

Retraction

Retracted: I3D: An Improved Three-Dimensional CNN Model on Hyperspectral Remote Sensing Image Classification

Security and Communication Networks

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Q. Haizhong, "I3D: An Improved Three-Dimensional CNN Model on Hyperspectral Remote Sensing Image Classification," *Security and Communication Networks*, vol. 2021, Article ID 5217578, 12 pages, 2021.

Research Article

I3D: An Improved Three-Dimensional CNN Model on Hyperspectral Remote Sensing Image Classification

Qian Haizhong 

Information Engineering College, Jiangsu Maritime Institute, Nanjing 211199, China

Correspondence should be addressed to Qian Haizhong; hzhqian@jmi.edu.cn

Received 12 October 2021; Revised 13 November 2021; Accepted 16 November 2021; Published 29 November 2021

Academic Editor: Jian Su

Copyright © 2021 Qian Haizhong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hyperspectral image data are widely used in real life because it contains rich spectral and spatial information. Hyperspectral image classification is to distinguish different functions based on different features. The computer performs quantitative analysis through the captured image and classifies each pixel in the image. However, the traditional deep learning-based hyperspectral image classification technology, due to insufficient spatial-spectral feature extraction, too many network layers, and complex calculations, leads to large parameters and optimizes hyperspectral images. For this reason, I proposed the I3D-CNN model. The number of classification parameters is large, and the network is complex. This method uses hyperspectral image cubes to directly extract spectral-spatial coupling features, adds depth separable convolution to 3D convolution to reextract spatial features, and extracts the parameter amount and calculation time at the same time. In addition, the model removes the pooling layer to achieve fewer parameters, smaller model scale, and easier training effects. The performance of the I3D-CNN model on the test datasets is better than other deep learning-based methods after comparison. The results show that the model still exhibits strong classification performance, reduces a large number of learning parameters, and reduces complexity. The accuracy rate, average classification accuracy rate, and kappa coefficient are all stable above 95%.

1. Introduction

The development of remote sensing technology [1] has promoted the improvement of the spatial, temporal, and spectral resolution of remote sensing images [2]. Among them, the spatial resolution refers to the range of the ground represented by a single pixel in the remote sensing image; the time resolution refers to the minimum time interval required for two adjacent observations at the same location in the remote sensing image; the spectral resolution [3] refers to the remote sensing image. From the perspective of spectral resolution, remote sensing images have gone through the development process from panchromatic images, multi-spectral images, to hyperspectral images [4, 5]. Hyperspectral remote sensing refers to the process of using hyperspectral sensors to obtain corresponding target data and applying them under certain environmental conditions in space [6, 7]. As an important hyperspectral sensor,

hyperspectral imaging spectrometers [8, 9] can effectively acquire images. Hyperspectral remote sensing images with rich spatial and spectral features provide strong support for characterization learning and discrimination learning. In view of this, researchers have carried out a lot of research work on hyperspectral images, including mixed pixel unmixing [10], noise evaluation [11], image classification [12], anomaly detection [13], and target detection [14]. Among them, hyperspectral feature classification is one of the key technical components of hyperspectral image processing system. The main purpose of hyperspectral image classification is to assign a category label to each pixel in the image. Its classification performance will affect the subsequent image processing process, so it is very important to achieve accurate classification of hyperspectral image features. At this stage, hyperspectral image classification has been widely used in the fields of natural environment monitoring [15], environmental change analysis [16],

natural resource exploration [17], military defense security [18], and natural disaster assessment [19]. Figure 1 shows the examples of remote sensing image.

The classification methods of traditional hyperspectral image, which is based on spectral information, generally include two important elements: feature engineering and classifiers [20]. Among them, the purpose of feature engineering is to reduce the dimensionality of hyperspectral images and get discriminative bands or features, which generally include feature extraction and feature selection. The purpose of feature extraction is to search a mapping from high-dimensional space to low-dimensional space so that different categories can be well distinguished in low-dimensional space. At this stage, feature extraction methods generally include linear discriminant analysis [21], independent component analysis [22], minimum noise separation transformation [23], and PCA [24]. Although the method of feature extraction is simple and intuitive, some key information may be lost or distorted. The purpose of feature selection is to retain the most representative spectral bands of the original hyperspectral image and discard the spectral bands with poor classification effects. Common feature selection methods include Bhattacharyya distance [25], J-M distance mutual information, and spectral angle mapping. Although the physical meaning of the feature selection method is very clear, it can retain the useful information of the hyperspectral image without spatial transformation, but these methods often need to be matched with the search algorithm to search for the most effective band or combination of bands. It takes a lot of time. The features obtained by feature engineering are sent to the classifier for classification [26].

At present, the main problems of hyperspectral image classification are as follows: (1) Due to the continuous transition of hyperspectral imaging from wideband imaging to narrow-band imaging, a large amount of redundant information is generated. (2) The storage capacity of current communication equipment is difficult to meet. In the process of transmitting hyperspectral image data, the demand for higher spatial resolution is maintained, so the spatial resolution of hyperspectral image data is very low. (3) Hughes phenomenon will appear in the process of hyperspectral image classification. The Hughes phenomenon refers to the classification accuracy of the hyperspectral image classification process and the classification accuracy is not proportional to the number of selected bands, but after reaching a critical value, continuing to increase the number of bands will actually lead to a decline in the classification accuracy.

In recent years, the research enthusiasm for deep learning methods has continued to rise. Deep learning has developed into a new field in machine learning research. When deep learning deals with classification problems, it does not rely on some previously assumed criteria, but different learning models are under different learning frameworks. For example, the convolutional neural network (CNN) [27] is a machine learning model and belongs to deep supervised learning [28]. The basic structure of a convolutional neural network includes a feature

extraction layer and a feature mapping layer. The feature extraction layer implicitly learns from the training data and discards the explicit feature extraction. The CCN does not need a tedious image preprocessing process. Because the network does not need a cumbersome image preprocessing process, the original data can be directly input into the model for training, so it can be widely used. In addition, the network structure is highly unchanged for general geometric transformations (such as translation and scaling).

The I3D-CNN model proposed in this paper uses the three-dimensional kernel function for hyperspectral image classification, thus making full use of the structural features of the three-dimensional hyperspectral image data. The three-dimensional convolutional neural network uses the learned local signal changes of the hyperspectral image as important information for judging category attributes. The network input proposed in this paper is the original spectral data cube, and the classifier model adopts an end-to-end approach. It can realize the pixel-level classification of hyperspectral images without any preprocessing and subsequent optimization processing. Because the pooling operation will further reduce the resolution of the feature map, that is, the pooling layer in the traditional neural network will reduce the spatial resolution of the hyperspectral image, so the pooling layer is not used in this model. At the same resolution, the three-dimensional convolutional neural network in this paper contains fewer parameters and is more suitable for the classification of hyperspectral images that lack high-quality training images.

2. Related Works

Deep learning algorithms are widely used by researchers in the classification of hyperspectral images, and good research results have been achieved. In 2014, the deep learning network SAE [29, 30] was applied to the classification of hyperspectral images, and a deep learning model fused with spectral and spatial features was proposed, which achieved high classification accuracy and more and more deep learning models. In 2015, the deep belief network [31] model was introduced into the classification of hyperspectral images; combined with the method of principal component analysis, hierarchical learning of features and logistic regression methods were used to extract the spatial spectrum features of hyperspectral images. The convolutional neural network (CNN) model was applied to hyperspectral image classification for the first time, but the established CNN model can only extract spectral features. In 2016, a CNN-based deep feature extraction [5, 32] method was proposed, and a deep finite element model based on a three-dimensional convolutional neural network was established to extract the spatial spectrum features of hyperspectral remote sensing images and obtain high classification accuracy. For general geometric transformations (such as translation and zooming), the height remains the same. Zhao et al. applied the multiscale two-dimensional CNN (2D-CNN) model [33] to the study of hyperspectral remote sensing image



FIGURE 1: Examples of remoting sensing image.

classification and realized the simultaneous use of multiple spectral features in the classification process but faced with the need to select different feature extraction for different feature categories. Mei [34] et al. found that the large number of parameters that emerged during the training of the 2D-CNN network easily caused the model to overfit, which greatly restricted the generalization ability of the model. In 2017, the Spectrum Spatial Residual Network (SSRN) [35] was proposed. The residual blocks in the SSRN use identity mapping to connect to other 3D convolutional layers, which facilitates the backpropagation of the gradient and extracts deeper spectral features at the same time; alleviating the accuracy degradation of other deep learning models is solved. In 2019, through adaptive dimensionality reduction, a semisupervised three-dimensional convolutional neural network (CNN) [36] for spectrum space HSIC was proposed to solve the dimensionality curse problem. These research results show that the method based on deep learning has achieved certain results in the classification of hyperspectral images. However, deep model-based methods usually have overfitting. This is because a large amount of labeled data is required when using the deep model method for training, but the labeled samples of hyperspectral images are insufficient. Therefore, in order to avoid such problems as much as possible, a suitable convolution model is required, which can not only give full play to the huge advantages of convolutional neural networks but also reduce the learnable parameters, thereby alleviating the overfitting problem and the demand for training sample data volume. The existing convolution model is more complicated, and the network parameters are large, which brings complicated calculation problems. A more lightweight convolutional network is needed to meet the requirements of computing time, efficiency, and memory.

3. Methodology

3.1. 3D Convolutional Neural Network. Aiming at the problem of insufficient utilization of the information of the three-dimensional hyperspectral data by the two-dimensional convolutional neural network, the three-dimensional convolutional neural network can be introduced to extract the spatial spectrum characteristics of the hyperspectral image at the same time. The network structure of the three-dimensional convolutional neural network (3D-CNN) and the two-dimensional convolutional neural network (2D-

CNN) is very similar, and both types of structures are composed of the basic convolutional layer and the pooling layer. The key difference is that the 3D-CNN structure uses a 3D convolution kernel to convolve the image. Figure 2 shows an example of 2D-CNN and 3D-CNN convolution operations. $N \times N$ represents the size of the convolution kernel, the three-dimensional is more than the two-dimensional by the spectral dimension of M , and L is the output channel of the convolutional layer. 3D-CNN performs operations on the spatial dimension and the spectral dimension at the same time so as to extract the spatial-spectral features of the image at the same time. It will not extract a certain type of feature separately, which leads to insufficient feature extraction, resulting in unsatisfactory classification results.

Among them, the convolution kernel of the three-dimensional convolutional neural network moves in the three directions of length, width, and channel, and the calculation formula for calculating the point value V of the j -th feature map of the i -th layer of the neural network at (x, y, z) is as follows:

$$V_{i,j}^{xyz} = f \left(\sum_m \sum_{l=0}^{L_i-1} \sum_{w=0}^{W_i-1} \sum_{h=0}^{H_i-1} W_{i,j,m}^{lwh} \cdot V_{i-1}^{(x+l)(y+w)(z+h)} + b_{i,j} \right). \quad (1)$$

In formula (1), m is the feature map connected to the current feature map in the $i-1$ th layer, l represents the length and width of the convolution kernel, and represents the size of the convolution kernel in the spectral dimension. W represents the connection weight of the m -th feature map connected to the $i-1$; b represents the bias of the j -th feature map in the i -th layer; f is the activation function.

3.2. Depth Separable Convolution. Deeply separable convolution (DSC) is a transformation form of the ordinary two-dimensional convolutional neural network, which can replace an ordinary two-dimensional convolutional neural network. The core idea is to split the convolution with ordinary N channels as M into 1 convolution with channel M . This convolution performs a single-channel filtering operation, which is different from the addition of channels after ordinary convolution filtering and $N \times 1 \times M$ convolution. Figure 3(a) is a common convolutional neural network, which is composed of convolutional layers, batch normalization

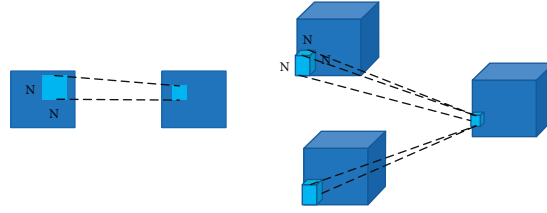


FIGURE 2: 2D-CNN versus 3D-CNN.

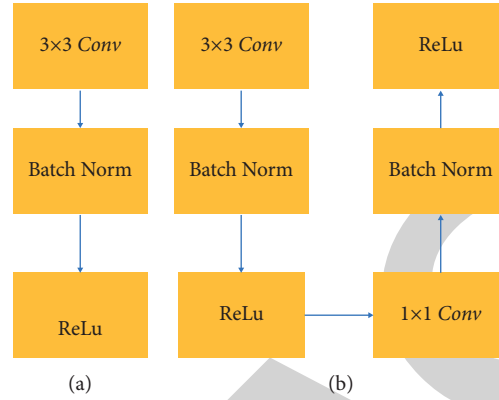


FIGURE 3: Ordinary convolution network and deep separable convolution. (a) Ordinary convolution; (b) deep separable convolution.

operations, and activation functions. Figure 3(b) shows the depth separable convolution, which is composed of a 3×3 convolution kernel size depth separable convolution layer, batch normalization, and activation function, and 1×1 convolution kernel size convolutional layer, batch normalization, and activation function composition. It is divided into two parts: depthwise convolution and pointwise convolution. In performing conventional 2D convolution on multiple input channels, the number of channels of the convolution kernel is the same as the number of input channels.

All channels are mixed to produce the final output. Deep convolution convolves each channel of the input feature map separately to capture the spatial characteristics of each channel. Point-by-point convolution integrates all the extracted spatial features, learns the channel-related information of the input feature map, and performs a channel fusion operation similar to ordinary convolution on the obtained feature map. The number of parameters and calculations can be reduced without much loss of accuracy.

3.3. 13D Convolution Kernel Separation. Through the above analysis, the network architecture of the model proposed in this paper is shown in Figure 4.

13D model consists of one input layer, three three-dimensional convolutional layers, two depth separable convolutional layers, which are, respectively, deep convolution and pointwise convolution, and fully connected layers, including a flatten smoothing layer, two dense layers, and two dropout layer compositions. To prevent overfitting, the convolutional layer uses the ReLU [37] activation function for nonlinear mapping.

The ReLU activation function converges faster than the traditional Sigmoid function and Tanh function. The form of the ReLU activation function is as follows:

$$f(x) = \max(0, x). \quad (2)$$

Finally, the soft-max classifier is used to classify the hyperspectral image features. The input and output dimensions and parameter sizes of each layer are shown in Table 1.

The soft-max loss to train the deep classifier is the same as the two-dimensional convolution model. It uses the random admiral descent of backpropagation to minimize the loss of the network. The kernel function is updated with the following formula:

$$m_{i+1} = 0.9 \times m_i - 0.005 \times w_i - \left(\frac{\partial L}{\partial w} | w_i \right), \quad (3)$$

$$w_{i+1} = w_i + \epsilon m_{i+1}.$$

The size of the 3D-CNN convolution kernel in the model is $3D_conv_layer1 = 8 \times 3 \times 3 \times 7 \times 1$. Among them, $K12 = 3$, $K2 = 3$, and $K32 = 5$. $3D_conv_layer3 = 32 \times 3 \times 3 \times 3 \times 16$, where $K3 = 3$, $K3 = 3$, and $K3 = 3$. Finally, reshape the three-dimensional output features to form two-dimensional data, extract the spatial features of the hyperspectral image, and add two depth separable convolutional layers $Separable_conv2d_1_layer4 = 3 \times 3 \times 64$ and $Separable_conv2d_1_layer5 = 1 \times 1 \times 128$.

In order to increase the number of spatial-spectral feature maps, three three-dimensional convolutional layers are deployed before the leveling layer. The spatial information of the hyperspectral image determines the spatial

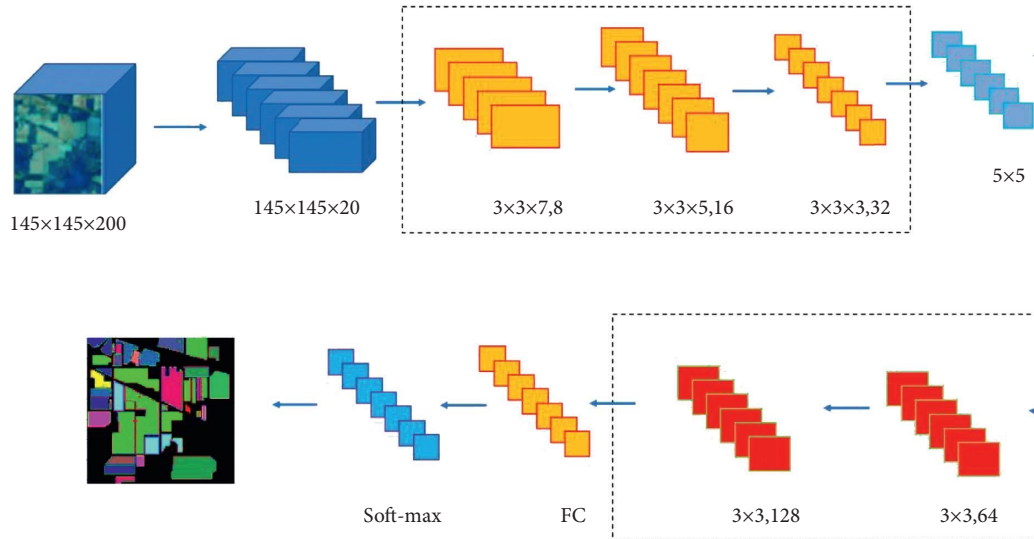


FIGURE 4: The framework of the I3D-CNN model.

TABLE 1: Units for magnetic properties and parameters of model.

Layer (type)	Output shape	Parameter
Input_1 (input layer)	(11, 11, 20, 1)	0
ConvI3d_1 (Cov3D)	(9, 9, 14, 8)	512
ConvI3d_2 (Cov3D)	(7, 7, 10, 16)	5776
ConvI3d_3 (Cov3D)	(5, 5, 8, 32)	13,856
Reshape_1 (reshape)	(5, 5, 256)	0
Separable_con2d_1 (separable)	(3, 3, 64)	18,752
Separable_con2d_1 (separable)	(3, 3, 128)	8384
Flatten_1 (flatten)	(128)	0
Dense_1 (dense)	(256)	295,168
Dropout_1 (dropout)	(256)	0
Dense_2 (dense)	(128)	32,896
Dropout_1 (dropout)	(128)	0

features between adjacent pixels in the spatial dimension, and the spatial features can compensate for the spectral features. The spatial feature is used to increase the features of the spectral spaces, and the classification accuracy of the hyperspectral image is improved. Therefore, after the three-dimensional convolutional layer, two depth separable convolutional layers are added, which can reduce the parameters while increasing the spatial features, extracting more abundant spatial spectrum features, and ensuring that the model can distinguish the spatial information of different bands without loss. The total parameters (i.e., the adjustable weight) of the proposed fast 3D-CNN and DSC combined model are 377,408, which is about half less than the parameters of the fast 3D-CNN alone. The filling method of the convolution is zero filling, which does not require batch normalization and data enhancement.

Figure 5 shows the learning framework of three-dimensional spatial-spectral features. This part consists of three three-dimensional convolutional layers and ReLU activation function, which extracts the spectral and spatial features of the hyperspectral image at the same time. The input data size of the network is $11 \times 11 \times 20$, and the size of the first layer of convolution kernel is $3 \times 3 \times 8$. After two

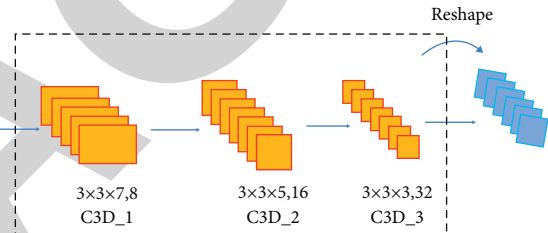


FIGURE 5: 3D spectral-spatial feature learning.

layers of three-dimensional convolution operation, the output is 32 feature maps of $5 \times 5 \times 8$ size. After completing the 3D convolution operation, perform spatial feature extraction again and use reshape to perform 3D to 2D transformation.

To learn the output features of the later two-dimensional space, reconstruct the three-dimensional features into 32 two-dimensional feature maps with a size of 5×5 . Only the two-dimensional spatial features need to be studied. Compared with the three-dimensional convolution, the network parameters and operating costs are reduced.

Figure 6 shows the learning of two-dimensional spatial features based on depth separable convolution. The depth separable convolution is used to extract the output two-dimensional features, and the spatial features can be extracted better without introducing additional parameters. Different from the traditional two-dimensional convolution, the depth separable convolution performs spatial convolution while maintaining channel independence and then performs deep convolution. After feature reshaping, the network input data is 256 feature maps with a size of 5×5 .

SeparableConv2D [38] implements the entire depth separation convolution process, that is, the depthwise spatial convolution and the point-by-point convolution in which the output channels are mixed together. The input data is convolved with $64 \times 3 \times 3$ convolution kernels, and the feature map of 64 channels is obtained. Each convolution kernel only convolves one channel of the input layer. The second step is

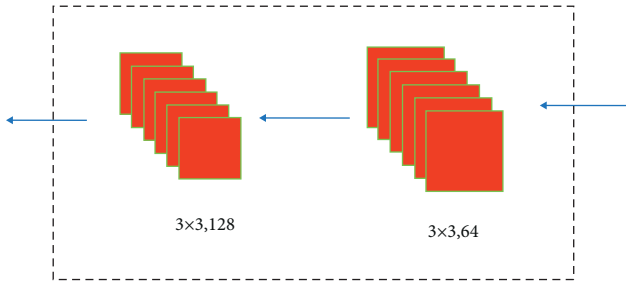


FIGURE 6: DSC spatial feature learning.

the pixel-by-pixel convolution operation. Use 128 1×1 size convolution kernels to perform convolution operations on these 64 feature maps to merge the information of different channels. After 1×1 convolution, the size depth is significantly reduced. The 64-channel output by the upper layer, 1×1 convolution, will embed these channels into a single channel. After 1×1 convolution, add batch normalization to improve the generalization ability of the model and add the nonlinear activation function ReLU, which allows the network to learn more complex functions. At the same time, each convolution kernel is convolved with the input image to obtain a spatial feature map. Deep separable convolution not only reduces the number of parameters and calculations in the network but also improves the network training speed and reduces the probability of overfitting in HSI classification. Use padding to ensure that the size of the output feature map is the same as the size of the input.

4. Experiments and Results

4.1. Dataset Description. Indian Pines (IP) is an image of the Indiana agricultural and forestry hyperspectral test site in northwest Indiana collected by the AVIRIS sensor [39]. The image consists of 145×145 pixels, of which 220 spectral bands range from 0.2 to 0.4 m, with a spatial resolution well. After removing 20 noise bands, there are 16 types of ground features in this dataset, which is shown in Table 2. In addition, the original image is Figures 7(a) and 7(b) is the feature image.

The Pavia University (PU) dataset was collected in Pavia, northern Italy, using a reflective optical system imaging spectrometer (ROSIS) optical sensor. The PU dataset consists of 610×610 space and 103 spectral bands. The dataset contains 9 types of ground features, which is also shown in Table 3.

Like the Indian Pines image, Salinas data is also captured by AVIRIS imaging spectrometer, which is an image of Salinas Valley in California, USA. Different from Indian Pines, it has a spatial resolution of 3.7 m. The image originally has 224 bands. Similarly, we generally use the image of 204 bands after excluding the 108-112154-167 band and the 224th band that cannot be reflected by water. The size of the image is $512 \times$. Therefore, it contains 111,104 pixels, of which 56,975 are background pixels, and 54,129 can be applied to classification. These pixels are divided into 16 categories, including fallow and celery. Table 4 shows the datasets of Salinas sense.

TABLE 2: Indian Pines datasets.

S/N	Name	Train	Test
1	Asphalt	25	35
2	Corn_notill	356	997
3	Coren_mintill	231	567
4	Corn	54	220
5	Grass_pasture	160	342
6	Grass_trees	232	567
7	Grass_pasture_moved	9	18
8	Hay_windrowed	130	341
9	Otas	4	14
10	Soybean_notill	299	656
11	Sobean_mintill	523	1783
12	Soybean_clean	100	500
13	Wheat	54	125
14	Woods	145	1243
15	Buildings_Grass_tree	28	59
16	Stone_Steel_Towel	28	55

4.2. Experiment Procedure and Environment. After continuous testing and adjustment during the experiment, the batch size is set to 256, the number of epoch iterations is set to 50, the Adam optimizer is used to train the network, and the initial learning rate is set to 0.001 (set decay = $1e - 06$ at the same time). The ReLU function is used as the activation function to improve the calculation efficiency and speed up the convergence of the function. Randomly select the IP, PU, and SA datasets with 60% training data and 40% test data for experiments. For the fairness of the experiment, the same spatial dimension is extracted from the three-dimensional patch of the input volume for different datasets. For example, the spatial dimensions of IP, SA, and UP are all $11 \times 11 \times 20$. OA, AA, and kappa (K) coefficients and confusion matrix are used to evaluate classification performance. Among them, OA is used to evaluate the classification accuracy rate of all samples, AA is the classification accuracy of each category, and the kappa coefficient is a commonly used method to calculate the classification accuracy, which represents the ratio of classification and completely random classification to the reduction of errors. Confusion matrix is to separately count the number of observations that are classified into the wrong class by the classification model and then display the results.

4.3. Classification Effect under Different Dimensionality Reduction Results. In this paper, IPCA is used to reduce the dimensionality of the data, and the first 75 principal components are selected by giving the percentage of the original data information required, down to 20 dimensions. However, if you need to retain more original data information, there will be many dimensions after the dimensionality reduction, resulting in an insignificant classification effect. Table 5 is based on the experimental analysis of the IP dataset. When the other hyperparameters are not changed, the dimensionality reduction parameter numComponents is taken as a single variable and the impact on the classification effect when the dimensionality is reduced to different

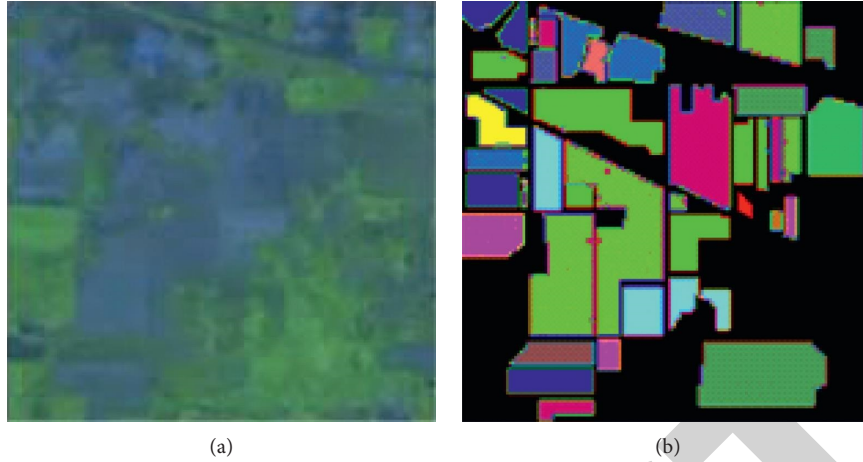


FIGURE 7: Data processing: (a) original image; (b) feature image.

TABLE 3: University of Pavia datasets.

S/N	Name	Train	Test
1	Asphalt	2133	3421
2	Meadows	6773	12,441
3	Gravel	421	1371
4	Trees	1120	2201
5	Painted metal sheets	234	1102
6	Bare soil	1423	4312
7	Bitumen	321	1212
8	Self-blocking bricks	888	2987
19	Shadowsl	299	799

TABLE 4: Salinas datasets.

S/N	Name	Train	Test
1	Brocoli_green_weeds_1	590	1345
2	Brocoli_green_weeds_22	989	2134
3	Fallow	598	1389
4	Fallow_rough_plow	212	1130
5	Fallow_smooth	897	1942
6	Stubble	1232	2567
7	Celery	1252	2745
8	Grapes_untrained	3012	8756
9	Soil_vinyard_develop	2011	4151
10	Corn_senesced_green_weeds	889	2451
11	Lettuee_romaine_4 wk	213	821
12	Lettuee_romaine_5 wk	812	1102
13	Lettuee_romaine_6 wk	231	771
14	Lettuee_romaine_7 wk	245	743
15	Vinyard_untrained	2228	3359
16	Vinyard_vertical_trellis	312	1550

dimensions. It can be seen from the table that when the dimensionality is reduced to 20, the three classification accuracy indicators of kappa, OA, and AA are all the highest. Therefore, under the condition that other parameters remain unchanged, this paper uses IPCA to reduce the dimensionality of the data to 20 dimensions. In the most suitable situation, it was found that IPCA preprocessed data faster during the experiment.

TABLE 5: Accuracy of different dimensionality reduction.

NumComponents	Kappa $\times 100$	OA	AA
15	98.91	99.12	96.54
20	99.56	99.49	99.87
25	98.66	98.65	98.12
30	99.47	99.45	99.59

4.4. The Impact of Different Spatial Dimensions on Classification Discussion. In a deep convolutional neural network, the larger the size of the input image, the larger the number of model convolution parameters and the higher the computational complexity. In addition, if the size of the input image is too small, the available fields received by the network will be too small, and a good classification result cannot be obtained. Table 6 shows the effect of different spatial neighborhood sizes on the performance of the proposed model. Set the spatial dimensions to $9 \times 9 \times 20$, $11 \times 11 \times 20$, $13 \times 13 \times 20$, $17 \times 17 \times 20$, $23 \times 23 \times 20$, and $25 \times 25 \times 20$, and pass the experiment on the three datasets. The training time of different windows and the classification accuracy of kappa, OA, and AA are obtained. The training time is highly dependent on the network speed, available memory, and the number of model parameters. When analyzing the accuracy of OA, AA, and K, it can be concluded that, with the gradual increase of the spatial dimension, the IP dataset basically shows an increasing trend, and the PU dataset decreases when the size is 23×23 , but the overall size is still increasing. The classification accuracy of the SA dataset is relatively stable. When the spatial input size reaches 11×11 , the classification accuracy begins to change slowly. The window size of 11×11 is sufficient for the three datasets of IP, PU, and SA in terms of accuracy and time. However, it is almost the same under the spatial dimensions of 13×13 , 17×17 , 23×23 , and 25×25 . Through experiments, it can be seen that, in the process of increasing spatial dimensions, the accuracy indicators of the model will increase significantly, but at the same time, the number of parameters increases and the calculation time increases. This method is mainly an improvement based on the fast 3D-CNN model. For the fairness of comparison, the network

TABLE 6: Impact of window size on I3D-CNN model in three datasets.

	IP				SA			
	K	OA	AA	Train_time (s)	K	OA	AA	Train_time (s)
9 × 9	99.07	99.12	98.78	19.65	99.78	99.89	99.87	199.1
11 × 11	99.3	99.34	99.90	45.80	99.95	99.98	99.95	243.5
13 × 13	99.88	99.87	99.76	36.78	99.98	99.78	99.98	314.5
17 × 17	99.54	99.5	99.85	65.15	99.87	99.99	99.99	412.6
23 × 23	99.78	99.79	99.78	147.78	100	100	100	1025.77
25 × 25	98.56	99.89	99.81	172.78	99.97	99.97	99.99	1451.54

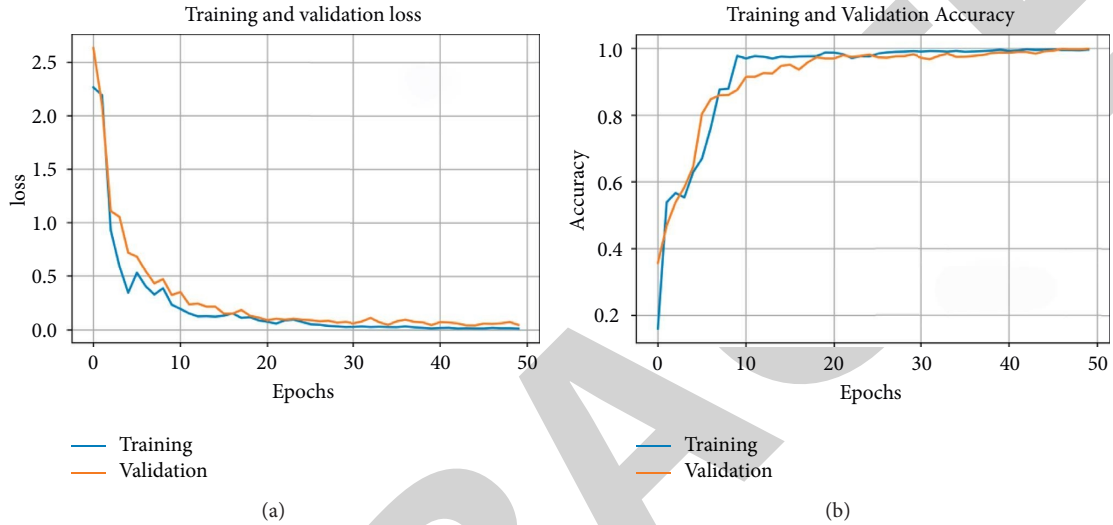


FIGURE 8: Loss curve and accuracy of space size of 9 × 9. (a) Training and validation loss. (b) Training and validation accuracy.

hyperparameters are unchanged, and the dimension size is also selected as $11 \times 11 \times 20$. Through comparison, it can be seen that the amount of model parameters is reduced, the accuracy of each classification index is also relatively high, and the training time is reduced.

5. Discussion

5.1. Classification Loss Rate and Accuracy Rate. The experiment analyzes the stability and fit of the network model by training and verifying the loss rate and accuracy rate. Mainly carry out experimental verification on the IP dataset. The curve in Figure 8 shows the loss rate and accuracy classification effect of the training and validation set when the window size of the IP dataset is 9×9 . Compared with the graph under the size of 11×11 , the model does not fit well. This article mainly focuses on the 11×11 size situation for training. As can be seen from the curve in Figure 9, from the left figure, it can be seen that the model starts to converge when the epochs reach about 15, and the loss rate of the training set and the validation set is close to 0. Figure 9(b) on the right also starts to converge when the epochs reach about 15, and the training set and validation set reach nearly 100% accuracy. It can be seen from the two figures that the curves of the training set and the validation set are basically the same, there is no large oscillation phenomenon, and the model has a fairly high degree of fit. Through experiments, we can see that the proposed model is relatively stable, converges very fast, and has high classification accuracy.

5.2. Comparison of Experimental Performance under Different Methods. In order to verify the correctness and effectiveness of the proposed network model method, finally compare the proposed convolutional network method with the traditional convolution model [40] and 2D-CNN [33], 3D-CNN [34], Multiscale-3D-CNN [41], and Hybrid SN [42] methods. In order to ensure the fairness of the experiment, the hyperparameters in all comparison networks are set the same; for example, the input data is reduced to 20 dimensions, the spatial dimension is set to $11 \times 11 \times 20$, the epoch period is 50, and the batch size is 256. As in the previous experiment, 60% of the training data and 40% of the test data were randomly selected from the three datasets of Indian Pines, Salinas scene, and Pavia University for verification, the experiment was repeated 30 times, and finally, the average value of these 30 times was taken. Table 7 shows the experimental results under different methods. It can be seen from the table that the proposed method is compared with Hybrid SN, which has the best classification performance among other methods. For the Indian Pines dataset, its OA is 1.86% higher, AA is 2.11% higher, and kappa coefficient is 2.34 higher. For the Salinas scene dataset, the OA is 1.89% higher, the AA is 1.16% higher, and the kappa coefficient is 2.09 higher. For the Pavia University dataset, its OA is 1.53% higher, AA is 1.9% higher, and the kappa coefficient is 2.02 higher. It can be seen that, on the basis of 3D-CNN, the combination of deep deconvolution has a better classification effect. Figure 10 shows the comparison of classification diagram between the I3D-CNN

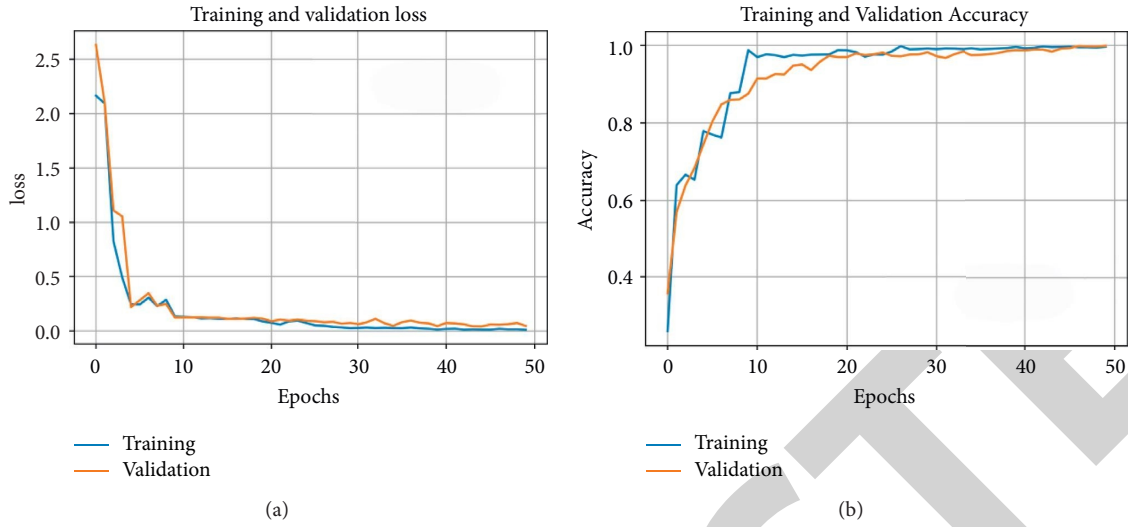


FIGURE 9: Loss curve and accuracy of space size of 9×9 . (a) Training and validation loss. (b) Train and validation accuracy.

TABLE 7: Comparison of experimental performance under difference methods.

Methods		2D-CNN	3D-CNN	Multi_scale-3D-CNN	Hybrid SN	I3D-CNN
Indian Pines	OA	80.11	82.12	98.78	97.34	99.98
	AA	65.42	75.12	98.90	97.46	99.97
	Kappa	73.22	80.12	96.76	97.44	99.99
Salinas scene	OA	96.56	89.12	95.85	98.62	99.99
	AA	94.36	83.45	95.78	97.43	98.78
	Kappa	95.93	83.43	99,581	95.87	99.78

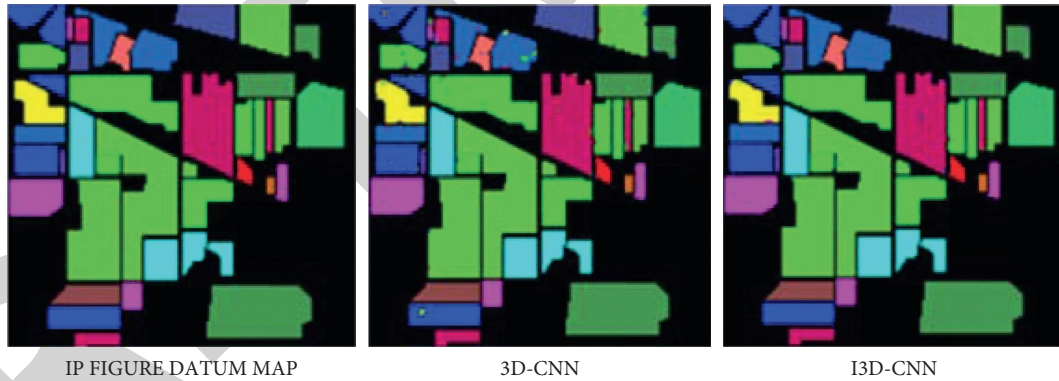


FIGURE 10: Comparison of classification of two models.

model and other CNN models. It can be seen from the figure that 3D-CNN with depth separable convolution has a better classification effect when other conditions are the same. Figure 11 shows the confusion matrix of the PU dataset. It can be seen that the classification accuracy of most of the features in the PU dataset has reached 100%, such as asphalt, meadows, and painted metal sheets. Only individual features are misclassified, such as 0.1% of gravel and divided into self-blocking bricks, showing a more obvious classification effect.

Figure 11 shows the effect of the hyperspectral image classification diagram under different methods, and the advantages of the proposed method can be seen from the effect diagram. It can be seen that the mentioned method, combined with the fast 3D-CNN model of deep separable convolution, has a better classification effect on the classification of hyperspectral images. It can be seen that there are a few misclassified classes, which are obvious in the confusion matrix.

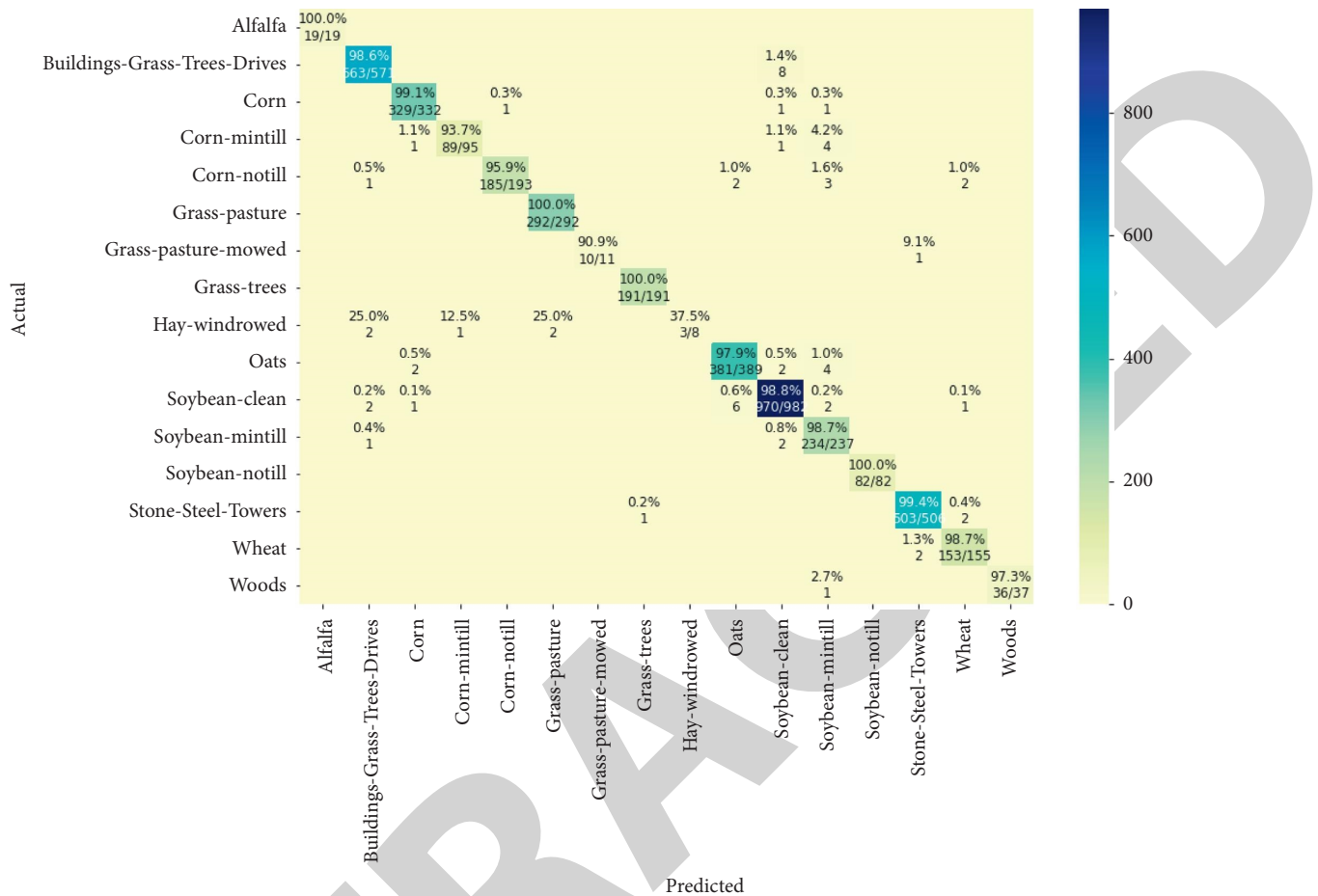


FIGURE 11: Confusion matrix of the PU dataset.

6. Conclusion

The I3D-CNN is proposed in this paper combined with the deep separable convolution of the hyperspectral image classification method. First, IPCA is used to preprocess the original hyperspectral image for dimensionality reduction, reducing the redundant spectrum, and reducing the number of image bands while maintaining the spatial dimension. We use a three-dimensional convolutional neural network to extract spectral and spatial features at the same time and then introduce deep separable convolution and design a new convolutional layer DSC layer. This layer gives full play to the advantages of deep separable convolution for spatial feature extraction and can greatly save learnable parameters; finally, based on two convolution methods, a network framework combining fast 3D-CNN and deep separable convolution is designed. Experiments show that this method not only shows better classification performance under limited label samples but also greatly reduces model complexity, reduces learnable parameters, and saves memory space compared with models based on standard convolutional layers.

Comparing the proposed model method with other traditional convolutional neural network methods, the classification performance is better, but there are still many shortcomings in this paper, for example, how to design a

more complete deep convolutional network model to solve the problem of network gradient decline, which will become the next research focus.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] J. Rogan and D. Chen, "Remote sensing technology for mapping and monitoring land-cover and land-use change," *Progress in Planning*, vol. 61, no. 4, pp. 301–325, 2004.
- [2] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2018.
- [3] R. N. Clark, T. V. V. King, M. Klejwa, G. A. Swayze, and N. Vergo, "High spectral resolution reflectance spectroscopy of minerals," *Journal of Geophysical Research*, vol. 95, no. B8, pp. 12653–12680, 1990.

- [4] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2012.
- [5] 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.
- [6] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 2020.
- [7] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, and Q. Du, "More diverse means better: multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, 2020.
- [8] I. Numata, D. Roberts, O. Chadwick, J. Schimel, L. Galvao, and J. Soares, "Evaluation of hyperspectral data for pasture estimate in the Brazilian Amazon using field and imaging spectrometers," *Remote Sensing of Environment*, vol. 112, no. 4, pp. 1569–1583, 2008.
- [9] L. Gaddis, J. Anderson, K. Becker, T. Becker, D. Cook, and K. Edwards, "An overview of the integrated software for imaging spectrometers (ISIS)," in *Proceedings of the Lunar and Planetary Science Conference*, vol. 28, p. 387, Houston, TX, USA, May 1997.
- [10] M. Petrou and P. G. Foschi, "Confidence in linear spectral unmixing of single pixels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 624–626, 1999.
- [11] N. Lawson, K. Thompson, G. Saunders et al., "Sound intensity and noise evaluation in a critical care unit," *American Journal of Critical Care*, vol. 19, no. 6, pp. e88–e98, 2010.
- [12] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823–870, 2007.
- [13] C. C. Noble and D. J. Cook, "Graph-based anomaly detection," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 631–636, Washington, DC, USA, August 2003.
- [14] I. Bekkerman and J. Tabrikian, "Target detection and localization using MIMO radars and sonars," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3873–3883, 2006.
- [15] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: common subspace learning from hyperspectral-multispectral correspondences," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4349–4359, 2019.
- [16] D. Hong, N. Yokoya, J. Chanussot, X. X. Zhu, and J. Xu, "Joint and progressive subspace analysis (JPSA) with spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction," *IEEE Transactions on Cybernetics*, vol. 51, no. 7, pp. 3602–3615, 2020.
- [17] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: a spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3791–3808, 2020.
- [18] M. Shimoni, R. Haelterman, and C. Perneel, "Hyperspectral imaging for military and security applications: combining myriad processing and sensing techniques," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 101–117, 2019.
- [19] I. Dópido, A. Villa, A. Plaza, and P. Gamba, "A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification," *IEEE Journal of Selected topics in applied earth observations and remote sensing*, vol. 5, no. 2, pp. 421–435, 2012.
- [20] T. Li, G. Kou, and Y. Peng, "Improving malicious URLs detection via feature engineering: linear and nonlinear space transformation methods," *Information Systems*, vol. 91, Article ID 101494, 2020.
- [21] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: a detailed tutorial," *AI Communications*, vol. 30, no. 2, pp. 169–190, 2017.
- [22] K. R. Kini and M. Madakyaru, "Improved process monitoring scheme using multi-scale independent component analysis," *Arabian Journal for Science and Engineering*, pp. 1–16, 2021.
- [23] Y. Wan, Y. Fan, and M. Jin, "Application of hyperspectral remote sensing for supplementary investigation of polymetallic deposits in Huaniushan ore region, northwestern China," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.
- [24] M. P. Uddín, M. A. Mamun, M. I. Afjal, and M. A. Hossain, "Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification," *International Journal of Remote Sensing*, vol. 42, no. 1, pp. 286–321, 2021.
- [25] I. Baidari and N. Honnikoll, "Bhattacharyya distance based concept drift detection method for evolving data stream," *Expert Systems with Applications*, vol. 183, Article ID 115303, 2021.
- [26] J. Pei, "Big data mining in the control of epidemic," *Basic and Clinical Pharmacology and Toxicology*, vol. 126, pp. 429–430, 2020.
- [27] T. P. Nguyen, S. Choi, S.-J. Park, S. H. Park, and J. Yoon, "Inspecting method for defective casting products with convolutional neural network (CNN)," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 8, no. 2, pp. 583–594, 2021.
- [28] X. Wang, S. Zhang, Z. Qing, Y. Shao, C. Gao, and N. Sang, "Self-supervised learning for semi-supervised temporal action proposal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1905–1914, Nashville, TN, USA, June 2021.
- [29] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.
- [30] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected topics in applied earth observations and remote sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.
- [31] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381–2392, 2015.
- [32] A. Ghosh and R. Chellappa, "Deep feature extraction in the DCT domain," in *Proceedings of the 2016 23rd international conference on pattern recognition (ICPR)*, pp. 3536–3541, IEEE, Cancun, Mexico, December 2016.
- [33] W. Zhao and S. Du, "Learning multiscale and deep representations for classifying remotely sensed imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 113, pp. 155–165, 2016.
- [34] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4520–4533, 2017.

- [35] 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.
- [36] C. Wu, F. Wu, S. Wu, Z. Yuan, J. Liu, and Y. Huang, "Semi-supervised dimensional sentiment analysis with variational autoencoder," *Knowledge-Based Systems*, vol. 165, pp. 30–39, 2019.
- [37] Z. Cai, J. Chen, and M. Liu, "Least-squares ReLU neural network (LSNN) method for linear advection-reaction equation," *Journal of Computational Physics*, vol. 443, Article ID 110514, 2021.
- [38] Y. Ma, S. Arshad, S. Muniraju et al., "Location- and person-independent activity recognition with WiFi, deep neural networks, and reinforcement learning," *ACM Transactions on Internet Technology*, vol. 2, no. 1, pp. 1–25, 2021.
- [39] C. J. Bruegge, G. T. Arnold, J. Czapla-Myers, R. A. Dominguez, and M. C. Helmlinge, "Vicarious calibration of eMAS, AirMSPI, and AVIRIS sensors during FIREX-AQ," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10286–10297, 2021.
- [40] S. Kumar, "McftCFT-CnnNN: Malware classification with fine-tune convolution neural networks using traditional and transfer learning in Internet of Things," *Future Generation Computer Systems*, vol. 125, 2021.
- [41] Y. Qing and W. Liu, "Hyperspectral image classification based on multi-scale residual network with attention mechanism," *Remote Sensing*, vol. 13, no. 3, p. 335, 2021.
- [42] C. Cai, J. Xu, H. Wang, and S. B. Park, "A comparative study of thermal fatigue life of Eutectic Sn-Bi, Hybrid Sn-Bi/SAC and SAC solder alloy BGAs," *Microelectronics Reliability*, vol. 119, Article ID 114065, 2021.