Signal, Image and Video Processing (2024) 18:3815–3827
https://doi.org/10.1007/s11760-024-03044-0

ORIGINAL PAPER

Hybrid 2D–3D convolution and pre-activated residual networks for hyperspectral image classification

Huanhuan Lv[1] · Yule Sun[1] · Hui Zhang[1] · Mengping Li[1]

Received: 31 October 2023 / Revised: 23 December 2023 / Accepted: 20 January 2024 / Published online: 20 February 2024

Abstract
The utilization of Convolutional Neural Networks (CNNs) in hyperspectral image (HSI) classification has become commonplace. However, traditional CNNs cannot fully extract the features of HSI and are prone to gradient vanishing when the network layer is deepened. We suggest a 2D–3D hybrid convolution and pre-activated residual networks-based HSI classification (HSIC) approach to tackle these problems. Firstly, the joint spatial–spectral features of HSI are extracted by a two-layer 3D convolution. Secondly, combining the advantages of 2D and 3D convolution to construct a spatial–spectral feature extraction module based on pre-activated residual networks, which can accelerate the convergence speed of the model while enhancing the capability of advanced spatial semantic feature extraction of HSI. Then, multiple residual modules are connected to take advantage of the different forms of features extracted by each convolutional layer, while multi-feature fusion is performed between blocks to achieve feature complementarity. Finally, a long-distance residual connection is introduced to fuse the shallow and deep features effectively, which further strengthens the expression ability of features. The results of the experiments conducted on three HSIs show that the overall classification accuracy of the model reaches 99.56%, 99.45% and 99.43%, respectively, when 10%, 1% and 1% of samples are randomly selected for training in each ground object class. Compared with other related CNN-based HSI classification models, our model can obtain higher classification accuracy. Consequently, the suggested method is capable of achieving feature reuse and obtaining deep high-level spatial–spectral features with superior discriminative and robustness, and its classification performance is superior to that of existing state-of-the-art methods.

Keywords Hyperspectral image classification · Hybrid convolution · Pre-activated residual networks · Feature fusion · Long-distance residual connection

1 Introduction

Hyperspectral Image (HSI) has the characteristics of map unity, rich spatial information, wide range of spectral bands, high resolution, etc., which enhances the ability of remote sensing (RS) to observe the ground and the ability of feature identification, and has been widely utilized in many fields, such as military exploration [1], environmental monitoring, precision agriculture and medical diagnosis [2], etc. Hyperspectral image classification (HSIC) is one of the basic problems in RS image processing, and it is also the basis and key of RS image analysis and interpretation. The main goal of HSIC is to recognize the actual ground object from the image, i.e., to give a unique category to each pixel in the image. In early HSIC, traditional machine learning algorithms such as support vector machine (SVM) [3], random forest (RF) and logistic regression (LR) mostly focus only on spectral information. However, the same ground object has spectral differences in different spaces, and different ground objects may also have similar spectral characteristics. Therefore, since such methods ignore the rich spatial structural features, resulting in classification results that often contain a large amount of noise, it is difficult to achieve accurate classification of complex features [4], so integrating spectral and spatial information is an effective way to improve the HSIC results. Considering that there is often interrelated information between spatially neighboring image elements, methods such as Markov Random Fields [5] and Morphological Attribute Profiles have been used to obtain the spectral

✉ Hui Zhang
03013@zjhu.edu.cn

[1] School of Information Engineering, Huzhou University, Huzhou 313000, China

and spatial information of HSI with good results. However, the above hand-crafted spatial–spectral features rely heavily on rich expertise, and the shallow features extracted have limited impact on the enhancement of classification precision.

Over the past few years, DL-based image classification techniques have become increasingly popular in HSIC [6]. By using these methods, it is possible to automatically extract abstract features from low-level semantics to high-level semantics in images with better representation performance, which makes the subsequent classification results more accurate. Convolutional Neural Network (CNN) is a prime example of a DL model that exhibits superior performance in feature extraction and classification. 1DCNNs are only able to identify the spectral characteristics of images for HSI pixel-level categorization [7]. In order to make the most of the spectral and spatial features of HSIs, scholars have proposed 2D and 3D CNN models in succession. Among them, Zhao et al. [8] used a two-dimensional CNN (2DCNN) for HSIC, which considered the important role of the spatial information of HSI. However, extracting spectral and spatial features individually does not take full advantage of the combination of spectral and spatial data and requires complex pre-processing. Considering that HSI has a 3D cubic structure, Li et al. [9] directly used 3D convolution to obtain the spatial–spectral features of HSI and achieved the improvement of HSIC accuracy. 3DCNN is more computationally intensive than 2DCNN and has higher memory requirements. Zheng et al. [10] sought to simplify the model while still maintaining a high level of classification accuracy. They created a mixed convolutions and covariance pooling model (MCNN-CP) by combining the advantages of 3DCNN and 2DCNN, and verified the potential of hybrid convolution in HSIC. Fırat et al. [11] introduced a depthwise separable convolution based on a 2D–3D hybrid convolution model, which effectively improves the accuracy of HSIC. However, the above models tend to ignore the variability in the importance of different features affecting the classification results. Considering the difference in the contribution of different types of features to the classification results [12], Shi et al. [13] suggested a 3D coordination attention mechanism network (3DCAMNet). Specifically, they used a combination of 3D convolution and attention mechanism to ascertain the disparity in significance between various spectral bands, which ultimately achieves the improving of the model performance. However, with the deepening of the network structure, the performance gradually decreases and the degradation problem easily occurs.

Residual connections in Residual Networks (ResNet) [14] can deepen the number of network layers and optimize the model structure. Qing et al. [15] introduced residual connections in 2DCNN to improve the HSIC accuracy. He et al. [16] combined 3DCNN and residual connection to construct a HSIC model, which still has some room for improvement in classification performance due to not fully utilizing the spatial–spectral information of HSI. To address the above problems, Cao et al. [17] built a comprehensive hybrid convolution residual network (BHModel) to improve the feature learning of HSI, which uses 2D–3D convolutional mixing to drastically decrease the amount of parameters, thus making the network architecture simpler. The single way of feature fusion and the underutilization of shallow features lead to certain limitations in the improvement of classification effect. Dang et al. [18] built a lightweight model (JPModel) for HSIC. The method reduces the network parameters and increases the paths for learning features by combining residual connection with depthwise separable convolution, which further improves the classification accuracy. He et al. [19] used a multi-scale residual network (SSMRN) to obtain the spectral-spatial information of HSI, which effectively learns the target features. Lei et al. [20] introduced capsule residual blocks to increase the depth of the network, and although more critical information was extracted, the complexity of the proposed MS-CapsNetW was relatively high.

The residual network described above can cope with the phenomenon of degradation, but it is also hindered by the slow network speed and underutilization of extracted features due to direct convolution operation on the data. Pre-activated residual connection can reduce the complexity of the model and make the model converge faster. So introducing the pre-activation mechanism [21, 22] into the residual network can improve the network structure of the original residual module, which can not only improve the training speed, but also obtain deeper features with stronger representativeness in the joint learning of spatial–spectral features. To address the problem of insufficient feature utilization, for deep network models, the long and short distance residual connection approach [23, 24] can solve the problem of gradient vanishing on the one hand, and on the other hand, for the loss of feature information caused by convolutional operation, this approach can also achieve the connection between the bottom and the top network, so as to ensure the stability of the training of the deep network. Consequently, the strategy partially compensates for the missing data and contributes to the enhancement of the feature fusion capability.

Inspired by the above research works, this study proposes a HSIC model PMCRNet on the basis of 2D–3D hybrid convolution and pre-activated residual network. PMCRNet can effectively compensate for the defect of incomplete feature extraction and enhance the computational performance. The major contributions of this study are as follows.

1. By utilizing a combination of hybrid convolution and a pre-activated residual module, we can obtain the deep

spatial–spectral joint features of HSI, which can accelerate the convergence speed of the model and minimize the amount of parameter computation while enhancing the ability of advanced spatial-semantic feature extraction.
2. We adopt the long and short distance residual connection to effectively fuse shallow and deep information to obtain advanced semantic features. This approach further increases the expressiveness of the features and addresses the gradient vanishing issue in deep network when the number of layers increases.
3. We experimentally validate and comparatively analyze the model of this study with seven other state-of-the-art related models on three publicly available HSI datasets. The HSIC experimental results show that the proposed PMCRNet outperforms other HSIC methods.

2 Residual neural network

In classification tasks, shallow networks have limited feature extraction capabilities, so more complex and rich features need to be learnt by building deeper networks. However, as the network layers continue to deepen and become saturated there will be a decrease in model performance. Therefore, relying solely on increasing the number of network layers will not necessarily improve classification accuracy. He et al. proposed Resnet to effectively avoid the problem of gradient degradation. The core of ResNet is Residual Building Block (RBB), which mainly consists of convolutional layers, batch homogenization layer, and activation function ReLU.

2.1 Convolutional layer

The convolutional layer is employed to acquire the feature information from the input, which is composed of multiple convolutional units. The back propagation algorithm optimizes the parameters of each convolutional unit. The features are extracted by regular shifting and convolution operations on the input image by different sized receptive fields. The computational process of the convolutional layer is defined as

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

where x_j^l represents the jth feature map of layer l, $f(\cdot)$ is the activation function, M_j denotes the set of input feature maps, x_i^{l-1} represents the ith feature map of layer l-1, k_{ij}^l denotes the weight from the ith convolution kernel in layer l-1 to the jth convolution kernel in layer l, * denotes a convolution operation, b_j^l denotes the bias term of the jth convolution kernel of layer l.

2.2 Batch normalization

For the network, Batch Normalization (BN) can speed up the convergence and improve the generalization ability. By deriving the mean and variance of each batch of data for normalization, it is possible to make each layer of information within the effective range that can be passed on to the next layer, and the process is computed through

$$w = \frac{v - c(v)}{\sqrt{d(v)^2 + \varepsilon}} \tag{2}$$

where w is the activation value normalized to the network, v is the activation value of a particular layer of the network, $c(v)$ and d(v) represent the mean and the variance, respectively. ε is a constant close to zero.

2.3 Activation function

Rectified Linear Unit (ReLU) is the most commonly used activation function in neural networks, which can help to reduce the issue of vanishing gradients and speed up the convergence of networks. The function is calculated as follows:

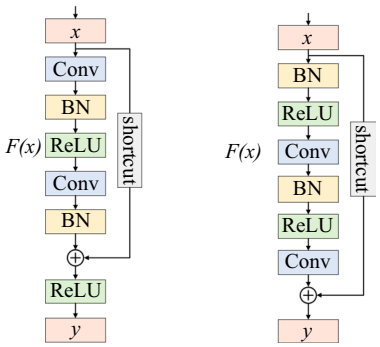
$$\mathrm{ReLU}(z) = \max(0, z) \tag{3}$$

where z and $\boldsymbol{ReLU}(z)$ are the input and output of the activation function, respectively.

2.4 Residual module

RBB uses shortcut connections to skip blocks of convolutional layers for efficient transfer of information, avoiding gradient explosion and vanishing, which helps construct deeper neural network structures and enhances the ultimate efficiency of the network.

Figure 1a shows the original RBB structure, and the execution path of residual is "Input $x \rightarrow$ Convolution layer $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Convolution layer $\rightarrow$ BN $\rightarrow$ Output $F(x)$". Directly performing convolution operation on the data will increase the computational complexity of network training and slow down the training network. He et al. [24] improved the original residual module and proposed Pre-activated Residual Building Block (PARBB), the specific structure is shown in Fig. 1b. The execution path of residual is "Input $x \rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Convolutional layer $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Convolutional layer $\rightarrow$ Output $F(x)$". The pre-activated residual approach puts the BN layer and the activation function before the convolution layer, which



(a) RBB unit structure (b) PARBB unit structure

Fig. 1 Different types of residual unit structures

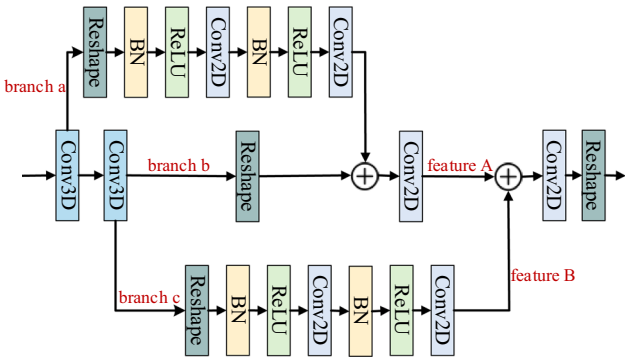


Fig. 2 Hybrid convolutional residual module (HCRM)

expands the range of constant mapping, makes the transmission of information smoother and the network more stable, and helps to improve the feature learning capacity of the network. The computation of the residual module is expressed as follows:

$$y = x + F(x) \tag{4}$$

where x denotes the input of RBB, y denotes the desired output of RBB. The residual, denoted as $F(x)$, signifies the disparity between the desired output and the input.

3 Proposed method

3.1 Hybrid convolutional residual module

HSI is a 3D cube image whose rich and fine spatial and spectral features can be extracted using CNN. When using 2DCNN as the HSIC model, the original HSI needs to be preprocessed, leading to a decrease in spectral dimensional information. 3D convolution can directly take HSI as the input to the network without complex preprocessing, and it can extract spatial and spectral features at the same time. However, 3DCNN needs more parameters to learn, which has the limitation of high computational complexity. And there is a considerable amount of redundant information and noise in HSI band, so the 3D convolution alone does not yield ideal HSIC results. To address these problems, we combine 2D convolution, 3D convolution and PARBB to design a hybrid convolution residual module (HCRM), which can effectively make up for the defects of incomplete feature extraction and improve the computational efficiency. The structure of HCRM is given in Fig. 2.

Conv3D and Conv2D denote 3D convolutional layer and 2D convolutional layer, respectively. BN denotes batch normalization. ReLU is activation function. The first Conv3D is used to implement spatial–spectral joint features learning on the input. Then, in branch a, to make the data dimensions satisfy the requirements of Conv2D, the feature map is concatenated in the spectral and channel dimensions and thus reconstructed (reshape) from a 4D tensor to a 3D tensor, which is fed into a PARBB for deep feature learning. In branch b, to enhance the information transfer, the further extracted spatial–spectral features from the second Conv3D are reshaped and element-wise added with the features obtained from branch a. The fused features are then fed into Conv2D to achieve spatial information enhancement after which we get feature A. On the other hand, to reduce the complexity and speed up the convergence, the features are first reshaped into a 3D tensor and then input into another PARBB to get the feature B in branch c. Finally, feature A and feature B are fused by element-wise adding and reshaped into a 4D tensor as output. HCRM is able to strengthen the ability to learn spatial–spectral joint features and features of different abstraction levels while ensuring the effective delivery of information, which in turn improves the HSCI effect.

3.2 HSIC model based on 2D–3D hybrid convolution and pre-activated residual

Figure 3 shows the basic framework of the proposed PMCR-Net. It mainly consists of five parts, feature extraction, HCRMs, feature fusion, full residual learning, and classification.

1. Feature extraction. This part is composed of two Conv3D. The first one is composed of 32 convolutional kernels of size 3×3 to extract shallow features, which is beneficial to retain more positional and detail information. The second one adopts 32 convolutional kernels of size 3×3 to capture deeper information of the HSI. The result of a single 3D convolution kernel on a block of 3D HSI data is a 3D tensor, whereas the feature map extracted by multiple 3D convolution kernels can be regarded as a 4D tensor, and thus the final output of the feature map in this part is a 4D tensor.

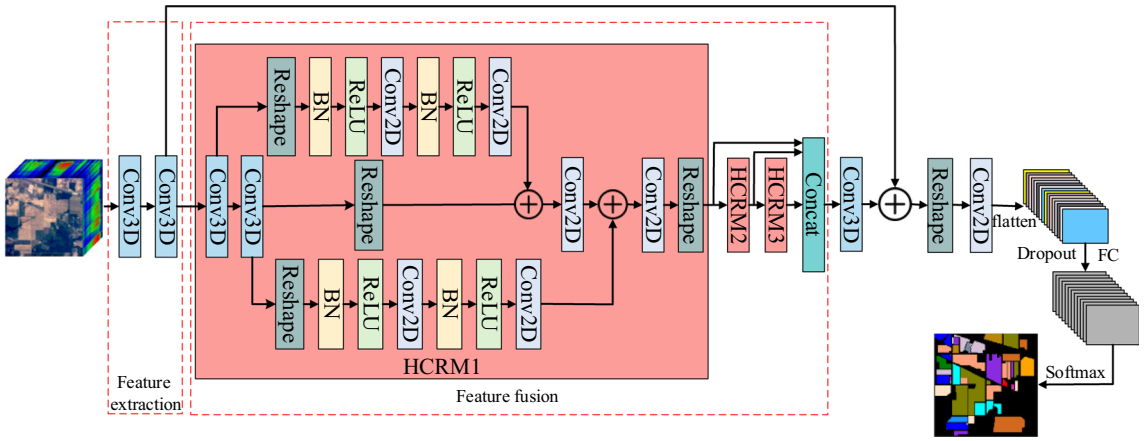


Fig. 3 The structure of PMCRNet

2. HCRMs. After the feature extraction part, higher-order features at different semantic levels are further extracted by three HCRMs. Each HCRM is a tandem structure consisting of two Conv3D and two Conv2D. Based on the PARBB, the first Conv3D span connects the first Conv2D and the second Conv3D span connects the second Conv2D. This design not only mines the useful spatial–spectral information in HSI at a deeper level, but also reduces the computational complexity.
3. Feature fusion. To reduce the loss of HSI features and enhance the comprehensive representation of features at different levels, the features output from each HCRM are input into the Concat layer for splicing. Then, a Conv3D consisting of 32 convolutional kernels of size 3×3 is used as a transition layer to ensure that the number of features outputted by the network is the same as the number of shallow features, so as to obtain fused features with stronger representation capability.
4. Full residual learning. It consists of both pre-activated short-range residual connections and long-range residual connections. The use of short-distance residual connections can alleviate the training problem of the deep network. The long-distance residual connection is used to fuse the output of the feature extraction part and the output of the feature fusion part to achieve the effective extraction of deep features and shallow information. The above structure can supplement the loss information to a certain extent and improve the fusion performance of PMCRNet, thus obtaining better HSIC results.
5. Classification. First, a Conv2D with 32 convolution kernels is used to convolve the output feature map to improve computational efficiency. Then, the output is subjected to a maximum pooling operation to retain the most influential factors in the feature region, thus effectively avoiding information loss. Finally, the output is processed through the fully connected layer to get the HSIC result by Softmax function. The function is obtained by

$$\text{soft max}(c_h) = \frac{e^{c_h}}{\sum_{d=1}^{D} e^{c_d}} \tag{5}$$

where c_h is the output of the hth node, D represents the count of output nodes, specifically the count of categories designated for classification. In addition, we also employ Dropout to effectively prevent the overfitting phenomenon, and at the same time, it can reduce the dependence on local features and enhance the generalization ability of PMCRNet.

4 Experiments and analysis

4.1 Dataset description

In order to validate the effectiveness of PMCRNet proposed in this paper, experiments are conducted using three publicly available HSI datasets, Indian Pines, Pavia University and Salinas. The details are shown in Table 1.

4.2 Experimental environment

The hardware environment utilized for the experiment is Intel Core i7-13700F processor, 32 GB RAM, RTX3090 24 GB graphics card. The software environment is based on Keras as the main deep learning framework. The compiler is Pycharm, and the compilation environment is Python3.6.

4.3 Selection of evaluation indicators and experimental data

Overall Accuracy (OA), Average Accuracy (AA) and Kappa coefficient are introduced as evaluation indicators. OA is the proportion of the number of samples in which the predicted

Table 1 Description of three datasets

Dataset	Spatial dimension	Spectral bands	Spatial resolution (m)	Wavelength (um)	Categories
IP	144 × 144	200	25	0.4–2.5	16
PU	610 × 340	103	1.3	0.43–0.86	9
SA	512 × 217	204	3.7	0.36–2.5	16

category is the same as the actual category to the total sample size and can be expressed as:

$$\mathrm{OA} = \frac{\sum_i n_{ii}}{\sum_i N_i} \tag{6}$$

where n_{ij} represents the number of samples of class i in the image that were incorrectly predicted as class j. n_{ii} represents the number of correctly classified samples in class i.$N_i = \sum_j n_{ij}$ is the total number of samples of class i to be classified. OA can be a good assessment of classification effectiveness, but for multi-category classification when there is an unequal distribution of categories in the dataset, those categories with more samples have a greater impact on the OA value.

AA is the ratio of correctly predicted samples in each category to the total number of samples, which reflects the classification of each category in an integrated way. It can be expressed as:

$$\mathrm{AA} = \frac{1}{k}\sum_i \frac{n_{ii}}{N_i} \tag{7}$$

where k represents the number of categories of the sample to be classified.

Kappa coefficient is a measure of consistency and can also be employed to evaluate the accuracy of classification. The degree of agreement between the actual classification results and the predicted results is what determines consistency. It can be expressed as:

$$\mathrm{Kappa} = \frac{\sum_i N_i \sum_i n_{ii} - \sum_i \left(\sum_j n_{ij} \cdot \sum_j n_{ji}\right)}{\left(\sum_i N_i\right)^2 - \sum_i \left(\sum_j n_{ij} \cdot \sum_j n_{ji}\right)} \tag{8}$$

where a higher Kappa coefficient represents a better classification effect of the model, i.e., the samples are less likely to be missed and misclassified.

The proportion of training set samples for IP is 5% and the remaining samples are the test set. The proportion of training set samples for PU and SA is 1% and the remaining samples are the test set. In order to ensure the randomness of the samples, the dataset division is to randomize each class of samples and then extract them according to the proportion. The experiment was repeated 10 times with a randomly divided dataset and then the average was taken as the final classification accuracy of our experiment.

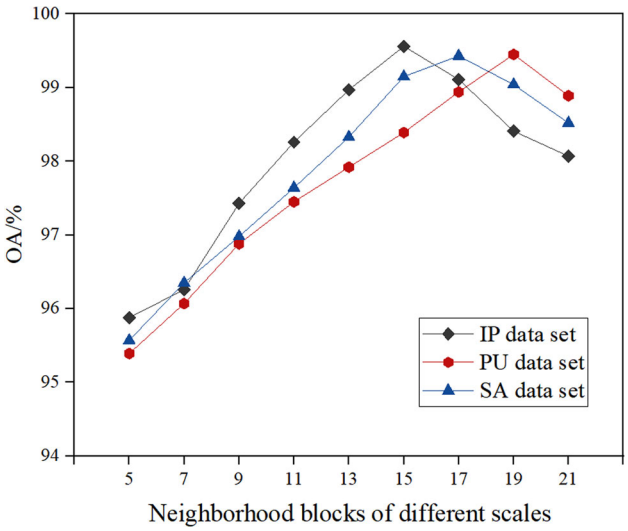


Fig. 4 Impact on spatial neighborhood blocks of different scales

4.4 Experimental parameter setting

We utilized Adam to optimize the loss function. Adam is an extension of stochastic gradient descent algorithm, which saves memory space and is computationally efficient. We chose a learning rate value of 0.001, batch size of 32, epoch of 600, and a ratio of 0.5 neurons removed from the Dropout layer in full connectivity.

Spatial neighborhood information has a significant impact on the HSIC accuracy. Choosing too small a spatial neighborhood can result in not obtaining discriminative features for key spectral bands, while too large a spatial neighborhood may introduce noise. The effect of spatial neighborhood size on the HSIC results is shown in Fig. 4. With the increase of spatial neighborhood size, the accuracy shows an increasing trend. However, when the size increases to a certain range, the accuracy shows a decreasing trend. For IP, PU and SV datasets, the optimal spatial neighborhood size is 15 × 15,

Table 2 Comparison of the structure and ablation results of the four models

Model	2D–3D-CNN	Long connection module	Pre-activated residual module	IP (OA/%)	PU (OA/%)	SA (OA/%)
Model A	√	×	×	94.78	95.12	95.2
Model B	√	√	×	96.12	95.97	96.02
Model C	√	×	√	98.63	98.59	98.64
Model D	√	√	√	99.56	99.45	99.43

19 × 19 and 17 × 17, respectively. The above results indicate that when the ground objects in a HSI occupy a larger area, it contains more information, so choosing a relatively large spatial neighborhood is more conducive to preserving the spatial information of each category of the ground object.

4.5 Ablation experiment

To validate the effectiveness of the 2D–3D-CNN long connectivity-based module and the pre-activation residual module, ablation experiments were performed on the three datasets. Using 2D–3D-CNN as the baseline model, the influence of the classification performance of different components in the proposed model is discussed by separately and fully introducing the long-connection module and the pre-activated residual module. The structure and experimental results of the four models are shown in Table 2, where model D is the proposed model.

From Table 2, it can be found that model D exhibits the best performance on all three datasets, with OA above 99%. Model B improves OA on all three datasets compared to model A, proving that long connection approach improves the sample classification performance. Model C achieves a small improvement in classification accuracy on the three datasets. The above experimental results illustrate that when the pre-activated residual module and the long-short connection are introduced into the 2D-3D hybrid model, the performance and classification ability of the model can be improved to a certain extent.

4.6 Experimental results and analysis

To verify the HSIC effect of PMCRNet, we conducted experimental comparisons using 2DCNN [8], 3DCNN [9], MCNN-CP [10], BHModel [17], JPModel [18], SSMRN [19], MS-CapsNetW [20] and PMCRNet. To ensure fairness, all experiments were conducted under the same settings, and the network parameters of the compared methods are kept consistent with the references.

Tables 3, 4 and 5 show the detailed classification results on the three datasets, respectively. It can be seen that the accuracy of 2DCNN is low, and there is a large gap between it and other methods on all three datasets due to the insufficient learning ability. The accuracy of 3DCNN is better than that of 2DCNN. In the three datasets, comparing the results of 3DCNN with 2DCNN, OA is 4.05%, 4.23% and 4.45% higher, AA is 1.85%, 5.89% and 4.29% higher, and the Kappa coefficient is 4.63%, 5.62% and 4.95% higher. This indicates that 3D convolution has a certain advantage in terms of spatial–spectral joint feature mining capability. Compared with the single 2DCNN and 3DCNN, MCNN-CP hybrid convolutional model also improves the HSIC effect on the three datasets, and the OA on the three dataset are 2.86%, 1.79% and 1.43% higher than that of the 3DCNN, the AA is improved by 5.25%, 1.22% and 1.58%, and the Kappa coefficient is improved by 3.24%, 2.28% and 1.59%, respectively, which verifying the potential of the hybrid model in mining HSI features. The OA, AA and Kappa coefficients of BHModel, JPModel, SSMRN, PMCRNet and MS-CapsNetW are higher than those of MCNN-CP, which proving the effectiveness of introducing residual structure.

Among the five residual models, BHModel has the lowest OA due to the simple structural design and limited ability to extract features. On the three datasets, compared to BHModel, the OA of JPModel improved by 0.86%, 1.24% and 0.43%, respectively, the AA decreased by 0.41%, improved by 1.67% and 0.27%, respectively, the Kappa coefficients improved by 0.99%, 1.67% and 0.48%, respectively. The reason is that JPModel combines residuals connection with depthwise convolution, using multiple residuals stacked to continuously extract spatial context features and spectral features of the data cube. Compared to JPModel, SSMRN has 0.99%, 0.8% and 0.76% higher OA, 2.01%, 1.22% and 1.11% higher AA and 1.12%, 1.05% and 0.85% higher Kappa coefficients on the three datasets, respectively. The reason is that SSMRN uses a multi-scale residual structure to capture spectral-spatial information, which can effectively learn the features of the target ground objects. Compared to SSMRN, MS-CapsNetW improves OA by 0.51%, 0.28% and 0.55%, AA by 0.45%, 0.44% and 0.27%, and Kappa coefficient by 0.58%, 0.38% and 0.62% on the three datasets, respectively. The above results indicate that the two-channel residual connection can better increase the depth of the model and extract more advanced and comprehensive features, which in turn

Table 3 Classification results (%) of different methods on IP dataset

Ground	2DCNN	3DCNN	MCNN-CP	BHModel	JPModel	SSMRN	MS-CapsNetW	PMCRNet
Alfalfa	93.93	86.66	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Corn-no-till	82.29	92.84	91.87	95.79	97.68	99.29	99.68	**99.84**
Corn-min-till	83.11	82.35	96.40	96.41	93.72	98.31	97.45	**100.00**
Corn	71.91	78.57	97.07	98.46	**100.00**	**100.00**	**100.00**	**100.00**
Grass-pasture	97.34	98.78	99.51	99.50	95.75	91.19	**100.00**	98.85
Grass-trees	83.71	94.26	99.08	96.32	97.29	**100.00**	97.91	99.39
Grass-pasture-mowed	96.00	**100.00**	96.15	92.85	**100.00**	**100.00**	**100.00**	**100.00**
Hay-windrowed	97.93	99.76	99.76	98.84	**100.00**	**100.00**	**100.00**	99.76
Oats	90.00	75.00	**100.00**	**100.00**	80.95	**100.00**	**100.00**	**100.00**
Soybeans-no-till	86.04	84.45	96.53	93.79	96.82	96.65	98.74	**100.00**
Soybeans-min-till	90.29	95.47	94.09	96.37	98.11	98.25	98.57	**99.10**
Soybeans-clean-till	77.69	86.09	81.17	94.03	96.99	**100.00**	98.52	98.52
Wheat	**100.00**	98.92	**100.00**	99.46	**100.00**	**100.00**	99.46	**100.00**
Woods	97.98	97.67	99.29	98.61	**100.00**	99.73	**100.00**	**100.00**
Bldg-grass-tree	84.09	91.11	94.97	99.39	93.66	99.71	**100.00**	**100.00**
Stone-steel-towers	**100.00**	**100.00**	**100.00**	97.64	**100.00**	**100.00**	**100.00**	**100.00**
OA (%)	88.22	92.27	95.13	96.67	97.53	98.52	99.03	**99.56**
AA (%)	89.52	91.37	96.62	97.34	96.93	98.94	99.39	**99.71**
Kappa (%)	86.57	91.20	94.44	96.20	97.19	98.31	98.89	**99.50**

The optimal results of the classification accuracies for each class of ground objects, OAs, AAs and Kappa coefficients corresponding to each method in the experiment are bolded

Table 4 Classification results (%) of different methods on PU dataset

Ground	2DCNN	3DCNN	MCNN-CP	BHModel	JPModel	SSMRN	MS-CapsNetW	PMCRNet
Asphalt	96.64	97.21	97.73	98.23	98.89	**99.84**	98.55	99.65
Meadows	96.89	98.61	98.88	99.37	99.56	99.69	99.77	**99.94**
Gravel	76.57	93.77	90.87	84.55	90.47	92.97	93.42	**96.93**
Trees	84.42	94.59	94.03	97.77	91.92	88.85	**98.57**	96.61
Painted metal sheets	99.23	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.64
Bare soil	76.24	80.74	90.54	90.23	97.16	99.28	99.83	**100.00**
Bitumen	65.16	90.68	80.57	91.82	99.56	99.91	99.73	**100.00**
Self-blocking bricks	85.89	88.67	93.04	93.26	94.18	**99.20**	94.65	**99.20**
Shadows	**100.00**	89.79	99.37	98.28	96.79	99.74	98.95	99.72
OA (%)	89.82	94.05	95.84	96.44	97.68	98.48	98.76	**99.45**
AA (%)	86.78	92.67	93.89	94.83	96.50	97.72	98.16	**99.08**
Kappa (%)	86.55	92.17	94.45	95.24	96.91	97.96	98.34	**99.26**

The optimal results of the classification accuracies for each class of ground objects, OAs, AAs and Kappa coefficients corresponding to each method in the experiment are bolded

improve the HSIC accuracy. Comparing the four residual models mentioned above, PMCRNet has the highest OA, AA and Kappa coefficients, suggesting that when introducing the pre-activated residual structure, more informative features can be utilized effectively.

Figures 5, 6 and 7 show the classification results obtained with different methods on the three datasets. As can be seen, compared with other comparative methods, the HSIC result map of PMCRNet has the least number of misclassified ground objects, and is overall smoother and has only a very few noise points, which is closer to the ground truth map.

Table 5 Classification results (%) of different methods on SA dataset

Ground	2DCNN	3DCNN	MCNN-CP	BHModel	JPModel	SSMRN	MS-CapsNetW	PMCRNet
Broccoli-green-weeds-1	99.03	**100.00**	**100.00**	99.88	99.49	**100.00**	**100.00**	**100.00**
Broccoli green-weeds-2	98.50	99.16	99.31	99.04	99.55	99.88	99.55	**100.00**
Fallow	97.27	97.54	99.09	98.26	98.92	99.20	99.65	**99.82**
Fallow-rough-plough	83.50	88.92	96.03	95.33	96.60	96.63	**97.74**	94.92
Fallow-smooth	89.51	94.03	96.93	96.43	95.32	99.53	98.70	**99.87**
Stubble	**100.00**	**100.00**	**100.00**	100.00	99.97	99.97	**100.00**	**100.00**
Celery	98.65	98.83	99.62	99.53	99.34	99.81	99.93	**99.96**
Grapes-untrained	85.81	92.00	94.29	95.51	95.65	97.12	98.37	**99.58**
Soil-vineyard-develop	98.29	98.21	99.91	99.96	99.64	99.98	99.96	**100.00**
Corn-senesced-green-weeds	90.23	96.17	95.98	97.30	98.79	99.55	**99.82**	99.11
Lettuce-romaine-4wk	78.67	89.57	96.43	98.14	96.94	99.24	98.83	**99.78**
Lettuce-romaine-5wk	92.83	97.95	98.64	98.10	98.60	99.24	99.82	**99.88**
Lettuce-romaine-6wk	80.00	95.03	94.51	96.72	97.65	99.87	**100.00**	99.87
Lettuce-romaine-7wk	95.18	96.04	94.33	96.13	95.40	**99.57**	99.47	98.34
Vineyard-untrained	75.09	86.95	88.09	93.13	95.66	94.64	96.64	**98.30**
Vineyard-vertical-trellis	95.76	96.47	98.99	98.67	98.93	99.93	99.93	**100.00**
OA (%)	90.32	94.77	96.20	97.23	97.66	98.42	98.97	**99.43**
AA (%)	91.14	95.43	97.01	97.63	97.90	99.01	99.28	**99.34**
Kappa (%)	89.23	94.18	95.77	96.91	97.39	98.24	98.86	**99.36**

The optimal results of the classification accuracies for each class of ground objects, OAs, AAs and Kappa coefficients corresponding to each method in the experiment are bolded

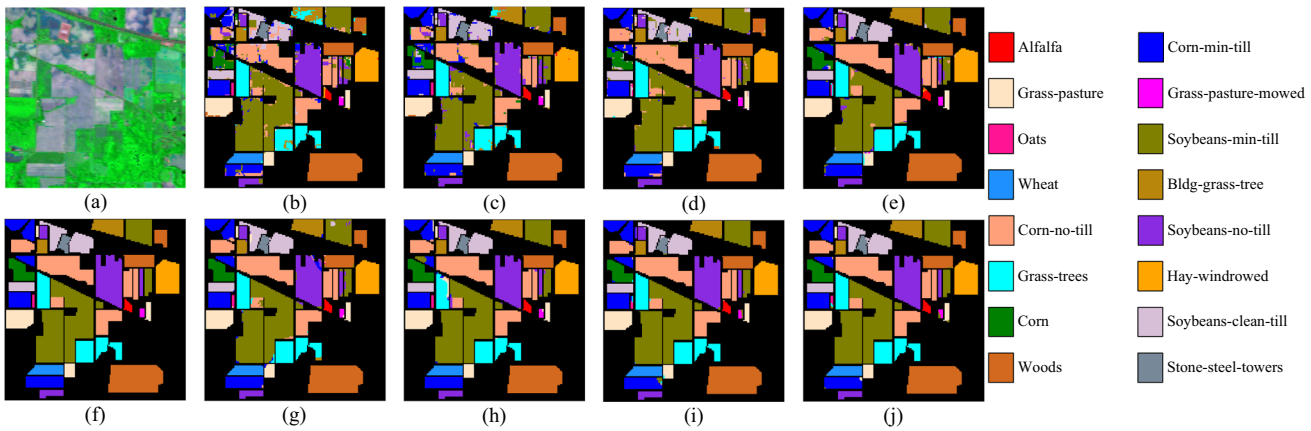


Fig. 5 Classification results of different methods for IP. **a** False color image **b** 2DCNN **c** 3DCNN **d** MCNN-CP **e** BHModel **f** real ground data **g** JPModel **h** SSMRN **i** MS-CapsNetW **j** PMCRNet

4.7 Runtime comparison

To better evaluate the HIC performance of the model, the training time, testing time and the number of parameters of different methods on the three datasets were analyzed through experiments. The number unit of model parameters is $M = 10^6$. The number of model parameters and comparison results are shown in Table 6

From Table 6, it can be seen from the results of training time and test time that PMCRNet takes more time to train than 2DCNN, 3DCNN, MCNN-CP and BHModel, mainly due to that it uses multiple hybrid convolutional blocks and adds long and short residual connections and a pre-activation mechanism to the network, thus increasing the training time. However, PMCRNet consumes less time than JPModel, SSMRN and MS-CapsNetW, mainly because of

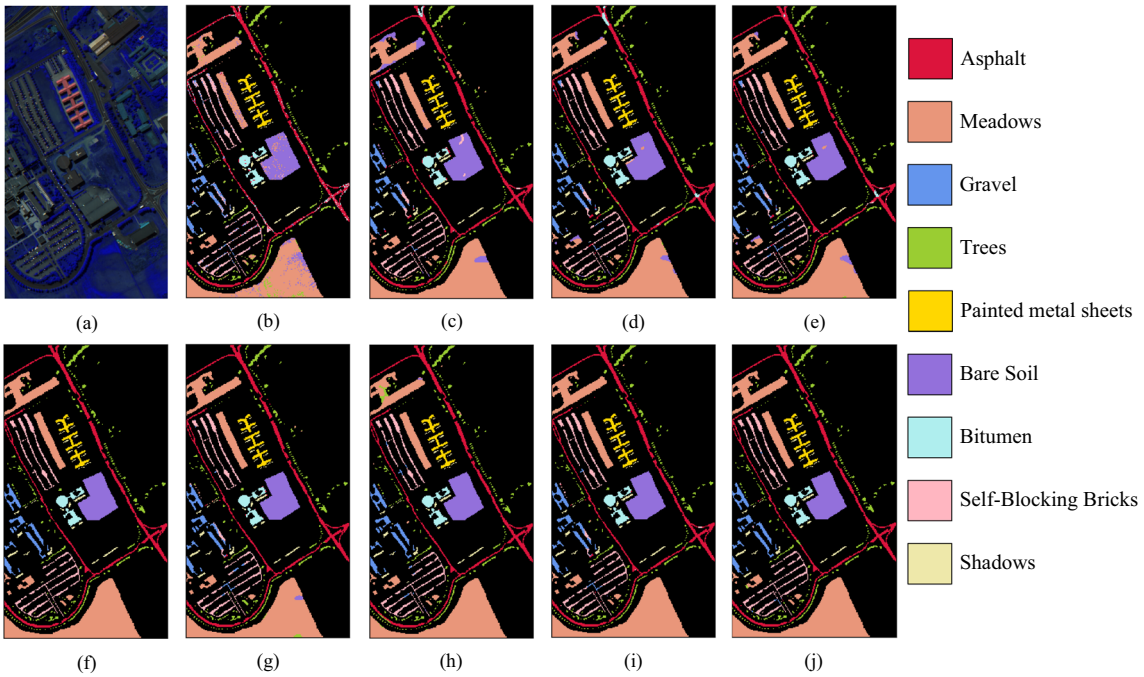


Fig. 6 Classification results of different methods for PU. **a** False color image **b** 2DCNN **c** 3DCNN **d** MCNN-CP **e** BHModel **f** real ground data **g** JPModel **h** SSMRN **i** MS-CapsNetW **j** PMCRNet

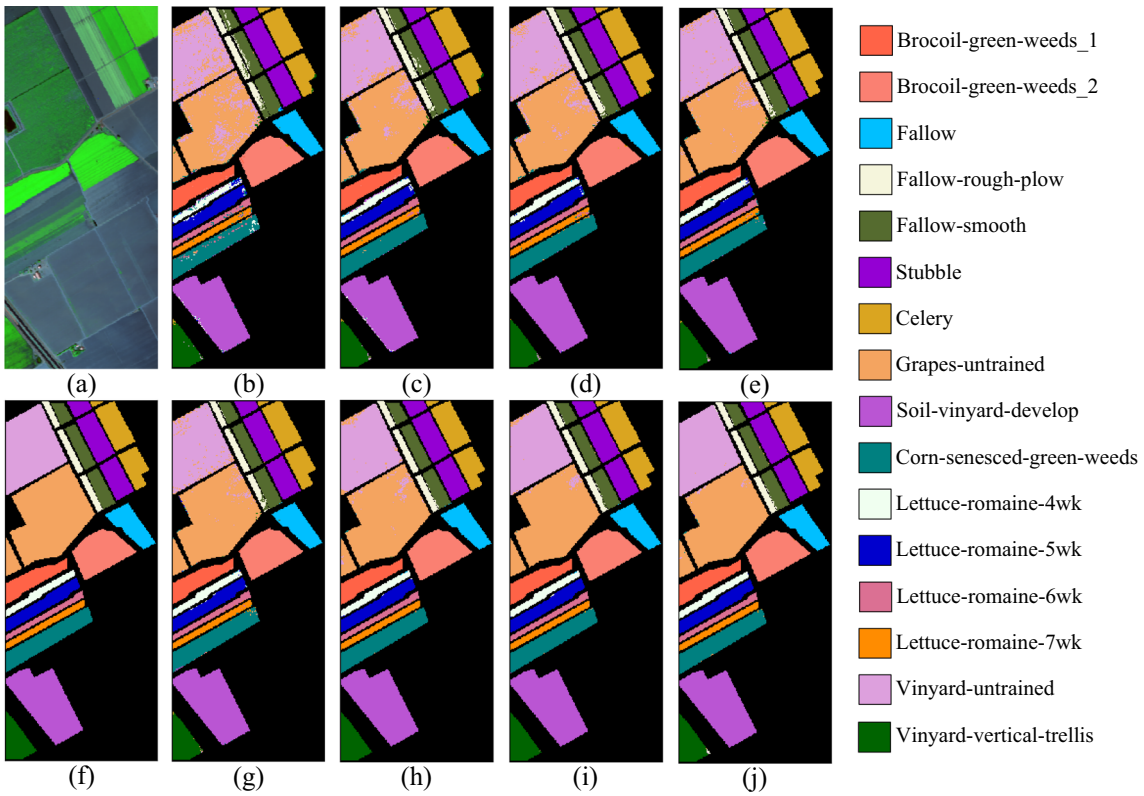


Fig. 7 Classification results of different methods for SA. **a** False color image **b** 2DCNN **c** 3DCNN **d** MCNN-CP **e** BHModel **f** real ground data **g** JPModel **h** SSMRN **i** MS-CapsNetW **j** PMCRNet

Table 6 Runtime (seconds) and the numbers of parameters ($M = 10^6$) of different methods on the three datasets

Methods	IP			PU			SA		
	Parameter (M)	Training time (s)	Test time (s)	Parameter (M)	Training time (s)	Test time (s)	Parameter (M)	Training time (s)	Test tim e(s)
2DCNN	0.12	89.7	7.3	0.11	72.1	6.6	0.12	80.3	6.9
3DCNN	0.77	183.2	15.2	0.85	140.9	10.1	0.79	143.9	10.6
MCNN-CP	2.79	140.7	10.1	2.68	90.5	7.2	2.74	102.9	8.1
BHModel	3.81	200.7	12.9	3.59	168.5	9.8	3.65	170.2	10.2
JPModel	5.97	326.2	20.3	5.63	264.7	17.1	5.92	272.9	17.8
SSMRN	5.22	277.7	15.2	5.14	225.5	15.2	5.18	231.2	15.9
MS-CapsNetW	6.06	330.1	21.7	5.85	270.3	17.6	5.76	265.1	17.2
PMCRNet	4.95	260.4	14.9	4.62	210.7	16.2	4.71	214.5	16.4

the higher utilization of the parameters of itself, which can better mine the rich spatial–spectral features in HSI for feature reuse and enhanced information transfer. In terms of the number of parameters, 2DCNN, 3DCNN, MCNN-CP and BHModel can quickly complete the model training due to the simple structure and small number of parameters, but this also leads to the model not fully extracting the features, and the HSIC accuracy is not good. Compared with JPModel, SSMRN and MS-CapsNetW, the proposed method reduces the number of parameters, reduces the computational complexity, and achieves better classification results.

Although PMCRNet does not achieve the optimal time consumption, comprehensive consideration of the HSIC accuracy and the number of parameters and other indicators can be found that the model has a better classification effect and is more suitable for a variety of practical engineering application scenarios.

4.8 Effectiveness of small sample sizes

To further demonstrate the effectiveness of the proposed method, 10,20,30,40,50 samples were randomly selected as training data for the experiments in each category of ground objects in IP, PU, and SA datasets, respectively. Figure 8 shows the comparison of OA of different methods under small sample conditions. It can be seen that our method still achieves the optimal classification results, which proves the robust effectiveness of this method in the case of small sample sizes.

5 Conclusion

This work combines 2D-3D hybrid convolution and pre-activated residual network to propose PMCRNet for HSIC. Compared with the traditional residual-based methods, PMCRNet can effectively accelerate the network training speed and reduce the parameter computation. In addition, the use of long and short distance residual connections solves the problem of gradient disappearance in the deep network, facilitates back propagation and better integrates the input information of the current layer. The method achieves the enhancement of feature reuse and information transfer, which helps to improve the ground object classification accuracy of HSI. By evaluating the classification effectiveness of the three datasets and comparing it with seven related state-of-the-art models, the results show that the model outperforms similar networks while ensuring higher classification accuracy.

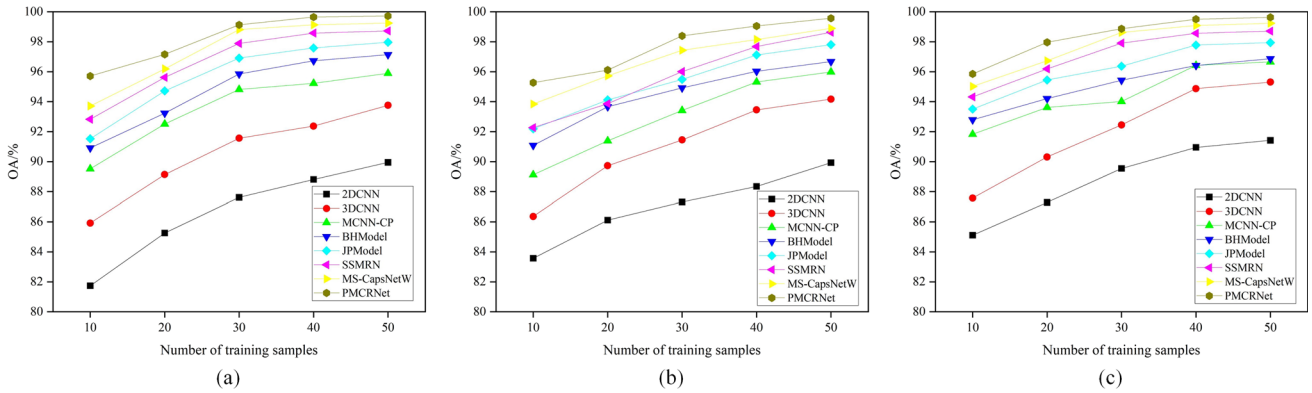


Fig. 8 Classification results of small sample sizes. **a** IP dataset **b** PU dataset **c** SA dataset

Author contributions HL conceptualized and designed the algorithm, contributed to algorithm improvements, and critically revised the manuscript for important intellectual content. YS built the model, verified and analyzed it experimentally, prepared the original manuscript draft. HZ assisted with manuscript writing and revisions, supervised the project, provided strategic direction in algorithm development and testing, and conducted a thorough review and final approval of the manuscript prior to submission. ML visualized experimental results. All authors reviewed the manuscript.

Funding This work was supported by Zhejiang Provincial Education Department General Research Project (No. Y202248546), Public Welfare Applied Research Project of Huzhou (No. 2023GZ29), Natural Science Foundation of Huzhou (No. 2023YZ55) and Zhejiang Provincial College Student Innovation and Entrepreneurship Training Program Project (No. S202310347089).

Data availability The data that support the findings of this study are openly available in http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval Not applicable.

References

1. Zhao, C., Wang, M., Feng, S.: A sparse and spectral smooth regularized low-rank tensor decomposition method for hyperspectral target detection. Int. J. Remote Sens. **43**(12), 4608–4629 (2022)
2. Gao, H., Wang, M., Sun, X., Cao, X., et al.: Unsupervised dimensionality reduction of medical hyperspectral imagery in tensor space. Comput. Methods Progr. Biomed. **240**, 107724 (2023)
3. Liu, G., Wang, L., Liu, D.: Hyperspectral image classification based on a least square bias constraint additional empirical risk minimization nonparallel support vector machine. Remote Sens. **14**(17), 4263 (2022)
4. Wang, H., Celik, T.: Sparse representation-based hyperspectral image classification. Sign. Image Video Process. **12**(5), 1009–1017 (2018)
5. Tan, X., Xue, Z., Yu, X., Sun, Y., et al.: Hyperspectral image classification with deep 3D capsule network and Markov random field. IET Image Process. **16**(1), 79–91 (2022)
6. Yang, L., Chen, J., Zhang, R., Yang, S., et al.: Precise crop classification of UAV hyperspectral imagery using kernel tensor slice sparse coding based classifier. Neurocomputing **551**, 126487 (2023)
7. Hu, W., Huang, Y., Wei, L., Zhang, F., et al.: Deep convolutional neural networks for hyperspectral image classification. J. Sensors **2015**, 258619 (2015)
8. Zhao, W., Du, S.: Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS J. Photogramm. Remote Sens. **113**, 155–165 (2016)
9. Li, Y., Zhang, H., Shen, Q.: Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sens. **9**(1), 67 (2017)
10. Zheng, J., Feng, Y., Bai, C., Zhang, J.: Hyperspectral image classification using mixed convolutions and covariance pooling. IEEE Trans. Geosci. Remote Sens. **59**(1), 522–534 (2021)
11. Fırat, H., Asker, M.E., Hanbay, D.: Classification of hyperspectral remote sensing images using different dimension reduction methods with 3D/2D CNN. Remote Sens. Appl.: Soc. Environ. **25**, 100694 (2022)
12. Liu, Z., Mao, X., Huang, J., Gan, M., et al.: Stratified attention dense network for image super-resolution. Sign. Image Video Process. **16**(3), 715–722 (2022)
13. Shi, C., Liao, D., Zhang, T., Wang, L.: Hyperspectral image classification based on 3D coordination attention mechanism network. Remote Sens. **14**(3), 608 (2022)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778; 2016.
15. Qing, Y., Liu, W.: Hyperspectral image classification based on multi-scale residual network with attention mechanism. Remote Sens. **13**(3), 335 (2021)
16. He, Z., Shi, Q., Liu, K., Cao, J., et al.: Object-oriented mangrove species classification using hyperspectral data and 3-D siamese residual network. IEEE Geosci. Remote Sens. Lett. **17**(12), 2150–2154 (2020)
17. Cao, F., Guo, W.: Deep hybrid dilated residual networks for hyperspectral image classification. Neurocomputing **384**, 170–181 (2020)
18. Dang, L., Pang, P., Lee, J.: Depth-Wise separable convolution neural network with residual connection for hyperspectral image classification. Remote Sens. **12**(20), 3408 (2020)

19. He, S., Jing, H., Xue, H.: Spectral-spatial multiscale residual network for hyperspectral image classification. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **43**, 389–395 (2022)
20. Lei, R., Zhang, C., Zhang, X., Huang, J., et al.: Multiscale feature aggregation capsule neural network for hyperspectral remote sensing image classification. Remote Sens. **14**(7), 1652 (2022)
21. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14, pp. 630–645: Springer, 2016
22. Gao, H., Yang, Y., Yao, D., Li, C.: Hyperspectral image classification with pre-activation residual attention network. IEEE Access **7**, 176587–176599 (2019)
23. Huan, H., Li, P., Zou, N., Wang, C., et al.: End-to-End super-resolution for remote-sensing images using an improved multi-scale residual network. Remote Sens. **13**(4), 666 (2021)
24. Wang, X., Xu, H., Yuan, L., Dai, W., et al.: A remote-sensing scene-image classification method based on deep multiple-instance learning with a residual dense attention convnet. Remote Sens. **14**(20), 5095 (2022)