Research Article

Composite Clustering Sampling Strategy for Multiscale Spectral-Spatial Classification of Hyperspectral Images

Chenming Li, Xiaoyu Qu, Yao Yang, Dan Yao, Hongmin Gao ⓘ, and Zaijun Hua

College of Computer and Information, Hohai University, Nanjing 211100, China

Correspondence should be addressed to Hongmin Gao; gaohongmin@hhu.edu.cn

In recent years, many high-performance spectral-spatial classification methods were proposed in the field of hyperspectral image classification. At present, a great quantity of studies has focused on developing methods to improve classification accuracy. However, some research has shown that the widely adopted pixel-based random sampling strategy is not suitable for spectral-spatial hyperspectral image classification algorithms. Therefore, a composite clustering sampling strategy is proposed, which can greatly reduce the overlap between the training set and the test set, while making sample points in the training set sufficiently representative in the spectral domain. At the same time, in order to solve problems of a three-dimensional Convolutional Neural Network which is commonly used in spectral-spatial hyperspectral image classification methods, such as long training time and large computing resource requirements, a multiscale spectral-spatial hyperspectral image classification model based on a two-dimensional Convolutional Neural Network is proposed, which effectively reduces the training time and computing resource requirements.

1. Introduction

Hyperspectral image (HSI) is acquired by dedicated hyperspectral cameras, which contains the spectral information of a same ground object in hundreds of continuous bands [1]. Compared with traditional remote sensing images such as RGB three-band remote sensing images and multispectral remote sensing images, the imaging bands of HSIs have greatly increased. Each band of HSIs maps a two-dimensional (2D) image with spatial geometric relationship, and each pixel has a spectral characteristic curve. Therefore, HSIs effectively combine spatial and spectral information and are widely used in the field of remote sensing. In a variety of application scenarios of HSIs, HSI classification technology is relatively mature and is widely used in urban research [2], marine disaster forecasting [3], and other tasks.

In recent years, some research in the field of deep learning has attracted the attention of many scholars. Deep learning is a new field of machine learning. With the improvement of computer processing capabilities and the emergence of excellent algorithms, the performance of various methods was considerably improved in the field of deep learning. At the

same time, deep learning is also widely combined with remote sensing. Many great models for HSI classification were proposed, such as stacked autoencoder (SAE) [4], deep belief networks (DBN) [5], recurrent neural network (RNN) [6], and Generative Adversarial Networks (GAN) [7].

In the field of deep learning, a Convolutional Neural Network (CNN) is much stronger than other deep learning models in feature selection and extraction of high-dimensional data. CNNs were widely used in HSI classification [8–10]. However, CNN models still have a great possibility to improve the performance of HSI classification. When CNNs were first used for HSI classification, models only used the spectral feature information of pixels [11, 12], which largely wastes an advantage of HSIs; that is, spatial information and spectral information are closely combined. In response to this problem, a spatial-spectral classification method was proposed [13], and the spatial position information and spectral information of pixels were full used during model training, which effectively improved classification performance of CNN models. Therefore, the spatial-spectral classification method has been widely used [14, 15]. In a large number of studies on spatial-spectral HSI classification

methods, it was found that multiscale spatial-spectral HSI classification methods can effectively enhance the classification ability and robustness of models [16]. At present, multiscale spatial-spectral HSI classification methods have very high classification performance [17].

Among many spatial-spectral HSI classification models, 3D-CNN models have achieved best performance, so most researchers choose 3D-CNN to build multiscale spatial-spectral HSI classification models [18, 19]. However, 3D-CNN has problems such as too many network parameters, too much computing resources for training, and too long training time, which limits the popularization and application of multiscale spatial-spectral HSI classification methods. Therefore, it is necessary to improve existing models.

In recent years, many methods were studied for spatial-spectral HSI classification. A large portion of the hyperspectral remote sensing community has focused their research on improving the classification accuracy by developing a variety of spectral-spatial methods [9, 20, 21], but little attention has been paid to experimental settings. In supervised deep learning methods, before models start training, labeled original dataset needs to be divided into a training set and a test set. Because it is difficult to obtain labeled HSI datasets, researchers generally use public HSI datasets such as Indian Pines. Therefore, in the research process, the training set and test set are often divided on a same HSI. In previous HSI classification methods that use only the spectral feature information of pixels, the most commonly used sampling strategy is to randomly select pixels on a HSI according to a predetermined proportion to form the training set, and remaining pixels constitute the test set. The random sampling strategy is consistent with people's intuition. It can select as representative pixels as possible to form the training set and make the training set and test set approximately meet the conditions of independent and identical distribution. Therefore, in research of spectral-spatial HSI classification methods, almost all studies have adopted the traditional random sampling strategy by default.

However, it is found that the use of the random sampling strategy in spectral-spatial HSI methods is unreasonable, as it will cause unfair performance evaluation [22]. In spectral-spatial HSI methods using the random sampling strategy, the correlation caused by an overlap between training samples and test samples will amplify classification accuracy, resulting in an improper evaluation of spectral-spatial HSI classification methods. The sampling problem was originally noticed by Friedl et al. [23], who referred to the overlap as autocorrelation. Geiß et al. [24] compared the effects of different sampling strategies and verified the necessity of using appropriate sampling strategies for model evaluation. Liang et al. [25] proved the influence of data dependence on the credibility of models by computational learning theory. Therefore, the widely adopted pixel-based random sampling strategy is not always suitable for spectral-spatial HSI classification algorithms, because it is difficult to determine whether the improvement of classification accuracy is caused by incorporating spatial information into a classifier or by increasing overlap between training and testing samples.

To solve this problem, several new sampling strategies were proposed. Liang et al. [25] proposed a new controlled random sampling strategy, which effectively enhanced the independence between the training set and the test set in spectral-spatial methods. Lange et al. [26] proposed two improved sampling strategies based on the density-based clustering algorithm (DBSCAN), which also enhanced the independence between the training set and the test set in spectral-spatial methods. However, the controlled random sampling strategy is to randomly select seed points in the entire dataset, without taking into account the problem of category imbalance; that is, the phenomenon that the number of pixels of one class is far more or less than that of other classes in the dataset. HSIs have the problem of the same spectrum from different materials and the same materials with different spectra. These sampling strategies are based on the spatial location of pixels, without considering the spectral domain representativeness of pixels.

In view of the above problems, this paper attempts to combine the excellent performance of DBSCAN on nonconnected regions and the good effect of the k-means clustering algorithm on connected regions and combines the spectral information of HSI datasets to propose a composite clustering sampling strategy. At the same time, an efficient multiscale spatial-spectral HSI classification model is proposed. It combines the advantages of 3D-CNN and two-dimensional Convolutional Neural Network (2D-CNN) to effectively extract multiscale spatial-spectral features while reducing the need for computing resources. The main contributions of this paper are as follows.

(1) A new composite clustering sampling strategy is proposed, which combines the respective characteristics of DBSCAN and k-means clustering algorithms, and uses the spectral domain average variance as a metric. Compared with other sampling strategies, the proposed method effectively improves the classification accuracy of spatial-spectral methods. Although it also reduces independence between the training set and the test set, the loss of independence is extremely limited and acceptable

(2) Combining the advantages of 3D-CNN and 2D-CNN, a new multiscale spatial-spectral HSI classification model is proposed. The proposed model not only can effectively extract multiscale spatial-spectral features of pixels but also overcomes problems of too many model parameters, long training time, and high computing resource requirements in existing methods

2. Related Work

2.1. DBSCAN. DBSCAN is a classic density-based clustering algorithm. Its basic principle is clustering by finding the largest set of densely connected points. It uses local density of points to divide clusters and does not need to set the number of clusters k in advance. Different from partition clustering methods and hierarchical clustering methods, it defines the

cluster as a largest set of densely connected points. It can find clusters of arbitrary shape in the dataset and can effectively identify connected areas in the dataset. Below, some key definitions [27] are given for dataset $D = \{x_1, x_2, \cdots, x_m\}$.

Definition 1. (*Eps-neighborhood*). For $x_i \in D$, the *Eps-neighborhood* of x_i contains points in dataset $D = \{x_1, x_2, \cdots, x_m\}$ whose distance from x_i is not more than *Eps*, that is, $N_{\text{Eps}}(x_i) = \{x_j \in D \mid \text{dist}(x_i, x_j) \leq \text{Eps}\}$.

Definition 2. (*core points*). For $x_i \in D$, if there are at least *MinPts* points in the *Eps-neighborhood* of x_i, then point x_i is called *core point*.

Definition 3. (*directly density-reachable*). For $x_i, x_j \in D$, if these conditions are satisfied: x_i is a core point and $x_j \in N_{\text{Eps}}(x_i)$, then x_j is said to be *directly density-reachable* from x_i.

Definition 4. (*density-reachable*). If there exists a series of points $y_1, y_2, \cdots, y_l$ that satisfy $x_i = y_1$, $x_j = y_l$, and y_{i+1} is *directly density-reachable* from y_i, then x_j is said to be *density-reachable* from x_i.

Definition 5. (*density-connected*). If there is a point y such that both x_i and x_j are *density-reachable* from y, then x_i and x_j are said to be *density-connected*.

DBSCAN defines the maximum set of a series of densely connected points as a cluster, and points that do not belong to any cluster are defined as noise. It works well on datasets with well-defined category boundaries.

2.2. k-Means Clustering Algorithm. The k-means clustering algorithm was proposed by MacQueen in 1967 [28]. Because of its advantages such as good effect and simple idea, the k-means clustering algorithm was widely used. The k-means clustering algorithm generally uses Euclidean distance as an index to measure the similarity between points. The similarity between points is inversely proportional to distance. The k-means clustering algorithm needs to set the number of clusters k in advance. The algorithm initially randomly selects k points as the center of clusters. Based on the similarity between points and the center of clusters, the position of the cluster center is continuously updated to reduce the Sum of Squared Error (SSE) of clusters. When SSE no longer changes or objective function converges, the algorithm ends and final result is obtained.

The formula for calculating the Euclidean distance between any point x and cluster center C_i in dataset $D = \{x_1, x_2, \cdots, x_m\}$ is

$$d(x, C_i) = \sqrt{\sum_{j=1}^{n} (x_j - C_{ij})^2}, \tag{1}$$

where x is a point, C_i is the ith cluster center, m is the dimension of points, and x_j and C_{ij} are the jth attribute values of x and C_{ij}.

The SSE calculation formula for dataset $D = \{x_1, x_2, \cdots, x_m\}$ is

$$\text{SSE} = \sum_{i=1}^{k} \sum_{x \in D} |d(x, C_i)|^2, \tag{2}$$

where k is the number of clusters.

2.3. CNN. CNN is a kind of feed-forward neural network. The main network layers of CNN include convolutional layers, pooling layers, and fully connected layers. Different from traditional neural networks, CNNs have the characteristics of sparse connections and weight sharing and have better stability and generalization capabilities [29]. According to different input signal dimensions, CNN can be divided into one-dimensional (1D) CNN, 2D-CNN, and 3D-CNN. In application, different network models are selected according to requirements. In these three CNNs, the structure of the convolution kernel is similar. Because 2D-CNN is most widely used, the formulas for 2D convolution kernels are listed below [30].

In 2D-CNN, the value $v_{i,j}^{x,y}$ of the jth feature map in the ith layer at point (x, y) is calculated as

$$v_{i,j}^{x,y} = f\left(\sum_m \sum_{h=0}^{H_i-1} \sum_{w=0}^{W_i-1} k_{i,j,m}^{h,w} \cdot v_{i-1,m}^{(x+h)(y+w)} + b_{i,j} \right), \tag{3}$$

where m represents the feature map connected to the current feature map in the $(i-1)$th layer, H_i and W_i represent the length and width of the convolution kernel, $v_{i,j}^{x,y}$ represents the value of position (x, y) on the jth convolution kernel in the ith layer, and $k_{i,j,m}^{h,w}$ represents the connection weight of the mth feature map connected to (h, w), and $b_{i,j}$ represents the bias of the jth feature map in the ith layer.

3. Proposed Method

3.1. Composite Clustering Sampling Strategy. In this paper, combining the advantages of DBSCAN and k-means clustering methods, a composite clustering sampling strategy is proposed. At the same time, the composite cluster sampling strategy uses the spectral domain average variance as a measure, which not only makes divided training set and test set maintain high independence but also makes the sample points in the training set have high spectral domain representativeness.

At the beginning of the composite clustering sampling strategy, DBSCAN is used for the first clustering of the HSI dataset. The purpose of this step is to divide the HSI dataset into multiple partitions. The partition referred here is a group of connected pixels with the same labels. For each class, there are usually several partitions distributed on the map, corresponding to the land cover of the same category in different locations. With the excellent performance of DBSCAN in identifying class boundaries, different partitions under the same category can be effectively identified.

For each partition identified by DBSCAN, the k-means algorithm is used for the second clustering. The k-means clustering algorithm performs well on the connected region; it can be used to divide each partition into k clusters, which are clustered spatially. For each spectral dimension of each cluster, the variance of all pixels in the cluster is calculated under this dimension. The variances of these different spectral dimensions are averaged, and it is called spectral domain average variance. The proposed spectral domain average variance can effectively evaluate differences among pixels within a cluster. Because pixels in a cluster have the same label, if the spectral domain average variance of a cluster is large, it means that pixels in the cluster have large differences. Obviously, for each class, it is desirable to choose as many different samples as possible to form the training set, so that models can be fully trained. Therefore, different clusters in a partition are sorted in a descending order according to spectral domain average variance, sampling points are acquired according to the sampling proportion to form the training set, and the remaining part constitutes the test set.

In some cases, there are few pixels in some partitions, which is not suitable for secondary clustering. Therefore, if the number of samples in a cluster is less than k (the number of clusters), the pixels in the partition are sorted according to spatial position, and then, sampling points are acquired according to a sampling ratio and incorporated into the training set. As shown in Figure 1, the composite clustering sampling strategy proposed in this paper mainly includes the following steps.

Step 1. For each class in the HSI dataset, use DBSCAN to perform the first clustering based on pixel coordinates to obtain all partitions in this class.

Step 2. Determine whether to perform secondary clustering based on the number of pixels in a partition. If the number of pixels is small, pixels are sorted according to spatial position, and then, sampling points are obtained according to a predetermined sampling rate and incorporated in the training set. Otherwise, the k-means method is used for the second clustering on the partition to form k clusters.

Step 3. Calculate the spectral domain average variance of each cluster obtained in Step 2; sort these clusters in a descending order according to spectral domain average variance.

Step 4. According to a predetermined sampling proportion, sample points are intercepted from the ordered arrangement obtained in Step 3 and incorporated in the training set.

3.2. New Multiscale Spatial-Spectral HSI Classification Model.
In traditional multiscale spatial-spectral HSI classification methods, a large-scale three-dimensional (3D) data block is often divided with each pixel as the center; then, convolution kernels of different sizes are used to extract multiscale spatial-spectral information. The extracted spatial-spectral features are integrated to form a spatial-spectral feature map, and the feature map is input to the subsequent network structure for training. As shown in Figure 2, traditional multiscale spatial-spectral classification methods are shown.

However, although traditional multiscale spatial-spectral classification methods can effectively extract multiscale
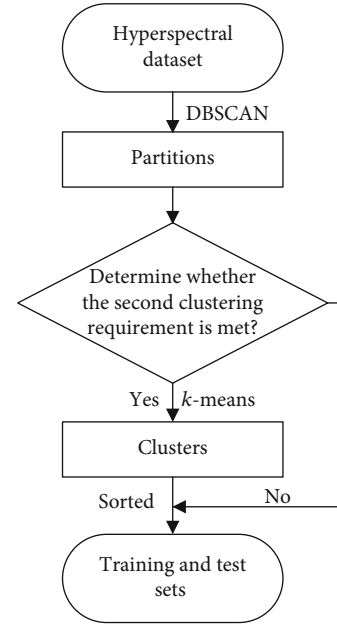

Figure 1: A flow chart of a composite clustering sampling strategy.

spatial-spectral features, noise information is often introduced. As shown in Figure 3, when the size of 3D data block is large, smaller convolution kernels will obtain spatial-spectral features that do not belong to the central pixel, which will bring noise to the extracted spatial-spectral feature map.

In view of this problem, based on the work of improving existing methods and models, a new multiscale spatial-spectral HSI classification model is proposed, which is called 1D-3D-2D-CNN model. In the 1D-3D-2D-CNN model, the spectral data of pixels and a plurality of 3D data blocks of different scales centered on pixels are extracted. For the extracted 1D spectral information, a 1D convolution kernel is used to extract spectral features; for extracted 3D data blocks, corresponding 3D convolution kernels are used to extract multiscale spatial-spectral features. Then, the extracted spectral features and multiscale spatial-spectral features are transposed into 2D feature maps and fused to obtain a multiscale spatial-spectral feature map and input it to a subsequent 2D-CNN for training. As shown in Figure 4, the new multiscale spatial-spectral HSI classification model proposed in this paper mainly includes the following steps.

Step 1. For each pixel, extracted spectral information corresponding to the pixel and a plurality of data blocks with different sizes centered on the pixel is, respectively, divided.

Step 2. The spectral data is convolved using a 1D convolution kernel to extract spectral features; 3D convolution kernels are used to convolve corresponding data blocks to extract multiscale spatial-spectral features.

Step 3. The extracted spectral features and multiscale spatial-spectral features are reshaped into 2D feature maps and fused to obtain a multiscale spatial-spectral map.

Step 4. The multiscale spatial-spectral feature map is input into 2D-CNN for training.
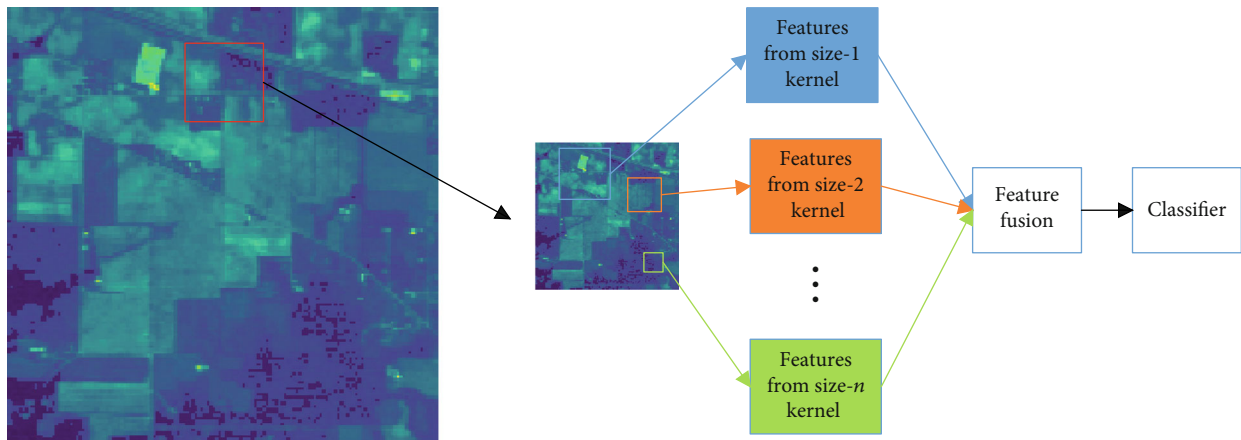
FIGURE 2: A schematic diagram of traditional multiscale spatial-spectral classification methods.
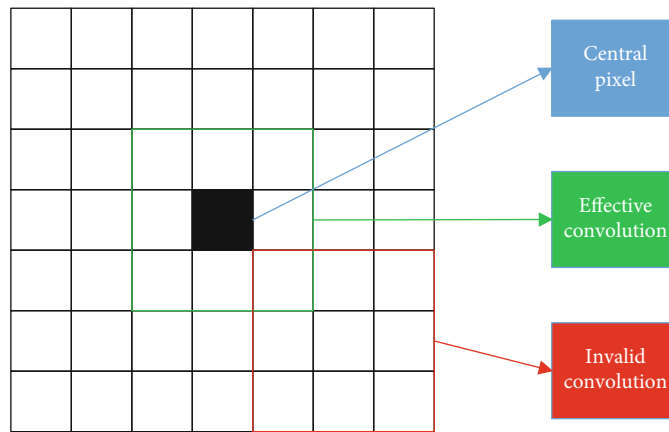

FIGURE 3: The disadvantage of traditional multiscale spatial-spectral classification methods.
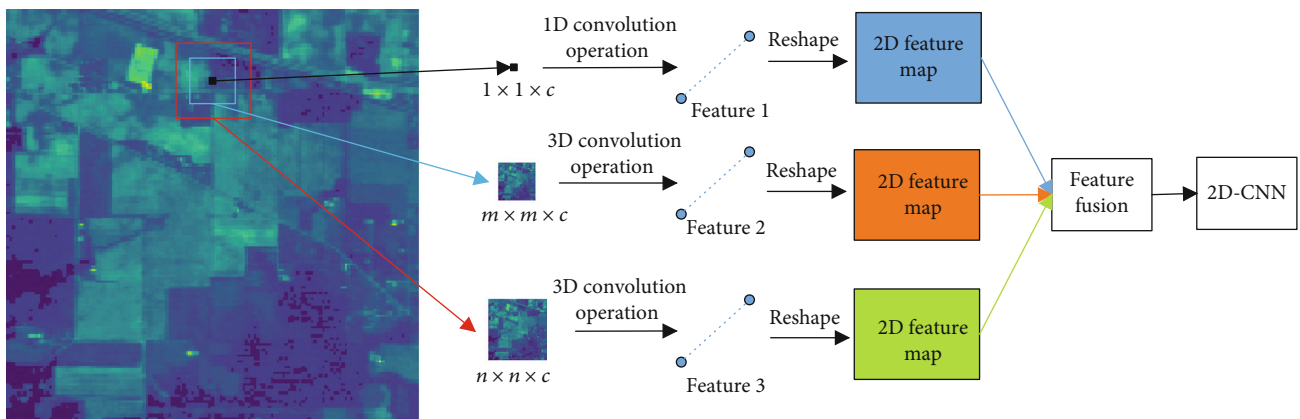
FIGURE 4: The new multiscale spatial-spectral HSI classification model.

3.3. Multiscale Spatial-Spectral HSI Classification Method Based on Composite Clustering Sampling Strategy.

Combining the work done in Sections 3.1 and 3.2, a multiscale spatial-spectral HSI classification method based on the composite clustering sampling strategy is proposed.

This method mainly includes the following steps.

Step 1. The composite clustering sampling strategy is used to divide the original HSI dataset into the suitable training set and test set.

Step 2. The training set is input to the 1D-3D-2D-CNN model shown in Figure 5 for training.

Step 3. The test set is input to the trained model for testing.

Figure 5: A flow chart of the 1D-3D-2D-CNN.

	Stone-steel-towers		Buildings-grass-trees-drives
	Woods		Wheat
	Soybean-clean		Soybean-mintill
	Soybean-notill		Oats
	Hay-windrowed		Grass-pasture-mowed
	Grass-trees		Grass-pasture
	Corn		Corn-mintill
	Corn-notill		Alfalfa

Figure 6: Band 11 pseudocolor map and category label map of Indian Pines dataset.

4. Experimental Results

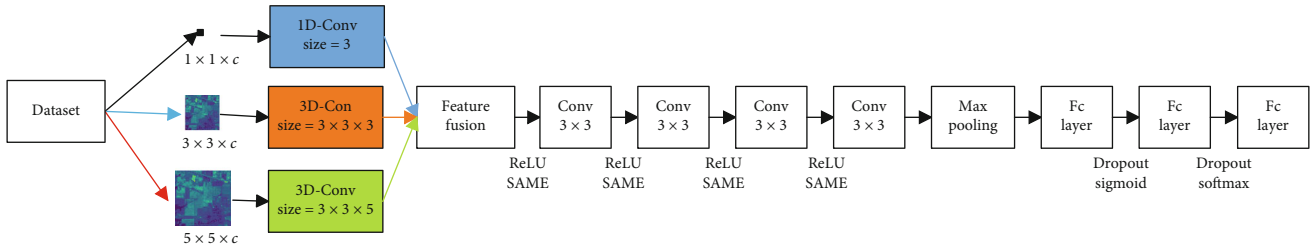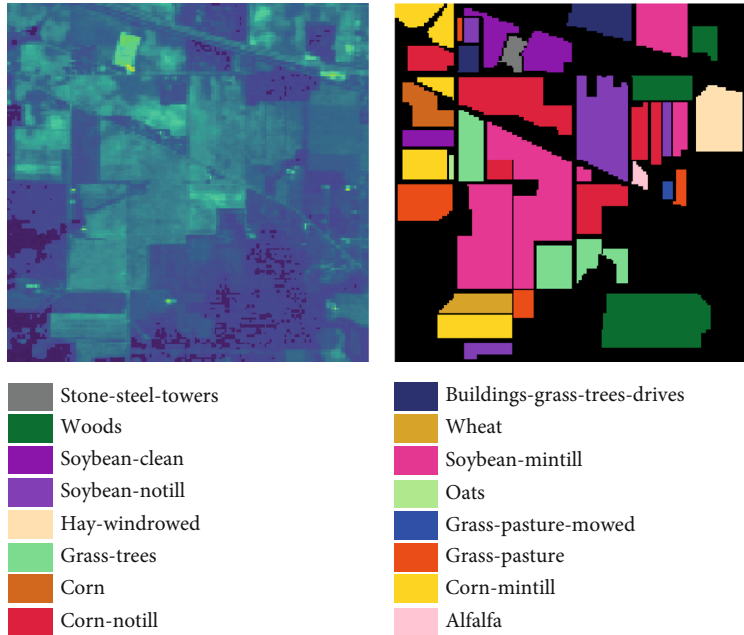In order to verify the effectiveness of the proposed multiscale spatial-spectral HSI classification method based on the composite clustering sampling strategy, experiments are performed on three commonly used public HSI datasets: Indian Pines, Pavia University, and Salinas. The experimental environment is the Google Colaboratory cloud computing platform. Google Colaboratory is provided by Google Inc. and provides free GPU acceleration services for artificial intelligence (AI) researchers.

In this paper, three indicators, Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient, are used as the evaluation criteria for model performance. Among them, OA points to each pixel to evaluate the classification accuracy of all sample points in the test set; AA points to the category, which means the average of classification accuracy of each class; Kappa considers the number of correctly classified pixels and misclassified pixels at the same time; it is an index to evaluate the consistency and credibility of classification results. In order to reduce the impact of random errors, all experimental data in this chapter are the average of five independent repeated experiments.

4.1. Dataset. The Indian Pines dataset is a HSI of the Indian Pines area in northwestern Indiana, USA, obtained by an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). It has a spatial resolution of 20 meters, is composed of 145 ×145 pixels, and contains 200 spectral bands (24 bands affected by water vapor and ozone are removed). Its wavelength range is between 0.4 and 2.5 microns. The ground object reference map of the Indian Pines dataset contains 16 different classes, including farmland, woods, grassland, and other vegetation. The number of pixels in different categories in the Indian Pines dataset is extremely uneven. Figure 6 is the 11th band pseudocolor map and the category label map of the Indian Pines dataset. As can be seen from Figure 6, the pixels of the Indian Pines dataset are clustered and the category boundaries are clear. Table 1 lists the category names of the Indian Pines dataset and the number of pixels in each category. As can be seen from Table 1, the Indian Pines dataset has a small number of pixels, with only 10249 labeled pixels. The number of pixels in some categories is too low. For example, category "alfalfa" has only 46 pixels, category "grass-pasture-mowed" has only 28 pixels, and category "oats" has only 20 pixels. When the number of sample points is too low, the effectiveness of deep learning methods

TABLE 1: Category name and the number of pixels in the Indian Pines dataset.

Label	Category name	The number of pixels
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-grass-trees-drives	386
16	Stone-steel-towers	93
Total number of pixels		10249

is not good, but for the sake of experimental consistency, experiments in this chapter did not exclude these small pixel categories.

The Pavia University dataset is a HSI of the Pavia University campus in northern Italy acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-3) developed in Germany. It has a spatial resolution of 1.3 meters, is composed of 610×340 pixels, and contains 103 spectral bands (12 bands affected by water vapor and ozone are removed). Its wavelength range is between 0.43 and 0.86 microns. The ground object reference map of the Pavia University dataset contains 9 different classes, such as lawn and macadam. Figure 7 is the 60th band pseudocolor map and the category label map of the Pavia University dataset. As can be seen from Figure 7, only in categories "bare land" and "lawn", pixels are clustered and category boundaries are clear. In other seven categories, pixels are scattered and category boundaries are blurred. Table 2 lists the category names of the Pavia University dataset and the number of pixels in each category. As can be seen from Table 2, the Pavia University dataset has a large number of labeled pixels, but only nine categories. The categories in the dataset are relatively balanced, and there is no phenomenon that the number of pixels in one category is too low.

The Salinas dataset is a HSI of Salinas Valley, California, USA, acquired by AVIRIS sensors. It has a spatial resolution of 3.7 meters, is composed of 512×217 pixels, and contains 204 spectral bands (20 bands affected by water vapor and ozone are removed). The ground object reference map of the Salinas dataset contains 16 different classes. Figure 8 is the 188th band pseudocolor map and the category label map of the Salinas dataset. As can be seen from Figure 8, labeled pixels of the Salinas dataset are clustered and category boundaries are clear. Table 3 lists the category names of the

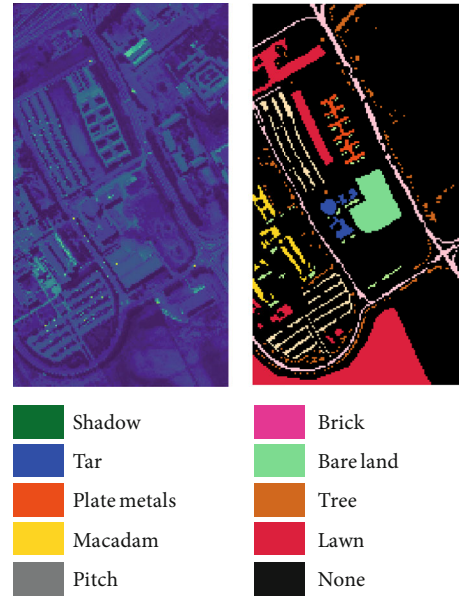|	Shadow		Brick
	Tar		Bare land
	Plate metals		Tree
	Macadam		Lawn
	Pitch		None

FIGURE 7: Band 60 pseudocolor map and category label map of the Pavia University dataset.

TABLE 2: Category name and the number of pixels in the Pavia University dataset.

Label	Category name	The number of pixels
1	Pitch	6631
2	Lawn	18649
3	Macadam	2099
4	Tree	3064
5	Plate metals	1345
6	Bare land	5029
7	Tar	1330
8	Brick	3682
9	Shadow	947
Total number of pixels		42776

Salinas dataset and the number of pixels in each category. As can be seen from Table 3, the Salinas dataset has a large number of pixels, with 16 categories. The categories in the dataset are relatively balanced, and there is no phenomenon that the number of pixels in one category is too low.

4.2. Experiments with the Proposed Method. In order to quantify the independence between the training set the and test set after using different sampling strategies, the test set independence rate is used to evaluate multiple sampling strategies. Test set samples that are not involved in the training process are called test set-independent sample points, and the test set independence rate refers to the ratio between the number of test set-independent sample points and the number of all samples in the test set. The k value of the composite clustering sampling strategy should be an integer greater than or equal to two. In the experiment, the k value of the composite clustering sampling strategy is from 2 to 16, and performance is observed when k value is small. If no conclusion can be
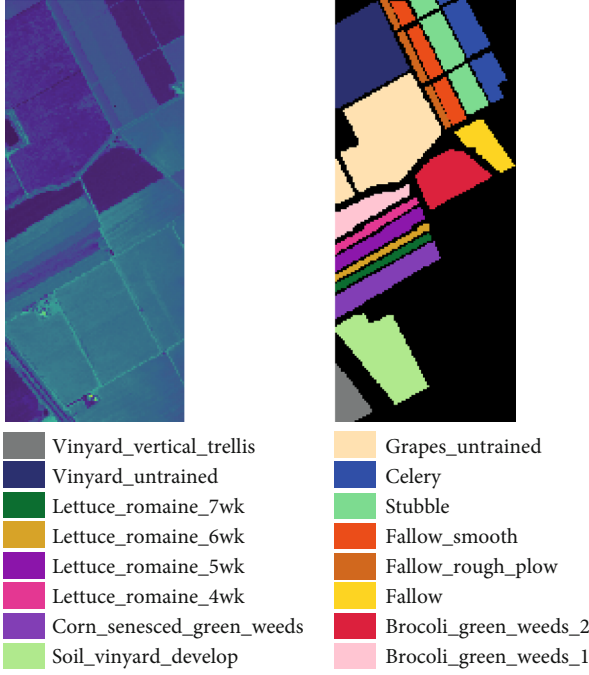
Figure 8: Band 188 pseudocolor map and category label map of the Salinas dataset.

Table 3: Category name and the number of pixels in the Salinas dataset.

Label	Category name	The number of pixels
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807
Total number of pixels		54129

drawn when k is not greater than 16, then the value of k is further expanded.

The 1D-3D-2D-CNN model proposed in this paper has five convolutional layers. In the first convolutional layer, a 1D convolution kernel of size 3 is used to extract spectral features for the spectral information of size $1 * 1 * c$, and 3D convolution kernels with sizes (3, 3, 3) and (5, 5, 3)

are used to extract multiscale spatial-spectral features for data blocks of size $3 * 3 * c$ and $5 * 5 * c$, where c is the spectral dimension size. The number of filters for the first convolution layer is two. After the first convolution layer, spectral features and multiscale spatial-spectral features are transposed into 2D feature maps and fused. The next four convolutional layers are 2D convolutional layers, convolution kernel sizes are all (3, 3), all use Rectified Linear Unit (ReLU) activation function, and the number of filters in each layer is 4, 16, 32, and 64. Among these 2D convolutional layers, the first two layers use "SAME" padding, and the last two layers use "VAILD" padding. A maximum pooling layer with a step size of 2 is used after the last convolutional layer. Three fully connected (FC) layers are used, of which the first two FC layers used dropout. The second FC layer uses sigmoid activation function, and the third FC layer uses softmax activation function. The model's training epoch is 1000, batch size is 128, loss function is cross entropy, and the model uses Adam optimizer. Using a hierarchical learning rate, when the batch is less than 400, the learning rate is 2×10^{-3}; when the batch is greater than 400 and less than 600, the learning rate is 1×10^{-3}; when the batch is greater than 600 and less than 800, the learning rate is 5×10^{-4}; and when the batch is greater than 800, the learning rate is 1×10^{-4}.

In order to verify the validity of the 1D-3D-2D-CNN model proposed in this paper, a 3D-CNN model is designed for comparison experiments. The 3D-CNN model differs from the 1D-3D-2D-CNN model only in convolutional and pooling layers. It has five 3D convolutional layers, the number of filters in each layer is 2, 4, 16, 32, and 64, and the size of convolution kernels are all (3, 3, 3). These 3D convolutional layers all use ReLU activation function. Among five 3D convolutional layers, the first three layers use "SAME" padding, and last two layers use "VAILD" padding. In the designed 3D-CNN model, no pooling layer is used. During model training, according to experience, a sampling rate of 10% to 30% is usually used to select the training set. In this section, the training set is selected at the most commonly used sampling ratio, which is 20%. In order to verify the effectiveness of the composite clustering sampling strategy, as there are few research results in this field, the paper chooses to realize the improved area-based sampling strategy proposed by Lange et al. [26] for comparison experiments.

As shown in Figure 9, in the Indian Pines dataset, although the test set independence rate of the composite clustering sampling strategy proposed in this paper is higher than that of the random sampling strategy, it is somewhat lower than that of the area-based sampling strategy. As the k value of the composite clustering sampling strategy increases, the test set independence rate gradually decreases. In the commonly used sampling ratio of 10% to 30%, when k is taken from 2 to 16, compared with the area-based sampling strategy, the composite clustering sampling strategy causes a decrease in the test set independence rate. When k is 2 and the sampling ratio is 10%, the test set independence rate decreases least, which is 6.4% lower than that of the area-based sampling strategy under the same conditions. When
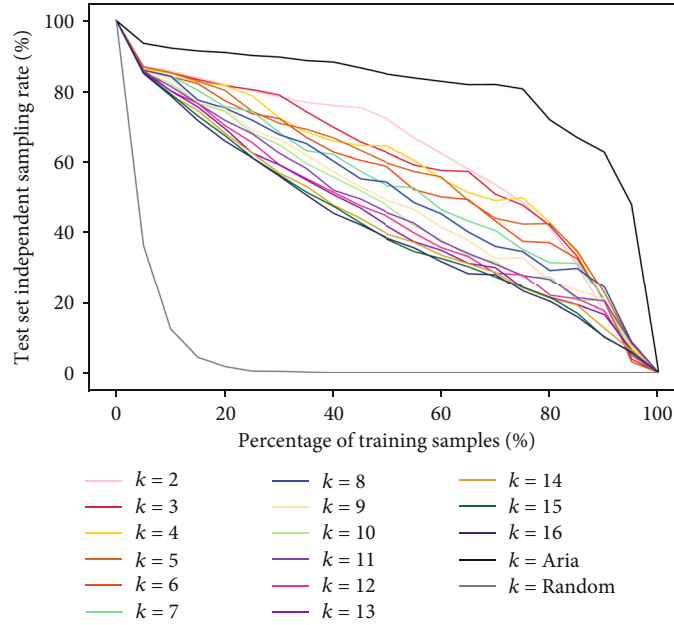
FIGURE 9: Test set independence rate for different sampling strategies with window tensor size w equal to five in the Indian Pines dataset.

TABLE 4: Classification accuracy and test set independence rate with different sampling strategies of the Indian Pines dataset.

Sampling strategy	3D-CNN				1D-3D-2D-CNN				Test set-independent sample rate (%)
	OA (%)	AA (%)	Kappa	Time (min.)	OA (%)	AA (%)	Kappa	Time (min.)	
Area	46.81	42.55	0.3950	9.24	49.76	51.12	0.4392	1.95	90.93
Random	90.33	91.69	0.8894	9.24	94.36	92.91	0.9356	1.87	1.75
$k = 2$	63.55	62.66	0.5808	10.15	77.19	70.83	0.7402	1.98	81.88
$k = 3$	58.76	58.84	0.5284	10.11	73.72	70.68	0.7007	2.01	81.48
$k = 4$	65.57	64.35	0.6059	10.10	76.06	72.53	0.7278	2.00	81.52
$k = 5$	66.24	66.20	0.6136	10.12	77.01	70.58	0.7385	2.04	80.17
$k = 6$	66.73	68.39	0.6161	10.21	78.60	77.01	0.7558	2.13	77.23
$k = 7$	63.57	65.34	0.5803	9.37	79.83	77.50	0.7699	2.10	75.51
$k = 8$	67.90	68.93	0.6295	9.41	78.27	77.38	0.7501	2.05	75.15
$k = 9$	66.19	68.59	0.6104	9.48	80.23	79.38	0.7742	2.12	75.02
$k = 10$	65.49	68.57	0.6019	9.42	78.62	79.64	0.7554	2.29	74.23
$k = 11$	65.08	69.98	0.5963	9.50	82.86	82.08	0.8039	2.34	71.68
$k = 12$	69.22	72.12	0.6443	9.51	80.45	78.71	0.7769	2.36	70.20
$k = 13$	69.73	73.32	0.6517	9.56	80.30	80.22	0.7743	2.40	69.34
$k = 14$	70.94	71.98	0.6650	9.61	80.09	78.36	0.7720	2.46	67.98
$k = 15$	69.40	73.11	0.6467	9.69	83.01	81.13	0.8059	2.44	67.34
$k = 16$	70.24	72.06	0.6566	9.25	81.94	80.28	0.7932	2.46	65.81

k is 16 and the sampling ratio is 30%, the test set independence rate decreases most, which is 33.7% lower than that of the area-based sampling strategy under the same conditions. The loss of independence is large between the training set and the test set divided by the composite clustering sampling strategy. This is due to the large number of categories and small number of pixels in the Indian Pines dataset. It can be seen from Figure 9 that in the composite clustering

sampling strategy, the independence of dataset has reached a lower level when k is 16. However, at low k values, the test set independence rate is still acceptable.

In the Indian Pines dataset, it can be seen from Table 4 that in both 3D-CNN and 1D-3D-2D-CNN models, the composite clustering sampling strategy can make model's classification performance better than the area-based sampling strategy. In the 3D-CNN model, when k is 14, it has
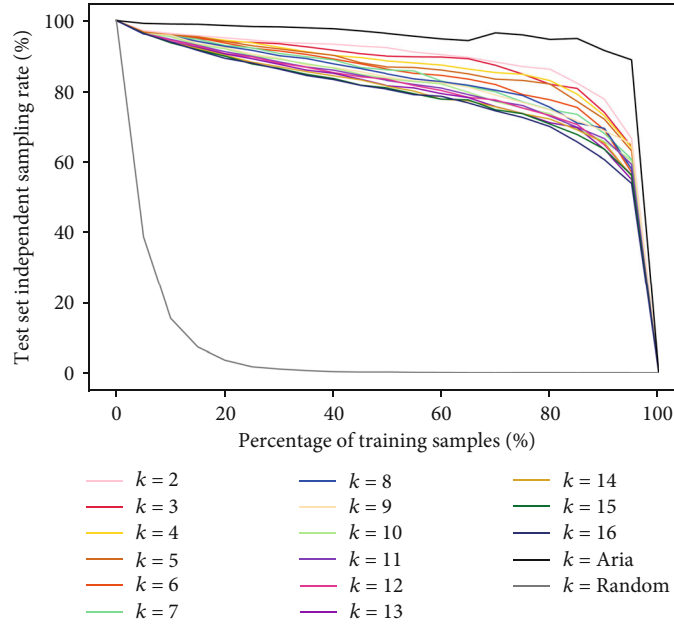
FIGURE 10: Test set independence rate for different sampling strategies with window tensor size w equal to five in the Pavia University dataset.

the highest OA and reached 70.94%, which is an increase of 24.13% compared with the area-based sampling strategy. When k is 13, it has the highest AA and reached 73.32%, which is an increase of 30.77% compared with the area-based sampling strategy. When k is 16, it has highest Kappa and reached 0.6566, which is an increase of 0.2616 compared with the area-based sampling strategy. In the model proposed in this paper, when k is 15, it has the highest OA and reached 83.01%, which is an increase of 33.25% compared with the area-based sampling strategy. When k is 11, it has highest AA and reached 82.08%, which is an increase of 30.96% compared with the area-based sampling strategy. When k is 15, it has highest Kappa and reached 0.8059, which is an increase of 0.3667 compared with the area-based sampling strategy. Experimental results on two models show that the composite clustering sampling strategy proposed in this paper can effectively improve the classification performance of models.

In the Indian Pines dataset, although the test set independence rate decreases greatly when the value of k is large, the composite clustering sampling strategy can greatly improve classification accuracy at a cost of smaller decreasing test set independence rate when the value of k is small. When k is 4 and using the 1D-3D-2D-CNN model, compared with the area-based sampling strategy, OA increased by 26.3%, AA increased by 21.41%, Kappa increased by 0.2886, and test set independence rate decreased by only 9.41%. Compared with the random sampling strategy, the composite clustering sampling strategy still has a certain gap in final classification accuracy. When k is 4 and using the 1D-3D-2D-CNN model, compared with the random sampling strategy, OA decreased by 18.3%, AA decreased by 20.38%, Kappa decreased by 0.2078, but test set independence rate increased by 79.77%. Considering both classification accuracy and test set independence rate, the performance of the composite clustering sampling strategy is acceptable. When k is large, the test set

independence rate decreases significantly and is no longer practical. For example, when k is 16 and the sampling rate is 20%, the test set independence rate is only 65.81%. As the value of k increases, the test set independence rate will further decrease; therefore, those cases where k is greater than 16 will not be discussed further.

Compared with the 3D-CNN model, the 1D-3D-2D-CNN model has the largest OA growth by 17.78% when k is 11 and has smallest OA growth by 2.95% in the area-based sampling strategy. It has the largest AA growth by 12.16% when k is 7 and has the smallest AA growth by 1.22% in the random sampling strategy. It has the largest Kappa growth by 0.2076 when k is 11 and has the smallest Kappa growth by 0.0442 in the area-based sampling strategy. When k is 2, training time is shortened the longest, shortened by 8.17 min; when k is 16, training time is shortened the shortest, shortened by 6.79 min. This is because the number of labeled pixels in the Indian Pines dataset is small and the 3D-CNN model cannot be sufficiently trained. In the 1D-3D-2D-CNN model proposed in this paper, although the use of 2D-CNN will reduce training accuracy, a multi-scale spectral-spatial method is used to effectively extract multiscale spectral-spatial features of pixels, which effectively compensates the accuracy loss caused by 2D-CNN. Therefore, in the Indian Pines dataset, the 1D-3D-2D-CNN model proposed in this paper can not only effectively improve classification accuracy but also effectively shorten training time of models.

As shown in Figure 10, in the Pavia University dataset, the test set independence rate of the composite clustering sampling strategy proposed in this paper is much higher than that of the random sampling strategy, and it is not much lower than that of the area-based sampling strategy. As the k value of the composite clustering sampling strategy increases, the test set independence rate gradually decreases. In the commonly used sampling ratio of 10% to 30%, when k

TABLE 5: Classification accuracy and test set independence rate with different sampling strategies of the Pavia University dataset.

Sampling strategy	3D-CNN				1D-3D-2D-CNN				Test set-independent sample rate (%)
	OA (%)	AA (%)	Kappa	Time (min.)	OA (%)	AA (%)	Kappa	Time (min.)	
Area	63.36	72.01	0.5581	17.94	66.12	79.12	0.5904	3.19	98.60
Random	97.54	96.90	0.9673	17.90	98.36	97.92	0.9782	3.44	3.31
$k = 2$	84.37	86.66	0.7979	17.80	84.91	88.71	0.8067	3.13	95.11
$k = 3$	87.31	85.70	0.8322	17.94	87.35	87.16	0.8337	3.65	94.10
$k = 4$	90.40	89.40	0.8710	18.05	92.12	91.21	0.8967	3.11	94.41
$k = 5$	88.78	89.73	0.8481	18.15	91.21	90.72	0.8843	3.28	94.08
$k = 6$	86.62	87.11	0.8236	17.99	88.83	88.56	0.8554	3.26	93.63
$k = 7$	88.35	87.32	0.8478	18.19	91.69	90.78	0.8912	3.34	92.94
$k = 8$	89.41	88.11	0.8594	17.71	91.27	89.94	0.8857	3.40	92.69
$k = 9$	90.52	88.15	0.8739	17.91	91.48	89.26	0.8876	3.52	92.27
$k = 10$	89.88	88.43	0.8646	17.97	92.84	91.35	0.9060	3.70	92.00
$k = 11$	89.52	88.59	0.8594	17.80	92.17	90.54	0.8968	3.70	91.16
$k = 12$	90.27	89.16	0.8701	17.60	92.45	91.46	0.9014	3.42	90.62
$k = 13$	92.06	90.95	0.8939	17.59	93.34	92.40	0.9124	3.92	90.52
$k = 14$	92.32	91.77	0.8973	17.76	93.60	93.29	0.9162	3.47	90.15
$k = 15$	90.54	89.55	0.8734	17.68	93.36	92.65	0.9127	3.87	89.87
$k = 16$	90.10	89.65	0.8689	17.87	90.41	90.88	0.8756	3.89	89.28

is taken from 2 to 16, compared with the area-based sampling strategy, the composite clustering sampling strategy causes a decrease in the test set independence rate. When k is 2 and sampling ratio is 10%, the test set independence rate decreases the least, which is 2.7% lower than that of the area-based sampling strategy under the same conditions. When k is 16 and the sampling ratio is 30%, the test set independence rate decreases the most, which is 12% lower than that of the area-based sampling strategy under the same conditions. In the Pavia University dataset, the composite clustering sampling strategy has a higher test set independence rate. This is due to the fewer categories and larger number of pixels in the Pavia University dataset.

In the Pavia University dataset, it can be seen from Table 5 that in both 3D-CNN and 1D-3D-2D-CNN models, the composite clustering sampling strategy can make the model's classification performance better than the area-based sampling strategy. In the 3D-CNN model, when k is 14, it has the highest OA and reached 92.32%, which is an increase of 28.96% compared with the area-based sampling strategy. When k is 14, it has highest AA and reached 91.77%, which is an increase of 19.76% compared with the area-based sampling strategy. When k is 14, it has highest Kappa and reached 0.8973, which is an increase of 0.3392 compared with the area-based sampling strategy. In the model proposed in this paper, when k is 14, it has the highest OA and reached 93.60%, which is an increase of 27.48% compared with the area-based sampling strategy. When k is 14, it has highest AA and reached 93.29%, which is an increase of 14.17% compared with the area-based sampling strategy. When k is 14, it has highest Kappa and reached 0.9127, which is an increase of 0.3258 compared with the area-based

sampling strategy. Experimental results on two models show that the composite clustering sampling strategy proposed in this paper can effectively improve the classification performance of models.

Although the test set independence rate decreases greatly when the value of k is large, the composite clustering sampling strategy can greatly improve classification accuracy at a cost of a smaller decreasing test set independence rate when the value of k is small. When k is 4 and using the 1D-3D-2D-CNN model, compared with the area-based sampling strategy, OA increased by 26%, AA increased by 12.09%, Kappa increased by 0.3063, and test set independence rate decreased by only 4.19%. Compared with the random sampling strategy, the composite clustering sampling strategy still has a certain gap in final classification accuracy. When k is 4 and using the 1D-3D-2D-CNN model, compared with the random sampling strategy, OA decreased by 6.24%, AA decreased by 6.71%, Kappa decreased by 0.0815, but test set independence rate increased by 90.88%. In the Pavia University dataset, considering both classification accuracy and test set independence rate, the composite clustering sampling strategy performs very well.

Compared with the 3D-CNN model, the 1D-3D-2D-CNN model has the largest OA growth by 3.34% when k is 7 and has the smallest OA growth by 0.04% when k is 3. It has the largest AA growth by 7.11% in the area-based sampling strategy and has the smallest AA growth by 0.99% when k is 5. It has the largest Kappa growth by 0.0434 when k is 7 and has the smallest Kappa growth by 0.0015 when k is 3. When k is 4, training time is shortened the longest, shortened by 14.94 min; when k is 13, training time is shortened the shortest, shortened by 13.67 min. This is because the number

TABLE 6: Supplementary experiments of the Pavia University dataset with high k value.

Sampling strategy	OA (%)	AA (%)	Kappa	Time (min.)	Test set-independent sample rate (%)
$k = 17$	94.13	93.55	0.9229	3.51	89.40
$k = 18$	93.41	92.67	0.9133	3.57	89.17
$k = 19$	93.36	92.76	0.9130	3.47	88.92
$k = 20$	93.14	92.36	0.9097	3.49	88.73
$k = 21$	91.87	92.00	0.8940	3.59	88.67
$k = 22$	91.77	90.90	0.8926	3.55	88.12
$k = 23$	91.11	90.44	0.8841	3.60	87.83
$k = 24$	92.36	93.09	0.9004	3.64	88.18
$k = 25$	90.89	90.50	0.8810	3.30	87.97
$k = 26$	91.25	91.56	0.8863	3.34	87.50
$k = 27$	88.34	89.75	0.8496	3.58	87.61
$k = 28$	92.96	92.51	0.9080	3.68	87.23
$k = 29$	88.13	88.99	0.8462	3.63	87.37
$k = 30$	91.20	90.72	0.8841	3.73	86.92
$k = 31$	91.79	91.88	0.8918	3.74	87.00
$k = 32$	90.16	89.65	0.8708	3.79	86.46

of labeled pixels in the Pavia University dataset is large, and 3D-CNN is fully trained. In the 1D-3D-2D-CNN model proposed in this paper, although the use of 2D-CNN will reduce training accuracy, a multiscale spectral-spatial method is used to effectively extract multiscale spectral-spatial features of pixels, which effectively compensates the accuracy loss caused by 2D-CNN. Therefore, in the Pavia University dataset, the 1D-3D-2D-CNN model can effectively shorten training time of models while ensuring classification accuracy.

Decrease in the test set independence rate has slowed down when k is greater than 6. When k is 16, the test set independence rate is 89.28%, and the independence between the training set and the test set remains at a high level. Therefore, the value of k greater than 16 should be further discussed.

The classification performance of the composite clustering sampling strategy with a k value of 17 to 32 is further discussed. As can be seen from Table 6, when k is 14, the classification accuracy of the 1D-3D-2D-CNN model has been further improved, OA reached 94.13%, AA reached 93.55%, Kappa reached 0.9229, and test set independence rate is still at a high level of 89.40%. However, when k is greater than 20, classification accuracy decreases in fluctuations, and the test set independence rate also decreases continuously. When k is 32, the test set independence rate is only 86.46%, which is already at a low level. As the value of k increases, the test set independence rate will further decrease. Therefore, those cases where k is greater than 32 will not be discussed further.

As shown in Figure 11, in the Salinas dataset, the test set independence rate of the composite clustering sampling strategy proposed in this paper is much higher than that of the random sampling strategy, and it is not much lower than

that of the area-based sampling strategy. As k value of the composite clustering sampling strategy increases, the test set independence rate gradually decreases. In the commonly used sampling ratio of 10% to 30%, when k is taken from 2 to 16, compared with the area-based sampling strategy, the composite clustering sampling strategy causes a decrease in the test set independence rate. When k is 9 and sampling ratio is 10%, the test set independence rate decreases the least, which is 1.9% lower than that of the area-based sampling strategy under the same conditions. When k is 15 and sampling ratio is 30%, the test set independence rate decreases the most, which is 11.4% lower than that of the area-based sampling strategy under the same conditions. The composite clustering sampling strategy has a higher test set independence rate.

In the Salinas dataset, it can be seen from Table 7 that in both 3D-CNN and 1D-3D-2D-CNN models, the composite clustering sampling strategy can make the model's classification performance better than the area-based sampling strategy. In the 3D-CNN model, when k is 12, it has the highest OA and reached 88.89%, which is an increase of 10.55% compared with the area-based sampling strategy. When k is 13, it has the highest AA and reached 93.97%, which is an increase of 14.72% compared with the area-based sampling strategy. When k is 12, it has highest Kappa and reached 0.8759, which is an increase of 0.118 compared with the area-based sampling strategy. In the model proposed in this paper, when k is 12, it has the highest OA and reached 90.09%, which is an increase of 7.41% compared with the area-based sampling strategy. When k is 14, it has the highest AA and reached 94.83%, which is an increase of 12.34% compared with the area-based sampling strategy. When k is 12, it has the highest Kappa and reached 0.8894, which is an increase of 0.0828 compared with the area-based sampling strategy. Experimental results on two models show that the composite clustering sampling strategy proposed in this paper can effectively improve classification performance of models.

Although the test set independence rate decreases greatly when the value of k is large, the composite clustering sampling strategy can improve classification accuracy at a cost of the smaller decreasing test set independence rate when the value of k is small. When k is 3 and using the 1D-3D-2D-CNN model, compared with the area-based sampling strategy, OA increased by 4.49%, AA increased by 8.51%, Kappa increased by 0.05, and test set independence rate decreased by only 3.36%. Compared with the random sampling strategy, the composite clustering sampling strategy still has a certain gap in final classification accuracy. When k is 3 and using the 1D-3D-2D-CNN model, compared with the random sampling strategy, OA decreased by 7.46%, AA decreased by 0.33%, Kappa decreased by 0.0835, but test set independence rate increased by 92.7%. In the Salinas dataset, considering both classification accuracy and test set independence rate, the performance of the composite clustering sampling strategy is acceptable.

The decrease in the test set independence rate has slowed down when k is greater than 5. When k is 16, the test set independence rate is 89.67%, and the independence between the training set and the test set remains at a high level. But when
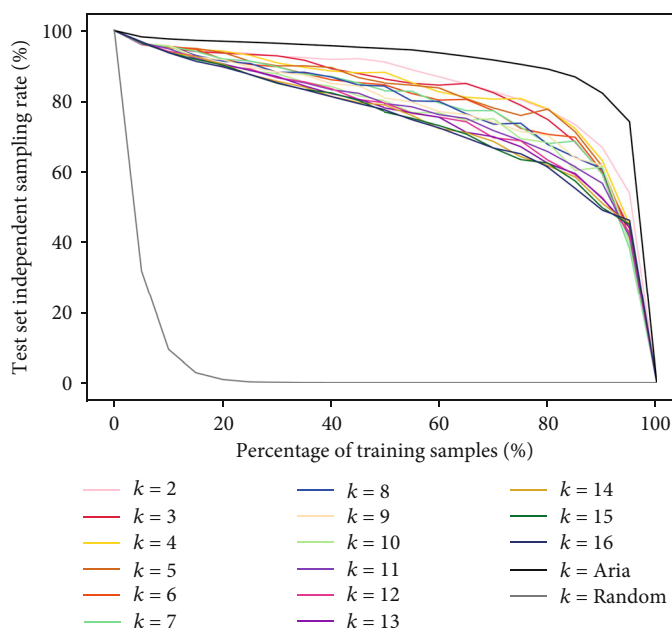
FIGURE 11: Test set independence rate for different sampling strategies with window tensor size w equal to five in the Salinas dataset.

k is greater than 12, the classification performance of models has decreased significantly. Therefore, those cases where k is greater than 16 will not be discussed further.

Compared with the 3D-CNN model, the 1D-3D-2D-CNN model has the largest OA growth by 4.36% in the random sampling strategy and has the smallest OA growth by 0.26% when k is 13. It has the largest AA growth by 3.24% in the area-based sampling strategy and has the largest AA reduction by 0.99% when k is 12. It has the largest Kappa growth by 0.049 in the random sampling strategy and has the smallest Kappa growth by 0.0031 when k is 13. In the area-based sampling strategy, training time is shortened the longest, shortened by 34.78 min; when k is 10, training time is shortened the shortest, shortened by 33.16 min. This is because the number of labeled pixels in the Salinas dataset is large, and 3D-CNN is fully trained. In the 1D-3D-2D-CNN model proposed in this paper, although the use of 2D-CNN will reduce training accuracy, a multiscale spectral-spatial method is used to effectively extract multiscale spectral-spatial features of pixels, which effectively compensates the accuracy loss caused by 2D-CNN. Therefore, in the Salinas dataset, the 1D-3D-2D-CNN model can effectively shorten training time while ensuring classification accuracy.

In summary, the composite clustering sampling strategy has performed well in Indian Pines, Pavia University, and Salinas datasets. The composite clustering sampling strategy has excellent performance in a dataset where category boundaries are fuzzy, the number of pixels is large, and the number of categories is small, such as the Pavia University dataset. The Indian Pines dataset has a small number of pixels and a large number of categories, the use of the composite clustering sampling strategy will cause the independence between the training set and the test set to a larger decrease, but it can greatly improve classification accuracy

at a cost of the smaller decreasing test set independence rate when the value of k is small. The Salinas dataset has a large number of pixels and clear category boundaries, the area-based sampling strategy already can perform well, and the performance improvement brought by the composite clustering sampling strategy is not obvious. However, using the 1D-3D-2D-CNN model proposed in this paper can effectively shorten training time of models while ensuring classification accuracy.

4.3. Comparative Experiments with Other Methods. The problem of sampling strategies in the spectral-spatial classification field has only been raised in recent years, and research is gradually carried out, so there are not many research results in this field. At present, the most widely accepted method in this field is the controlled random sampling strategy [25, 31]. Liang et al. [25] conducted experiments using a controlled random sampling strategy by support vector machine (SVM) and random forest (RF) models combined with 3D discrete wavelet (3D-DWT) and morphological profile (EMP) spectral-spatial feature extraction methods. In order to verify the effectiveness of the proposed method, the classification performance of the controlled random sampling strategy and composite clustering sampling strategy is compared using OA as the metric. The proposed 1D-3D-2D-CNN with the four space-spectrum classification methods of SVM-3D-DWT, SVM-EMP, RF-3D-DWT, and RF-EMP was compared. The experimental data of these four models are quoted from the research of Liang et al. [25].

It can be found from Table 8 that under different sampling ratios, the number of pixels in the training set is not strictly a multiple relationship. This is because whether it is a controlled random sampling strategy or a composite cluster sampling strategy, training samples are selected for each partition according to a predetermined sampling ratio.

TABLE 7: Classification accuracy and test set independence rate with different sampling strategies of the Salinas dataset.

Sampling strategy	3D-CNN				1D-3D-2D-CNN				Test set-independent sample rate (%)
	OA (%)	AA (%)	Kappa	Time (min.)	OA (%)	AA (%)	Kappa	Time (min.)	
Area	78.34	79.25	0.7579	42.17	82.68	82.49	0.8066	7.39	96.96
Random	90.27	90.46	0.8911	41.21	94.63	91.33	0.9401	7.61	0.90
$k = 2$	77.70	82.77	0.7515	41.18	79.41	82.81	0.7710	7.68	93.99
$k = 3$	83.26	89.99	0.8135	41.25	87.17	91.00	0.8566	7.86	93.60
$k = 4$	78.84	83.04	0.7649	41.50	80.08	83.98	0.7788	8.03	94.18
$k = 5$	74.19	75.63	0.7123	41.44	77.18	81.63	0.7459	7.49	93.79
$k = 6$	83.79	85.33	0.8193	41.41	85.41	85.69	0.8373	7.75	91.80
$k = 7$	83.61	85.64	0.8167	41.84	84.99	85.99	0.8324	7.40	91.92
$k = 8$	86.44	86.27	0.8489	41.85	86.73	87.67	0.8522	8.03	91.31
$k = 9$	85.97	86.75	0.8437	41.19	87.91	87.56	0.8653	7.54	91.68
$k = 10$	87.32	91.21	0.8588	41.23	87.63	92.28	0.8624	8.07	90.60
$k = 11$	87.69	88.46	0.8629	41.48	89.10	91.12	0.8787	7.91	90.38
$k = 12$	88.89	93.07	0.8759	41.31	90.09	92.24	0.8894	7.84	89.98
$k = 13$	87.82	93.97	0.8644	41.54	88.08	94.29	0.8675	7.55	90.48
$k = 14$	86.07	91.76	0.8450	41.78	87.79	94.83	0.8643	7.68	90.42
$k = 15$	85.05	86.08	0.8331	41.46	86.05	87.43	0.8450	7.70	90.20
$k = 16$	85.86	93.77	0.8431	41.17	87.92	94.37	0.8660	7.97	89.67

TABLE 8: Number of pixels in the training set at different sampling ratios.

	5%	10%	25%
Indian Pines	492	1004	2549
Pavia University	2032	4155	10592
Salinas	2694	5041	13523

The number of samples used to compose the training set in each partition does not contain decimals. When the number of partitions in the dataset is large, this will cause a significant gap.

First, observe the performance of the five models when using a controlled random sampling strategy. As can be seen from Tables 9–11, in the Indian Pines dataset, the classification accuracy of 1D-3D-2D-CNN is better than that of the other four models when the sampling ratio is 10% and 25%, but when the sampling ratio is 5%, the classification accuracy of 1D-3D-2D-CNN is not the best. In the Pavia University dataset, the classification accuracy of 1D-3D-2D-CNN is better than that of the other four models when the sampling ratio is 25%, but when the sampling ratio is 5% and 10%, the classification accuracy of 1D-3D-2D-CNN is not the best. In the Salinas dataset, when the sampling ratio is 5%, 10%, and 25%, the classification accuracy of 1D-3D-2D-CNN is better than that of the other four models. In general, the classification ability of 1D-3D-2D-CNN is acceptable.

Second, the experimental results of 1D-3D-2D-CNN under the controlled random sampling strategy and composite clustering sampling strategy are compared. As can be seen from Tables 9–11, in the Indian Pines dataset, the composite

TABLE 9: Comparison experiment of 5%, 10%, and 25% sampling rates in the Indian Pines dataset with controlled random sampling and composite clustering sampling ($k = 11$) (∗).

	5%	10%	25%
SVM-3D-DWT	65.2	69.9	79.1
SVM-EMP	64.8	69.2	77.2
RF-3D-DWT	57.4	61.4	67.5
RF-EMP	64.4	69.6	76.2
1D-3D-2D-CNN	63.2	72.0	79.6
1D-3D-2D-CNN∗	64.1	72.3	82.4

TABLE 10: Comparison experiment of 5%, 10%, and 25% sampling rates in the Pavia University dataset with controlled random sampling and composite clustering sampling ($k = 14$) (∗).

	5%	10%	25%
SVM-3D-DWT	84.8	86.4	89.2
SVM-EMP	87.4	89.5	91.7
RF-3D-DWT	75.8	79.0	83.4
RF-EMP	78.2	80.9	88.2
1D-3D-2D-CNN	85.4	88.6	92.2
1D-3D-2D-CNN∗	85.8	87.4	93.8

cluster sampling strategy is superior to the controlled random sampling strategy. In the Pavia University dataset, when the sampling ratio is 10%, the controlled random sampling strategy performs better, and when the sampling ratio is 5% and 25%, the composite clustering sampling strategy

Table 11: Comparison experiment of 5%, 10%, and 25% sampling rates in the Salinas dataset with controlled random sampling and composite clustering sampling ($k = 12$) (*).

	5%	10%	25%
SVM-3D-DWT	80.9	82.2	83.4
SVM-EMP	83.5	85.0	84.8
RF-3D-DWT	77.6	80.6	83.8
RF-EMP	82.0	84.4	86.6
1D-3D-2D-CNN	83.7	86.1	87.7
1D-3D-2D-CNN*	80.1	86.4	92.4

performs better. In the Salinas dataset, when the sampling ratio is 5%, the controlled random sampling strategy performs better, and when the sampling ratio is 10% and 25%, the composite clustering sampling strategy performs better. This is caused by the different advantages and disadvantages of those two sampling strategies. The working principle of the controlled random sampling strategy is to randomly select seed points in each partition and use a region growing algorithm to select a training set of sufficient size around the seed points. Although this method can enhance the independence between the training set and the test set, it cannot guarantee the spectral representativeness of pixels in the training set. When the number of pixels in the partition is small and the sampling ratio is small, the controlled random sampling strategy can be used to obtain a training set with high spectral domain representativeness, but when the number of pixels in the partition is large or the sampling ratio is large, it is difficult to obtain a training set with high spectral domain representativeness using a controlled random sampling strategy. Contrary to the controlled random sampling strategy, when the sampling ratio is large, the composite clustering sampling strategy performs better, but when the sampling ratio is relatively small, the composite clustering sampling strategy does not perform satisfactorily. This is because when the sampling ratio is relatively small, for each partition, the pixels used for training models in the partition obtained by using the composite clustering sampling strategy will be distributed in a cluster, which makes the spectral domain representativeness of the training set unable to be guaranteed. The Indian Pines dataset has the characteristics of fewer pixels and fewer partitions, and the pixels are unevenly distributed; there are both partitions with a small number of pixels and partitions with a large number of pixels. Therefore, in the Indian Pines dataset, when the sampling ratio is small, the controlled random sampling strategy performs well, but when the sampling ratio is large, the composite clustering sampling strategy is superior to the controlled random sampling strategy. The Pavia University dataset has the characteristics of a large number of pixels and partitions, so that each partition has a small number of pixels. Therefore, in the Pavia University dataset, the performance of the controlled random sampling strategy and composite cluster sampling strategy is very close. The Salinas dataset has the characteristics of a large number of pixels and a small number of partitions, so the number of pixels in each partition is large. Therefore, in the Salinas dataset, when the sampling ratio is small, the controlled random

Table 12: Comparative experiments with different network depths in the Indian Pines dataset ($k = 11$).

	2	3	4	5	6	7
OA (%)	79.65	81.52	81.95	82.78	81.33	80.87
AA (%)	78.6	79.49	82.24	83.92	80.23	79.41
Kappa	0.7672	0.7891	0.7929	0.8075	0.7875	0.7875

Table 13: Comparative experiments with different network depths in the Pavia University dataset ($k = 14$).

	2	3	4	5	6	7
OA (%)	91.74	92.39	93.41	93.74	92.95	92.47
AA (%)	90.77	92.06	93.48	93.19	94.03	92.61
Kappa	0.8910	0.8996	0.9137	0.9183	0.9068	0.9010

Table 14: Comparative experiments with different network depths in the Salinas dataset ($k = 12$).

	2	3	4	5	6	7
OA (%)	87.83	88.18	89.31	90.19	89.72	89.03
AA (%)	93.02	93.35	93.06	93.91	93.05	91.58
Kappa	0.8639	0.8679	0.8805	0.8906	0.8853	0.8781

sampling strategy performs better than the composite clustering sampling strategy, but as the sampling ratio increases, the classification accuracy of the controlled random sampling strategy only increases slightly, and the classification accuracy of the composite clustering sampling strategy has increased significantly. Generally, when the sampling rate is low and the number of pixels in each partition of the data set is small, the performance of the controlled random sampling strategy will be better; when the sampling rate is large, the performance of the composite cluster sampling strategy is better. In general, when the sampling rate is low and the number of pixels in each partition of the dataset is small, the controlled random sampling strategy performs better; when the sampling rate is larger, the composite clustering sampling strategy performs better.

4.4. Influence of the Number of Convolutional Layers on Model Performance. In order to verify the effect of the number of different convolutional layers on the performance of 1D-3D-2D-CNN models, comparative experiments were conducted at the sampling rate of 20% in the above three datasets.

As can be seen from Tables 12–14, in the three datasets of Indian Pines, Pavia University, and Salinas, as the number of convolution layer increases, the classification accuracy of the 1D-3D-2D-CNN model gradually increases. However, the effect of improving the classification performance by increasing the depth of models is limited. When the depth of the model reaches a critical value, continuing to increase the depth of the model will cause the classification accuracy to decrease. This is due to the problem of gradient disappearance caused by too many convolutional layers. Therefore,

this paper chooses to use five convolutional layers to build the 1D-3D-2D-CNN model.

5. Conclusion

This paper proposes a composite clustering sampling strategy for spectral-spatial HSI classification methods, which not only maintains a high independence between the training set and the test set but also makes the sample points in the training set have a higher spectral domain representation. At the same time, a new multiscale spectral-spatial HSI classification model is proposed, which can effectively shorten training time and reduce computing resource requirements while maintaining or slightly reducing classification accuracy. However, at smaller sampling ratios such as 5%, the performance of the proposed method is poor. In the future, a sampling strategy for spectral-spatial HSI classification methods will continue to be improved to enhance their performance at smaller sampling ratios. Although the classification performance of the proposed method is higher than other existing methods, it still has a gap compared with the random sampling method. In the future, better models will continue to be proposed to enhance classification capabilities.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

C.L., X.Q., and H.G. conceived and designed the experiments; Y.Y., D.Y., and Z.H. presented tools and carried out the data analysis; C.L. and X.Q wrote the paper; H.G. guided and revised the paper.

Acknowledgments

References

[1] H. Gao, Y. Yang, X. Zhang, C. Li, Q. Yang, and Y. Wang, "Dimension reduction for hyperspectral remote sensor data based on multi-objective particle swarm optimization algorithm and game theory," *Sensors*, vol. 19, no. 6, p. 1327, 2019.

[2] Q. Cao, A. L. Ma, Y. F. Zhong, J. Zhao, B. Zhao, and L. P. Zhang, "Urban classification by multi-feature fusion of hyperspectral image and LiDAR data," *Journal of Remote Sensing*, vol. 23, no. 5, pp. 892–903, 2019.

[3] Z. C. Jiang, Y. Ma, T. Jiang, and C. Chen, "Research on the extraction of red tide hyperspectral remote sensing based on the deep belief network (DBN)," *Journal of Ocean Technology*, vol. 38, no. 2, pp. 1–7, 2019.

[4] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4823–4833, 2019.

[5] A. Mughees and L. Tao, "Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images," *Tsinghua Science and Technology*, vol. 24, no. 2, pp. 183–194, 2019.

[6] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5384–5394, 2019.

[7] H. Gao, D. Yao, M. Wang et al., "A hyperspectral image classification method based on multi-discriminator generative adversarial networks," *Sensors*, vol. 19, no. 15, p. 3269, 2019.

[8] Y. Liu, S. Zhou, W. Han, W. Liu, Z. Qiu, and C. Li, "Convolutional neural network for hyperspectral data analysis and effective wavelengths selection," *Analytica Chimica Acta*, vol. 1086, pp. 46–54, 2019.

[9] H. Gao, Y. Yang, C. Li, X. Zhang, J. Zhao, and D. Yao, "Convolutional neural network for spectral–spatial classification of hyperspectral images," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8997–9012, 2019.

[10] H. Gao, Y. Yang, S. Lei, C. Li, H. Zhou, and X. Qu, "Multibranch fusion network for hyperspectral image classification," *Knowledge-Based Systems*, vol. 167, pp. 11–25, 2019.

[11] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, 2017.

[12] P. Ghamisi, Y. Chen, and X. X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 10, pp. 1537–1541, 2016.

[13] G. Y. Fu, H. Y. Gu, and H. Q. Wang, "Spectral and spatial classification of hyperspectral images based on convolutional neural networks," *Science Technology and Engineering*, vol. 17, no. 21, pp. 268–274, 2017.

[14] L. Zhi, X. C. Yu, B. Zou, and B. Liu, "A multi-layer binary pattern based method for hyperspectral imagery classification using combined spatial-spectral characteristics," *Geomatics and Information Science of Wuhan University*, vol. 44, no. 11, pp. 1659–1666, 2019.

[15] Y. Wang, L. Liu, X. H. Shen, J. Hou, N. Zhang, and Z. W. Shi, "Active learning based on spatial-spectral feature for hyperspectral image classification," *Aerospace Shanghai*, vol. 36, no. 5, pp. 50–56, 2019.

[16] J. Feng, L. Wang, H. Yu, L. Jiao, and X. Zhang, "Divide-and-conquer dual-architecture convolutional neural network for classification of hyperspectral Images," *Remote Sensing*, vol. 11, no. 5, p. 484, 2019.

[17] Z. Tian, J. Ji, S. Mei, J. Hou, S. Wan, and Q. Du, "Hyperspectral classification via spatial context exploration with multi-scale CNN," in *IGARSS 2018-2018 IEEE International Geoscience*

and Remote Sensing Symposium, pp. 2563–2566, IEEE, Valencia, Spain, 2018.

[18] C. Wang, N. Ma, Y. Ming, Q. Wang, and J. Xia, "Classification of hyperspectral imagery with a 3D convolutional neural network and J-M distance," *Advances in Space Research*, vol. 64, no. 4, pp. 886–899, 2019.

[19] S. Mirzaei and S. Khosravani, "Hyperspectral image classification using non-negative tensor factorization and 3D convolutional neural networks," *Signal Processing: Image Communication*, vol. 76, pp. 178–185, 2019.

[20] Q. Gao and S. Lim, "Classification of hyperspectral images with convolutional neural networks and probabilistic relaxation," *Computer Vision and Image Understanding*, vol. 188, p. 102801, 2019.

[21] Y. Liu, S. Zhou, W. Han, C. Li, W. Liu, and Z. Qiu, "Hyperspectral classification using deep fusion spectral–spatial features," *Journal of Applied Remote Sensing*, vol. 13, no. 3, 2019.

[22] J. Zhou, J. Liang, Y. Qian, Y. Gao, and L. Tong, "On the sampling strategies for evaluation of joint spectral-spatial information based classifiers," in *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, IEEE, Tokyo, Japan, 2015.

[23] M. A. Friedl, C. Woodcock, S. Gopal, D. Muchoney, A. H. Strahler, and C. Barker-Schaaf, "A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data," *International Journal of Remote Sensing*, vol. 21, no. 5, pp. 1073–1077, 2010.

[24] C. Geib, P. Aravena Pelizari, H. Schrade, A. Brenning, and H. Taubenbock, "On the effect of spatially non-disjoint training and test samples on estimated model generalization capabilities in supervised classification with spatial features," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2008–2012, 2017.

[25] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 862–880, 2017.

[26] J. Lange, G. Cavallaro, M. Götz, E. Erlingsson, and M. Riedel, "The influence of sampling methods on pixel-wise hyperspectral image classification with 3D convolutional neural networks," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2087–2090, IEEE, Valencia, Spain, 2018.

[27] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Kdd*, vol. 96no. 34, pp. 226–231, 1996.

[28] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1no. 14, pp. 281–297, 1967.

[29] H. Gao, S. Lin, C. Li, and Y. Yang, "Application of hyperspectral image classification based on overlap pooling," *Neural Processing Letters*, vol. 49, no. 3, pp. 1335–1354, 2019.

[30] H. Gao, D. Yao, Y. Yang, C. Li, H. Liu, and Z. Hua, "Multiscale 3-D-CNN based on spatial–spectral joint feature extraction for hyperspectral remote sensing images classification," *Journal of Electronic Imaging*, vol. 29, no. 1, 2020.

[31] X. Cao, M. Ren, J. Zhao, H. Lu, and L. Jiao, "Non-overlapping classification of hyperspectral imagery based on set-to-sets distance," *Neurocomputing*, vol. 378, pp. 422–434, 2020.