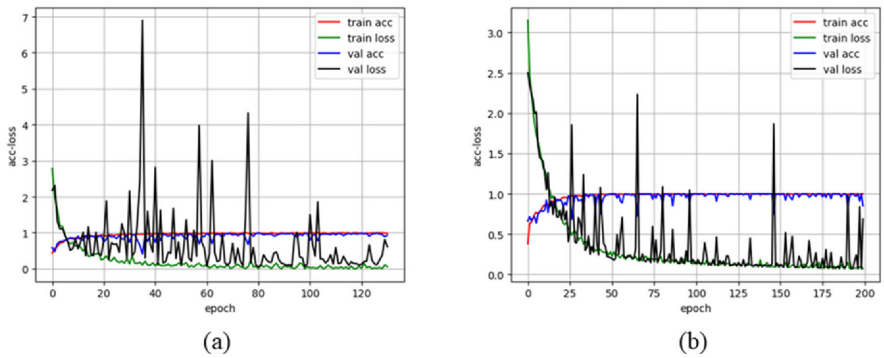

Fig. 10. Loss and accuracy changes in training and validation in the Kennedy Space Center dataset of (a) the DFDN ($M \times N = 17 \times 17$; $c = 3$; ratio = 1:1:8) and (b) the DFRN ($M \times N = 19 \times 19$; $c = 3$; ratio = 1:1:8).

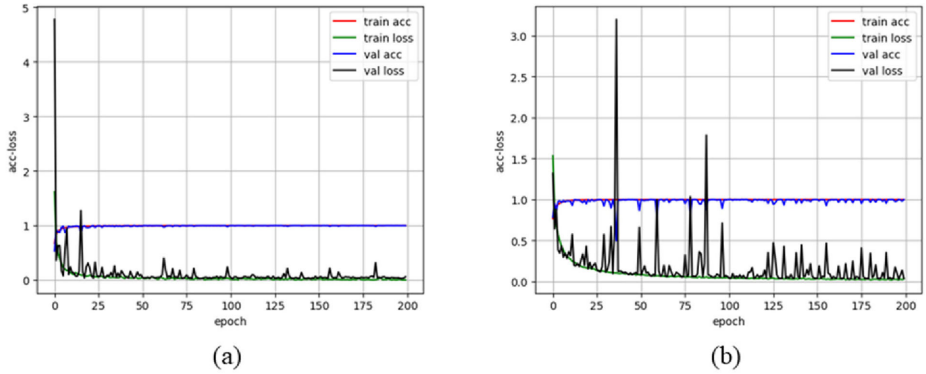

Fig. 11. Loss and accuracy changes in training and validation in the Pavia University dataset of (a) the DFDN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8) and (b) the DFRN ($M \times N = 19 \times 19$; $c = 5$; ratio = 1:1:8).

tends to converge quickly. From the perspective of aggregation view, whether it was the DFRN or the DFDN, although different aggregation methods were adopted, the network model can be further transmitted because of the aggregation itself, thereby extracting the depth features.

The DFAN model performs well in Indian Pines, Kennedy Space Center, and Pavia University datasets, effectively improving the classification accuracy of HSIs. For the Indian Pines and Kennedy Space Center datasets, the DFRN model was chosen to be more accurate, whereas on the Pavia University dataset, the accuracy of the DFDN is basically the same as that of the DFRN, but choosing the DFDN was more conducive to verifying the stability of the validation set. Therefore, the accuracy of the DFRN is higher than the DFDN when the training samples are limited. On the contrary, if the training samples are sufficient, the performance of the DFDN is better. Moreover, the DFDN has a faster convergence speed during training. When faced with different HSI dataset, the corresponding aggregation method can be chosen more flexibly, according to the number of training samples and the convergence speed requirement. This is very advantageous in the classification of HSIs.

V. Conclusion

The DFRN and the DFDN proposed in this study are typical extensions of DFAN. The model performs well in Indian Pines, Pavia University and Kennedy Space Center datasets, effectively improving the classification accuracy of HSIs. The DFRN is based on residual learning and uses the summation aggregation method. When the training samples are limited, it has a higher classification accuracy than the DFDN. The DFDN is based on dense connectivity and uses concatenation aggregation. The performance of the DFDN is better than the DFRN while the training samples are sufficient. Moreover, the convergence speed of the DFDN is always faster than the DFRN during training process. The most direct effect of aggregation is to deepen the network, strengthen the flow of information, and achieve the feature extraction of spectral–spatial information of different HSIs. In addition, considering that different layers can extract features of different sizes and provide complementary information, a fusion mechanism is adopted to utilize multilayer features. There are several ways to feature aggregation. We can design a more reasonable model for classifying HSIs from the perspective of multiple aggregation views.

In future, we will focus on using different aggregation patterns to extract different forms of features, discuss the application of deep networks in HSI classification, enhance feature extraction, and design more easily generalized models.

Acknowledgment

The authors would like to thank the Editors and three anonymous Reviewers for their constructive comments and suggestions, which greatly helped to improve the quality of this article.

References

[1] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Plenum, 2003.
[2] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Apr. 2019.
[3] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.
[4] H. Su, B. Zhao, Q. Du, P. Du, and Z. Xue, "Multifeature dictionary learning for collaborative representation classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2467–2484, Apr. 2018.
[5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remotesensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[6] Y. Zhong and L. Zhang, "An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*,vol. 50, no. 3, pp. 894–909, Aug. 2012.
[7] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Aug. 2010.
[8] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
[9] G. Camps-Vallset, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[10] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, Jan. 2012.

[11] H. Su, B. Zhao, Q. Du, and P. Du, "Kernel collaborative representation with local correlation features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1230–1241, Feb. 2019.
[12] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3778–3790, Jun. 2020.
[13] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[14] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state-of-the-art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
[15] 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.
[16] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.
[17] J. Yue *et al.*, "Spectral–spatial classification of hyperspectral images using deep convolutional neural networks," *Remote Sens. Lett.*, vol. 6, no. 6, pp. 468–477, Dec. 2014.
[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.*, Jul. 2015, pp. 4959–4962.
[19] 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.
[20] Z. Zhong, J. Li, Z. Luo, and M. Zhong, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
[21] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Nov. 2016.
[22] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3137–3184, Jun. 2018.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
[24] G. Huang *et al.*, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*.
[25] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
[26] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14h Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[27] Z. Wu, C. Shen, and A. Hengel, "Wider or deeper: Revisiting the ResNet model for visual recognition," *Pattern Recognit.*, vol. 90, pp. 119–133, 2019.
[28] L. Zhu *et al.*, "Sparsely connected convolutional networks," 2018, *arXiv:1801.05895*.
[29] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," 2016, *arXiv:1605.06431*.
[30] D. M. Hawkins, "The problem of overfitting," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, Dec. 2004.
[31] C. Zhang, G. Li, S. Du, and X. Zhang, "Three-dimensional densely connected convolutional network for hyperspectral remote sensing image classification," *Appl. Remote Sens.*, vol. 13, no. 1, pp. 1–22, 2019.
[32] G. Hinton, N. Srivastava, and K. Swersky, "RMSProp: Divide the gradient by a running average of its recent magnitude," Neural Netw. Mach. Learn., Coursera Lect., Oct. 2012, pp. 26–31.
[33] C. Zhang, G. Li, and S. Du, "Multi-scale dense networks for hyperspectral remote sensing image classification, " *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9201–9222, Nov. 2019.
[34] Y. Jiang, Y. Li, and H. Zhang, "Hyperspectral image classification based on 3-D separable ResNet and transfer learning," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1949–1953, Dec. 2019.
[35] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral–spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, p. 1068, Jul. 2018.

Chunju Zhang received the Ph.D. degree in cartography and geographic information system from the School of Geographical Science, Nanjing Normal University, Nanjing, China, in 2013.

She is currently an Associate Professor with the School of Civil Engineering, Hefei University of Technology, Hefei, China. Her research interests include remote sensing data processing and machine learning.

Guandong Li was born in Anhui, China, in 1993. He is currently working toward the master's degree with the School of Civil Engineering, Hefei University of Technology, Hefei, China.

His research interests include hyperspectral data analysis, high-resolution image processing, and deep learning techniques.

Shihong Du received the B.S. and M.S. degrees in cartography and geographic information system from Wuhan University, Hubei, China, in 1998 and 2001, respectively, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2004.

He is currently an Associate Professor with Peking University, Beijing, China. His research interests include qualitative knowledge representation, reasoning and its applications, and semantic understanding of spatial data including GIS and remote sensing data.