CMR-CNN: Cross-Mixing Residual Network for Hyperspectral Image Classification

Zhen Yang, *Member, IEEE*, Zhipeng Xi, Tao Zhang, *Member, IEEE*, Weiwei Guo *Member, IEEE*, Zenghui Zhang, *Senior Member, IEEE*, and Heng-Chao Li, *Senior Member, IEEE*

***Abstract*—With the development of deep learning, various convolutional neural network (CNN) based methods have been proposed for the hyperspectral image (HSI) classification. Although most of them achieve good classification performance, there are still more misclassifications in the prediction map with fewer training samples. In order to address this shortcoming, this paper proposes to simultaneously use pixels' spatial information and spectral information for HSI classification. Briefly speaking, a new cross-mixing residual network denoted by CMR-CNN is developed, wherein one three-dimensional (3D) residual structure responsible for extracting the spectral characteristics, one two-dimensional (2D) residual structure responsible for extracting the spatial characteristics, and one assisted feature extraction (AFE) structure responsible for linking the first two structures are respectively designed. With respect to experiments performed on five different datasets Indian Pines, University of Pavia, Salinas Scene, KSC, and Xuzhou in the case of different numbers of training samples show that, compared to some state-of-the-art methods, CMR-CNN can achieve higher overall accuracy (OA), average accuracy (AA), and Kappa values. Particularly, compared with the newly proposed HSI classification methods OCT-MCNN, CMR-CNN respectively improves OA, AA and kappa by 4.13%, 3.67%, and 2.75% on average.**

***Index Terms*—HSI, classification, spatial information, spectral information, AFE, CMR-CNN.**

I. Introduction

DIFFERENT from conventional optical images [1], infrared images [2] or synthetic aperture radar images [3], HSI has higher spectral resolution, meaning a larger number of spectral bands [4]. Also, this implies, much more information of scenes can be obtained from HSI. By virtue of this advantage, the related works, for example, object detection [5] and geological exploration [6], has been done and achieved some progress so far. Particularly, how to use HSI for classification becomes one hotspot in recent years.

In the early works, machine learning-based methods were often used for HSI classification, such as support vector machine [7], logistic regression [8], random forest [9], k-means clustering [10], and kernel-based method [11]. However, these traditional techniques easily yield more misclassifications, having the unsatisfactory classification accuracy. Deep learning (DL) [12] can extract more relevant features compared to manually designed features. In this regard, how to use convolutional neural network (CNN) for HSI classification becomes one research hotspot due to its strong ability to extract high-level semantic features of HSIs.

Up to now, various CNN-driven HSI classification methods have also been proposed. For example, Cheng $et\ al.$ directly explored hierarchical convolutional features for HSI classification in [13]. Lee $et\ al.$ [14] proposed to extract contextual information contained in HSI for classification, and He $et\ al.$ [15] used transfer learning methods based on CNN for HSI classification. Xu $et\ al.$ [16] proposed an unsupervised method to realize HSI classification. Marinoni $et\ al.$ [17] developed an information maximization method to find the most relevant features among pixels for HSI classification. In [18], Marinoni $et\ al.$ further made use of mutual information to retrieve the most relevant features for HSI classification. Zhang $et\ al.$ [19] proposed a deep convolutional neural network CloudNet for HSI cloud classification. Yang $et\ al.$ [20] used a two-channel deep CNN for HSI classification. Gong $et\ al.$ [21] used the multi-scale feature map obtained from CNN for HSI classification. Makantasis $et\ al.$ [22] utilized a supervised learning-based CNN for HSI classification. Xu $et\ al.$ made use of a fully CNN for HSI classification in [23]. In recently, Xu $et\ al.$ [24] used a self-attention network (SAC-NET) to address the threat of adversarial attacks on HSI classification. Duan $et\ al.$ [25] proposed the method of fusing dual spatial information to classify HSIs. In the latest literature [26], the author thought that more attention should be paid to the relationship between pixels in the feature map. With this guidance, they constructed a network named ENL-FCN. Lin $et\ al.$ [27] used generative adversarial networks for HSI classification. In a recent HSI classification task, Le $et\ al.$ [28] proposed to use a spectral-spatial feature label converter, by which an improved transformer (a densely connected transformer, namely Dense-Transformer) was developed to capture sequence spectral relations. Bhatti $et\ al.$ [29] proposed a local similarity-based spatial-spectral fusion method for HSI classification.

This work was supported in part by the National Natural Science Foundation of China under Grant 62261026, 62201343, 62071333, 62271311, and 62271418, and in part by the General Project of Jiangxi Nature Science Foundation under Grant 20212BAB202013, and in part by the Double First-Class Construction Foundation under Grant WH220503031, and in part by the ESA-Most China Dragon 5 Programm under Grant 58190. *(Corresponding author: Tao Zhang.)*

Z. Yang is with the School of Communication and Electronics, Jiangxi Science and Technology Normal University, Nanchang 330000, and also with the Guangdong Atv Academy For Performing Arts, Dongguan 523000, China.

Z. P. Xi is with the School of Communication and Electronics, Jiangxi Science and Technology Normal University, Nanchang 330000, China.

T. Zhang and Z. H. Zhang are with the Shanghai Key Laboratory of Intelligent Sensing and Recognition, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhangtao8902@sina.cn).

W. W. Guo is with the Center for Digital Innovation, Tongji University, Shanghai 200092, China.

H.-C. Li is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China, and also with the National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China.

In general, above methods are mainly built on the 2-dimensional (2D) convolution. Actually, in recent years, researchers are gradually turning their attention from feature extraction with 2D convolution to that with 3D convolution [30], which enables us to attain the spectral characteristics of HSIs. For instance, in [31], He *et al.* used a 3D deep CNN to obtain the multi-scale features for HSI classification. Using multi-scale feature maps can greatly obtain the information in the feature maps, but it also brings about the problem of information redundancy. In [32], a feature funsion 3D deep CNN was proposed for HSI classification. Chen *et al.* [33] directly used 3D-CNN for HSI classification. However, both 2D convolution and 3D convolution have intrinsic shortcomings in feature extraction. For example, in [22], while contextual information was obtained through the multi-scale method with 2D convolution, the spectral information of image were lost. In 3D-CNN [30], only the 3D convolution kernel was used to extract spectral information, yet the spatial information in the pixels was lost. Without adding other strategies and structures, a single network model is always unable to extract more effective information. According to the uniqueness of hyperspectral data and previous works, some scholars tried to fuse 3D convolution and 2D convolution together for HSI classification. In a recently presented work [34], a mixed CNN named MCNN-CP with the covariance pooling was proposed for HSI classification. Covariance pooling techniques are used to extract second-order information from the spectral-spatial feature maps, and channel shifts and weighting are used to highlight the importance of different spectral bands. Besides, Feng *et al.* [35] proposed a hybrid convolutional neural network (OCT-MCNN) using 3D Octave and 2D Vanilla for HSI classification. In brief, the authors first utilized the spectral 3D convolution and the spatial 2D convolution to obtain hybrid feature maps, and then, employed covariance pooling to extract second-order information from the spectral-spatial feature maps for HSI classification. Another recent work constructed a HybridSN model for HSI classification [36], where the 3D and 2D convolution operations were also applied together. Besides, the impact of combining convolution kernels of different dimensions on HSI classification was explored as well. However, its convolution layers is limited so that it cannot obtain the satisfactory classification performance.

Recently, the residual module [37] was adopted to increase the number of the layers of networks so as to extract more discriminative features for HSI classification. In particular, Zhong *et al.* proposed the Spectral-Spatial Residual Network (SSRN) in [38], where the residual block was used to connect each 3D convolutional layer for improving the classification accuracy. Paoletti *et al.* [39] developed a network named DPRN for HSI classification, wherein the residual block was utilized as well. Inspired by these methods, here we subtly design the cross-mixing framework of 3D residual and 2D residual structures, and develop a new HSI classification network (named by CMR-CNN). In this way, CMR-CNN can extract much deeper and more discriminative spatial-spectral features for HSI classification. Overall, the contributions of this paper are as follows.

1) In order to further improve the classification accuracy, the 3D residual structure and 2D residual structure are respectively designed based on SSRN and DPRN. The former is responsible for extracting the spectral features, while the latter is applied to the extraction of spatial features.

2) An assisted feature extraction (AFE) module is constructed with two convolutional layers, which goal is to bridge the 3D and 2D residual structures together. By AFE, it enables us to extract the spectral-spatial information simultaneously.

3) An end-to-end convolutional neural network named CMR-CNN is proposed for HSI classification via fusing the 3D and 2D residual structures with AFE. Experiments carried out on five different HSI datasets verify its effectiveness.

The remainder of this paper consists of the following parts: Section II introduces the proposed method, and Section III presents the experiments and discussions about experiments. Section IV is the conclusion of this paper.

II. METHODOLOGY

Traditional neural networks, like 3D-CNN [30] and 2D-CNN [22], easily loss the structural features of HSIs due to their limited layers. To deal with this problem, HybridSN combines 3D convolution and 2D convolution together for HSI classification, and DPRN introduces the residual block into the 3D convolution for HSI classification. Inspired by them, this section proposes a new network CMR-CNN, wherein one 3D residual structure, one 2D residual structure, and one AFE module are respectively designed, as shown in Fig. 1. Briefly, the 3D residual and 2D residual structures are respectively designed for extracting spectral and spatial information. Then, to bridge these two structures together, a module named AFE is further developed. Finally, the network CMR-CNN is proposed based on these three frameworks for classifying HSIs. Note that, in CMR-CNN, the principal component analysis (PCA) [40] is also adopted to reduce the redundant spectral information for the purpose of decreasing the computational complexity and avoiding the curse of dimensionality.

3D residual structure: After removing some unnecessary spectral features by PCA, we propose to use 3D convolution to extract the spectral information of feature maps. However, ordinary 3D convolution structure could have more training errors as the depth of network increases, which is also described as network degradation. So, to cure this disadvantage, we here introduce the residual structure into it, which is correspondingly shown in Fig. 2. Note that, the difference between the proposed 3D residual structure and that in [38] is, an additional convolution layer is performed here. Actually, in this way, the new 3D residual structure can allow us to extract more effective and diverse features for HSI classification. The details are given below.

The sizes of the convolution kernels in Fig. 2 are respectively set to (8×3×3×3), (16×3×3×3), and (32×3×3×3), where (8×3×3×3) represents 8 3D convolution kernels of dimensions (3×3×3). Then, the steps used in [38] are adopted. That is, the convolution kernel $(k_i \times s_i \times s_i)$ sequentially acts on the input feature maps to perform dot product with their weights and deviations. The corresponding output is the

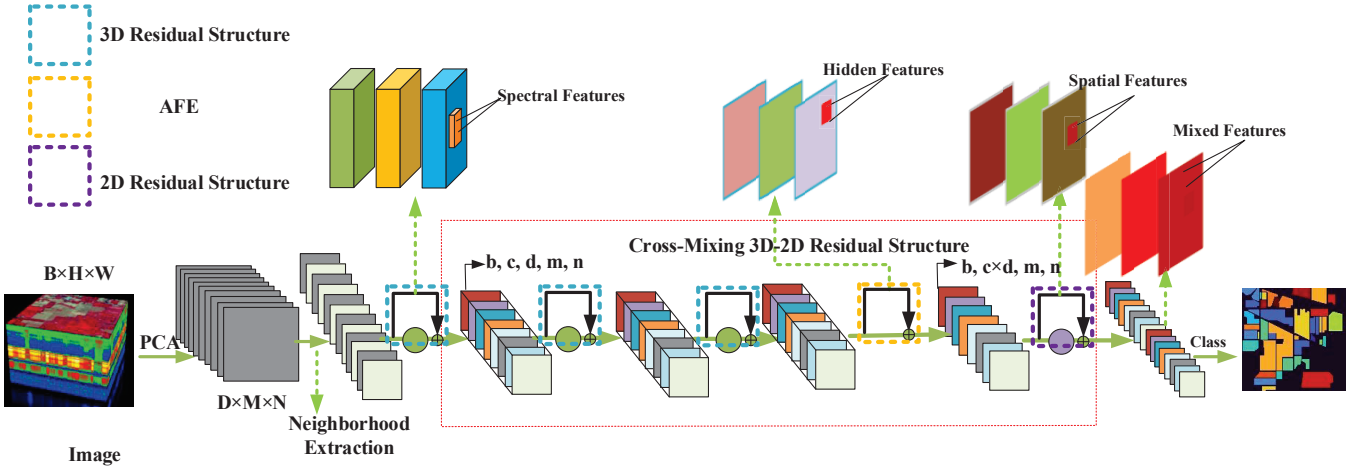


Fig. 1. CMR-CNN: 3D-2D Cross-Mixing Residual Network for HSI Classification.

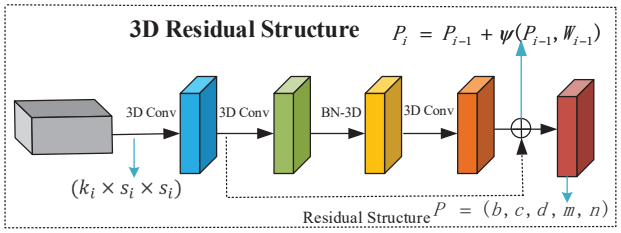


Fig. 2. The proposed 3D residual structure.

feature map P with b the batchsize, c the number of channels, d the number of spectral, m is the width, and n is the length.

$$\Psi(P_i, W_i) = \Phi(W_{i-1} \cdot BN(P_{i-1})) + b_{i-1} \tag{1}$$

$$P_{i+1} = P_i + \Psi(P_i, W_i) \tag{2}$$

where P_{i+1} is the output with the i^{th} layer feature map,$\Psi(\cdot)$ is the 3D residual part, and W_{i-1} is determined by the convolution kernel of residual module. And the activation value v at the position (x, y, z) in the j^{th} feature map of the i^{th} layer can be expressed as:

$$v_{i,j}^{x,y,z} = \Phi\sum_{\tau=1}^{r_{l-1}}\sum_{\lambda=-\eta}^{\eta}\sum_{\rho=-t}^{t}\sum_{\sigma=-s}^{s} W_{i,j,\tau}^{\sigma,\rho,\lambda} \times v_{i-1,\tau}^{x+\sigma,y+\rho,z+\lambda} + B_{i,j} \tag{3}$$

here, $\Phi(\cdot)$ is a nonlinear activation function, $B_{i,j}$ is the bias of the feature map of layer, r_{l-1} is the number of feature maps in the $(l-1)^{th}$ layer and the depth of the convolution kernel, t is the width of the convolution kernel, s is the length of the convolution kernel, $W_{i,j}$ is the weight of the j feature maps of the i^{th} layer, and η represents the spectral band.

Next, we analyze the differences between the proposed 3D residual structure and the traditional 3D convolution structure. For the latter, it first extracts the spectral information from the network based on a three-dimensional convolution kernel operation, and then directly sends the extracted features to the classification network. Since the problem of gradient degradation is prone to occurring when the ordinary network is too deep, it cannot extract enough deep spectrum features for HSI classification. However, for the former, it is based on the residual structure, thereby solving the problem of gradient degradation well. Moreover, it also holds an ability to deepen the depth of network so as to further ensure that the model maximizes the extraction of semantic information.

2D residual structure: [30] had pointed out that, 3D convolution was unable to extract effective spatial characteristics of HSIs. To tackle this problem, here, we design a 2D residual structure similar to [39], which is shown in Fig. 3. Note that, the difference between the proposed 2D residual structure and that designed in [39] is, the classical residual block is used in this paper.

In detail, we first transform $P = (b, c, d, m, n)_{i,j}$ into a 2D feature map by reshaping the tensor operation, which corresponding size is $(b, c \times d, m, n)_{i+1,j+1}$. Then, the 2D residual structure is expressed as

$$\psi(I_i, W_i) = \phi(W_{i-1} \cdot BN(I_{i-1})) + b_{i-1} \tag{4}$$

$$I_{i+1} = I_i + \psi(I_i, W_i) \tag{5}$$

where I_{i-1} represents the data input of the 2D residual module $\psi(\cdot)$, W_{i-1} is the weight determined by the convolution kernel of 2D residual module. $\phi(\cdot)$ is a nonlinear activation function.

Different from the common residual structure, we keep the number of the channels of feature map unchanged for reducing the amount of calculation. The activation value v at the position (x, y) in the j^{th} feature map of the i^{th} layer is expressed as

$$v_{i,j}^{x,y} = \Phi\sum_{\tau=1}^{r_{l-1}}\sum_{\rho=-t}^{t}\sum_{\sigma=-s}^{s} W_{i,j,\tau}^{\sigma,\rho} \times v_{i-1,\tau}^{x+\sigma,y+\rho} + B_{i,j} \tag{6}$$

Noticeably, the parameters of Eq. (6) are the same as those of Eq. (3). The only difference is that Eq. (6) does not have the parameters of spectral dimension. The original hyperspectral image dataset is a three-dimensional structure, which has one more spectral dimension than common optical images. In order to facilitate the 2D convolution operation, we fuse the spectral

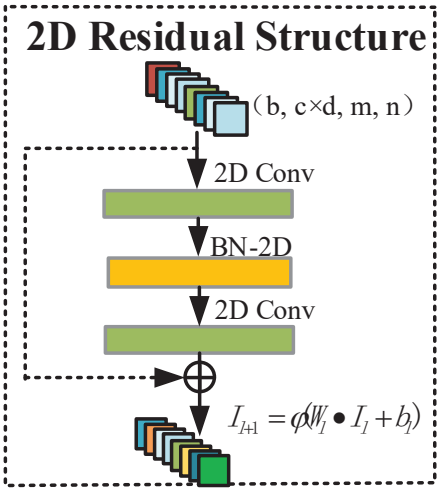


Fig. 3. The 2D residual structure.

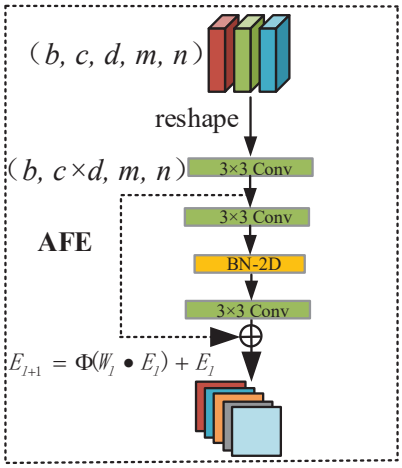


Fig. 4. The AFE structure.

information and spatial information from the feature map to form new feature information, which is the reason why Eq. (6) has one less parameter than Eq. (3).

Without loss of generality, hereinafter, we depict the advantage of the proposed 2D convolution residual structure against the traditional 2D convolution structure. That is, the latter only processes HSIs with simple convolution operations, which easily causes the ignorance of discriminative information. However, for the former, it can help the network obtain stronger spatial features. Therefore, compared to the traditional 2D convolution, the proposed 2D residual structure enables us to get higher classification accuracy.

AFE: The feature maps obtained by the 3D residual structure cannot be directly used as the input of 2D residual structure due to their different dimensions. To bridge them together, we here propose an assisted feature extraction (AFE) structure, by which a cross-mixing 3D-2D residual structure can be correspondingly formed, as shown in Fig. 1. Note that, different from existing networks that mainly achieve the fusion at the feature level, AFE is put forward for the first time to achieve the fusion at the structure level.

In detail, AFE is mainly composed of two convolutional layers, each of which contains a 3×3 convolution kernel, as shown in Fig. 4. Its goal is to decrease the number of channel input. That is, it uses the reshaping tensor to turn the output of 3D residual structure into the format of the 2D residual structure input. Mathematically, the relationship between the input and output of these two structures can be expressed as

$$E_{l+1} = \Phi(W_l \cdot E_l) + E_l \tag{7}$$

whose tensor format is

$$\begin{bmatrix} y_{1,1} & \cdots \\ \vdots & \vdots \\ \vdots & y_{m,n} \end{bmatrix} = \Phi\left(\begin{bmatrix} w_{1,1} & \cdots \\ \vdots & \vdots \\ \cdots & w_{m,n} \end{bmatrix} \begin{bmatrix} x_{1,1} & \cdots \\ \vdots & \vdots \\ \cdots & x_{m,n} \end{bmatrix} \right) + \begin{bmatrix} x_{1,1} & \cdots \\ \vdots & \vdots \\ \cdots & x_{m,n} \end{bmatrix} \tag{8}$$

Among them, E_{l+1} is the output with the E^{th} layer feature map, the W_l weight matrix is defined by the convolution kernel 3×3 and E_l is the output X of the 3D residual module obtained by the reshape operation,$\Phi(\cdot)$is a nonlinear activation function, y is the parameter in E_{l+1}, x is the parameter in E^{th}, w is the parameter in W_l, m, n are the width and height of the E^{th} layer feature map. It should be noted that, AFE here adopts the additive fusion operation to avoid losing the information of feature map.

Generally, the reasonability behind AFE is that, using the 3×3 convolution is able to reduce the number of the channels of feature maps. On the other hand, the AFE structure can also help the network extract more relevant feature information, wherein the additive feature fusion method is adopted to avoid the turbulence problem of network.

Batch normalization: To further solve the problem of gradient disappearance and gradient explosion during the backpropagation of residual network, we here introduce the BN layer [41] into CMR-CNN, viz,

$$BN(x) = \frac{x - mean[x]}{\sqrt{[Var[x]] + \in}} \cdot \gamma + \beta \tag{9}$$

wherein x represents the parameter of feature map, $\in$ represents the set constant, γ and β represent the parameter vector of sustainable learning. By converting the data of each layer to a state where the mean is zero and the variance is one, the distribution of data in each layer is the same. In the forward propagation, changing the value of hidden unit can reduce the covariance shift so that each layer can learn more independently. It should be pointed out that, BN-3D and BN-2D in Fig. 2 and Fig. 4 mean that the BN operation is respectively performed in a three-dimensional way and a two-dimensional way.

Based on the three new structures, next, we build the network CMR-CNN. Detailedly, in Fig. 1, the main spectral

information of input HSI is first obtained through PCA, and then the obtained cube is input into the 3D residual structure. In the 3D residual structure, the spectral information of feature map is extracted, which is subsequently input to AFE. AFE first reshapes the input into a feature map that can be operated by 2D convolution, then performs dimensionality reduction processing to reduce the number of channels, and uses the addition operation to fuse spectral information and spatial information together to avoid information loss. After this, the feature map is input to the 2D residual structure for further extracting the spatial information of feature map. Finally, we downsample the feature map with global average pooling and reshape it into a vector in order to feed it into a fully connected layer for classification. It should be noted that, we choose the stochastic gradient method to optimize the network model. The commonly used cross entropy loss function is here utilized as the classification function, which is defined as

$$L = \sum_{\imath=1}^{N} log(1 + e^{-ys}) \tag{10}$$

where L is the sum of the loss function, N is the number of sample classes, y is the label, and s is the score for each class.

III. Experiments

A. Datasets

In this paper, we adopt five publicly available HSI datasets to test the proposed method, including the Indian Pines, the University of Pavia, the Salinas Scene, KSC, and Xuzhou. Particularly, 10% and 90% are usually selected as the training and testing percentages, such as in [36] and [34]. Different from this setting, in this paper, we respectively select 5% and 95%, 1% and 99%, 0.5% and 99.5%, 20% and 80%, 1% and 99% as the training and testing set of these five datasets.

1) Indian Pines [42] was recorded by the AVIRIS sensor, which size is 145×145. There are 224 spectral bands in the wavelength range from 400 to 2500 nanometers and the effective spectral band is 200, as 24 of them with moisture and noise interference are discarded. This data has a total of 16 crop categories and 110,366 labeled pixels. In our experiment, we randomly select 5% of each class as the training set. The actual data is shown in Fig. 5(a).

2) Pavia University [43] was acquired by the ROSIS sensor, owning 103 bands after removing 12 noisy bands. Its size is 615×345 with 9 categories, as shown in Fig. 5(b). For this dataset, 1% of each class are used as the training set.

3) Salinas locating in Salinas Valley, California, was taken by the AVIRIS sensor. The spatial resolution of this dataset is 3.7 meters and the size is 512×217. After removing the bands with severe water vapor absorption, only 204 bands are remained. There exists 16 crop categories in this dataset, as shown in Fig. 5(c). For this dataset, 0.5% of each class are used as the training set.

4) The KSC data imaginged at the Kennedy Space Center was also captured by the AVIRIS sensor, which size was 512×614. After neglecting the bands related to water vapor noise, only 176 bands are remained. The spatial resolution is 18 meters, and there are 13 categories in total, as shown in Fig. 5(d). 20% of each class are chosen as the training set due to its smaller sample numbers.

5) The Xuzhou dataset was collected by an airborne HYSPEX hyperspectral camera over the Xuzhou peri-urban site in November 2014 [44]. It consists of 500 × 260 pixels, with a very high spatial resolution of 0.73 m/pixel. The number of spectral bands used in the experiment was 436, after removing the noisy bands ranging from 415 nm to 2508 nm. The scene is peri-urban and is characterized by nine categories, including crops, vegetation, man-made structures and so on, as shown in Fig. 5(e). For this dataset, 1% of each class are used as the training set.

B. Experimental Setup

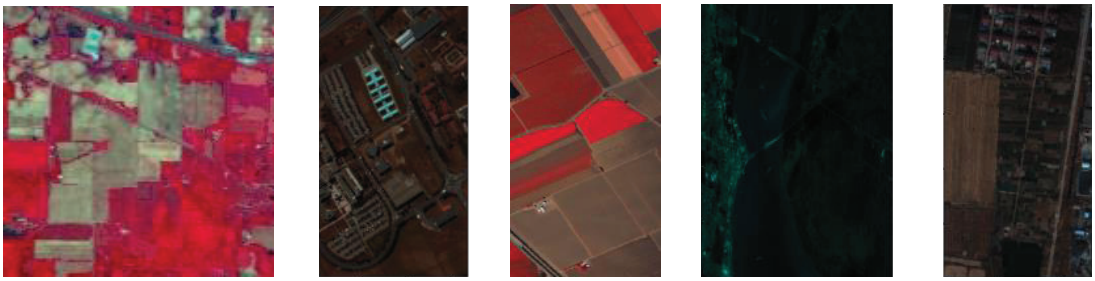
To prove the effectiveness of our network, we select the following methods for comparing, including SVM [7], 2D-CNN [22], 3D-CNN [30], SSRN [38], DPRN [39], HybridSN [36], and the recently proposed methods OCT-MCNN [35], SAC-NET [24], MCNN-CP [34]. Meanwhile, for fairly comparing the performance of each method, we select OA, AA, and kappa coefficients as the evaluation criteria in experiments. OA represents the overall accuracy which is the ratio of correctly classified pixels to the total pixels; AA represents the average accuracy of each class; and Kappa represents the ratio of error reduction between classification and completely random classification, which combines the diagonal of the confusion matrix wherein line and off-diagonal terms are a robust consistency measure [33]. All experiments are performed on Tesla V-100 with Pytorch environment. Note that, the learning rate is set to 0.001, and Epoch is set to 100 so as to compare the convergence speed of different network models.

C. Hyperparameters Setting

For CMR-CNN, three hyperparameters, i.e., the principal component value of PCA, the number of 3D residual layer, and the number of 2D residual layer, directly affect its performance. Thus, it is necessary for us to show how to optimally set them. To this goal, in the following, we give the detailed decision process of hyperparameters. We randomly select 1% as the validation set and the rest as training and testing sets when tuning hyperparameters for each dataset.

First, we fix the principal component value of PCA, and design the number of 3D residual layer which is selected from {1, 2, 3, 4} and corresponding number of output channels is respectively 8, 16, 32, and 64. Table II lists the OA values obtained by different layer numbers. Through experiments, we found that when the number of 3D residual layer is 4, the OA value begins to decrease, indicating that the best number of 3D residual layer is 3. For the number of 2D residual layer, obviously, when it is 2, the OA value begins to decrease. So, this means, when the number of 2D residual layer is 1, we are able to extract more effective classification features.

Second, after determining the number of residual layers, we begin to set the principal component value of PCA, which is here selected from {10, 20, 30, 40, 50, 60}. Table III displays the OA values obtained under different principal component values. From Table III, we can easily see that when the



(a) (b) (c) (d) (e)

Fig. 5. Datasets. (a) Indian Pines; (b) Pavia University; (c) Salinas; (d) KSC; (e) Xuzhou.

TABLE I
LAND COVER CLASSES AND NUMBERS OF SAMPLES FOR INDIA PINES, PAVIA UNVERSITY, SALINAS, KSC, AND XUZHOU DATASETS

Indian Pines			Pavia Unversity			Salinas			KSC			Xuzhou		
Land-cover type	color	Samples	Land-cover type	color	Samples	Land-cover type	color	Samples	Land-cover type	color	Samples	Land-cover type	color	Samples
Background		10776	Background		164624	Background		56975	Background		56975	Background		61123
Alfalfa		46	Asphalt		6631	Brocoli-green-weeds-1		2009	Scrub		761	Bareland-1		26396
Corn-notill		1428	Meadows		18649	Brocoli-green-weeds		23726	Willow-swamp		243	Bareland-2		4027
Corn-min		830	Gravel		2099	Fallow		1976	CP-hammock		256	Crops-1		2783
Corn		237	Trees		3064	Fallow-rough-plownn		1394	Slash-pine		252	Crops-2		5214
Grass/Pasture		483	Painted metal sheets		1345	Fallow-smooth		2678	Oak/Broadleaf		161	Coals		13184
Grass/Trees		730	Bare Soil		5029	Stubble		3959	Hardwood		229	Concrete		2436
Grass/pasture-mowed		28	Bitumen		1330	Celery		3579	Swap		105	Trees		6990
Hay-windrowed		478	Self-Blocking Bricks		3682	Grapes-untrained		11271	Graminoid-marsh		431	Lake		4777
Oats		20	Shadows		947	Soil-vinyard-develop		6203	Spartina-marsh		520	House-Roof		3070
Soybeans-notill		972				Corn-senesced-green-weeds		3278	Cattail-marsh		404			
Soybeans-min		2455				Letture-romaine-4wk		1068	Salt-marsh		419			
Soybeans-clean		593				Letture-romaine-5wk		1927	Mud-flats		503			
Wheat		205				Letture-romaine-6wk		916	Water		927			
Woods		1265				Letture-romaine-7wk		1070						
Bldg-Grass-Tree-Drivers		386				Vinyard-untrained		7268						
Stone-steel towers		93				Vinyard-vertical-trellis		1807						
Total samples		21025	Total samples		207400	Total samplis		111104	Total samples		314368	Total samples		130000

TABLE II
THE CLASIFICATION RESULTS OA UNDER DIFFERENT NUMBER (K) OF RESIDUAL LAYERS

Conv-layer \ Datasets		Indian Pines	University of Pavia	Salinas Scene	KSC	Xuzhou
3D Res-layer	1	75.88	67.49	72.90	70.43	72.64
	2	77.63	77.51	84.96	71.12	84.79
	3	78.04	91.75	96.88	75.72	96.58
	4	70.91	84.30	86.49	66.47	86.62
2D Res-layer	1	**84.61**	**96.03**	**99.25**	**76.91**	**98.52**
	2	76.43	64.59	92.97	66.71	92.67

TABLE III
THE CLASIFICATION RESULTS OA UNDER DIFFERENT NUMBER (K) OF PCA

Datasets \ PCA	10	20	30	40	50	60
Indian Pines	75.73	80.05	**83.91**	80.16	79.64	76.03
University of Pavia	87.03	89.53	94.77	**95.01**	93.75	92.59
Salinas Scene	93.35	95.22	97.51	**98.13**	96.80	96.79
KSC	75.416	78.34	**80.59**	79.58	76.25	72.56
Xuzhou	94.02	96.82	97.33	97.19	**98.74**	97.91

principal component value is 30, Indian Pines and KSC can get the best results. When it is 40, University of Pavia and Salinas Scene hold the best results. The most suitable value for Xuzhou is 50. So, Table III proves, from another aspect, that reducing redundant spectral bands can really allow us to achieve better classification results. In this paper, we set the value of PCA to 30.

D. Experimental Results

Tables IV-VIII and Figs. 6-10 are the results of different methods on these five datasets.

1) ***The experimental analyses on Indian Pines***: Table IV shows the quantitative results of different methods on Indian Pines. Clearly, our method CMR-CNN performs best on the OA, AA, and Kappa. Particularly, compared with the method HybridSN, CMR-CNN respectively improves by 4.19%, 8.53%, and 4.56% on these three metrics. Moreover, Fig. 6 exhibits the classification results of each network on this dataset. Through visual analyses, it is easily found that CMR-CNN has the least areas of prediction error. In detail, the classification result of SVM in Fig. 6(c) is the worst among these methods since lots of misclassifications are present,

TABLE IV
CLASSIFICATION ACCURACY (%) AND SCORE OF OA, AA, KAPPA FOR INDIAN PINES(USING 5% TRAINING SAMPLES)

Class	SVM [7]	2D-CNN [22]	3D-CNN [30]	SSRN [38]	DPRN [39]	HybridSN [36]	OCT-MCNN [35]	SAC-NET [24]	MCNN-CP [34]	CMR-CNN
1	61.00	97.00	100.00	100.00	88.36	62.00	59.00	100.00	100.00	100.00
2	77.00	91.00	74.00	94.00	95.14	86.00	91.00	92.00	98.00	95.00
3	65.00	90.00	82.00	87.00	96.20	87.00	71.00	91.00	96.00	95.00
4	79.00	97.00	100.00	87.00	86.22	87.00	88.00	100.00	99.00	99.00
5	90.00	90.00	92.00	87.00	98.69	96.00	96.00	99.00	98.00	100.00
6	80.00	94.00	98.00	97.00	99.13	90.00	93.00	98.00	100.00	100.00
7	83.00	97.00	100.00	100.00	51.85	100.00	60.00	94.00	74.00	93.00
8	92.00	89.00	90.00	96.00	99.78	93.00	70.00	100.00	95.00	100.00
9	60.00	97.00	100.00	87.00	99.74	100.00	50.00	100.00	100.00	100.00
10	71.00	87.00	86.00	79.00	94.04	93.00	80.00	91.00	92.00	96.00
11	81.00	93.00	73.00	88.00	97.34	86.00	85.00	96.00	98.00	96.00
12	79.00	88.00	87.00	87.00	84.55	77.00	64.00	92.00	99.00	97.00
13	81.00	97.00	100.00	99.00	98.46	100.00	74.00	83.00	100.00	95.00
14	94.00	94.00	91.00	99.00	97.67	91.00	98.00	95.00	95.00	98.00
15	86.00	93.00	100.00	96.00	92.22	96.00	54.00	91.00	100.00	99.00
16	80.00	92.00	93.00	87.00	96.59	94.00	69.00	94.00	95.00	82.00
OA	76.72 ± 2.44	90.54 ± 2.84	89.56 ± 1.45	90.84 ± 3.21	96.27 ± 1.67	92.53 ± 3.64	90.02 ± 3.82	94.68 ± 1.56	96.24 ± 1.34	**96.72** ± 0.44
AA	66.58 ± 1.40	88.59 ± 1.57	82.72 ± 2.54	92.68 ± 2.16	92.53 ± 1.54	86.14 ± 4.14	83.69 ± 3.76	75.56 ±1.34	89.98 ± 1.25	**94.67** ± 0.54
Kappa	73.41 ± 2.51	90.76 ± 1.85	88.02 ± 1.52	91.25 ± 3.04	95.75 ± 1.28	91.70 ± 3.53	88.90 ± 3.93	93.76 ± 1.61	95.71 ± 1.33	**96.26** ± 0.57

TABLE V
CLASSIFICATION ACCURACY (%) AND SCORE OF OA, AA, KAPPA FOR PAVIA UNIVERSITY(USING 1% TRAINING SAMPLES)

Class	SVM [7]	2D-CNN [22]	3D-CNN [30]	SSRN [38]	DPRN [39]	HybridSN [36]	OCT-MCNN [35]	SAC-NET [24]	MCNN-CP [34]	CMR-CNN
1	75.00	91.00	83.00	93.00	94.00	89.00	96.00	97.00	94.00	92.00
2	92.00	95.00	93.00	97.00	99.88	96.00	98.00	87.00	97.00	100.00
3	64.00	94.00	48.00	53.00	67.18	57.00	79.00	98.00	98.00	86.00
4	95.00	93.00	83.00	96.00	93.18	93.00	95.00	97.00	90.00	91.00
5	84.00	97.00	97.00	100.00	97.45	100.00	100.00	100.00	100.00	92.00
6	100.00	98.00	82.00	99.00	96.79	97.00	99.00	100.00	100.00	100.00
7	92.00	96.00	73.00	88.90	78.06	90.00	95.00	99.00	89.00	97.00
8	52.00	85.00	88.00	68.00	89.05	74.00	89.00	97.00	92.00	94.00
9	96.00	94.00	94.00	83.00	97.12	84.00	98.00	96.00	86.00	91.00
OA	87.94 ± 1.17	92.67 ± 1.89	84.36 ± 1.48	93.47 ± 1.34	94.77 ± 1.59	90.12 ± 2.83	95.46 ± 2.36	93.08 ± 1.23	95.28 ± 1.80	**96.21** ± 1.46
AA	88.00 ± 1.56	90.86 ± 2.04	83.39 ± 1.45	92.79 ± 1.70	90.30 ± 1.77	83.37 ± 2.96	**94.83** ± 2.37	92.01 ± 1.25	92.85 ± 1.28	91.44 ± 1.14
Kappa	84.14 ± 1.23	91.57 ± 2.62	84.01 ± 1.67	92.31 ± 1.43	93.04 ± 1.82	86.84 ± 5.12	94.35 ± 2.88	90.77 ± 1.33	94.24 ± 1.55	**94.97** ± 1.24

TABLE VI
CLASSIFICATION ACCURACY (%) AND SCORE OF OA, AA, KAPPA FOR SALINAS(USING 0.5% TRAINING SAMPLES)

Class	SVM [7]	2D-CNN [22]	3D-CNN [30]	SSRN [38]	DPRN [39]	HybridSN [36]	OCT-MCNN [35]	SAC-NET [24]	MCNN-CP [34]	CMR-CNN
1	95.00	99.00	96.00	99.00	98.30	97.00	100.00	100.00	100.00	100.00
2	100.00	97.00	99.00	100.00	99.89	99.00	100.00	98.00	100.00	100.00
3	99.00	96.00	100.00	100.00	95.83	94.00	100.00	100.00	99.00	100.00
4	97.00	95.00	89.00	100.00	99.64	98.00	100.00	100.00	100.00	94.00
5	88.00	81.00	87.00	95.00	99.21	94.00	96.00	98.00	95.00	99.00
6	98.00	100.00	94.00	100.00	98.78	100.00	100.00	99.00	100.00	100.00
7	99.00	98.00	100.00	100.00	100.00	100.00	100.00	98.00	98.00	100.00
8	79.00	79.00	80.00	91.00	90.26	96.00	95.00	99.00	87.00	100.00
9	90.00	100.00	100.00	99.00	100.00	100.00	99.00	95.00	99.00	100.00
10	97.00	92.00	94.00	93.00	97.30	98.00	99.00	75.00	100.00	100.00
11	94.00	90.00	83.00	100.00	96.14	100.00	100.00	99.00	100.00	92.00
12	95.00	100.00	99.00	98.00	99.63	97.00	89.00	99.00	100.00	95.00
13	92.00	61.00	74.00	91.00	99.56	66.00	100.00	100.00	94.00	78.00
14	98.00	99.00	100.00	82.00	97.09	93.00	99.00	100.00	85.00	93.00
15	73.00	73.00	82.00	86.00	88.32	98.00	83.00	99.00	89.00	95.00
16	100.00	100.00	100.00	100.00	95.49	100.00	100.00	100.00	98.00	93.00
OA	89.00 ± 2.38	88.59 ± 2.68	90.37 ± 1.56	94.86 ± 1.37	95.58 ± 1.08	96.69 ± 2.91	95.00 ± 1.41	91.07 ± 1.38	94.54 ± 1.39	**97.95** ± 0.52
AA	90.29 ± 2.10	89.17 ± 3.55	92.26 ± 3.00	95.45 ± 1.94	**97.22** ± 1.39	94.89 ± 4.04	95.51 ± 1.86	94.77 ± 1.50	95.52 ± 1.03	97.14 ± 0.60
Kappa	87.72 ± 2.45	87.29 ± 2.94	89.27 ± 1.74	94.28 ± 1.53	95.08 ± 1.10	96.31 ± 4.29	95.92 ± 0.84	90.01 ± 1.22	93.91 ± 1.59	**97.71** ± 0.34

TABLE VII
CLASSIFICATION ACCURACY (%) AND SCORE OF OA, AA, KAPPA FOR KSC(USING 20% TRAINING SAMPLES)

Class	SVM [7]	2D-CNN [22]	3D-CNN [30]	SSRN [38]	DPRN [39]	HybridSN [36]	OCT-MCNN [35]	SAC-NET [24]	MCNN-CP [34]	CMR-CNN
1	91.00	97.00	89.00	90.00	98.00	99.00	90.0	100.0	99.0	99.00
2	83.00	74.00	93.00	89.00	97.00	96.00	88.0	92.0	93.00	99.00
3	94.00	93.00	85.00	100.00	99.00	99.00	77.0	99.0	80.00	100.00
4	73.00	69.00	89.00	94.00	98.00	97.00	86.0	97.0	88.00	100.00
5	82.00	87.00	88.00	95.00	99.00	95.00	68.0	98.0	89.00	100.00
6	70.00	84.00	79.00	96.00	98.00	99.00	92.0	95.0	79.00	96.97
7	77.00	80.00	86.00	89.00	96.00	97.00	80.0	89.0	88.00	100.00
8	99.00	92.00	93.00	97.00	96.00	96.00	85.0	96.0	96.00	100.00
9	76.00	76.00	87.00	92.00	95.00	93.00	97.0	89.0	95.00	97.00
10	90.00	85.00	84.00	91.00	95.00	98.00	85.0	99.0	97.00	100.00
11	100.00	96.00	99.00	98.58	100.00	100.00	98.0	99.0	100.00	100.00
12	92.00	84.00	87.00	94.00	100.00	97.00	94.0	92.0	998.00	5.00
13	98.00	90.00	95.00	97.00	99.00	99.00	98.0	99.0	99.00	100.00
OA	91.21 ± 2.23	91.82 ± 2.04	91.98 ± 1.42	94.06 ± 0.70	97.82 ± 0.41	97.51 ± 1.06	89.34 ± 3.35	96.33 ± 0.23	95.01 ± 1.21	**99.02** ± 0.25
AA	84.86 ± 2.21	87.96 ± 2.37	86.93 ± 1.36	91.83 ± 1.20	96.50 ± 0.85	95.84 ± 1.84	90.44 ± 3.38	94.96 ± 1.54	92.41 ± 1.03	**98.29** ± 0.43
Kappa	90.21 ± 2.13	90.88 ± 2.29	90.84 ± 1.60	93.37 ± 0.79	97.57 ± 0.44	97.22 ± 1.39	90.01 ± 3.39	95.91 ± 0.27	86.69 ± 1.13	**98.90** ± 0.51

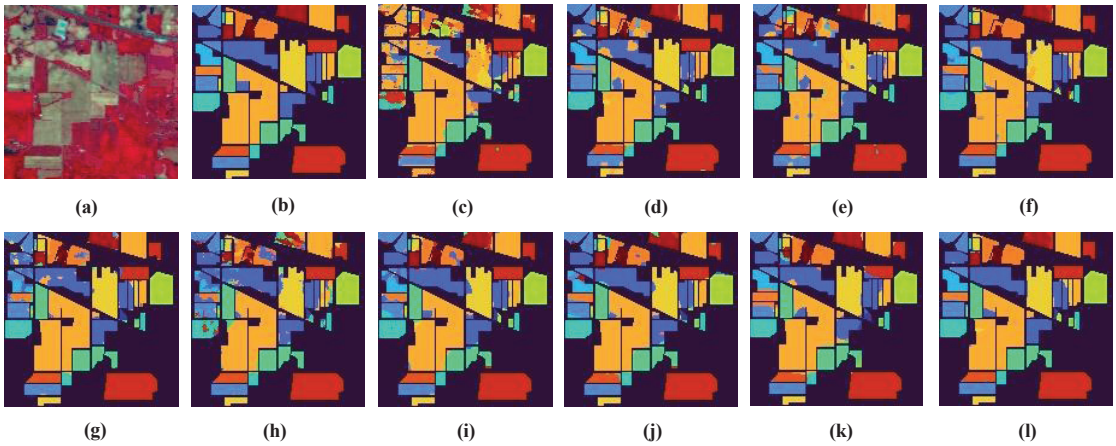


Fig. 6. The classification results of Indian Pines. (a) False color image; (b) Ground truth; (c) SVM; (d) 2D-CNN; (e) 3D-CNN; (f) SSRN; (g) DPRN; (h) HybridSN; (i) OCT-MCNN; (j) SAC-NET; (k) MCNN-CP; (l) CMR-CNN.

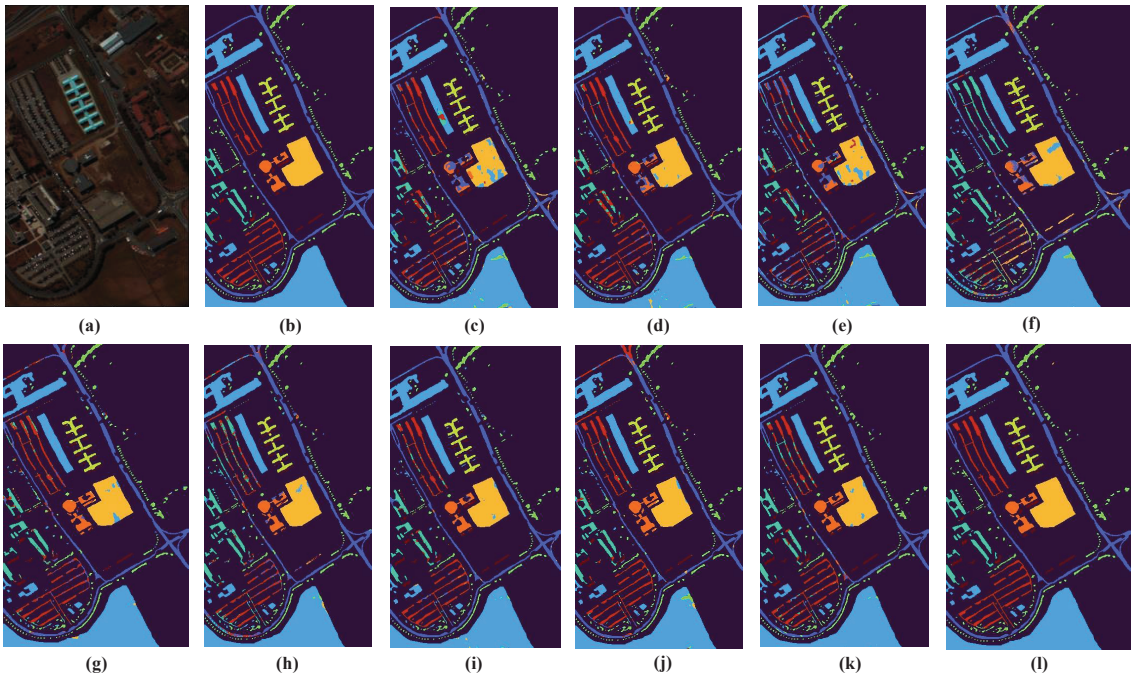


Fig. 7. The classification results of Pavia University. (a) False color image; (b) Ground truth; (c) SVM; (d) 2D-CNN; (e) 3D-CNN; (f) SSRN; (g) DPRN; (h) HybridSN; (i) OCT-MCNN; (j) SAC-NET; (k) MCNN-CP; (l) CMR-CNN.

which Kappa value is also the lowest (73.41%) in Table IV. Compared to it, the classification results of 2D-CNN and 3D-CNN are better in Figs. 6(d) and 6(e). Different from these methods, SSRN has fewer misclassifications in Fig. 6(f). The classification result of DPRN is shown in Fig. 6(g). Obviously, in comparison with SSRN, DPRN has a better classification performance. Unfortunately, OCT-MCNN's classification result in Fig. 6(i) is unsatisfactory, which can also be verified by its Kappa value (88.90%) in Table IV. Fig. 6(k) displays the classification result of MCNN-CP, from which one can see that most of the categories are correctly classified. Compared with other methods, the classification effect of OCT-MCNN is poor in the case of fewer training samples. From Fig. 6(j), it can see that the recently proposed method SAC-NET also performs better. Compared to MCNN-CP, the misclassifications caused by CMR-CNN in Fig. 6(l) is a little fewer, and much fewer than HybridsSN. Therefore, it directly demonstrates that the strategy used to construct CMR-CNN is effective.

2) ***The experimental analyses on Pavia University***: Table V lists the quantitative results of different methods on the Pavia University dataset, and Fig. 7 is the prediction maps corresponding to these methods. It can be seen from Table V that the proposed method CMR-CNN has achieved the best classification results on the evaluation indicators OA and Kapaa. In detail, compared with the HrbridSN method, CMR-CNN improves by 6.09% and 8.07% on OA and AA, and 8.13% on Kappa, respectively. Through observing Fig. 7(h) and Fig. 7(l), we can also demonstrate that CMR-

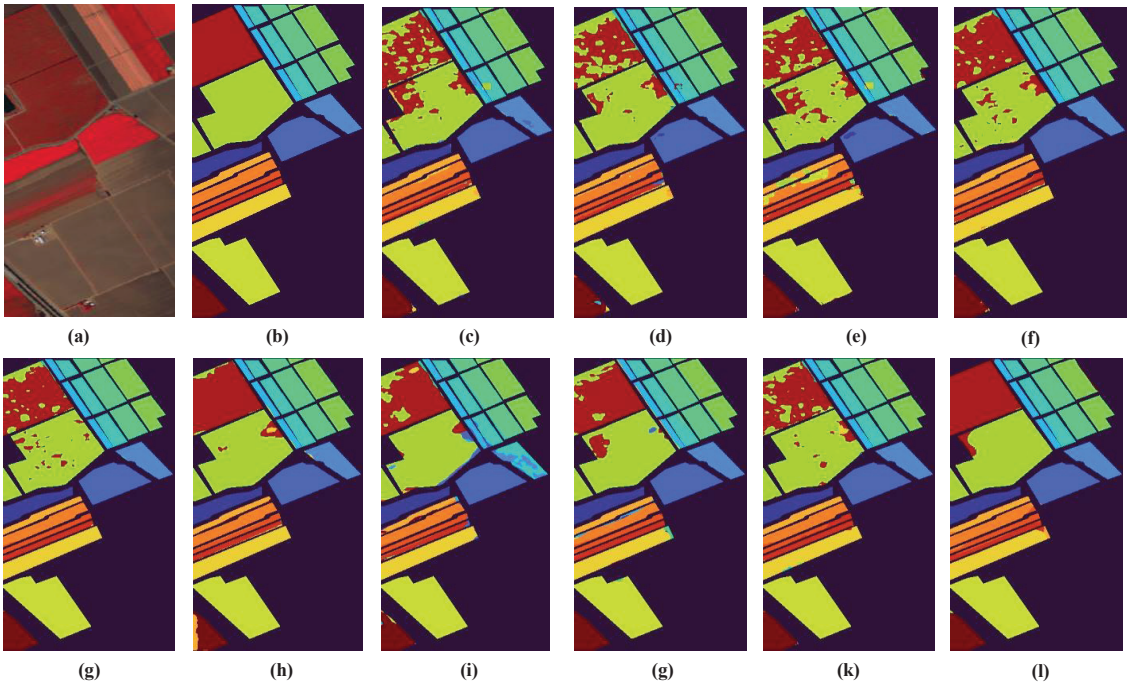


Fig. 8. The classification results of Salinas. (a) False color image; (b) Ground truth; (c) SVM; (d) 2D-CNN; (e) 3D-CNN; (f) SSRN; (g) DPRN; (h) HybridSN; (i) OCT-MCNN; (j) SAC-NET; (k) MCNN-CP; (l) CMR-CNN.

TABLE VIII
CLASSIFICATION ACCURACY (%) AND SCORE OF OA, AA, KAPPA FOR XUZHOU(USING 1% TRAINING SAMPLES)

Class	SVM [7]	2D-CNN [22]	3D-CNN [30]	SSRN [38]	DPRN [39]	HybridSN [36]	OCT-MCNN [35]	SAC-NET [24]	MCNN-CP [34]	CMR-CNN
1	83.00	100.00	98.00	99.00	99.00	99.00	98.00	98.00	97.00	100.00
2	100.00	100.00	99.00	92.00	97.00	100.00	100.00	99.00	98.00	100.00
3	56.00	97.00	96.00	99.00	100.00	100.00	98.00	98.00	99.00	99.00
4	63.00	95.00	91.00	99.00	100.00	99.00	97.00	99.00	93.00	100.00
5	85.00	96.00	88.00	99.00	99.00	99.00	99.00	99.00	98.00	100.00
6	69.00	89.00	94.00	95.00	98.00	99.00	99.00	98.00	99.00	100.00
7	83.00	97.00	88.00	99.00	97.00	98.00	99.00	100.00	91.00	99.00
8	63.00	96.00	99.00	98.00	99.00	99.00	93.00	99.00	98.00	99.00
9	77.00	94.00	99.00	99.00	99.00	100.00	98.00	95.00	92.00	99.00
OA	95.45 ± 2.27	94.74 ± 1.47	94.20 ± 2.34	97.65 ± 0.43	98.05 ± 0.63	97.99 ± 0.32	98.12 ± 0.37	98.12 ± 0.24	95.73 ± 0.55	**98.70** ± 0.19
AA	93.80 ± 3.14	95.55 ± 1.51	91.15 ± 2.57	98.25 ± 0.37	97.45 ± 0.59	97.63 ± 0.58	97.35 ± 0.36	97.27 ± 0.33	94.10 ± 0.65	**98.67** ± 0.23
Kappa	94.21 ± 2.42	95.90 ± 1.59	92.84 ± 2.53	97.04 ± 2.35	97.31 ± 0.54	97.45 ± 0.69	97.89 ± 0.50	98.01 ± 0.27	94.58 ± 0.47	**98.14** ± 0.30

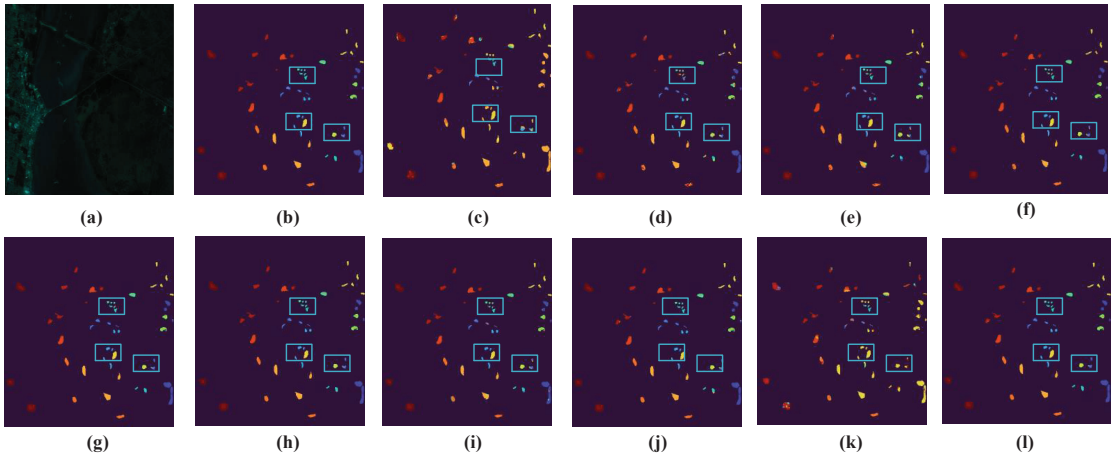


Fig. 9. The classification results of KSC. (a) False color image; (b) Ground truth; (c) SVM; (d) 2D-CNN; (e) 3D-CNN; (f) SSRN; (g) DPRN; (h) HybridSN; (i) OCT-MCNN; (j) SAC-NET; (k) MCNN-CP; (l) CMR-CNN.

CNN is more useful for HSI classification than HrbridSN. It should be pointed out that, among the five datasets, the

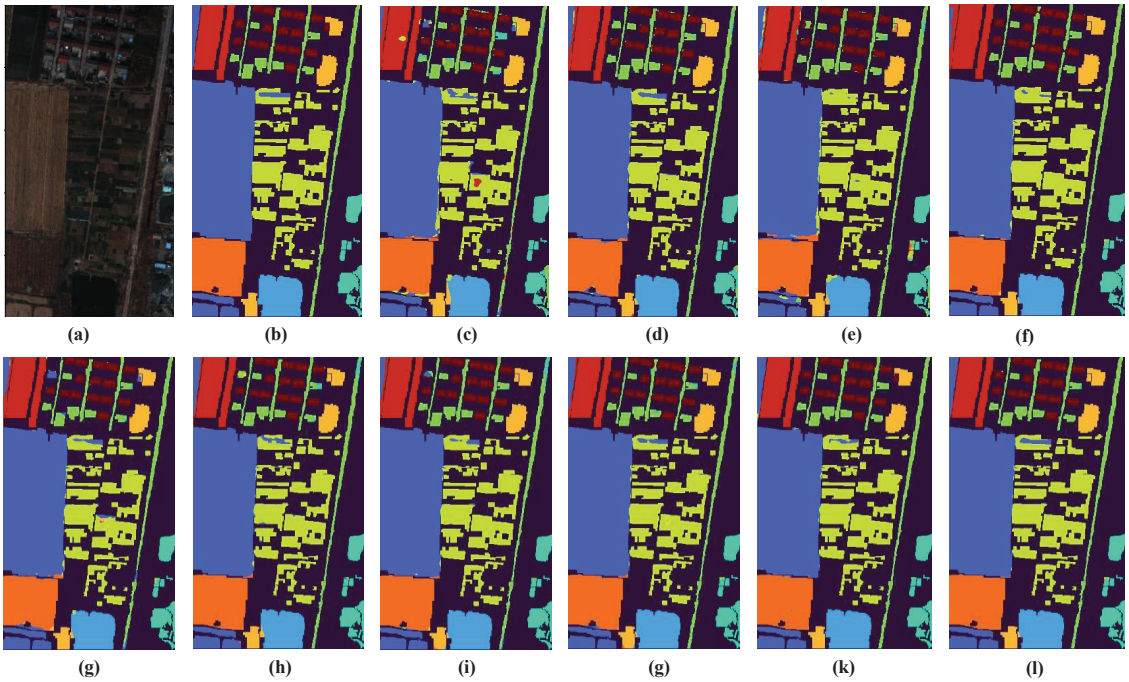


Fig. 10. The classification results of Xuzhou. (a) False color image; (b) Ground truth; (c) SVM; (d) 2D-CNN; (e) 3D-CNN; (f) SSRN; (g) DPRN; (h) HybridSN; (i) OCT-MCNN; (j) SAC-NET; (k) MCNN-CP; (l) CMR-CNN.

TABLE IX
THE AVERAGE VALUE OF THE EIGHT METHODS ON THESE FIVE DATASETS

Methods	SVM [7]	2D-CNN [22]	3D-CNN [30]	SSRN [38]	DPRN [39]	HybridSN [36]	OCT-MCNN [35]	SAC-NET [24]	MCNN-CP [34]	CMR-CNN
OA	84.58 ± 2.10	92.38 ± 2.18	90.29 ± 1.65	94.18 ± 1.41	96.50 ± 1.08	94.27 ± 2.15	93.59 ± 1.46	94.49 ± 0.93	95.36 ± 1.26	**97.72** ± 0.57
AA	84.94 ± 2.08	91.11 ± 2.21	87.29 ± 2.18	94.20 ± 1.47	94.80 ± 1.23	91.56 ± 2.71	92.37 ± 2.01	90.91 ± 1.19	92.97 ± 1.05	**96.04** ± 0.59
Kappa	85.65 ± 2.15	92.14 ± 2.26	89.40 ± 1.81	93.65 ± 1.86	95.75 ± 1.04	93.90 ± 3.00	94.54 ± 1.51	93.50 ± 0.94	93.02 ± 1.21	**97.29** ± 0.59

TABLE X
THE TRAINING/TESTING TIME CONSUMED BY DIFFERENT METHODS TO COMPLETE ONE EPOCH ON THESE FIVE DATASETS(UNIT: SECOND)

Train(s)/Test(s)	SVM [7]	2D-CNN [22]	3D-CNN [30]	SSRN [38]	DPRN [39]	HybridSN [36]	OCT-MCNN [35]	SAC-NET [24]	MCNN-CP [34]	CMR-CNN
Indian Pines	0.12/0.64	2.06/6.62	16.67/120.06	4.45/31.47	0.47/2.03	4.75/6.05	4.66/23.23	0.25/1.85	18.30/33.19	2.27/5.15
University of Pavia	0.07/0.68	1.76/27.10	3.73/134.89	0.92/41.58	0.52/4.03	5.49/26.64	3.69/29.22	0.3/2.76	2.60/20.35	3.49/18.63
Salinas Scene	0.05/1.25	1.48/31.91	2.67/170	0.84/42.50	0.57/8.76	10.95/171	3.00/37.38	0.41/3.06	3.45/51.82	4.26/25.26
KSC	0.03/0.18	3.04/2.49	8.16/12.62	1.24/4.32	0.67/0.85	7.84/2.58	6.72/2.21	0.76/0.65	8.49/2.71	2.95/3.02
xuzhou	0.12/5.78	2.24/40.02	5.70/201.96	2.47/100.91	1.18/21.96	6.58/43.57	4.51/49.19	1.02/19.45	5.88/61.83	5.64/29.93
Average	0.08/1.71	2.12/21.63	7.39/127.91	1.98/44.15	0.68/7.53	7.12/49.97	4.52/28.25	0.55/5.55	7.74/33.98	3.72/16.40

TABLE XI
ABLATION TEST: THE OA, AA, AND KAPPA VALUES OF DIFFERENT COMBINATIONS ON INDIAN PINES (%)

Method	3D-Conv+2D-Conv	3D-Res	2D-Res	3D-Res+2D-Res	3D-Res+AFE	AFE+2D-Res	CMR-CNN Non-Res	CMR-CNN
OA	85.93	97.83	95.45	98.26	98.33	98.01	89.29	**99.08**
AA	79.66	95.99	95.19	97.35	97.37	97.01	79.26	**98.25**
Kappa	84.01	97.53	94.81	98.66	98.10	97.39	87.76	**98.96**

University of Pavia dataset contains more outliers and more indistinguishable small regions. For some HSI classification methods proposed earlier, i.e., SVM, 2D-CNN, and 3D-CNN, the values of OA, AA, and Kappa are all lower in Table V. In addition, they all get more misclassified regions on the prediction maps in Figs. 7(c)-7(e) when the training ratio is lower. At the same time, compared with these three methods and SSRN, the method DPRN with more complex network structure achieves better classification performance in Table V, and the misclassification area caused by it is also less in Fig. 7(g). In Table V, OCT-MCNN achieves the highest value on the AA evaluation metric among all methods. In the case of few training samples, the recently proposed method SAC-NET achieves better classification performance in Fig. 7(j). Compared to the result of MCNN-CP in Fig. 7(k), CMR-CNN achieves a better visual result in Fig. 7(l). This is in agreement with the quantitative result in Table V, that is, CMR-CNN has the greatest OA value.

3) ***The experimental analyses on Salinas***: Table VI shows the quantitative results of different methods on the Salinas dataset. Fig. 8 is the corresponding prediction graphes. Compared with other datasets, the sample distribution of this

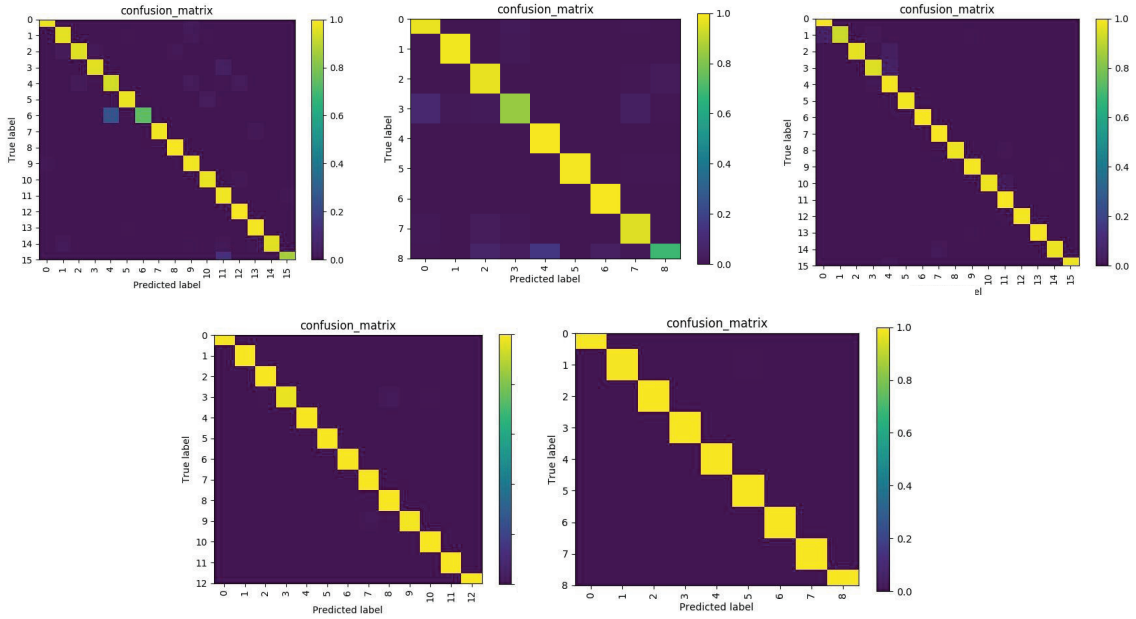


Fig. 11. The confusion matrix using the proposed method on Indian Pines, University of Pavia, Salinas Scene, KSC, and Xuzhou in 1^{st},2^{nd}, 3^{rd}, 4^{th}, and 5^{th} matrix, respectively.

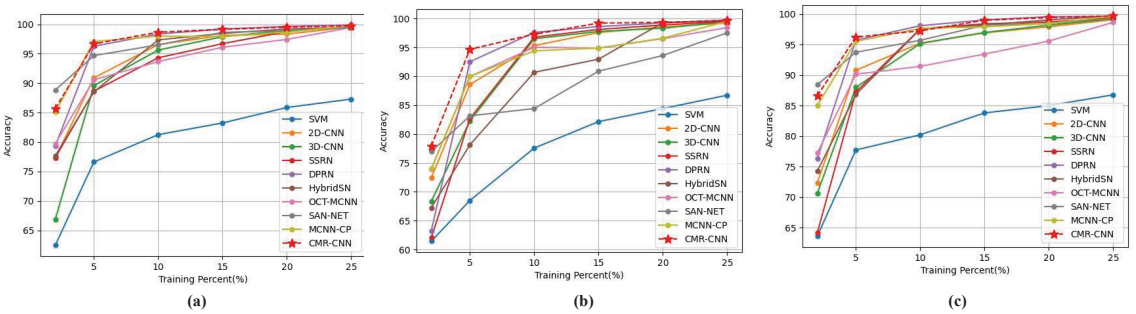


Fig. 12. Accuracy of different methods with different numbers of training samples on Indian Pines. (a) OA; (b) AA; (c) Kappa.

dataset is more regular. In order to better reflect the classification performances of different methods, we here choose 0.5% as the training ratio. In comparison with other methods in Table VI, the proposed method CMR-CNN achieves the best classification results on OA and Kappa. However, on AA, it is 0.08% lower than DPRN. Detailedly, compared with the method HybridSN, the proposed network CMR-CNN improves by 1.26%, 2.25%, 1.40% on OA, AA, and Kappa, respectively in Table VI. The values obtained by the three methods SVM, 2D-CNN, and 3D-CNN in Table VI are not quite different from each other. So, their classification results are also similar to each other in Figs. 8(c)-8(e). Compared with the first three methods, both SSRN and DPRN achieve better classification results with fewer misclassified regions in Figs. 8(f)-8(g). Obviously, compared to SSRN and DPRN, the visual result of HybridSN is better in Fig. 8(h). In Figs. 8(i) and 8(j), in the case of fewer training samples, the classification effect of OCT-MCNN is better than that of SAC-NET. Unfortunately, for the recently proposed method MCNN-CP, the values in Table VI are not ideal, and there also exist many misclassified regions in Fig. 8(k).

4) ***The experimental analyses on KSC***: Table VII reports the experimental results of different methods on this dataset. To visualize the performances of different methods, we further zoom in the rectangles of prediction maps in Fig. 9, where the classes are harder to distinguish than others. Compared with the other HSI classification methods, the proposed method CMR-CNN achieves the highest scores on the three indicators OA, AA, and Kappa. Besides, the prediction map obtained by CMR-CNN in Fig. 9(l) is more accurate in visual performance. In detail, compared with HybridSN, OA, AA, and Kappa are improved 1.51%, 2.45% and 1.68% by CMR-CNN, respectively. Similarly, the classification results of 2D-CNN in Table VII are the worst among these methods, and its prediction in Fig. 9(d) also has a large number of misclassifications. In contrast, the classification results of SVM and 3D-CNN are better in Figs. 9(c) and 9(e). Compared with 3D-CNN, SSRN used the residual structure in the network architecture, and obtained better classification performance in Table VII. With the same training samples, the classification result of OCT-

MCNN is worse than that of MCNN-CP. On the contrary, SAC-NET performs better than OCT-MCNN. It is worth noting that compared to the other eight methods except CMR-CNN, DPRN achieves better classification performance in Table VII, and there are fewer misclassifications in the area framed in Fig. 9(g).

5) ***The experimental analyses on Xuzhou***: To save space, hereinafter, we just briefly analyze the results of these ten methods. Table VIII reports the quantitative results of different methods on this dataset. Compared to the other methods, the proposed method CMR-CNN still achieves the highest scores on the three evaluation metrics. So, once again, the effectiveness of our method is verified. In addition, Fig. 10 shows the visual results of different methods on the dataset. Clearly, the proposed method CMR-CNN achieves the least classification error in Fig. 10(l).

6) ***Confusion matrix and network performance under different training ratios***: Fig. 11 is the confusion matrix related to CMR-CNN on Indian Pines, Pavia University, Salinas, KSC, and Xuzhou respectively. According to the distribution of confusion matrix, we can easily see the proposed method suffers individual prediction errors on the first two datasets, but the prediction results on the latter three datasets are better. Fig. 12 shows the test results of different methods on the India Pines dataset at different training ratios. Obviously, when the training ratio increases, the accuracy of different HSI classification methods increases as well. However, no matter whether the training ratio is high or low, the accuracy of traditional methods such as SVM, 2D-CNN, and 3D-CNN is always lower than that of the methods proposed in recent years, like SSRN. When the training rate is high (such as 20%), the methods have little difference in experimental results. But, SSRN and DPRN show a large drop as the training ratio decreases. When the training rate is low (5%), the proposed method CMR-CNN shows a better classification performance, which indirectly reflects that CMR-CNN can still extract effective discriminant information with few training samples.

Table IX further lists the average values of OA, AA, and Kappa of these ten methods on the five datasets. Overall, from this table we can intuitively see that the proposed method achieves the best classification result. Particularly, compared with 2D-CNN and DPRN wherein only the spatial information is used, the OA, AA, and Kappa values are respectively increased by 5.34%, 4.93%, 5.15% and 1.22%, 1.24% and 1.54%. Compared with 3D-CNN and SSRN wherein only the spectral information is used, the OA, AA, and Kappa values are respectively increased by 7.43%, 8.75%, 7.89% and 3.54%, 1.84% and 3.64%. So, this directly verifies that, in comparison with the strategy that only adopts the spatial or spectral information, the strategy using spatial and spectral information at the same time is more appropriate for HSI classification.

Detailedly, in Table IX, HybridSN has higher OA, AA, and Kappa values than 2D-CNN and 3D-CNN. It directly proves that without adding other strategies, a single convolutional network framework cannot fully extract the discriminative information in the feature map effectively. Compared with 2D-CNN and 3D-CNN, SSRN and DPRN using residual structures also achieve better quantitative results, which proves that residual structures enable us to further extract more effective classification features. The classification performance of CMR-CNN is superior to that of SSRN and DPRN, due to the simultaneous utilization of 3D and 2D residual structures. Compared with MCNN-CP, OCT-MCNN, the classification result of CMR-CNN is better as well, which proves the effectiveness of proposed method again. Even so, CMR-CNN is more time-consuming than some methods, such as SVM, DPRN, and SAC-NET. Table X lists the training/testing time-consumption (unit: second/(s)) that one epoch is completed by different methods. Averagely, the proposed method takes 16.40s for training and 3.72s for testing one epoch. The reason is that, the entire model of CMR-CNN is composed of convolutional layers and three 3D residual structures, which expends a certain amount of computing resources.

E. Discussion

1) ***Model convergence rate***: For saving space, we here take MCNN-CP for example. Figs. 13 and 14 respectively show the convergence speed of MCNN-CP and CMR-CNN on Indian Pines, University of Pavia, Salinas Scene, KSC and Xuzhou five datasets. From Figs. 13 and 14, we can see that in the same dataset, the proposed method converges faster than MCNN-CP; for the same epoch, the OA value of CMR-CNN is higher, and the inflection point in the curve is more few. In this regard, we conduct an analysis. In convolutional neural networks, there are many factors that affect the speed of network convergence and robustness. On the actual loss surface, some local minimas slow down the convergence. Another situation affecting the speed of convergence is the saddle point. The shape of saddle point is similar to that of a saddle. The gradient is the smallest in one direction and the largest in the other. It is easy to oscillate back and forth in the direction of maximum value, slow down the convergence speed, and even cause incorrect convergence. Residual structure actually provides a "shortcut" for gradient propagation, allowing gradients to skip intermediate layers and pass directly to deeper layers. In fact, it uses the recommended skip connection. This alleviates the problem of vanishing gradients, which speeds up convergence. The proposed network model requires less training time to reach the desired value, i.e., CMR-CNN just takes almost half as long as MCNN-CP. Moreover, the curve in Fig. 13 has fewer inflection points than the curve in Fig. 14, which further verifies that the residual structure can not only ensure the accuracy of the network model, but also improve the convergence speed and robustness of the model.

2) ***Ablation experiments***: The proposed network CMR-CNN is mainly constructed on SSRN and HybridSN. To better find the reason behind CMR-CNN, in the following, we do some ablation experiments.

Table XI shows the performance of each module in our CMR-CNN on the India Pines dataset when the training ratio is 10%. It should be noted that, in order to detect the role of each module in CMR-CNN, we only compare and analyze the overall network performance of CMR-CNN, and do not

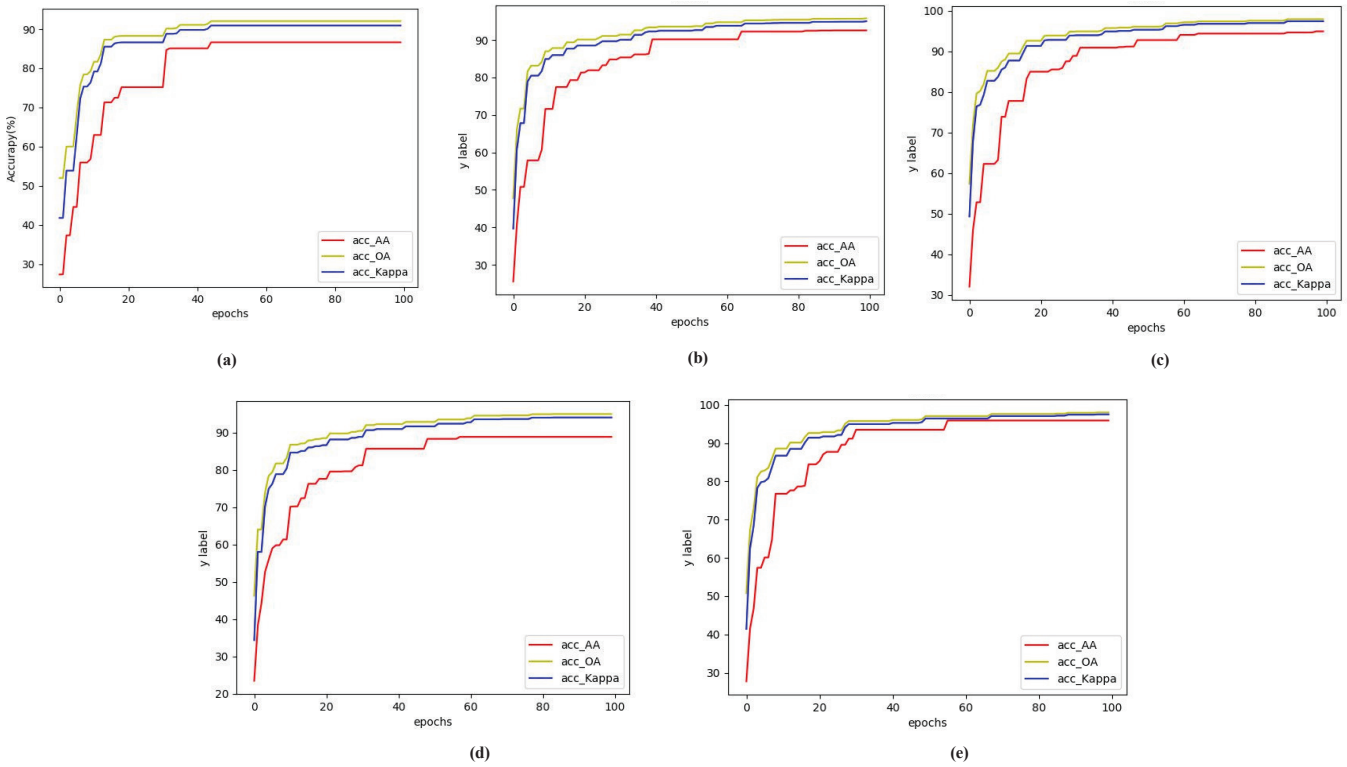


Fig. 13. The curves of MCNN-CP between OA, AA, Kappa, and epochs used to reach the maximum value. (a) Indian Pines; (b) University of Pavia; (c) Salinas Scene; (d) KSC; (e) Xuzhou.

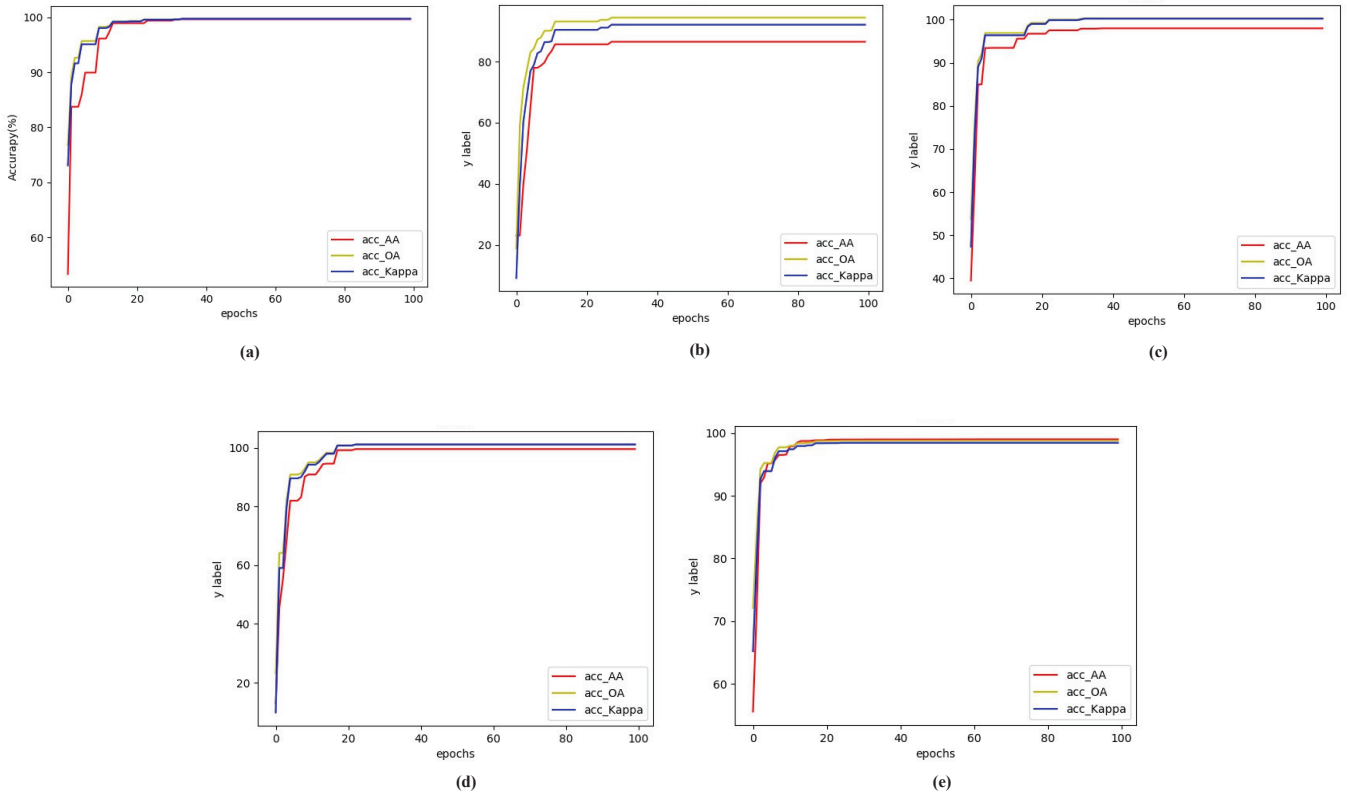


Fig. 14. The curves of CMR-CNN between OA, AA, Kappa, and epochs used to reach the maximum value. (a) Indian Pines; (b) University of Pavia; (c) Salinas Scene; (d) KSC; (e) Xuzhou.

compare and analyze with other methods. The first column 3D-Conv+2D-Conv is the classification result obtained by CMR-CNN that removes the residual structure and AFE that just leave the 3D convolution and 2D convolution; the second column 3D-Res removes the 2D residual structure and AFE, leaving only the 3D residual structure; the third column 2D-Res removes the 3D residual structure and AFE, leaving only the 2D residual structure; the fourth column 3D-Res+2D-Res means the 3D residual structure and 2D residual structure after removing AFE; the fifth column 3D-Res+AFE is the combination of the 3D residual structure and AFE; the sixth column 3D-Res+AFE is the combination of the 3D residual structure and AFE; the seventh column CMR-CNN Non-Res is the result obtained by removing the residual structure from the network; the nighth column is the result obtained by the overall network CMR-CNN.

In Table XI, by respectively comparing the first and second columns, the first and fourth columns, it is not difficult for us to find out that the classification performance of network is very poor without the residual structure: 3D-Conv+2D-Conv in OA, AA, and Kappa are 85.93%, 79.66%, and 84.01%, respectively, but 3D-Res+2D-Res is correspondingly 98.26%, 97.35%, and 98.66%, reflecting that the residual structure is beneficial to help the network extract deep information and improve the classification performance. The same conclusion can be obtained by analyzing CMR-CNN Non-Res and CMR-CNN. Comparing 3D-Conv+2D-Conv and CMR-CNN Non-Res, we can also easily demonstrate the effectiveness of AFE on HSI classification. Compared to CMR-CNN, when only the 3D-Res structure is retained, OA and Kappa are reduced by nearly 1.5%, and OA is reduced by nearly 2.3%. Comparing the 2D-Res structure (only the spatial information is used) and the complete CMR-CNN, we can also find that, CMR-CNN has higher classification accuracies which OA increased by 3.63%, AA increased by 4.06%, and kappa increased by 4.15%. So, using spectral and spatial information together is more apt for HSI classification.

IV. Conclusion

In this paper, we proposed a novel convolutional neural network named CMR-CNN for HSI classification. First, we used the 3D residual structure to extract the spectral information of HSI and the 2D residual network to extract the spatial information of HSI. Subsequently, two layers of 3×3 convolution kernels were used to form AFE for bridging 3D and 2D residual structures together, which also allows us to further extract more hidden features of pixels. After that, CMR-CNN was proposed for HSI classification via fusing them. Experiments show that, i) the classification accuracy can be significantly improved when the spectral and spatial information are simultaneously used; ii) residual structures enable the network to extract more effective classification features; iii) the proposed method CMR-CNN has a better classification performance than the other SOTA methods. In spite of this, future work still need to be done on how to remove the influence of noise, as the spectrometer is easily affected by factors such as weather and light when collecting images. Besides, we will also try to further optimize CMR-CNN for reducing the time-consumption.

References

[1] Ding. Y, Ma. Z, Wen. S, Xie. J, Chang. D., Si. Z, and Ling. H. "AP-CNN: Weakly Supervised Attention Pyramid Convolutional Neural Network for Fine-Grained Visual Classification," in *IEEE Transactions on Image Processing*, vol. 30, pp. 2826-2836, 2021.

[2] X. Han, H. Zhang, and J. Ma. "Classification Saliency-Based Rule for Visible and Infrared Image Fusion," in *IEEE Transactions on Computational Imaging*, vol. 7, pp. 824-836, 2021.

[3] S. Zhao, Z. Zhang, T. Zhang, W. Guo and Y. Luo, "Transferable SAR Image Classification Crossing Different Satellites Under Open Set Condition," in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022.

[4] D. Landgrebe, "Hyperspectral Image Data Analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002.

[5] W. Lin, C. I. Chang, L. C. Lee, Y. Wang, X. Bai, M. Song, C. Yu, and S. Li, "Band Subset Selection for Anomaly Detection In Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 4887–4898, 2017.

[6] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral Remote Sensing Data Analysis and Future Challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[7] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images with Support Vector Machines," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[8] J. Haut, M. Paoletti, A. Paz-Gallardo, J. Plaza, and A. Plaza, "Cloud Implementation of Logistic Regression for Hyperspectral Image Classification," in *International Conference Computsyional and Mathematical Methods in Science and Engineering (CMMSE)*, vol. 3, pp. 1063–2321, 2017.

[9] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of The Random Forest Framework for Classification of Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005.

[10] J. M. Haut, M. Paoletti, J. Plaza, and A. Plaza, "Cloud Implementation of The K-Means Algorithm for Hyperspectral Image Analysis," *The Journal of Supercomputing*, vol. 73, no. 1, pp. 514–529, 2017.

[11] G. Camps-Valls and L. Bruzzone, "Kernel-Based Methods for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[13] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring Hierarchical Convolutional Features for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6712–6722, 2018.

[14] H. Lee and H. Kwon, "Going Deeper with Contextual CNN for Hyperspectral Image Classification," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, 2017.

[15] X. He, Y. Chen, and P. Ghamisi, "Heterogeneous Transfer Learning For Hyperspectral Image Classification Based on Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3246–3263, 2019.

[16] X. Xu, Z. Shi, and B. Pan, "A New Unsupervised Hyperspectral Band Selection Method Based on Multiobjective Optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2112–2116, 2017.

[17] A. Marinoni, G. C. Iannelli, and P. Gamba, "An Information Theory-Based Scheme for Efficient Classification of Remote Sensing Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5864–5876, 2017.

[18] A. Marinoni and P. Gamba, "Unsupervised Data Driven Feature Extraction by Means of Mutual Information Maximization," *IEEE Transactions on Computational Imaging*, vol. 3, no. 2, pp. 243–253, 2017.

[19] J. Zhang, P. Liu, F. Zhang, and Q. Song, "Cloudnet: Ground-Based Cloud Classification With Deep Convolutional Neural Network," *Geophysical Research Letters*, vol. 45, no. 16, pp. 8665–8672, 2018.

[20] J. Yang, Y. Zhao, J. C.-W. Chan, and C. Yi, "Hyperspectral Image Classification Using Two-Channel Deep Convolutional Neural Network," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 5079–5082.

[21] Z. Gong, P. Zhong, Y. Yu, W. Hu, and S. Li, "A CNN With Multiscale Convolution and Diversified Metric for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3599–3618, 2019.

[22] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep Supervised Learning for Hyperspectral Data Classification Through Convolutional Neural Networks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015, pp. 4959–4962.

[23] Y. Xu, B. Du, and L. Zhang, "Beyond The Patchwise Classification: Spectral-Spatial Fully Convolutional Networks for Hyperspectral Image Classification," *IEEE Transactions on Big Data*, vol. 6, no. 3, pp. 492–506, 2019.

[24] Y. Xu, B. Du, and L. Zhang, "Self-Attention Context Network: Addressing The Threat of Adversarial Attacks for Hyperspectral Image Classification," *IEEE Transactions on Image Processing*, vol. 30, pp. 8671–8685, 2021.

[25] P. Duan, P. Ghamisi, X. Kang, B. Rasti, and R. Gloaguen, "Fusion of Dual Spatial Information for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–13, 2020.

[26] Y. Shen, S. Zhu, C. Chen, Q. Du, L. Xiao, J. Chen, and D. Pan, "Efficient Deep Learning of Nonlocal Features for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 2031-2042, 2020.

[27] Z. Lin, Y. Chen, P. Ghamisi, and JA. Benediktsson "Generative Adversarial Networks for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5046-5063, 2018.

[28] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral–Spatial Feature Tokenization Transformer for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022.

[29] U. A. Bhatti, Z. Yu, J. Chanussot, Z. Zeeshan, L. Yuan, W. Luo, S. A. Nawaz, M. A. Bhatti, Q. U. Ain, and A. Mehmood, "Local Similarity-Based Spatial–Spectral Fusion Hyperspectral Image Classification with Deep CNN and Gabor Filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022.

[30] Y. Li, H. Zhang, and Q. Shen, "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network," *Remote Sensing*, vol. 9, no. 1, pp. 67, 2017.

[31] M. He, B. Li, and H. Chen, "Multi-Scale 3D Deep Convolutional Neural Network for Hyperspectral Image Classification," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3904–3908.

[32] Z. Li, L. Huang, and J. He, "A Multi-Scale Deep Middle-Level Feature Fusion Network for Hyperspectral Classification," *Remote Sensing*, vol. 11, no. 6, pp. 695, 2019.

[33] 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[34] J. Zheng, Y. Feng, C. Bai, and J. Zhang, "Hyperspectral Image Classification Using Mixed Convolutions and Covariance Pooling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 522–534, 2020.

[35] Y. Feng, J. Zheng, M. Qin, C. Bai, and J. Zhang, "3D Octave and 2D Vanilla Mixed Convolutional Neural Network for Hyperspectral Image Classification with Limited Samples," *Remote Sensing*, vol. 13, no. 21, p. 4407, 2021.

[36] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "Hybridsn: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2019.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[38] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2017.

[39] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep Pyramidal Residual Networks for Spectral–Spatial Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 740–754, 2018.

[40] X. Kang, X. Xiang, S. Li, and J. A.Benediktsson, "PCA-Based Edge-Preserving Features for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7140–7151, 2017.

[41] Z. Zhang and M. M. Crawford, "A Batch-Mode Regularized Multimetric Active Learning Framework for Classification of Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6594–6609, 2017.

[42] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, "Imaging Spectroscopy and The Airborne Visible/Infrared Imaging Spectrometer (Aviris)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.

[43] B. Kunkel, F. Blechinger, R. Lutz, R. Doerffer, H. Van der Piepen, and M. Schroder, "Rosis (Reflective Optics System Imaging Spectrometer)-A Candidate Instrument for Polar Platform Missions," in *Optoelectronic Technologies for Remote Sensing From Space*, vol. 868, pp.134–141, 1988.

[44] K. Tan, F. Wu, and X. Wang, "Xuzhou Hyspex Dataset," 2018. [Online]. Available: https://dx.doi.org/10.21227/t3c9-h862

Zhen Yang received the B.S.degree in Automation from Changchun Institute of Technology in 2008, the M.S. degree in Automation from Qingdao University of Science & Technology in 2011, and the Ph.D. Degree in Shanghai Jiao Tong University in 2016. He is an Associate Professor of the School of Communication and Electronics in Jiangxi Science and Technology Normal University. His research interests include computer vision, machine learning, object recognition, and image classification.

Zhipeng Xi received the B.S.degree in Automation from Taiyuan Institute of Technology in 2016. He is pursuing the School of Communication and Electronics, Jiangxi Science and Technology Normal University. His research interests include computer vision, machine learning, and image classification.

Tao Zhang (Member, IEEE) received the B.Sc. degree in electronic information engineering from Huainan Normal University, Huainan, China, in 2011, and the M.Sc. degree in communication and information system from Sichuan University, Chengdu, China, in 2014, and the Ph.D. degree in control science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019.

From 2017 to 2018, he was with the Department of Geoinformatics. The KTH Royal Institute of Technology, Stockholm, Sweden, as a joint Ph.D. Student. From 2019 to 2021, he worked as a Post-Doctoral with the Department of Electronics, Tsinghua University, Beijing, China. He is currently working as an Assistant Professor with the Shanghai Key Laboratory of Intelligent Sensing and Recognition, Shanghai Jiao Tong University. His research work focuses on object recognition, SAR/PolSAR image processing, and machine learning.

Weiwei Guo (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2005, 2007, and 2014, respectively. He has been a Visiting Ph.D. Student in the area of multimedia and computer vision at Electrical Engineering and Computer Science (EECS), Queen Mary University of London, London, U.K., between 2008 and 2010, and a Post-Doctoral Researcher with Shanghai Jiao Tong University, Shanghai, China, in the area of remote sensing image interpretation.

He is currently an Associate Professor with the Center of Digital Innovation, Tongji University. His main research interests lie in pattern recognition and machine learning and their applications in the fields of computer vision, remote sensing image understanding, and Human–Computer Interaction (HCI).

Zenghui Zhang (Senior Member, IEEE) received the B.Sc. degree in applied mathematics, the M.Sc. degree in computational mathematics, and the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2001, 2003, and 2008, respectively.

He is currently a Professor with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include radar signal processing, microwave imaging, SAR image interpretation, target detection, and recognition.

Heng-Chao Li (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008, all in information and communication engineering. Currently, he is a Full Professor with the School of Information Science and Technology, Southwest Jiaotong University. From November 2013 to October 2014, he has been a Visiting Scholar working with Prof. William J. Emery with the University of Colorado Boulder, Boulder, CO, USA. His research interests include statistical analysis of SAR images, remote sensing image processing, and pattern recognition. He has authored over 100 JCR journal papers, such as published in the ISPRS Journal of Photogrammetry and Remote Sensing, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and the Pattern Recognition.

Prof. Li was a recipient of the 2018 Best Reviewer Award from IEEE Geoscience and Remote Sensing Society for his service to the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), Sichuan Natural Science Award, CSIG Natural Science Award, CSGPC Science and Technology Progress Award. He is serving as an Associate Editor for the IEEE JSTARS, and is an Editorial Board Member of the Journal of Southwest Jiaotong University and the Journal of Radars. Moreover, he has served as a Guest Editor of Special Issues of the Journal of Real-Time Image Processing, the IEEE JSTARS, and the IEEE JMASS, a Program Committee Member for the 26th International Joint Conference on Artificial Intelligence (IJCAI-2107) and the 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MULTITEMP-2019), and the Session Chair for the 2017 International Geoscience and Remote Sensing Symposium (IGARSS-2017), the 2019 Asia-Pacific Conference on Synthetic Aperture Radar (APSAR-2019), and the 2021 CIE International Conference on Radar (RADAR-2021). He is also an Active Reviewer of over 30 international journals.