remote sensing

Article

Hyperspectral Image Classification Based on Multi-Scale Residual Network with Attention Mechanism

Yuhao Qing * and Wenyi Liu

School of Instrument and Electronics, North University of China, Taiyuan 030000, China; liuwenyi@nuc.edu.cn
* Correspondence: s2006262@st.nuc.edu.cn; Tel.: +86-185-9547-3046

Abstract: In recent years, image classification on hyperspectral imagery utilizing deep learning algorithms has attained good results. Thus, spurred by that finding and to further improve the deep learning classification accuracy, we propose a multi-scale residual convolutional neural network model fused with an efficient channel attention network (MRA-NET) that is appropriate for hyperspectral image classification. The suggested technique comprises a multi-staged architecture, where initially the spectral information of the hyperspectral image is reduced into a two-dimensional tensor, utilizing a principal component analysis (PCA) scheme. Then, the constructed low-dimensional image is input to our proposed ECA-NET deep network, which exploits the advantages of its core components, i.e., multi-scale residual structure and attention mechanisms. We evaluate the performance of the proposed MRA-NET on three public available hyperspectral datasets and demonstrate that, overall, the classification accuracy of our method is 99.82 %, 99.81%, and 99.37, respectively, which is higher compared to the corresponding accuracy of current networks such as 3D convolutional neural network (CNN), three-dimensional residual convolution structure (RES-3D-CNN), and space–spectrum joint deep network (SSRN).

Keywords: hyperspectral image classification; convolutional neural networks; attention mechanism; multi-scale; residual network

Citation: Qing, Y.; Liu, W. Hyperspectral Image Classification Based on Multi-Scale Residual Network with Attention Mechanism. *Remote Sens.* **2021**, *13*, 335. https://doi.org/10.3390/rs13030335

Academic Editor: Hossein M. Rizeei
Received: 21 December 2020
Accepted: 16 January 2021
Published: 20 January 2021

1. Introduction

A hyperspectral image presents a target region in a spectrum of continuous and narrow bands, containing both spatial and spectral feature information at a pixel-level resolution [1]. Hyperspectral imagery is widely used in various applications such as urban planning, agricultural development, and environmental testing [2]. Technically, a hyperspectral image is a three-dimensional image composed of an image along several spectral dimensions. The analysis of a hyperspectral image's spatial and spectral characteristics can effectively contribute to the classification of ground objects. However, hyperspectral images are prone to the Hughes phenomenon due to the complexity of their structure and suffer from a small number of labeled samples affecting the overall performance of hyperspectral image classification. Due to these deficiencies, hyperspectral image classification still poses a hot research topic.

Thus, the literature presents several attempts toward hyperspectral image classification. For example, traditional classifiers are utilized such as the support vector machine (SVM), K-nearest neighbor algorithm, and multinomial logistic regression (MLR) [3–6]. These algorithms mainly exploit the spectral information of the image due to its high dimensionality. Principal component analysis (PCA), independent component analysis (ICA), and image sparse representation (SR) methods are also used to process the spectral information, by extracting its main features and reducing computational complexity [7–11]. However, as opposed to deep learning networks, these traditional methods are not able to extract deep-level features, imposing a relatively low classification accuracy. The deep learning method can extract the abstract-level information of the image, so it is more effective to deal with the hyperspectral classification problem. Indeed, literature offers several deep-learning-based hyperspectral image-classification solutions. In [12], the authors propose a stacked autoencoder (SAE), while the work of [13] presents a convolutional neural network (CNN) architecture appropriate for hyperspectral image classification. Both methods have improved the accuracy of hyperspectral classification compared to the classic computer vision type of approaches presented earlier. Further attempts utilizing deep networks involve a two-dimensional CNN (Conv2D-CNN) to extract spectral–spatial feature information, combined with an image dimensionality reduction algorithm [14,15]. To extract better spatial and spectral features, [16–18] use 3D convolution to extract hyperspectral image features, and [19–21] fuse a residual with an attention deep network. Methods such as attention mechanism [22] and joint space–spectral features [23] that improve the convolutional neural network classification accuracy and also increase the number of the network parameters have also been used. Finally, in [24] the authors utilize the method of generative adversarial networks (GANs), in [25] the transfer learning technique, in [26] 2D and 3D convolutions, respectively, in [27–29] multi-scale adaptive feature extraction, combine these methods with hyperspectral classification problems to further improve the hyperspectral image classification performance.

However, even though all these methods have achieved an acceptable classification performance, the classification accuracy still needs to be further improved. Spurred by that, this paper proposes a multi-scale residual convolutional neural network that integrates an attention mechanism, embeds the lightweight and improved channel attention mechanism proposed by Wang et al. [30], i.e., the efficient channel attention network (ECA-NET), and uses local cross-channel interaction without dimensionality reduction. This strategy, through the adaptive one-dimensional convolution, effectively extracts the spatial and spectral features of the image and reduces the redundancy of the training sample information. Convolutional kernels with different scales and sizes are used along with a novel residual structure, and the space–spectrum features of the image are extracted multiple times. Overfitting is effectively prevented by employing within our deep learning model regularization and smoothing normalization strategies on multiple layers. Experiments on the Pavia University (UP), Kennedy Space Center (KSC), and Indian Pines(IN) datasets highlight the appealing classification performance of our method.

The remaining part of the paper is organized as follows: Section 2 presents the suggested deep learning hyperspectral image-classification strategy, while Section 3 demonstrates the classification capabilities of our method on three datasets. Finally, Section 4 concludes this paper.

2. Multi-Scale Residual Network Model Integrating Attention Mechanism

Hyperspectral images present high spectral resolution along with a large amount of redundant information such as spatial, inter-spectrum, and band correlation. Thus, in this work, we suggest a multi-scale residual convolutional neural network model to extract image features from different scales and multiple times, which greatly improves the classification accuracy on hyperspectral images. The core components of the suggested deep network are presented in the following paragraphs.

2.1. ECA-NET Block

The efficient channel attention network (ECA-NET) is an improved attention mechanism network for the squeeze and excitation network (SE-NET) [31]. Its core strategy is to use non-dimensionality reduction local cross-channel interaction and use one-dimensional convolution instead of SE-NET's dimensionality reduction/upgrading mechanism to reduce the number of network parameters. ECA-NET reduces the model complexity, manages output feature channels with different weights, and achieves the extraction of important features within the image. The ECA-NET network structure is shown in Figure 1.

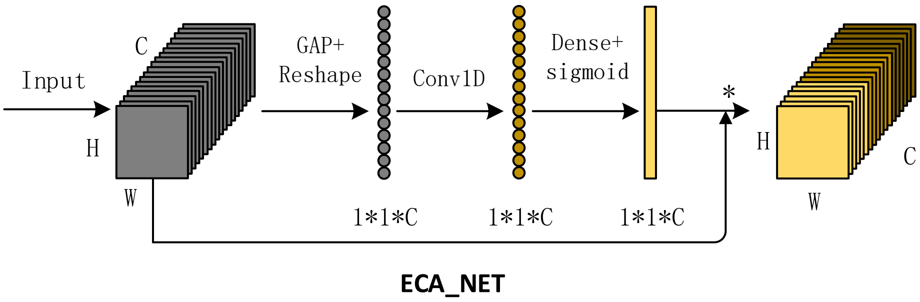


Figure 1. Efficient channel attention network (ECA-NET) block.

For the feature map with an input dimension of {W, H, C} along the corresponding dimension, ECA-NET initially performs a global average pooling (GAP) [32] operation, which reduces the number of parameters and integrates the spatial information of the feature map. A reshape operation is carried out to obtain a matrix of size {1, 1, c}. The second matrix with the shape {1, 1, c} is obtained via a one-dimensional convolution process, and the last matrix is operated by the fully connected layer and the sigmoid activation function and then output. Finally, the ECA-NET output feature is the multiplication of the original feature map {W, H, C} with the feature map {1, 1, c}. In this work, we used the ECA-NET with the improved attention mechanism to extract useful features that are associated with the target feature classes and ultimately produce output feature information with more characterization capabilities that fully combine the space–spectrum features of the hyperspectral images.

2.2. S2A Block

The Spectral-Spatial Attention (S2A) block [33] combines the SE-Net structure with the residual structure and uses two convolution kernels of different sizes to perform deep separable convolution on the input feature map. The feature map obtained by the convolution kernel is transposed and multiplied. Finally, the latter feature map is connected with the residual network that has an attention mechanism to realize the extraction of spatial and spectral feature information of the image. The S2A network structure diagram is shown in Figure 2.

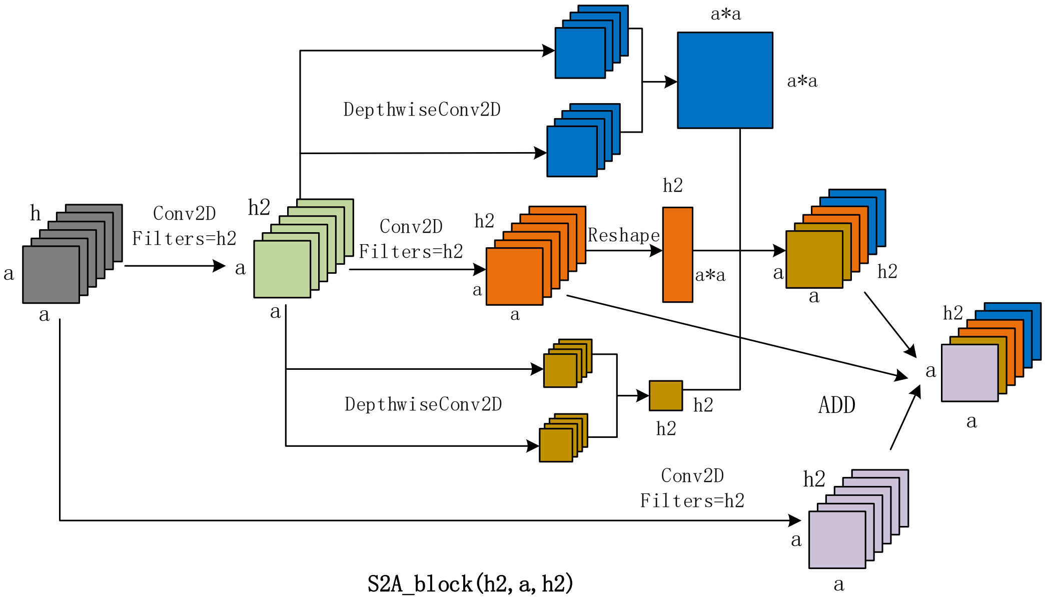


Figure 2. S2A block.

Specifically, the SA2 block comprises three parallel processing subnetworks, the output of which is ultimately fused into a single three-dimensional feature map. Given the feature map with dimensions {a, a, h}, it is input in the first subnetwork after two-dimensional convolution, and it undergoes two depth-separable convolutions with a convolution kernel of (1*1), producing two feature maps of constant size. The first two channels of the two feature maps are merged to obtain two matrices of shape {a*a, h}, then, the latter matrix is transposed and multiplied with the former matrix to obtain the {a*a, a* a} matrix, which is output from the first subnetwork through a Softmax activation function. Additionally, in the second subnetwork, the input feature map is subjected to two-dimensional convolution twice to obtain the shape {a, a, h2} feature map, and the first two channels are merged to obtain the shape {a*a, h2} matrix. In the third subnetwork, the operation is the same as the first subnetwork, the input matrix {a, a, h} is subjected to a two-dimensional convolution and then two depth-separable convolutions to obtain two features. The difference from the first subnet is that this time the former matrix is transposed, and the latter is multiplied to obtain a matrix of shape {h2, h2}, which is then output by the Softmax activation function. Finally, the three distinct outputs of the corresponding subnetworks are multiplied forming a new matrix {a*a, h2}, which is then reshaped to an {a, a, h2} matrix. After passing the input feature map through the two-dimensional convolution with filters h2 again, a new feature map with shape {a, a, h2} is obtained. Ultimately, the newly obtained feature map, the feature map output by the second subnetwork after the second convolution and feature map of three subnetworks' multiplied output are added together to obtain the final feature map and is input to a batch standard normalization (BN) [34] function, a ReLU [35] activation function, and a MaxPooling operation to produce the final output.

The SA2 block employs several convolutions with different kernel sizes, and also transpose multiplication operations, to obtain matrices containing spatial and spectral features that are connected by a residual structure to better capture the relationship between the classified features and the spectral information.

2.3. Residual Convolutional Layer Structure (ECA_Residual_NET)

Considering that a deeper network may cause gradient dispersion, an improved residual structure was used to build a neural network model, i.e., the ECA residual net. This structure can retain weak image information, effectively deepen the neural network, extract high-level abstract feature information of the image without increasing the number of network parameters, and solve the problem of network degradation. This network structure is presented in Figure 3. Given an input feature of {c, c, d}, it undergoes a batch normalization (BN) operation, which greatly improves the processing speed of the subsequent data information. The output is then input to a two-dimensional convolutional layer utilizing a kernel of (3, 3) with a step size of a, and a ReLU activation function. The output then passes through a BN process. The latter convolutional BN process is repeated, and then, the output passes through an ECA-NET feature extraction module. The ECA-NET output result is added to the feature map that passed the BN layer for the first time, and ultimately the output is provided by a ReLU activation function. It is worth noting that the output result assigns lower weights to insensitive features in the hyperspectral domain and higher weights to abstract deep features.

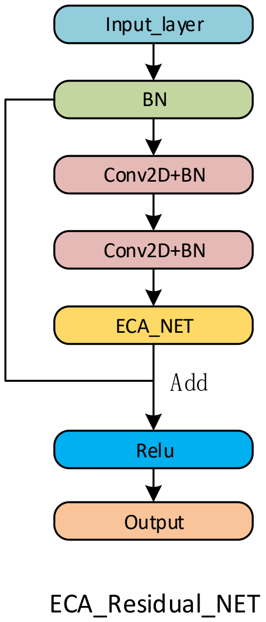


Figure 3. ECA residual net.

2.4. CRE_Block

The CRE_Block comprises a two-dimensional convolution process, a ReLU activation function, a batch normalization, and an attention mechanism (ECA_NET) to aggregate the spatial spectrum characteristics of the image. The module structure is depicted in Figure 4.

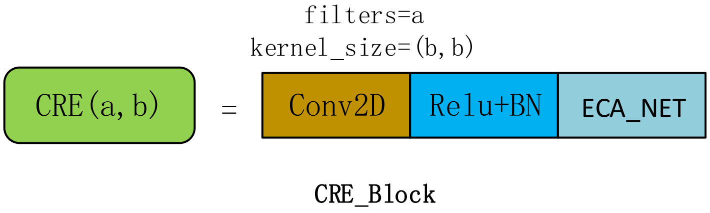


Figure 4. CRE block.

2.5. Overall Network Structure of the Suggested Deep Classification Technique

The proposed deep network architecture uses the PCA algorithm to separate and extract the image space–spectrum features for the first time, and then uses the S2A module, which contains the attention mechanism and the improved residual network structure, to extract the spatial and spectral features of the image multiple times. Additionally, in the middle of the model, we also exploit the ECA-NET network, and finally perform a feature fusion process to merge the features from the distinct subnetwork, e.g., S2A module, ECA-NET, etc., and input the fused feature to a fully connected layer for classification. The proposed model structure is shown in Figure 5.

We utilize the PCA method to reduce the image spectral dimensions to 3 and 20 and select patches of different sizes to create two inputs (Input_1, Input_2) with size {27,27,3} and {7,7,20}, respectively. The feature map with the large patch size and the small spectral dimension, i.e., Input_2, contains more spatial feature information, while the feature map with the small patch size but large spectral dimension contains more spectral feature information, i.e., Input_1.

Initially, we input to our architecture the Input_1 feature map and extract the space–spectrum feature information of the image through an S2A block module with 128 filters and convolution kernels of (1,1) and (5,5). The output tensor of the S2A block is {13, 13, 128}, which is then input to the subnetworks Net_1 and Net_2, respectively. The Net_1 network first consists of an S2A block module with 64 filters and convolution kernels of (1, 1), (3, 3), and comprises of two CRE modules followed by MaxPooling operations. Ultimately, the output shape is of this subnetwork is {3, 3, 64}. Regarding the Net_2 subnetwork, the Input_1 feature tensor is input to two ECA_Residual_NET modules and then passes through two CRE modules and MaxPooling operations to obtain an output feature map (map2) of size {3,3,64}. Considering the Input_2 feature map, it is initially input to the two CRE modules to create a feature map of {7, 7, 192} that is sent to Net_3 and Net_4, respectively. The Net_3 network contains two CRE modules with different parameters followed by a MaxPooling operation that outputs a feature map (map3) with shape {3, 3, 64}. The Net_4 network considers an S2A block module with 64 filters and (1, 1), (3, 3) convolution kernels followed by a MaxPooling layer, to ultimately create an output feature map (map4) with shape {3, 3, 64}. Finally, the four output feature maps (maps) are concatenated to realize the multi-scale feature extraction of the hyperspectral image, first through the global average pooling layer (GAP), and then through the two fully connected layers with two parameters (the variable g is used to represent the number of layers in the legend) respectively of 200, 100. The connection layer adds the sigmoid activation function, and finally, the hyperspectral image classification information is obtained through a fully connected layer (the variable h represents the classification type) plus the Softmax activation function output. This paper extracts hyperspectral image information from four different scales for many times, and finally connects the extracted features, which effectively improves the accuracy of the hyperspectral image classification problem.

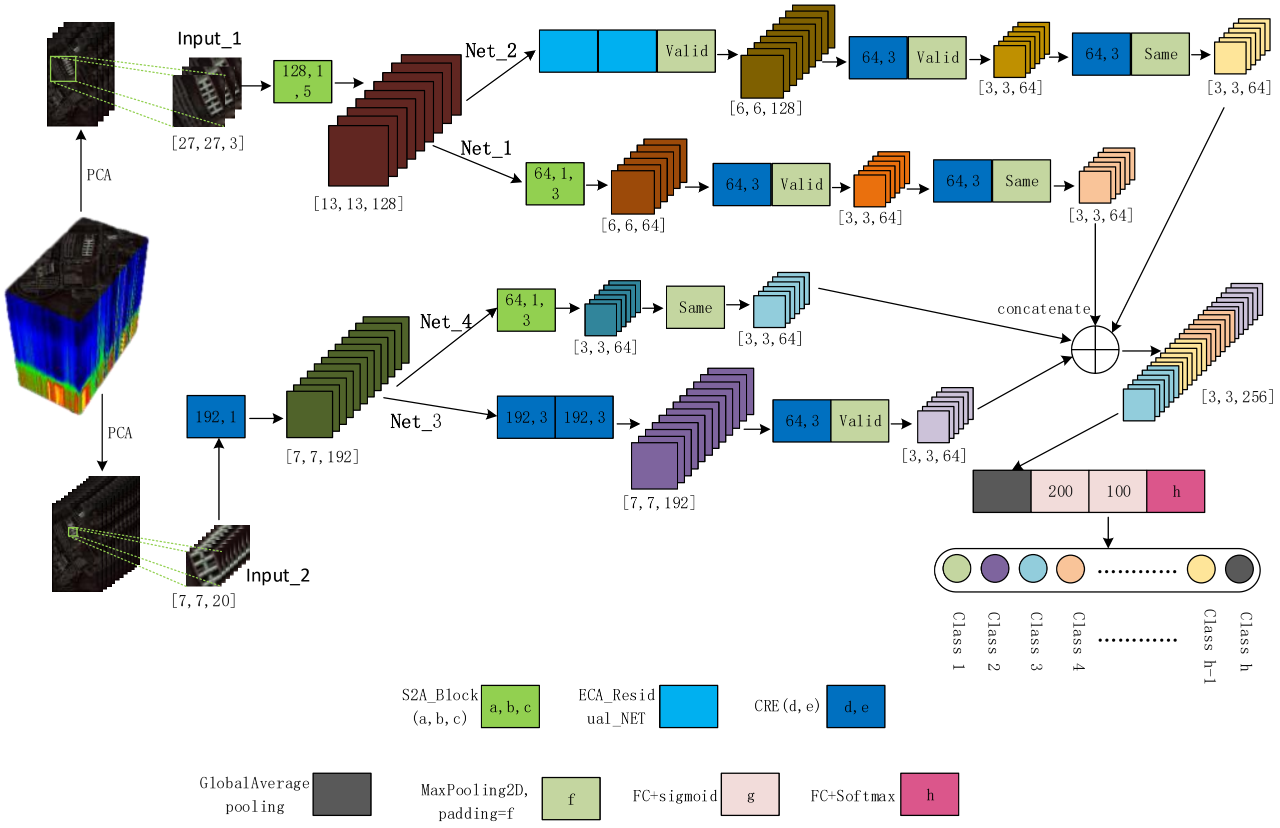


Figure 5. Overall network structure.

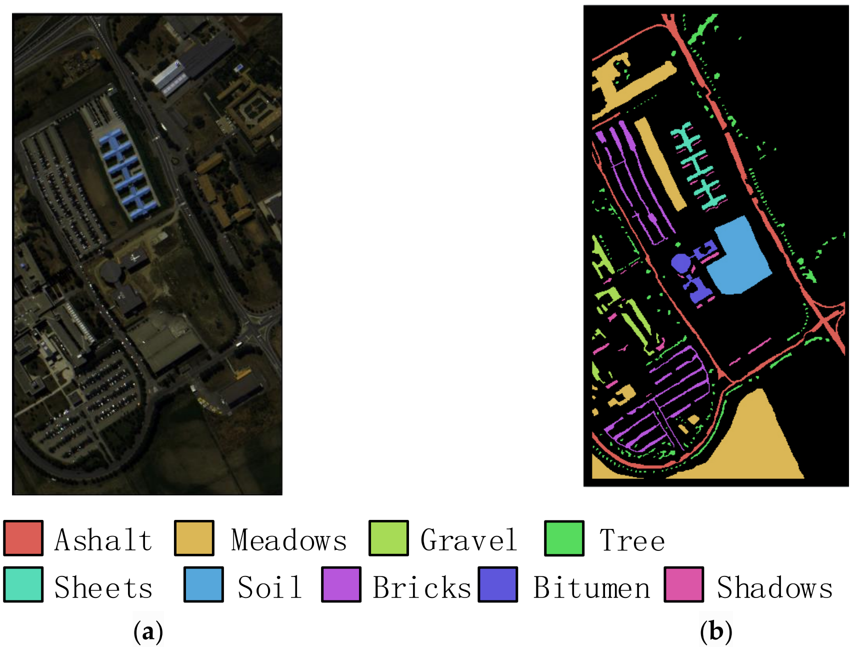
3. Experimental Platform and Experimental Result

All experiments are performed on a Windows 10 system, with an Intel Core i7-9600 CPU, an Nvidia GeForce GTX2060S GPU with 8 GB video memory, using the TensorFlow2.3 deep learning framework and a Python 3.7 compiler. During trials, we challenge the classification performance of the proposed network model against current hyperspectral classification models on various datasets. Additionally, we also analyze the influence of various hyperparameters on the classification performance of our model. The evaluation metrics used are the overall accuracy (OA), the average classification (AA) accuracy, and the Kappa coefficient.

3.1. Introduction of the Dataset

3.1.1. Pavia University Dataset

The Pavia University dataset (UP) is a collection of hyperspectral images obtained from Pavia, Italy, with an example presented in Figure 6a. The spatial image size is 610 × 340 pixels, while the spectral information has 103 effective bands, and the wavelength range is 430~860 nm. The spatial resolution is 1.3 m, including nine types of ground features such as grass, asphalt, bricks, etc. The real ground feature map is shown in Figure 6b, with 42,776 pixels marked in total. During trials, as in work [20], we randomly select 10%, 10%, and 70% of the whole labeled samples as training, validation, and testing sets for the datasets. The dataset feature types, along with the training and test set sample information are shown in Table 1.



(**a**) (**b**)

Figure 6. Pavia University dataset, (**a**) original image, (**b**) mark map of real ground objects.

Table 1. Training validation and testing sample numbers for the Pavia University dataset.

Number	Class	Training	Validation	Test	Total Samples
1	Asphalt	663	663	5305	6631
2	Meadows	1864	1864	14,921	18,649
3	Gravel	209	209	1681	2099
4	Trees	306	306	2452	3064
5	Sheets	134	134	1077	1345
6	Bare Soil	502	502	4025	5029
7	Bitumen	133	133	1064	1330
8	Bricks	368	368	2946	3682
9	Shadows	94	94	759	947
Total		4273	4273	34,230	42,776

3.1.2. KSC Dataset

The KSC dataset is a hyperspectral imagery collection obtained from the Kennedy Space Center. An example is shown in Figure 7a. The spectral information has a total of 176 effective bands, the spatial image size is 614 × 512 pixels, and the wavelength range is 400~2450 nm, including scrub, oak hammock, slash pine, and 13 other class categories. The ground truth feature map is shown in Figure 7b, where a total of 5211 pixels are labeled. During trials, as in work [20], we randomly select 20%, 10%, and 70% of the whole labeled samples as training, validation, and testing sets for the datasets. The dataset feature types, along with the training and test set sample quantity information are shown in Table 2.

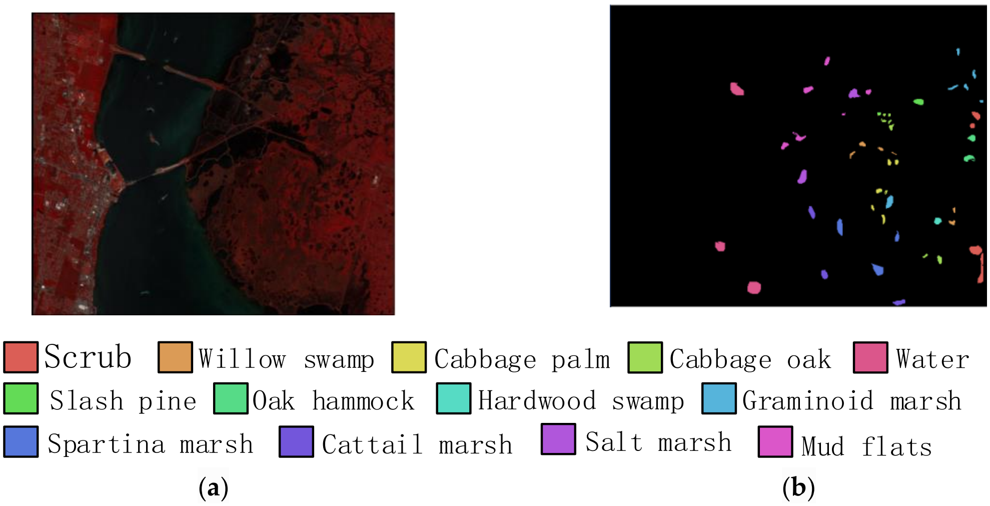


Figure 7. KSC dataset, (**a**) original image, (**b**) mark map of real ground objects.

Table 2. Training validation and testing sample numbers in KSC.

Number	Class	Training	Validation	Test	Total Samples
1	Scrub	152	76	533	761
2	Willow swamp	48	24	171	243
3	Cabbage palm	50	25	181	256
4	Cabbage oak	50	25	177	252
5	Slash pine	32	16	113	161
6	Oak hammock	46	23	160	229
7	Hardwood swamp	20	10	75	105
8	Graminoid marsh	86	43	302	431
9	Spartina marsh	104	52	364	520
10	Cattail marsh	80	40	284	404
11	Salt marsh	84	42	293	419
12	Mud flats	100	50	353	503
13	Water	186	93	648	927
Total		1038	519	3654	5211

3.1.3. Indian Pines Dataset

The Indian Pines dataset (IN) was collected in Indiana, USA. An example of that dataset is shown in Figure 8a. The spectral information has a total of 200 effective bands, the spatial image size is 145 × 145 pixels, the wavelength range is 400~2500nm, and the spatial resolution is 20 m. This dataset includes 16 feature categories such as alfalfa, corn, oats, etc. The ground truth feature map is shown in Figure 8b, with 10,249 pixels labeled. Similarly, to the previous datasets, during our experiments, as in work [20], we randomly select 20%,10%, and 70% of the whole labeled samples as training, validation, and testing sets for the datasets. Table 3 shows the types of object classes in the dataset, the number of samples in the training set, and the test set.

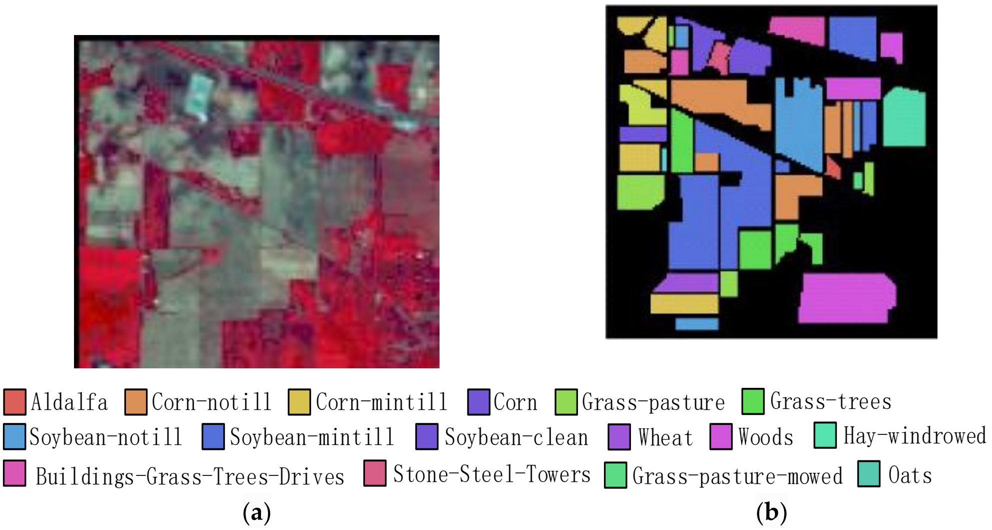


Figure 8. Indian Pines dataset, (**a**) original image, (**b**) mark map of real ground objects.

Table 3. Training validation and testing sample numbers in Indian Pines.

Number	Class	Training	Validation	Test	Total Samples
1	Alfalfa	8	4	34	46
2	Corn—no till	284	142	1002	1428
3	Corn—min till	166	83	581	830
4	Corn	46	23	168	237
5	Grass/pasture	146	73	511	730
6	Grass/tress	96	48	339	483
7	Grass/pasture—mowed	6	3	19	28
8	Hay—windrowed	94	47	337	478
9	Soybeans—no till	194	97	681	972
10	Soybeans—min till	490	245	1720	2455
11	Soybeans—clean till	118	59	416	593
12	Wheat	40	20	145	205
13	Woods	252	126	887	1265
14	Buildings–grass–trees	76	38	272	386
15	Stone–steel towers	18	9	66	93
16	Oats	4	2	14	20
Total		2038	1019	7192	10,249

3.2. *Analysis of Experimental Results*

During trials, we challenge our proposed deep network architecture against PCA [7], SVM [5], two-dimensional convolutional neural network (2D-CNN) [14], three-dimensional convolutional neural network (3D-CNN) [18], three-dimensional residual convolution structure (RES-3D-CNN) [36], and the space–spectrum joint deep network (SSRN) [20].

3.2.1. Parameter Setting

In this section, we analyze the interplay between the parameter setup and the classification performance of the proposed model. The tuned parameters include learning rate, batch_size, and training sample ratio. Regarding learning rate, it controls the speed of the gradient descent during the training process, with the appropriate learning rate parameters effectively controlling the convergence ability and speed of the model. We evaluate our network by using six learning rates with different sizes, i.e., 0.00005, 0.0001, 0.0003, 0.0005, 0.001, and 0.005. The test results are shown in Figure 9, from which we observe that when the learning rate is 0.0003, the classification performance on the three datasets is better. Additionally, tuning the learning rate parameter has less impact on the accuracy of the Pavia University dataset and a greater impact on the Indian Pines dataset.

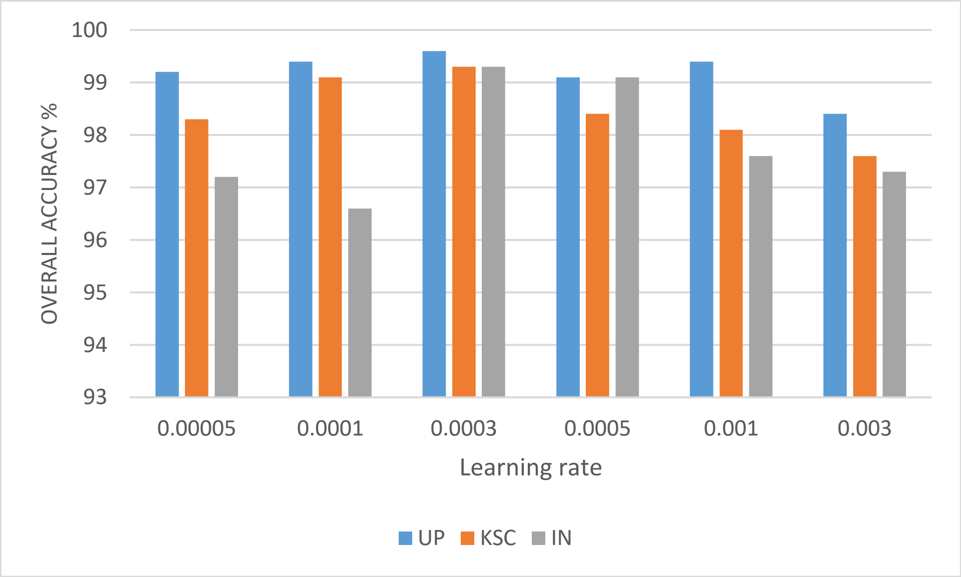


Figure 9. Overall accuracy (OA, %) of the proposed network with different learning rates in Indian Pines (IN), Pavia University (UP), and KSC dataset.

The next trial investigates how the batch size affects the overall accuracy of our method. The batch size refers to the number of samples selected during training. Choosing a suitable batch_size can effectively improve the memory utilization and improve the convergence accuracy of the model. We challenge the performance on batch_size of 16, 32, 64, and 128, with the corresponding results presented in Figure 10. Our trials demonstrate that when the batch_size is 16, the classification attained on the three datasets is better. However, in the case of fewer training samples, a smaller batch_size will perform better.

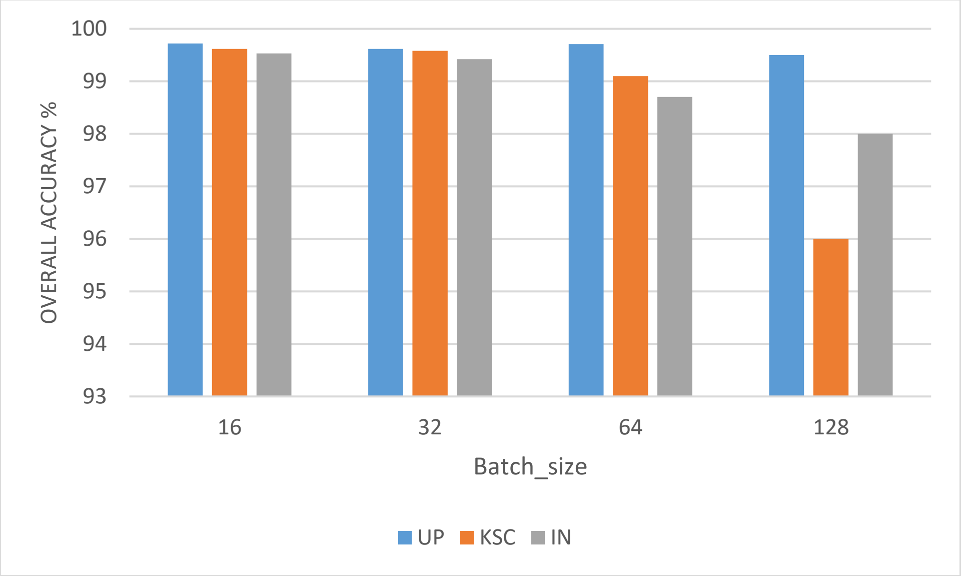


Figure 10. OA (%) of the proposed network with different batch_size in IN, UP, and KSC dataset.

Our final trial considers utilizing 5%, 10%, 20%, 30%, and 40% of the sample data as the training set. After our model being trained on the respective sample dataset, we test our network, with the corresponding results presented in Figure 11. From the latter figure, we conclude that as the training samples increase, the overall accuracy of our model is increasing. To compare with other networks, we adopt the strategy of [20], and we randomly select 20%, 10%, and 70% of the whole labeled samples as training, validation, and testing sets, respectively, for the Indian Pines and KSC datasets. For the Pavia University dataset, we employ a 10%, 10%, and 80% strategy.

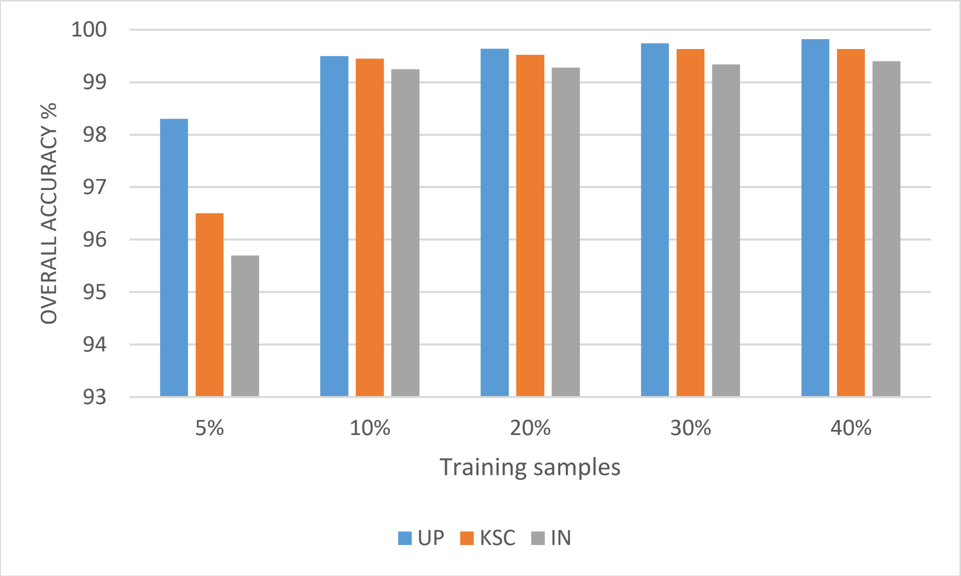


Figure 11. OA (%) of the proposed network with different training samples in IN, UP, and KSC dataset.

3.2.2. Pavia University Dataset

Figure 12 shows the accuracy curve of each classification model in the Pavia University dataset. It can be seen that our deep network model converges quickly, and the classification accuracy is higher compared to the competitor techniques. The model is trained on the Pavia University dataset in just 3 minutes and 31 seconds. Table 4 shows the classification accuracy of each model for all nine object classes. From that table, we observe that our classification network manages 99.67% OA, 99.21% AA, and the Kappa coefficient is 0.9971. Compared to the competitor algorithms, our method attains higher classification results. Specifically, the OA of the PCA algorithm is 87.23%, the AA is 88.15%, and the Kappa coefficient is 0.85. Under the same conditions, the overall classification accuracy of the SVM algorithm compared to PCA is increased by 3.3%, and the average classification accuracy is increased by 2.1%, but still inferior to our method. The OA of 2D-CNN is 93.33%, AA is 94.17%, and the Kappa coefficient is 0.92, while the OA of the 3D-CNN is 94.68%, AA is 95.37%, and the Kappa coefficient is 0.94. Compared to the traditional algorithms, the classification accuracy of both CNN methods is greatly improved reflecting the superiority of deep learning in the hyperspectral classification problem. Compared with 3D-CNN, the overall classification accuracy of the 3D residual network is increased by 3.1%, and the average classification accuracy is increased by 2.8%. SSRN network presents an appealing classification performance attaining 98.17% OA, 98.64% AA, and a Kappa coefficient of 0.98. We bolded the data with the highest accuracy of feature classification. The experimental results show that the overall performance of our proposed network model is better than other models.

Figure 13 depicts some classification examples per method. It can be seen that the PCA and SVM classification algorithms have poor accuracy and present more misclassifications. The 2D-CNN classification results are slightly improved but still contain many misclassifications. The 3D-CNN, RES-3D, and SSRN models attain improved classification accuracy. However, the classification results of the suggested deep network architecture show an even more accurate classification, as they do not contain salt and pepper noise, and the boundaries are smooth and fit.

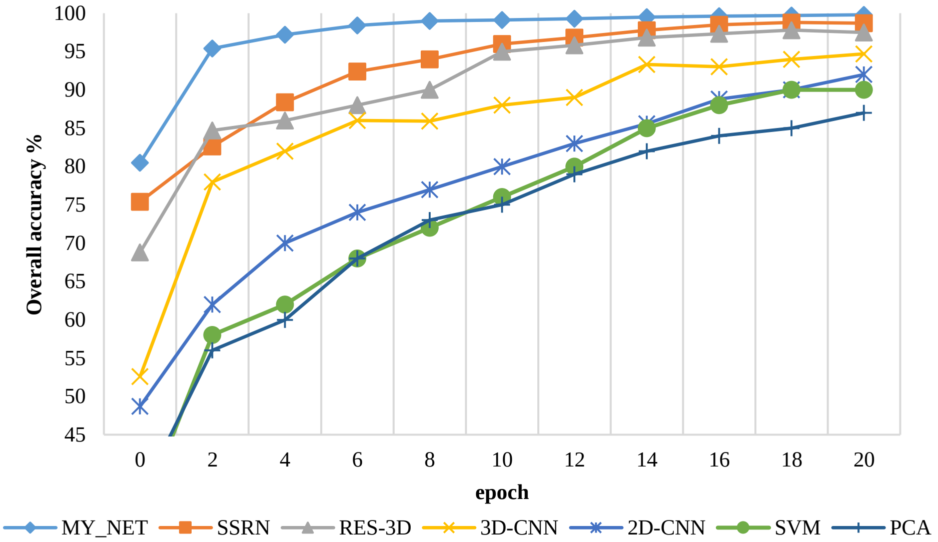


Figure 12. Overall accuracy curve of different models in Pavia University dataset.

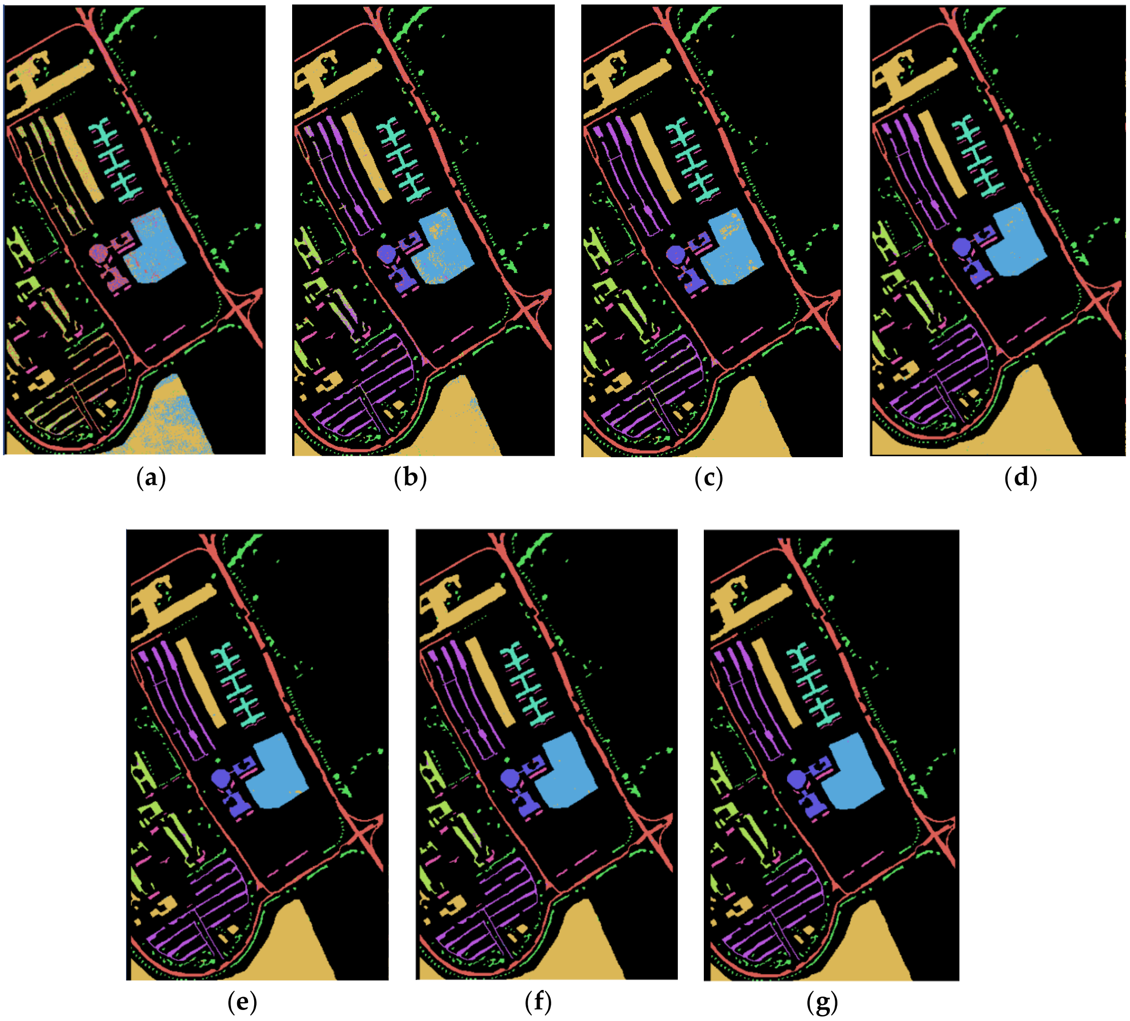


Figure 13. (**a**–**g**) Classification map of principal components analysis (PCA), support vector machine (SVM), 2D convolutional neural network (CNN), 3D-CNN, RES-3D, space–spectrum joint deep network (SSRN), and proposed on Pavia University dataset.

Table 4. Classification results of different models in Pavia University.

Class	PCA	SVM	2D-CNN	3D-CNN	RES-3D	SSRN	My_Net
Asphalt	76.65	85.57	93.80	0.97.15	96.18	**98.84**	98.74
Meadows	88.86	86.56	89.43	93.43	98.10	97.23	**100.00**
Gravel	92.24	89.43	92.27	94.48	97.21	**99.82**	99.56
Trees	90.38	92.17	94.13	92.39	95.31	98.15	**100.00**
Metal	87.56	85.68	90.51	94.37	**99.64**	99.43	98.83
Soil	85.52	94.79	92.34	96.68	97.72	96.17	**100.00**
Bitumen	88.67	90.65	94.72	93.48	96.31	96.56	**99.32**
Bricks	91.73	92.43	95.78	97.76	98.48	**100.00**	**100.00**
Shadows	89.86	92.79	91.56	94.34	**100.00**	98.34	99.67
OA (%)	87.23	90.53	93.33	94.68	97.78	98.17	**99.82**
AA (%)	88.15	90.25	94.17	95.37	98.17	98.64	**99.59**
Kappa × 100	85.23	89.24	92.48	94.46	97.25	98.76	**99.71**

3.2.3. KSC Dataset

The accuracy curve of each classification model for the KSC dataset are presented in Figure 14. It can be seen that as the epoch increases, the classification accuracy rate is increasing. The model is trained on the KSC dataset in just 1 minute and 43 seconds. Table 5 shows the precise classification indicators of each method, where the proposed technique attains the highest metrics compared to the competitor methods, i.e., 99.81% OA, 99.74% AA, and 0.9952 Kappa metric. Additionally, from this table we observe that the overall accuracy of PCA, SVM, 2D-CNN, 3D-CNN, RES-3D, and SSRN has improved by 17.72%, 11.05%, 9.56%, 6.29%, 2.34%, and 1.65%, respectively, in the KSC dataset examined here compared to the Pavia University dataset evaluated in Section 3.2.2. Accordingly, the average accuracy has increased by 18.18%, 10.57%, 8.42%, 5.54%, 2.49%, 1.61%. This trial shows that RES-3D and the suggested network manage a better classification and highlights that our method is more suitable for a real landmark map. Classification examples per method are shown in Figure 15. Our network has started to converge in the 4th Epoch, and its performance is better than that of other models. Our network model shows better performance on the problem of hyperspectral image classification.

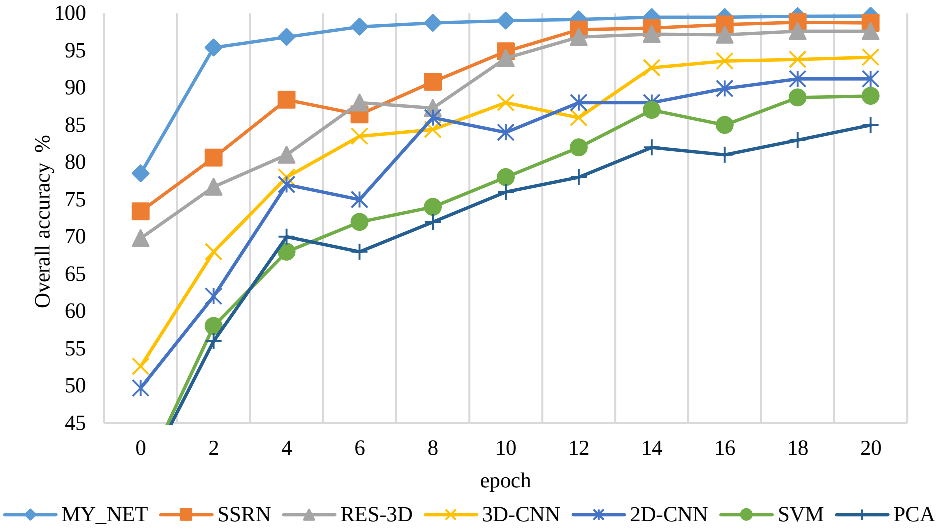
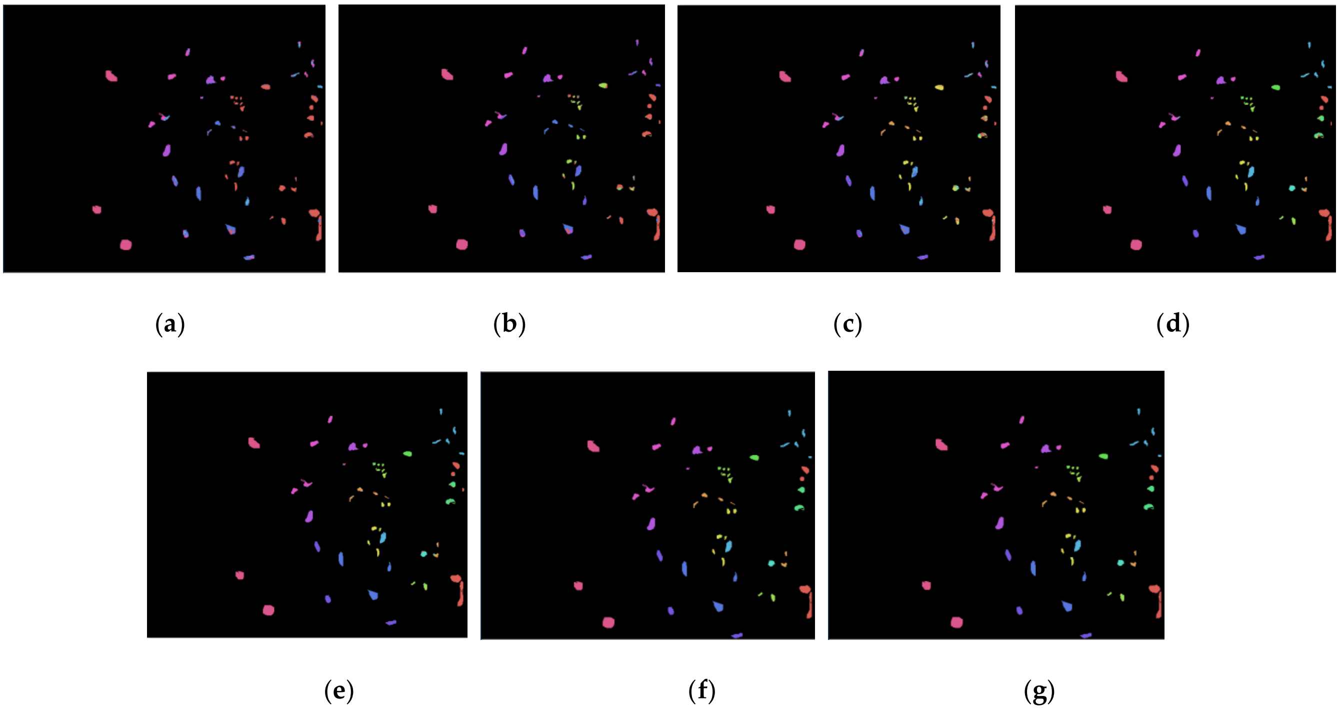


Figure 14. Overall accuracy curve of different models in KSC dataset.

Table 5. Classification results of different models in KSC.

Class	PCA	SVM	2D-CNN	3D-CNN	RES-3D	SSRN	My_Net
Scrub	66.75	90.28	84.42	97.19	97.96	98.46	**99.37**
Willow swamp	93.51	81.96	79.61	82.38	96.27	96.17	**99.82**
Cabbage palm	74.43	75.65	89.34	94.24	94.17	98.42	**99.86**
Cabbage oak	92.38	81.35	94.91	90.18	96.76	99.49	**100.00**
Slash pine	83.63	92.83	66.48	70.64	98.32	96.34	**99.58**
Oak hammock	69.43	74.51	73.34	70.37	94.51	**100.00**	**100.00**
Hardwood swamp	77.28	79.62	69.64	74.15	99.34	99.37	**100.00**
Graminoid marsh	76.72	95.83	81.72	90.46	98.94	**99.82**	99.57
Spartina marsh	82.79	91.92	90.16	95.37	97.85	97.71	**100.00**
Cattail marsh	81.56	88.14	91.87	98.48	100.00	99.42	**100.00**
Salt marsh	73.92	91.42	93.16	99.16	98.42	99.45	**100.00**
Mud flats	69.34	84.48	88.64	98.49	**100.00**	97.18	99.75
Water	85.16	86.94	95.54	**100.00**	**100.00**	**100.00**	**100.00**
OA (%)	82.09	88.76	90.25	93.52	97.47	98.16	**99.81**
AA (%)	81.56	89.17	91.32	94.20	97.25	98.13	**99.74**
Kappa × 100	81.72	88.56	91.92	94.77	97.25	98.64	**99.52**



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**)

Figure 15. (**a**–**g**) Classification map of PCA, SVM, 2D-CNN, 3D-CNN, RES-3D, SSRN, and proposed on KSC dataset.

3.2.4. Indian Pines Dataset

Figure 16 shows the accuracy curve of each classification model in the Indian Pines dataset. Overall, the proposed deep classification network has the fastest convergence and the highest accuracy, and its classification performance is better compared to the competitor models. The model is trained on the Indian Pines dataset in just 1 minute and 37 seconds. Table 6 shows the precise classification index of each model for 13 classes of ground objects. From the latter table, we observe that the overall accuracy value of the proposed network is 24.05%, 18.81%, 13.22%, 5.09%, 1.79%, and 1.10% compared to PCA, SVM, 2D-CNN, 3D-CNN, RES-3D, and SSRN, respectively. Accordingly, the average accuracy attained by our model is higher by 24.02%, 18.02%, 14.22%, 5.48%, 2.09%, 1.17%, respectively, reaching a classification of 99.45% average accuracy and a Kappa coefficient of

0.9961. The classification results of each model are shown in Figure 17. It can be seen that the classification effect of PCA, SVM, and 2D-CNN models is poor, with more noise and speckles. 3D-CNN and RES-3D model classification results are less noisy, which improves the classification accuracy of these models. SSRN and the suggested network both attain an appealing classification accuracy. It can be seen from Table 6 that the classifier shows that the classification accuracy of Corn—min till is slightly lower, while the classification accuracy of the other classes is higher. In addition, it can be seen from Figure 16 that our network converges extremely fast and achieved good classification accuracy in the 6th Epoch, which is also due to other models.

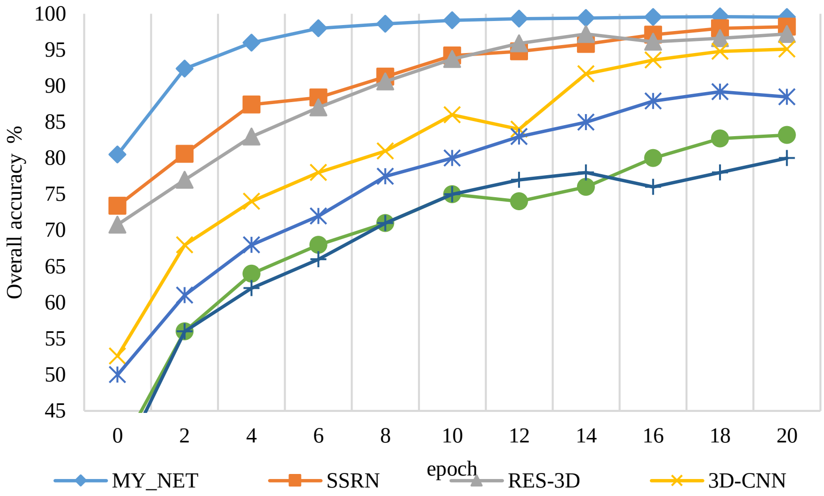


Figure 16. Overall accuracy curve of different models in Indian Pines dataset.

Table 6. Classification results of different models in Indian Pines.

Class	PCA	SVM	2D-CNN	3D-CNN	RES-3D	SSRN	My_Net
Alfalfa	72.46	78.53	74.37	91.62	97.38	97.48	**100.00**
Corn—no till	69.35	82.61	86.61	89.33	95.37	**100.00**	**100.00**
Corn—min till	71.58	74.84	91.49	93.97	98.64	**99.46**	99.25
Corn	81.62	79.71	90.82	95.94	97.92	96.56	**100.00**
Grass—pasture	67.34	72.67	73.58	82.34	97.97	97.76	**100.00**
Grass—tress	81.29	85.42	82.65	96.72	**99.48**	**99.48**	98.56
Grass—pasture	75.52	81.19	79.37	81.61	95.19	**100.00**	**100.00**
Hay—windrowed	77.43	83.51	87.51	79.37	99.41	99.56	**100.00**
Oats	86.96	83.37	90.43	93.19	97.76	99.12	**100.00**
Soybeans—no till	80.84	88.28	93.16	97.64	98.84	**100.00**	99.17
Soybeans—min till	82.61	74.76	95.19	93.28	97.19	99.14	**100.00**
Soybeans—clean till	76.39	86.51	94.72	98.76	**100.00**	98.08	100.00
Wheat	85.64	82.43	92.49	99.24	96.15	97.58	**100.00**
Woods	77.52	75.97	90.84	94.49	98.76	**100.00**	99.38
Buildings–grass –trees	84.48	88.91	87.37	89.19	96.14	**99.67**	98.89
Stone–steel towers	73.47	71.58	86.64	94.64	**99.39**	98.84	99.38
OA (%)	75.32	80.56	86.15	94.28	97.58	98.27	**99.37**
AA (%)	75.43	81.43	85.23	93.97	97.36	98.28	**99.45**
Kappa×100	75.87	80.26	85.64	94.15	97.44	98.52	**99.61**

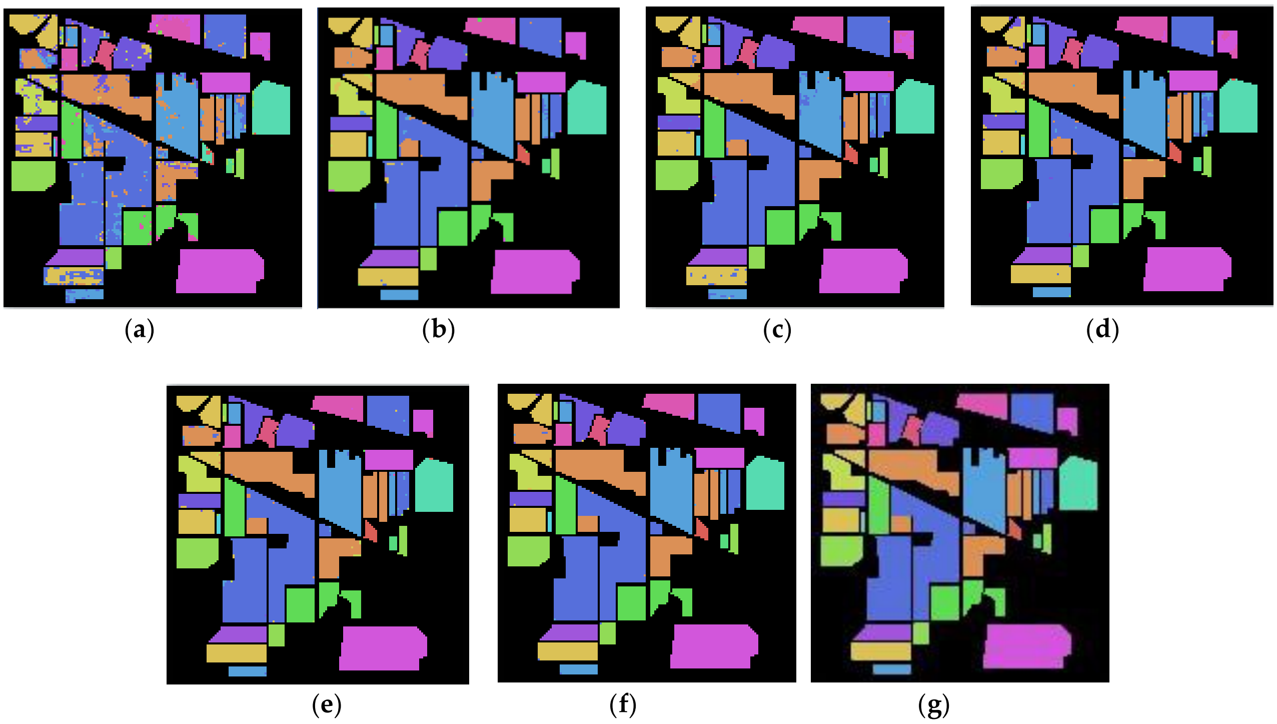


Figure 17. (**a**–**g**) Classification map of PCA, SVM, 2D-CNN, 3D-CNN, RES-3D, SSRN, and proposed on Indian Pines dataset.

4. Conclusions

This paper studies the application of deep learning for hyperspectral image classification. Aiming at the characteristics of a wide range of hyperspectral images, high spectral resolution, and a large amount of redundant information, a multi-scale residual convolution with ECA-NET is designed. The neural network model extracts image feature information. The model uses ECA-Net, improved residual network, and other structures to extract hyperspectral image information multiple times from different scales, can fully fuse and extract the space–spectrum characteristics of the image and effectively solve the problems of gradient dispersion and sample information redundancy. We challenge our suggested classification model against six current classification models, i.e., PCA, SVM, 2D-CNN,3D-CNN, RES-CNN, SSRN, on the Pavia University, KSC, and Indian Pines datasets, and demonstrate that our algorithm can effectively classify various object classes and has certain advantages in dealing with hyperspectral classification problems. Our proposed method successfully attains 99.82%, 99.81%, and 99.37% overall accuracy, respectively, on the three different free datasets. All trials demonstrate the superiority of our method against the competitor ones attaining a high classification accuracy. Future work shall focus on studying spatial and spectral feature fusion methods to enhance the feature extraction process, improve the network structure and parameters, accelerate the model convergence, and reduce network training time.

Author Contributions: Conceptualization, Y.Q. and W.L.; methodology, Y.Q.; software, Y.Q.; validation, Y.Q. and W.L.; formal analysis, Y.Q.; writing—original draft preparation, Y.Q. and W.L.; writing—review and editing, Y.Q.; visualization, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bioucas-dias, J.M.; Plaza, A.; Camps-valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
2. Li, W.; Wu, G.; Zhang, F.; Du, Q. Hyperspectral image classification using deep pixel-pair features. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 844–853. [CrossRef]
3. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]
4. Maghsoudi, Y.; Zoej, M.J.V.; Collins, M. Using class-based feature selection for the classification of hyperspectral data. *Int. J. Remote Sens.* **2011**, *32*, 4311–4326. [CrossRef]
5. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
6. Li, J.; Bioucas-Dias, J.; Plaza, A. Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 809–823. [CrossRef]
7. Licciardi, G.; Marpu, P.; Chanussot, J.; Benediktsson, J. Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 447–451. [CrossRef]
8. Deng, Y.J.; Li, H.C.; Pan, L.; Shao, L.Y.; Du, Q.; Emery, W.J. Modified tensor locality preserving projection for dimensionality reduction of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 277–281. [CrossRef]
9. Villa, A.; Benediktsson, J.A.; Chanussot, J.; Jutten, C. Hyperspectral Image Classification With Independent Component Discriminant Analysis. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4865–4876. [CrossRef]
10. Liu, J.; Wu, Z.; Wei, Z.; Xiao, L.; Sun, L. Spatial-spectral kernel sparse representation for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2462–2471. [CrossRef]
11. Gao, L.; Hong, D.; Yao, J.; Zhang, B.; Gamba, P.; Chanussot, J. Spectral Superresolution of Multispectral Imagery with Joint Sparse and Low-Rank Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, 1–12. [CrossRef]
12. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [CrossRef]
13. Hu, W.; Huang, Y.Y.; Wei, L.; Zhang, F.; Li, H.C. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *J. Sens.* **2015**, *2015*, 1–12. [CrossRef]
14. Hao, S.; Wang, W.; Ye, Y.; Nie, T.; Bruzzone, L. Two-stream deep architecture for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2349–2361. [CrossRef]
15. Mei, S.; Ji, J.; Bi, Q.; Hou, J.; Du, Q.; Li, W. Integrating spectral and spatial information into deep convolutional neural networks for hyperspectral classification. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5067–5070.
16. Chen, C.; Zhang, J.; Zheng, C.; Yan, Q.; Xun, L. Classification of hyperspectral data using a multi-channel convolutional neural network. In Proceedings of the 14th International Conference on Intelligent Computing (ICIC), Wuhan, China, 15–18 August 2018; pp. 81–92.
17. Gao, F.; Huang, T.; Sun, J.; Wang, J.; Hussain, A.; Yang, E. A New Algorithm of SAR Image Target Recognition Based on Improved Deep Convolutional Neural Network. *Cogn. Comput.* **2019**, *11*, 809–824. [CrossRef]
18. Rao, M.; Tang, P.; Zhang, Z. A Developed Siamese CNN with 3D Adaptive Spatial-Spectral Pyramid Pooling for Hyperspectral Image Classification. *Remote Sens.* **2020**, *12*, 1964. [CrossRef]
19. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef]
20. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 847–858. [CrossRef]
21. Wu, P.; Cui, Z.; Gan, Z.; Liu, F. Residual Group Channel and Space Attention Network for Hyperspectral Image Classification. *Remote Sens.* **2020**, *12*, 2035. [CrossRef]
22. Mou, L.; Zhu, X.X. Learning to Pay Attention on Spectral Domain: A Spectral Attention Module-Based Convolutional Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 110–122. [CrossRef]
23. Wang, L.; Peng, J.T.; Sun, W.W. Spatial–Spectral Squeeze-and-Excitation Residual Network for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 884. [CrossRef]
24. Zhong, Z.; Li, J.; Clausi, D.A.; Wong, A. Generative Adversarial Networks and Conditional Random Fields for Hyperspectral Image Classification. *IEEE Trans. Cybern.* **2020**, *50*, 3318–3329. [CrossRef] [PubMed]
25. Yuan, Y.; Zheng, X.T.; Lu, X.Q. Hyperspectral image superresolution by transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1963–1974. [CrossRef]
26. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 277–281. [CrossRef]
27. Niu, R. HMANet: Hybrid Multiple Attention Network for Semantic Segmentation in Aerial Images. *arXiv* **2020**, arXiv:2001.02870.
28. Zhang, M.; Li, W.; Du, Q. Diverse region-based CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2018**, *27*, 2623–2634. [CrossRef] [PubMed]
29. Fang, L.Y.; LI, S.T.; Kang, X.D.; Jon, A. Spectral spatial classification of hyperspectral images with a superpixel-based discriminative sparse model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4186–4201. [CrossRef]

30. Wang, Q.L.; Wu, B.G.; Zhu, P.F.; Li, P.H.; Zuo, W.M.; Hu, Q.H. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.
31. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E.H. Squeeze-and-Excitation Networks. *arXiv* **2018**, arXiv:1709.01507v4.
32. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
33. Li, L.; Yin, J.H.; Jia, X.P.; Li, S.; Han, B.N. Joint Spatial-Spectral Attention Network for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *130*, 38–45. [CrossRef]
34. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, Paris, France, 6–11 July 2015; pp. 448–456.
35. Nair, V.; Hinton, G.E. Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML), Haifa, Israel, 21–24 June 2010; Omnipress: Madison, WI, USA, 2010; pp. 807–814.
36. Liu, B.; Yu, X.; Zhang, P.; Tan, X. Deep 3D convolutional network combined with spatial-spectral features for hyperspectral image classification. *Acta Geod. Cartogr. Sin.* **2019**, *48*, 53–63.