



Article

Mixed Structure with 3D Multi-Shortcut-Link Networks for Hyperspectral Image Classification

Hui Zheng ^{1,2}, Yizhi Cao ^{1,*}, Min Sun ², Guihai Guo ¹, Junzhen Meng ¹, Xinwei Guo ¹ and Yanchi Jiang ³

¹ Ural Institute, North China University of Water Resource and Electric Power, Zhengzhou 450045, China; zhenghui@ncwu.edu.cn (H.Z.); ggh@ncwu.edu.cn (G.G.); mengjunzhen@ncwu.edu.cn (J.M.); guoxinwei@ncwu.edu.cn (X.G.)

² School of Earth and Space Sciences, Peking University, Beijing 100871, China; sunmin@pku.edu.cn

³ Department of Applied Chemistry, Chubu University, 1200 Matsumoto-cho, Kasugai 487-8501, Japan; jiangyc@isc.chubu.ac.jp

* Correspondence: caoyizhi@ncwu.edu.cn

Abstract: A hyperspectral image classification method based on a mixed structure with a 3D multi-shortcut-link network (MSLN) was proposed for the features of few labeled samples, excess noise, and heterogeneous homogeneity of features in hyperspectral images. First, the spatial-spectral joint features of hyperspectral cube data were extracted through 3D convolution operation; then, the deep network was constructed and the 3D MSLN mixed structure was used to fuse shallow representational features and deep abstract features, while the hybrid activation function was utilized to ensure the integrity of nonlinear data. Finally, the global self-adaptive average pooling and L-softmax classifier were introduced to implement the terrain classification of hyperspectral images. The mixed structure proposed in this study could extract multi-channel features with a vast receptive field and reduce the continuous decay of shallow features while improving the utilization of representational features and enhancing the expressiveness of the deep network. The use of the dropout mechanism and L-softmax classifier endowed the learned features with a better generalization property and intraclass cohesion and interclass separation properties. Through experimental comparative analysis of six groups of datasets, the results showed that this method, compared with the existing deep-learning-based hyperspectral image classification methods, could satisfactorily address the issues of degeneration of the deep network and “the same object with distinct spectra, and distinct objects with the same spectrum.” It could also effectively improve the terrain classification accuracy of hyperspectral images, as evinced by the overall classification accuracies of all classes of terrain objects in the six groups of datasets: 97.698%, 98.851%, 99.54%, 97.961%, 97.698%, and 99.138%.

Keywords: hyperspectral image classification; 3D multi-shortcut-link networks; large softmax; SELU; PReLU



Citation: Zheng, H.; Cao, Y.; Sun, M.; Guo, G.; Meng, J.; Guo, X.; Jiang, Y. Mixed Structure with 3D Multi-Shortcut-Link Networks for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 1230. <https://doi.org/10.3390/rs14051230>

Academic Editors: Yangquan Chen, Subhas Mukhopadhyay, Nunzio Cennamo, M. Jamal Deen, Junseop Lee and Simone Morais

Received: 28 January 2022

Accepted: 28 February 2022

Published: 2 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hyperspectral images (HSIs) contain rich spatial and spectral information [1,2] and are widely applied in the fields of precision agriculture [3], urban planning [4], national defense construction [5], and mineral exploitation [6], among other fields. It has been very successful in allowing active users to participate in collecting, updating, and sharing the massive amounts of data that reflect human activities and social attributes [7–10].

The terrain classification of hyperspectral images is a fundamental problem for various applications, where the classification aims to assign a label with unique class attributes to each pixel in the image based on the sample features of HSI. However, an HSI is high-dimensional with few labeled samples, images between wavebands have a high correlation, and terrain objects with heterogeneous structures that may be homogeneous present the terrain classification of HSIs with huge challenges.

To this end, scholars have put forward many terrain classification algorithms for HSIs, such as support vector machine (SVM) [11–14], sparse representation [15], semi-supervised classification [16,17], extreme learning machine (ELM) [18], and artificial neural network (ANN) [19]. Unfortunately, these classification methods cannot fully utilize the spatial features of HSI data and have poor expressiveness and generalization ability, leading to relatively low classification accuracy [20–22].

Over the recent years, deep-learning-based image classification has achieved substantial development. Compared with traditional algorithms, it mainly utilizes a complex neural network structure to extract deep abstract features, thereby improving the accuracy of image classification. Hence, it is widely applied in target detection [23], image classification [24,25], pattern recognition [26–28], and other fields. In HSI classification, spatial and spectral information can be extracted and transformed via channel mapping into one-dimensional vectors for classification by using stack autoencoding (SAE) [29,30], a deep belief network (DBN) [31], a recurrent neural network (RNN) [32,33], etc. However, these methods require a large number of parameters for expression at the loss of spatial 2D features and relevance between different wavebands, resulting in low classification accuracy. As the most representative network model in deep learning, a convolutional neural network (CNN) can extract spatial and spectral features simultaneously and exhibits good performance in HSI classification [34–41]. However, with the increase in the number of network layers, CNN is prone to losing detailed information that is useful for data fitting during the training process, hence producing the phenomenon of gradient disappearance.

Although many current deep learning network structures have already achieved satisfactory classification results, it remains very difficult for them to achieve perfect classification results in the face of issues such as the curse of dimensionality, image noise, too few labeled samples, and “the same object with distinct spectra, and distinct objects with the same spectrum” [42–44].

Against the above-mentioned issues, this study fused shallow features and deep features by constructing a mixed structure with 3D multi-shortcut-link networks, using hybrid functions to activate neurons; utilized the global self-adaptive average pooling layer to eliminate noises; and finally implemented the terrain classification of HSI via the L-softmax loss function. The validity of the algorithm in this study was verified using six groups of hyperspectral datasets.

2. Related Work

2.1. Shortcut Link

Theories and practices using short links have been studied for a long time [45–47]. In the early stages, training a multi-layer perceptron was mainly achieved by adding an input linear layer to the output layer of the network [46,47]. To solve the gradients vanishing/exploding issue, Szegedy [48] and Lee [49] connected some intermediate layers to auxiliary classifiers directly. Some researchers [50–53] introduced the centralization of layer responses, gradients, and propagation errors to the shortcut links.

In [54,55], Ioffe and Szegedy composed an inception structure with a shortcut branch and a few deeper branches. He [56,57] did a series of studies on residuals and shortcuts and deduced the mathematical formula of residual networks (ResNet) in detail simultaneously. Then, Szegedy introduced ResNet to an inception structure [58]; through a large amount of parameter tuning, the training speed and performance were greatly improved.

2.2. ResNet in HSI Classification

To address the issue of model degeneration, Lu et al. [59] introduced the ResNet into HSI classification methods; the identity mapping of residual modules can be utilized to effectively address the issues of gradient disappearance and overfitting while increasing the network depth. Liu et al. [60] adopted the multilevel fusion structure to extract multiscale spatial–spectral features and introduced ResNet to map shallow features into the space of deep features, promoting the model’s classification accuracy. Meng et al. [61] made

hybrid use of the dense network and ResNet to extract deeper features to enhance the expressiveness of a CNN. Zhong et al. [62] proposed an end-to-end spatial–spectral residual network to learn the continuous features of HSI data, thereby improving the classification accuracy. Cao et al. [63] used hybrid dilated convolution in a ResNet to extract deep features of a vast receptive field without increasing the computational load. Xu et al. [64] used the anti-noise loss function to improve the model’s robustness and used a single-layer dual-channel residual network to classify HSIs.

Gao et al. [65] combined multiple filters to obtain a multiscale residual network to extract multiscale features from HSIs with different receptive fields while reducing the computational load of the network. Dang et al. [66] combined depth-separable convolution with a ResNet structure to shorten the model training time while ensuring high classification accuracy. Paoletti et al. [67] proposed a deep residual network with a pyramid structure, which consisted of multiple residual modules of pyramid bottlenecks that can extract more abstract spatial–spectral features with an increase in the number of layers. Wang et al. [68] introduced the compression incentive mechanism into ResNet and utilized the attention mechanism to extract strongly discriminative features and inhibit nonessential features to enhance the classification accuracy. Zhang et al. [69] designed a fractal residual network to extract spatial–spectral information, enhancing the model’s classification ability. Xu et al. [70] introduced multiscale feature fusion and dilated convolution into a ResNet, which could fit in better with the cubic structure of HSI data and make effective use of the spatial–spectral joint information.

In this study, we constructed 3D multi-shortcut-link networks (MSLNs) on the basis of a 2D ResNet, analyzed its theoretical support in detail, and aimed to solve the problem of matter spectrum disorder and deep network model degradation, where the effectiveness of the MSLN model was demonstrated using experiments.

3. Methods

3.1. Three-Dimensional CNN

A CNN is a feed-forward neural network with a deep structure that includes convolution computation and is one of the most representative algorithms for deep learning. It usually comprises an input layer, a convolution layer, a pooling layer, a fully connected layer, and an output layer. Combining the properties of local connection and weight sharing, it can go through the end-to-end processing procedure, and in 2D images, it can utilize the deep network structure to solve complex classification problems, exhibiting an outstanding performance. However, for 3D data of HSIs, not only is the spatial–spectral information unable to be fully utilized but dimension reduction processing is typically required when using a 2D CNN for HSI classification. This would even add a large number of parameters, which dramatically decreases the computational efficiency and training speed, and more likely causes overfitting for HSIs with fewer labeled samples.

Modified on the basis of a 2D CNN, a 3D CNN [39,70] is mainly applied in video classification, action recognition, and other fields and can perform simultaneous convolution operations in three directions—height, width, and depth—for HSIs without undergoing dimension reduction processing (Figure 1), which can immediately extract spatial–spectral joint high-order features and make full use of spatial–spectral correlation information.

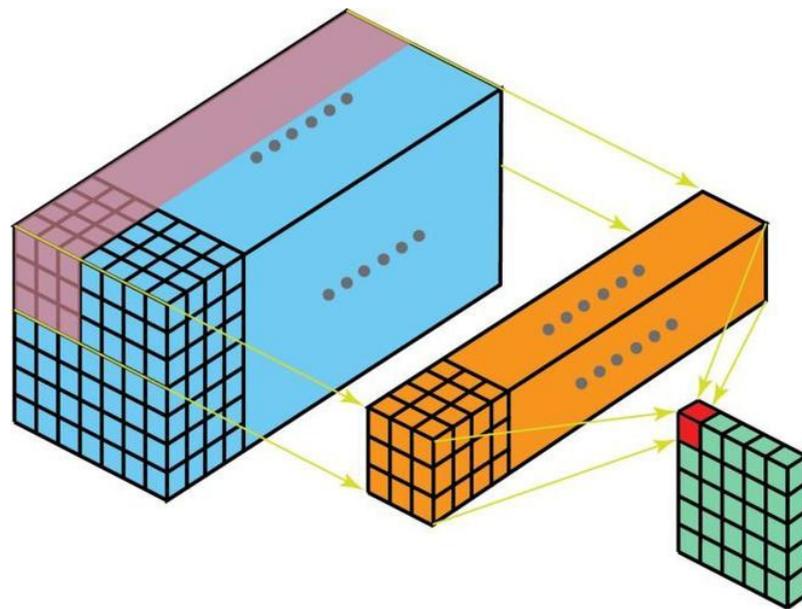


Figure 1. The theory of a 3D CNN.

3.2. Residual Networks (ResNets)

With the deepening of network layers, a 3D CNN is prone to gradient dispersion or gradient explosion. Proper use of regularity initialization and the intermediate regularization layer can deepen the network, but the training set accuracy will be saturated or even decreased. A ResNet [56] alleviates the gradient disappearance problem that occurs in deep neural networks by skipping a connection in the hidden layer.

A ResNet is a hypothesis raised on the basis of identity mapping: assuming a network n with K layers is the currently optimal network, then several layers of a built deeper network should be the identity mapping from the outputs at the K th layer of the network n . The deeper network should not underperform relative to the shallower network. If the input and output dimensions of the network's nonlinear units are consistent, each unit can be expressed as a general formula:

$$y = F(x, \{W_i\}) + x \quad (1)$$

Here, x and y are input and output vectors of layers considered, respectively, and $F(\cdot)$ is a residual function. In Figure 2, there are two layers, i.e., $F = W_2\sigma(W_1x)$, in which σ denotes the ReLU and the biases are omitted for simplifying notations. The process of $F + x$ is an operation of a shortcut connection and element-wise addition.

The residual learning structure (Figure 2) functions by adding the output(s) from the previous layer(s) and the output computed at the current layer and inputting the result of the summation into the activation function as the output of the current layer, which addresses the degeneration of neural networks satisfactorily. A ResNet converges faster under the precondition of the same number of layers.

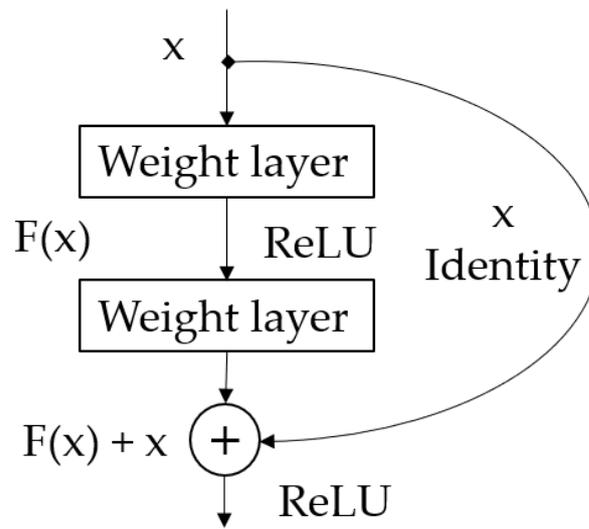


Figure 2. The structure of residual learning.

3.3. Activation Function

Since the gradient of the activation function (Figure 3) of the rectified linear unit (ReLU) [71] is always 0 when the input is a negative value, the ReLU neurons will not be activated after parameters are updated, leading to the “death” of some neurons during the training process.

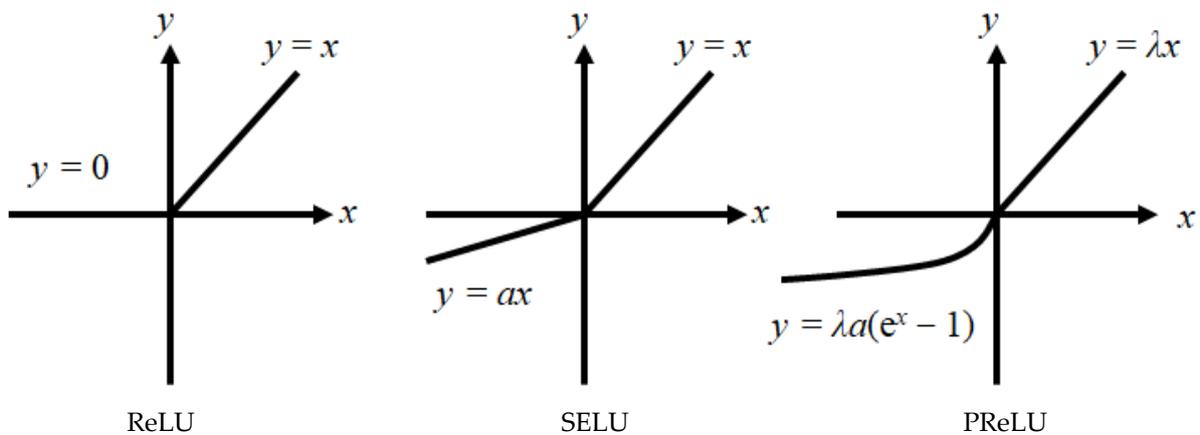


Figure 3. Activation function.

The parametric rectified linear unit (PReLU) [72] is used to address the issue of neuronal death brought about by the ReLU function, where the parameters in it are learned through backpropagation.

The activation function of the self-exponential linear unit (SELU) [73] demonstrates high robustness against noise and enables the mean value of activation of neurons to tend to 0 so that the inputs become fixedly distributed after a certain number of layers.

Therefore, the algorithm in this study used PReLU as the activation function after the shallow multi-shortcut-link networks convolution operation, and SELU as the activation function in the deep residual structure of the block, which can make full use of hyperspectral 3D cube data.

3.4. Loss Function

In deep learning, the softmax function (Equation (2)) is usually used as a classifier, which maps the outputs of multiple neurons into the interval (0, 1). Define the *i*th input feature x_i with label y_i , f_j denotes the *j*th data point ($j \in [1, N]$, N is the number of classes)

of the vector of class scores f , and M is the number of training data. In Equation (2), f is the activations of a fully connected layer W ; $f_{y_i} = W_{y_i}^T x_i$ in which W_{y_i} is the y_i th column of W . Take a dichotomy as an example, i.e., $\|W_1\| \|x\| \cos(\theta_1) > \|W_2\| \|x\| \cos(\theta_2)$, and thus the correct classification result of x is obtained. However, its learning ability is relatively weak for strongly discriminative features. This study adopted large-softmax as the loss function to upgrade the classification accuracy of HSI datasets.

$$L = \frac{1}{M} \sum_i L_i = \frac{1}{M} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right) \tag{2}$$

Large-softmax [74] is a margin-based softmax loss function, i.e., $\|W_1\| \|x\| \cos(\theta_1) \geq \|W_1\| \|x\| \cos(m\theta_1) \geq \|W_2\| \|x\| \cos(\theta_2)$. By adding a positive integer variable m to adjust the required margin, a decision-making margin constraint is added to endow the learned features with the properties of intraclass compactness and interclass separation, as well as to effectively avoid overfitting.

The L-softmax loss function can be defined by the following expression:

$$L_i = -\log \left(\frac{e^{\|W_{y_i}\| \|x_i\| \varphi(\theta_{y_i})}}{\|W_{y_i}\| \|x_i\| \varphi(\theta_{y_i}) + \sum_{j \neq y_i} e^{\|W_j\| \|x_i\| \cos(\theta_j)}} \right) \tag{3}$$

where $\varphi(\theta)$ can be expressed as:

$$\varphi(\theta) = \begin{cases} \cos(m\theta), & 0 \leq \theta \leq \frac{\pi}{m} \\ D(\theta), & \frac{\pi}{m} \leq \theta \leq \pi \end{cases} \tag{4}$$

Experiments demonstrated that the features acquired by L-softmax have more distinctive distinguishability [74,75] and achieve better results than using softmax in both classification and verification tasks.

3.5. Multi-Shortcut-Link Networks (MSLNs)

3.5.1. Analysis of a Multi-Shortcut Link

In a ResNet, the author develops an architecture that stacks building blocks of the same shortcut connecting pattern called “residual units (ResUs).” The original ResU can be computed using these formulas:

$$y_l = h(x_l) + F(x_l, W_l) \tag{5}$$

$$x_{l+1} = f(y_l) \tag{6}$$

Here, x_l is the input for the l th ResU. $W_l = \{W_{l,k} \mid 1 \leq k \leq K\}$ is a set of weights and bias corresponding to the l th ResU, and K is the number of layers in the ResU. F denotes the residual function, e.g., a stack of two 3×3 convolutional layers in a ResNet. This study expanded the convolutional dimension to $3 \times 3 \times 3$. Function f is the process after element-wise addition, then operates the ReLU activation function. Function h is set as the identity mapping: $h(x_l) = x_l$.

This study mainly focused on creating a multi-shortcut-link path for propagating tensor information, not only within ResU but also through the entire network model. As mentioned above, we denote $s(x_0)$ as a shortcut link of the original ResU, which suggests $y_0 = h(x_0) + F(x_0, W_0)$. If f was also used as an identity mapping, that meant $x_{l+1} \equiv y_l$; putting $s(x_0)$ and Equation (2) into Equation (1), and adding a multi-shortcut link S gave

$$x_{l+1} = x_l + F(x_l, W_l) + x_0 + F(x_0, W_0) \tag{7}$$

After recursion, we obtained

$$x_{l+2} = x_{l+1} + F(x_{l+1}, W_{l+1}) = x_l + F(x_l, W_l) + F(x_{l+1}, W_{l+1}) + x_0 + F(x_0, W_0) \quad (8)$$

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) + L(x_0 + F(x_0, W_0)) \quad (9)$$

Equation (9) indicates that for any deeper (L) and shallower (l) unit, the feature x_L of L can be represented as the feature x_l of l plus a residual operation, which is between any L and l presenting as a residual function in an MSLN.

Denoting a loss function as ξ , according to the chain rule of backpropagation, we obtained

$$\frac{\partial \xi}{\partial x_L} = \frac{\partial \xi}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \xi}{\partial x_L} \left(1 + \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F(x_i, W_i) \right) \quad (10)$$

Equation (10) exhibits that the gradient $\frac{\partial \xi}{\partial x_l}$ can be decomposed into two additive terms: $\frac{\partial \xi}{\partial x_L}$ and $\frac{\partial \xi}{\partial x_L} \left(\frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F \right)$ which propagates information directly without any concerned weight layers and with weight layers. The additive term of $\frac{\partial \xi}{\partial x_L}$ guarantees that information is propagated back to any shallower unit l directly. In another way, it implies that because the term $\frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F$ cannot always be -1 , the gradient $\frac{\partial \xi}{\partial x_l}$ is unlikely to be canceled out for a mini-batch. This indicates that even when the weights are extremely small, the gradient of a layer also does not vanish.

This derivation reveals that if we add a shortcut link before a residual block and both $h(x_l)$ and $f(y_l)$ are identity mappings, the feature map signal could be propagated both forward and backward. This indicates that fused shallow features and deep features via multi-shortcut-link networks can be certain to obtain strongly discriminative features, which was also shown in the experiments in Section 4.2.

3.5.2. Structure of an MSLN

Based on the network structure of ResNet 18 (Table 1), this study added a convolution layer preceding each of the 2nd, 6th, 10th, and 14th layers, which is spliced in depth with the output result of the previous layer and as the input of the next convolution layer. Meanwhile, the 3D convolution operation of the original HSI cube block ($H \times W \times C$) contains the 1st, 2nd, 7th, 12th, and 17th layers in the MSLN, which implies the input shape of all these five layers are (batch size, input data, channels of HSI, kernel size, stride), and the numbers of convolution kernels were set according to Table 1, which were 16, 16, 16, 32, and 64 (Figure 4), respectively.

In order to make the MSLN more convenient and minimize the conflicts when splicing channels as much as possible, the number of convolution kernels in the block was set to 16, 32, 64, and 128, respectively. Compared with the number of channels in the ResNet (64, 128, 256, 512), the total size of parameters in the ResNet (130.20 ± 0.65 MB) was substantially decreased (10.87 ± 0.27 MB) and were greatly improved regarding the convergence speed.

As shown in Figure 5, conv i_{-1} ($i = 1, 2, 3, 4$) output were shallow features, and the feature graph had a higher resolution, which could retain more feature information and better describe the overall characteristics of the data. As the depth of the network increased, the deep features became more and more abstract. Fusing shallow features and deep features via multi-shortcut-link networks could reduce the loss of shallow features and correlation decay of gradients, boost the use ratio of features, and enhance the network's expressiveness.

Table 1. Convolutional layer of MSLN 22 and ResNet 18.

Layer_Name	18-Layer ResNet Kernel_Size Kernel_Number Stride	Layer_Name	22-Layer MSLN Kernel_Size Kernel_Number Stride
conv 1	7 × 7 64 S2	conv 1	3 × 3 × 3 16 S1
Block1	$\begin{bmatrix} 3 \times 3 & 64 & S1 \\ 3 \times 3 & 64 & S1 \end{bmatrix} \times 2$	conv 1_1	3 × 3 × 3 16 S1
Block2	$\begin{bmatrix} 3 \times 3 & 128 & S1 \\ 3 \times 3 & 128 & S1 \end{bmatrix} \times 2$	Block1	$\begin{bmatrix} 3 \times 3 \times 3 & 16 & S1 \\ 3 \times 3 \times 3 & 16 & S1 \end{bmatrix} \times 2$
Block3	$\begin{bmatrix} 3 \times 3 & 256 & S1 \\ 3 \times 3 & 256 & S1 \end{bmatrix} \times 2$	conv 2_1	3 × 3 × 3 16 S1
Block4	$\begin{bmatrix} 3 \times 3 & 512 & S1 \\ 3 \times 3 & 512 & S1 \end{bmatrix} \times 2$	Block2	$\begin{bmatrix} 3 \times 3 \times 3 & 32 & S1 \\ 3 \times 3 \times 3 & 32 & S1 \end{bmatrix} \times 2$
		conv 3_1	3 × 3 × 3 32 S1
		Block3	$\begin{bmatrix} 3 \times 3 \times 3 & 64 & S1 \\ 3 \times 3 \times 3 & 64 & S1 \end{bmatrix} \times 2$
		conv 4_1	3 × 3 × 3 64 S1
		Block4	$\begin{bmatrix} 3 \times 3 \times 3 & 128 & S1 \\ 3 \times 3 \times 3 & 128 & S1 \end{bmatrix} \times 2$
	Average pool 1000-d fc l-softmax		Average pool 128-d fc l-softmax

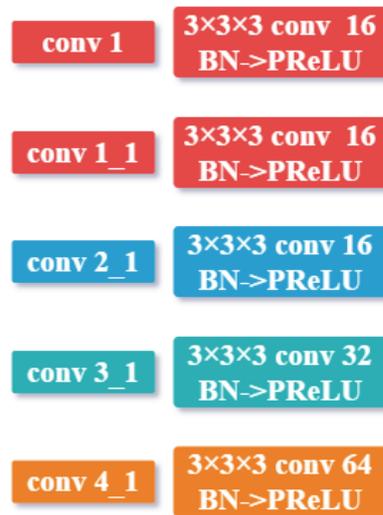


Figure 4. Parameter settings of the multi-shortcut-link convolution kernel.

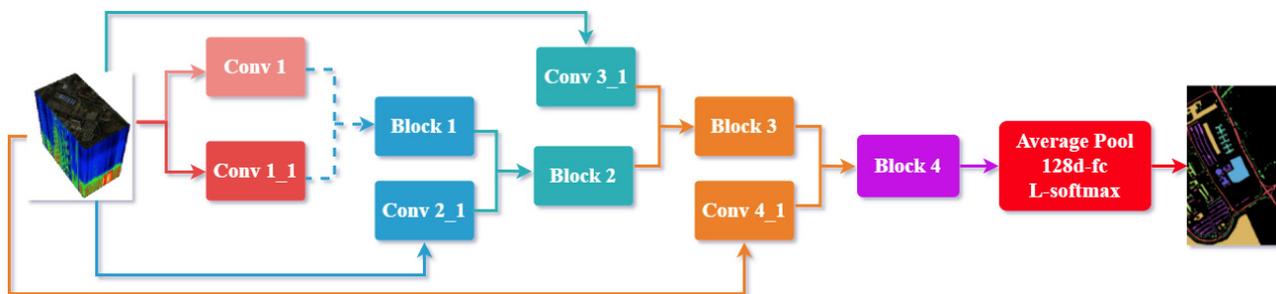


Figure 5. The flowchart of the MSLN 22 HSI classification framework.

Therefore, splicing the shallow feature $conv\ i_1$ ($i = 1, 2, 3, 4$) in depth with the output of each residual block ($conv\ j_1$ ($j = 1, 2, 3, 4$)) to implement the multi-shortcut-link network fusion of features across different network layers could better alleviate gradient dispersion (explosion) and even network degeneration.

Figure 4 displays the overall process of the HSI classification framework of the MSLN. It can be seen from this figure that the multi-shortcut-link network's structure is bridged to four residual blocks (Figure 6) for a total of four times (Figure 7); then, the last output layer of the third residual block is spliced with conv4_1 as the input tensor for the first layer of the fourth residual block; after the fourth residual block is processed, the global self-adaptive average pooling downsampling is used to expand the output tensor into one-dimensional vectors via the fully connected layer to map the learned distributed features into the space of sample labels; finally, the large-softmax loss function is used for classification.

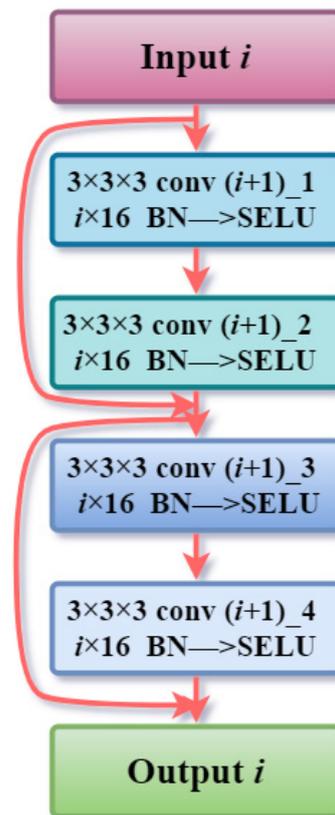


Figure 6. The structure of Block i ($i = 1, 2, 3, 4$).

In this study, all the convolution kernels adopted the uniform size $3 \times 3 \times 3$, which could both reduce the computational load and enlarge the receptive field of convolution operation [56].

Figure 8 shows the visualization of the MSLN structure and training process using the Botswana dataset, which covers 145 wavebands. The gray elements in the graph indicate the node is a backward operation, and the light blue elements indicate the node is a tensor that is input/output.

At the top of Figure 8, there is a tensor whose shape is $(64, 1145, 3, 3)$, which means the batch size processed in the model was 64 and the input 3D tensor of the MSLN model was $(145, 3, 3)$. There were five sets of bias matrices and weight matrices whose background color is light blue in the second row, corresponding to the first convolutional layer and four shortcut link convolutional layers, and the names of the rectangles are arranged in order conv1, add 1.0, add 2.0, add 3.0, and add 4.0. There are four residual blocks in the MSLN structure referred to as layer x ($x = 2, 3, 4$) and each layer contains four convolution layers, which were named layer x .0.conv1, layer x .0.conv2, layer x .1.conv1, and layer x .1.conv2 ($x = 1, 2, 3, 4$), and the downsampling operation was mainly done to manage the problem of an inconsistent number of convolution kernel channels. At the bottom of this figure, there is a green element, which is the final output of the results of the HSI classification.

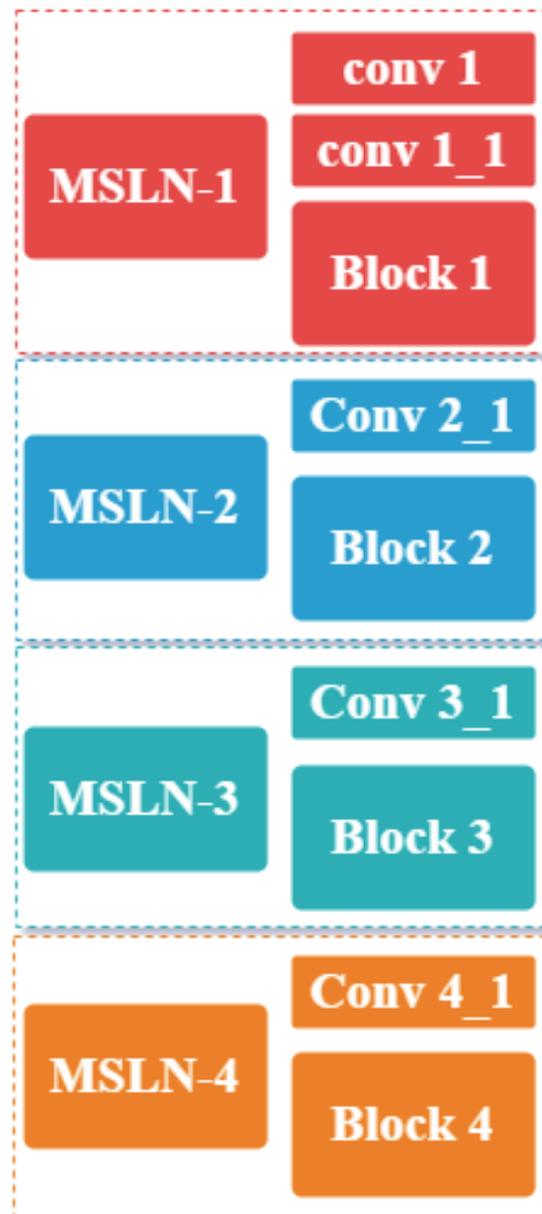


Figure 7. The structure of the MSLN.

Deep features are abstract and associated with a small receptive field. When the shallow features of a vast receptive field are mapped into the space of abstract features of a small receptive field, the number of parameters will grow with the increase in the number of layers, which will lead to an increased computational load, as well as a large loss of shallow representation information. The multi shortcut link networks structure proposed in this study, combined with the ResNet, can make up for the information loss of the deep network's shallow features very well and better address the difficulty in learning deep abstract features.

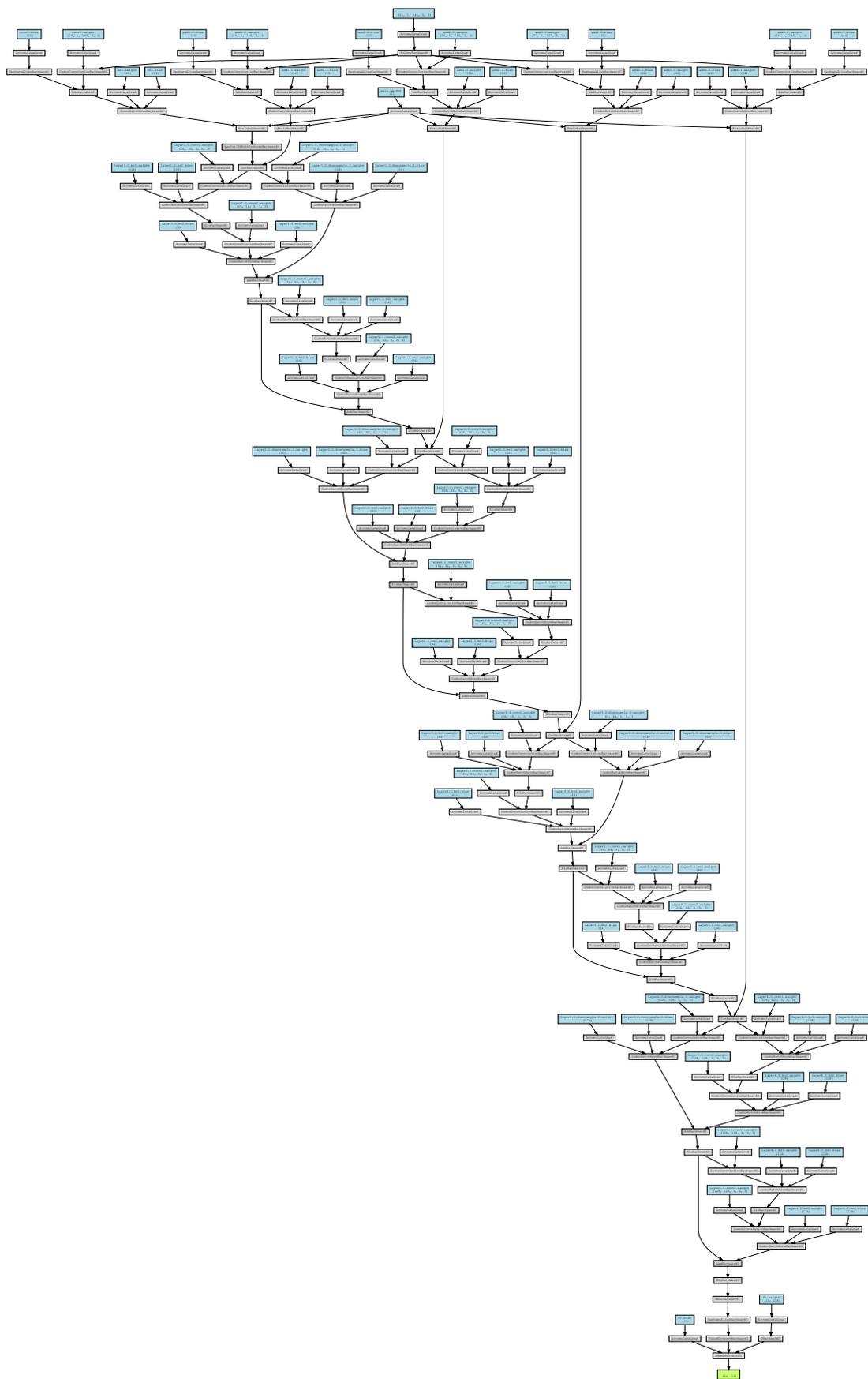


Figure 8. Network structure/training process visualization.

4. Datasets Results and Analysis

The MSLN proposed in this study was based on the Python language and PyTorch deep learning framework, with the test environment being a Windows 10 OS with 32 GB RAM, an Intel i7-8700 CPU, and an NVIDIA Quadro P1000 4 GB GPU.

4.1. Hyperspectral Test Datasets

To validate the robustness and generalization property of the proposed algorithm, six groups of opensource datasets collected by M Graña et al. were used [76] to learn all types of labeled terrain objects without undergoing any human screening, and the ratio of training sets to validation sets to test sets was 0.09:0.01:0.9 (Tables 2–7 for details).

Table 2. The information of samples in the Indian Pines dataset.

Sample No.	Class	Train	Validation	Test
1	Alfalfa	4	1	41
2	Corn-notill	129	14	1285
3	Corn-mintill	75	8	747
4	Corn	22	2	213
5	Grass-pasture	43	5	435
6	Grass-trees	43	7	680
7	Grass-pasture-mowed	3	1	24
8	Hay-windrowed	43	5	430
9	Oats	3	1	16
10	Soybean-notill	87	10	875
11	Soybean-mintill	220	25	2210
12	Soybean-clean	53	6	534
13	Wheat	18	2	185
14	Woods	113	13	1139
15	Buildings-grass-trees-drives	35	4	347
16	Stone-steel-towers	8	1	84
Total		899	105	9245

Table 3. The information of samples in the Salinas dataset.

Sample No.	Class	Train	Validation	Test
1	Brocoli_green_weeds_1	181	20	1808
2	Brocoli_green_weeds_2	355	37	3334
3	Fallow	177	20	1779
4	Fallow_rough_plow	125	14	1255
5	Fallow_smooth	241	27	2410
6	Stubble	357	39	3563
7	Celery	322	36	3221
8	Grapes_untrained	1014	113	10,144
9	Soil_vinyard_develop	558	62	5583
10	Corn_senesced_green_weeds	292	33	2953
11	Lettuce_roumaine_4wk	96	11	961
12	Lettuce_roumaine_5wk	171	19	1737
13	Lettuce_roumaine_6wk	81	9	826
14	Lettuce_roumaine_7wk	96	11	963
15	Vineyard_untrained	646	71	6551
16	Vineyard_vertical_trellis	157	17	1633
Total		4869	539	48,721

Table 4. The information of samples in the Pavia Centre dataset.

Sample No.	Class	Train	Validation	Test
1	Water	5925	657	59,251
2	Trees	684	76	6838
3	Asphalt	277	31	2736
4	Self-blocking bricks	241	27	2417
5	Bitumen	592	66	5924
6	Tiles	833	92	8317
7	Shadows	656	73	6558
8	Meadows	3842	425	38,415
9	Bare soil	257	29	2577
	Total	13,307	1476	133,033

Table 5. The information of samples in the Pavia University dataset.

Sample No.	Class	Train	Validation	Test
1	Asphalt	593	65	5973
2	Meadows	1674	186	16,789
3	Gravel	188	21	1890
4	Trees	275	31	2758
5	Painted metal sheets	121	13	1211
6	Bare Soil	453	50	4526
7	Bitumen	120	13	1197
8	Self-blocking bricks	331	37	3314
9	Shadows	85	10	852
	Total	3840	426	38,510

Table 6. The information of samples in the Kennedy Space Center dataset.

Sample No.	Class	Train	Validation	Test
1	Scrub	68	8	685
2	Willow-swamp	22	2	219
3	Cabbage palm hammock	23	3	230
4	Cabbage palm/oak hammock	22	3	227
5	Slash pine	14	2	145
6	Oak/broadleaf hammock	21	2	206
7	Hardwood swamp	10	1	94
8	Graminoid marsh	39	4	388
9	Spartina marsh	47	5	468
10	Cattail marsh	36	4	364
11	Salt marsh	38	4	377
12	Mud flats	45	5	453
13	Wate	83	10	834
	Total	468	53	4690

Table 7. The information of samples in the Botswana dataset.

Sample No.	Class	Train	Validation	Test
1	Water	24	3	243
2	Hippo grass	9	1	91
3	Floodplain grasses 1	23	2	226
4	Floodplain grasses 2	19	2	194
5	Reeds	24	3	242
6	Riparian	24	3	242
7	Firescar	23	3	233
8	Island interior	18	2	183
9	Acacia woodlands	28	3	283
10	Acacia shrublands	23	2	233
11	Acacia grasslands	27	3	275
12	Short mopane	16	2	163
13	Mixed mopane	24	3	241
14	Exposed soils	9	1	85
	Total	291	33	2934

(1) Indian Pines (IP) Dataset

The IP dataset was acquired using AVIRIS sensors on the test site in Indiana in 1996 with an image resolution of 145×145 pixels, a spectral range of 0.4–2.45 μm , and a spatial resolution of 20 m. There remained 200 effective wavebands for classification after the wavebands affected by noises and suffering severe water vapor absorption were eliminated, and a total of 16 crops were labeled. This dataset was shot in June, when some crops, such as corn and soybean, were in the early growth stage. The coverage rate of less than 5% was prone to pixel mixture, leading to a significant increase in the difficulty of vegetative terrain classification.

(2) Salinas (S) Dataset

The S dataset was shot in Salinas Valley, California, using AVIRIS sensors with an image resolution of 512×217 pixels, a spectral range of 0.43–0.86 μm , and a spatial resolution of 3.7 m. There remained 204 effective wavebands for classification after the wavebands affected by noises and suffering severe water vapor absorption were eliminated, and a total of 16 crops were labeled, covering vegetables, bare soils, vineyards, etc.

(3) Pavia Centre (PC) and Pavia University (PU) Datasets

The PC and PU datasets stemmed from two scenes were captured using ROSIS sensors during a flight over Pavia in northern Italy, with 102 and 103 wavebands remaining, respectively, after the noise-affected wavebands and information-free regions were eliminated; with image resolutions of 1096×715 pixels and 610×340 pixels, respectively; and a spatial resolution of 1.3 m; both images have nine classes of labeled terrain objects, though the categories are not fully congruent.

(4) Kennedy Space Center (KSC) Dataset

The KSC dataset was shot using AVIRIS sensors at Kennedy Space Center, Florida, on March 23, 1996, with an image resolution of 512×614 pixels, a spectral range of 0.4–2.5 μm , and a spatial resolution of 18 m. There remained a total of 176 wavebands for analysis after the wavebands suffering water vapor and noises were eliminated, and a total of 13 classes of terrain objects. The extremely low spatial resolution plus the similarity in the spectral signatures of some vegetation types provide considerably increased difficulty regarding terrain classification.

(5) Botswana (B) Dataset

The B dataset was shot using Hyperion sensors at Okavango Delta in Botswana with an image resolution of 1476×256 pixels, a spectral range of 0.4–2.5 μm , and a spatial resolution of 30 m, covering 242 wavebands in total. The UT Space Research Center eliminated the uncalibrated and noise-affected wavebands covered with moisture absorption features, leaving a total of 145 wavebands for classification, including 14 observation data whose categories were determined, and these categories include the seasonal and sporadic swamps, dry forests, and other land cover types in the delta.

The IP dataset and S dataset both contain 6 major categories and 16 sub-categories; it is necessary to improve the discrimination between classes to improve the classification accuracy. However, the IP dataset has a lower resolution; if the training data is selected according to the above ratio, four terrain objects have no training data (alfalfa, grass-pasture-mowed, oats, and stone-steel-towers). The same situation is also reflected in the KSC dataset (hardwood swamp) and B dataset (hippo grass and exposed soils), mainly due to too few labeled samples. To reduce the validation error and test error, one sample was randomly selected among the training samples for validation.

The PC dataset and PU dataset have fewer categories, with a higher resolution and richer labeled samples, but have the issue of “distinct objects with the same spectrum.” There are 7 swamp types out of the 13 classes of terrain objects in the KSC dataset, which has the issue of “the same object with distinct spectra” and considerably increased difficulty regarding terrain classification.

In this study, the MSLN structure in the ResNet and some hyperparameter settings could resolve the above problems and improve the classification accuracy of the six groups of the datasets.

4.2. Results and Analysis

To validate the effectiveness and classification results of the MSLN 22 network structure, an ordinary 3D CNN [37] was selected as a baseline network for comparative analysis with an RNN [32], a multiscale 3D CNN (MC 3D CNN) [44], a 3D CNN residual (3D CNN Res) [41], and a 3D ResNet. The settings of the hyperparameters were as follows: batch size was set to 64, which not only mitigated gradient oscillation but also allowed for better performance of the GPU; the initial learning rate was set to 0.01 and the learning rate dropped to 0.001 after the loss function was stabilized; the maximum count of iterations epoch was set to 600; the cross-entropy was selected as the loss function in the comparison algorithm, whereas L-softmax was adopted as the loss function in the network produced in this study; the dropout was set to 0.5 after the global self-adaptive average pooling and before the fully connected layer; and 50% of the neurons were discarded at random to address the overfitting issue and enhance the model’s generalization property.

Tables 8–13 present the classification results corresponding to the six datasets. The multi-shortcut-link networks structure of the MSLN proposed in this study could extract spatial–spectral joint features and fuse shallow and deep features, and the evaluation criterion kappa coefficient, average accuracy (AA), and overall accuracy (OA) were all the highest among the networks tested. Figures 9–14 each represent the comparison of the HSI classification results among all network structures.

Tables 8–10 and Figures 9–11 indicate that the MSLN, which had 22 convolution layers, could extract rich deep features, and the multi-shortcut-link network’s structure also fused the shallow features of a vast receptive field and mitigated the effect of noises via the global self-adaptive average pooling layer, achieving a significant improvement and exhibiting good robustness in the classification results, with the average accuracy of classification across all terrain objects reaching 98.1% (IP dataset), 99.4% (S dataset), and 99.3% (B dataset), including the terrain objects that share similar attributes (untrained grapes, untrained vineyard, and vineyard vertical trellis) and relatively few samples in the IP dataset (alfalfa, grass-pasture-mowed, oats, and stone-steel-towers) and the B dataset (hippo grass and exposed soils).

Table 8. The classification results from using different methods on the Indian Pines dataset.

Indian Pines	3D CNN	RNN	MC 3D CNN	3D CNN Res	3D ResNet	MSLN
1	0.656	0.436	0.543	0.000	0.086	0.978
2	0.563	0.646	0.815	0.454	0.491	0.982
3	0.626	0.424	0.591	0.555	0.313	0.961
4	0.543	0.362	0.825	0.416	0.269	0.947
5	0.710	0.821	0.868	0.361	0.664	0.996
6	0.916	0.894	0.969	0.857	0.878	0.995
7	0.864	0.323	0.700	0.343	0.077	1.000
8	0.907	0.939	0.968	0.939	0.912	0.996
9	0.000	0.538	0.944	0.000	0.329	1.000
10	0.586	0.566	0.762	0.640	0.068	0.979
11	0.433	0.670	0.808	0.709	0.175	0.986
12	0.535	0.599	0.731	0.339	0.500	0.981
13	0.952	0.956	1.000	0.921	0.930	1.000
14	0.924	0.928	0.952	0.843	0.847	0.987
15	0.582	0.575	0.665	0.532	0.429	0.912
16	0.794	0.852	0.903	0.839	0.880	0.989
Kappa	0.601	0.655	0.789	0.610	0.449	0.974
AA	0.662	0.658	0.815	0.547	0.491	0.981
OA (%)	64.466	69.908	81.615	66.016	50.331	97.698

Table 9. The classification results from using different methods on the Salinas dataset.

Salinas	3D CNN	RNN	MC 3D CNN	3D CNN Res	3D ResNet	MSLN
1	0.983	0.997	0.976	0.995	0.991	0.993
2	1.000	0.998	0.999	0.995	1.000	1.000
3	0.997	0.983	0.988	0.985	0.995	1.000
4	0.997	0.988	0.992	0.996	0.996	0.999
5	0.996	0.981	0.996	0.989	0.994	1.000
6	0.996	0.999	0.992	0.999	1.000	1.000
7	0.993	0.998	0.993	0.996	0.998	1.000
8	0.920	0.778	0.905	0.878	0.911	0.985
9	0.997	0.987	0.996	0.997	0.998	1.000
10	0.977	0.963	0.955	0.955	0.973	0.995
11	0.972	0.975	0.957	0.946	0.978	0.998
12	0.985	0.985	0.978	0.991	0.987	0.997
13	0.987	0.976	0.988	0.977	0.992	0.998
14	0.971	0.948	0.977	0.965	0.987	0.994
15	0.875	0.239	0.826	0.798	0.869	0.975
16	0.940	0.994	0.913	0.986	0.964	0.972
Kappa	0.947	0.858	0.929	0.933	0.950	0.987
AA	0.974	0.924	0.964	0.966	0.977	0.994
OA (%)	95.225	87.401	93.643	94.012	95.503	98.851

Table 10. The classification results from using different methods on the Botswana dataset.

Botswana	3D CNN	RNN	MC 3D CNN	3D CNN Res	3D ResNet	MSLN
1	0.998	1.000	1.000	1.000	1.000	1.000
2	0.989	0.978	0.984	0.906	0.984	1.000
3	0.998	0.971	0.980	0.915	0.736	1.000
4	0.956	0.904	0.950	0.754	0.831	1.000
5	0.852	0.836	0.876	0.719	0.786	0.985
6	0.808	0.760	0.795	0.610	0.844	0.974
7	0.994	0.991	0.996	0.987	0.969	1.000
8	0.986	0.958	0.984	0.893	0.894	1.000
9	0.896	0.760	0.898	0.765	0.906	0.987
10	0.796	0.942	0.984	0.866	0.332	0.972
11	0.862	0.955	0.995	0.965	0.975	0.980
12	0.905	1.000	0.991	0.981	0.832	1.000
13	0.897	0.998	0.994	0.919	0.673	1.000
14	0.988	0.951	0.944	0.982	0.944	1.000
Kappa	0.907	0.914	0.948	0.855	0.822	0.991
AA	0.923	0.929	0.955	0.876	0.836	0.993
OA (%)	91.382	92.100	95.212	86.662	83.550	99.138

Table 11. The classification results from using different methods on the Pavia Centre dataset.

Pavia Centre	3D CNN	RNN	MC 3D CNN	3D CNN Res	3D ResNet	MSLN
1	0.999	0.996	0.994	0.999	0.997	0.998
2	0.961	0.985	0.972	0.980	0.990	0.996
3	0.893	0.950	0.911	0.941	0.961	0.979
4	0.834	0.980	0.965	0.950	0.977	0.999
5	0.949	0.993	0.991	0.985	0.990	0.999
6	0.961	0.988	0.985	0.982	0.988	0.997
7	0.945	0.988	0.979	0.978	0.988	0.999
8	0.996	0.996	0.994	0.999	0.998	0.998
9	0.990	0.998	0.998	1.000	1.000	1.000
Kappa	0.976	0.985	0.977	0.989	0.990	0.993
AA	0.948	0.986	0.977	0.979	0.988	0.996
OA (%)	98.285	98.944	98.376	99.250	99.259	99.540

Table 12. The classification results from using different methods on the Pavia University dataset.

Pavia University	3D CNN	RNN	MC 3D CNN	3D CNN Res	3D ResNet	MSLN
1	0.900	0.974	0.972	0.969	0.983	0.992
2	0.952	0.968	0.956	0.975	0.976	0.982
3	0.678	0.940	0.948	0.938	0.943	0.983
4	0.938	0.983	0.981	0.936	0.981	0.994
5	0.999	0.995	0.999	1.000	0.999	0.999
6	0.880	0.991	0.987	0.962	0.985	0.998
7	0.728	0.961	0.964	0.911	0.964	0.992
8	0.771	0.961	0.978	0.948	0.964	0.991
9	0.994	0.995	0.995	0.995	0.999	1.000
Kappa	0.865	0.945	0.933	0.949	0.958	0.973
AA	0.871	0.974	0.976	0.959	0.977	0.992
OA (%)	89.839	95.784	94.852	96.151	96.803	97.961

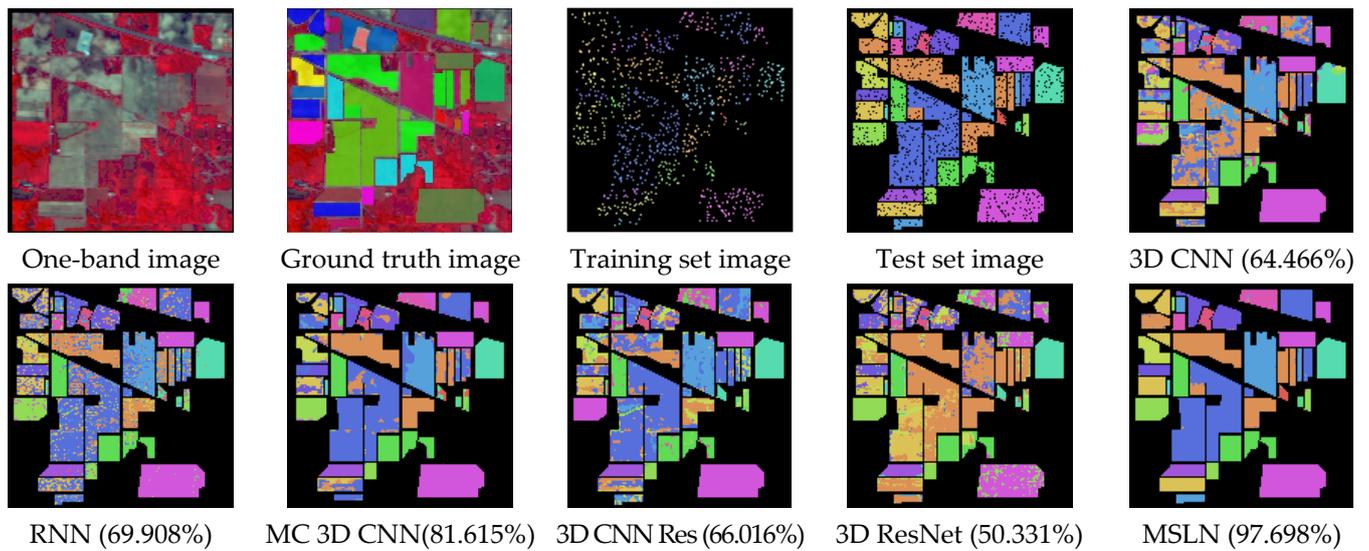


Figure 9. Classification maps and overall accuracy from using different methods on the Indian Pines dataset.

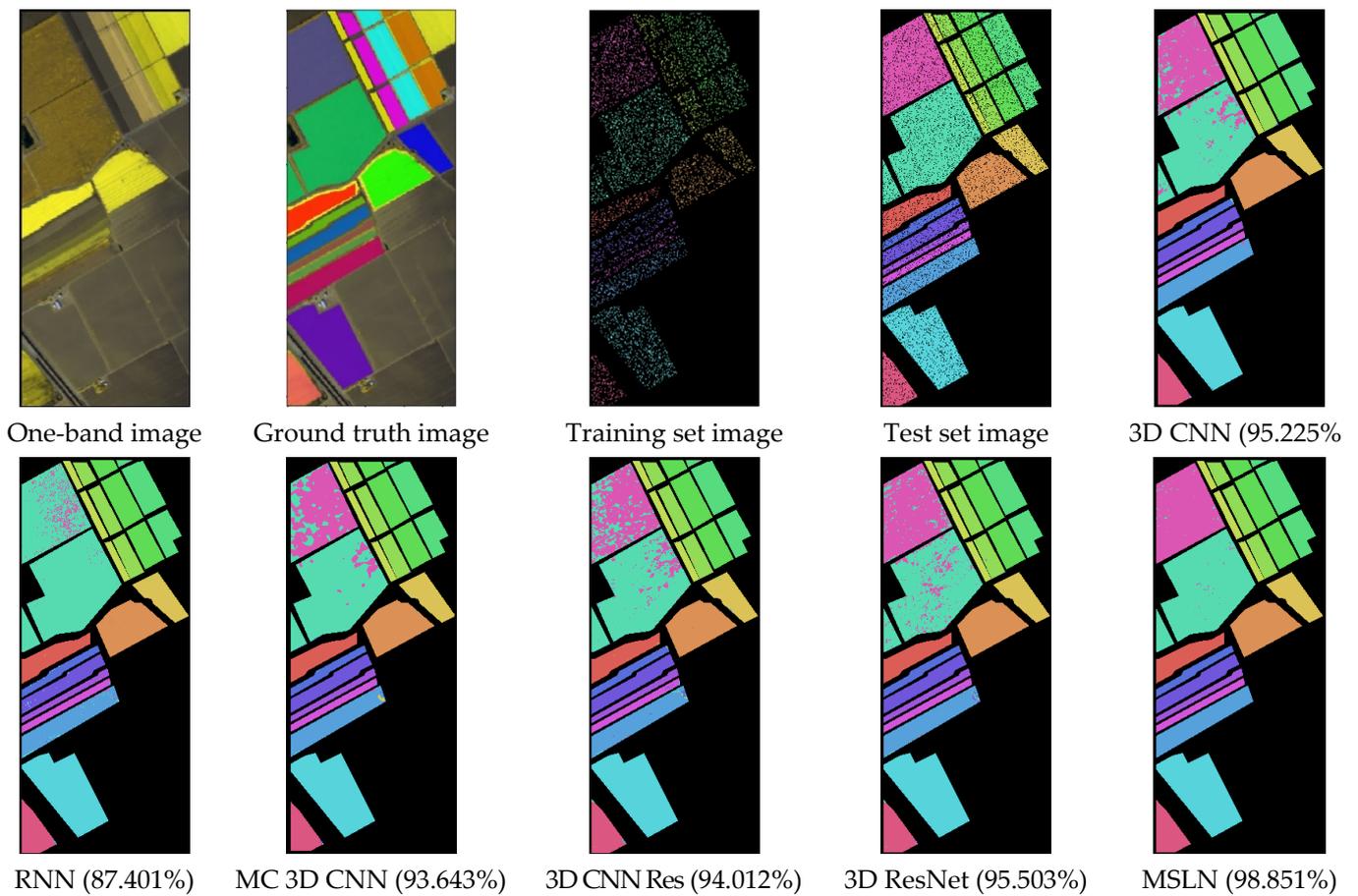
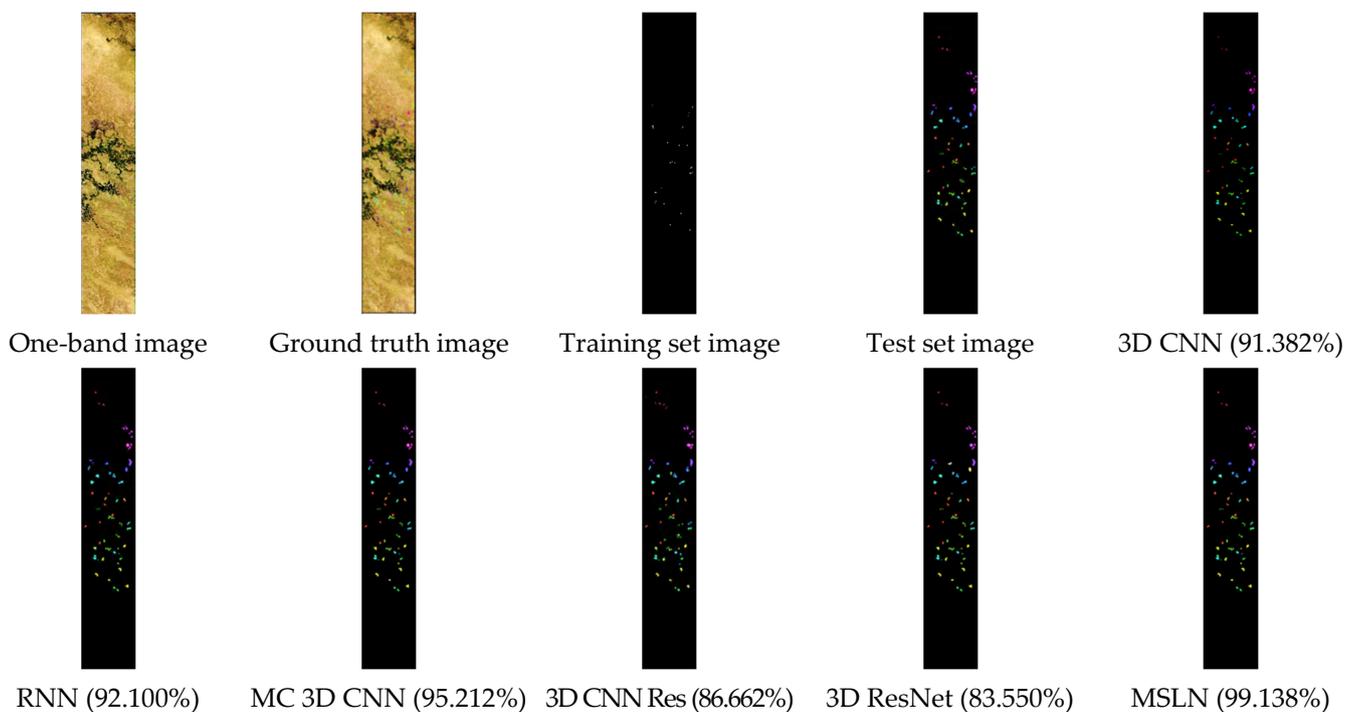


Figure 10. Classification maps and overall accuracy from using different methods on the Salinas dataset.

Table 13. The classification results from using different methods on the KSC dataset.

KSC	3D CNN	RNN	MC 3D CNN	3D CNN Res	3D ResNet	MSLN
1	0.065	0.536	0.935	0.917	0.894	0.988
2	0.000	0.000	0.885	0.636	0.822	0.980
3	0.526	0.000	0.779	0.539	0.947	0.976
4	0.000	0.009	0.504	0.277	0.562	0.886
5	0.000	0.000	0.671	0.400	0.460	0.928
6	0.000	0.000	0.694	0.567	0.503	0.894
7	0.000	0.000	0.842	0.000	0.662	0.960
8	0.000	0.224	0.868	0.833	0.889	0.979
9	0.000	0.000	0.935	0.922	0.966	0.998
10	0.275	0.187	0.853	0.785	0.848	0.988
11	0.000	0.582	0.971	0.901	0.924	1.000
12	0.588	0.361	0.829	0.740	0.770	0.978
13	0.409	0.743	0.965	0.931	0.954	0.999
Kappa	0.200	0.360	0.852	0.759	0.827	0.974
AA	0.143	0.203	0.825	0.650	0.785	0.966
OA (%)	31.748	44.542	86.652	78.401	84.478	97.698

**Figure 11.** Classification maps and overall accuracy from using different methods on the Botswana dataset.

The PC and PU datasets each have nine disjoint classes of terrain objects with excellent connectivity and whose attributes are simple, while the connectivity is slightly inferior for the terrain objects asphalt, shadows, gravel, and bare soil, which are exposed to more noise, highlighting the issue of “distinct objects with the same spectrum” and even resulting in as low as a 67.8% classification accuracy for some algorithms. Through the double constraints of shallow and deep features, the algorithm produced in this study could improve the classification results so that the classification accuracies for the above terrain objects were 97.9%, 99.9%, 98.3%, and 99.8%, respectively, and that the object–spectrum confounding issue was addressed effectively.

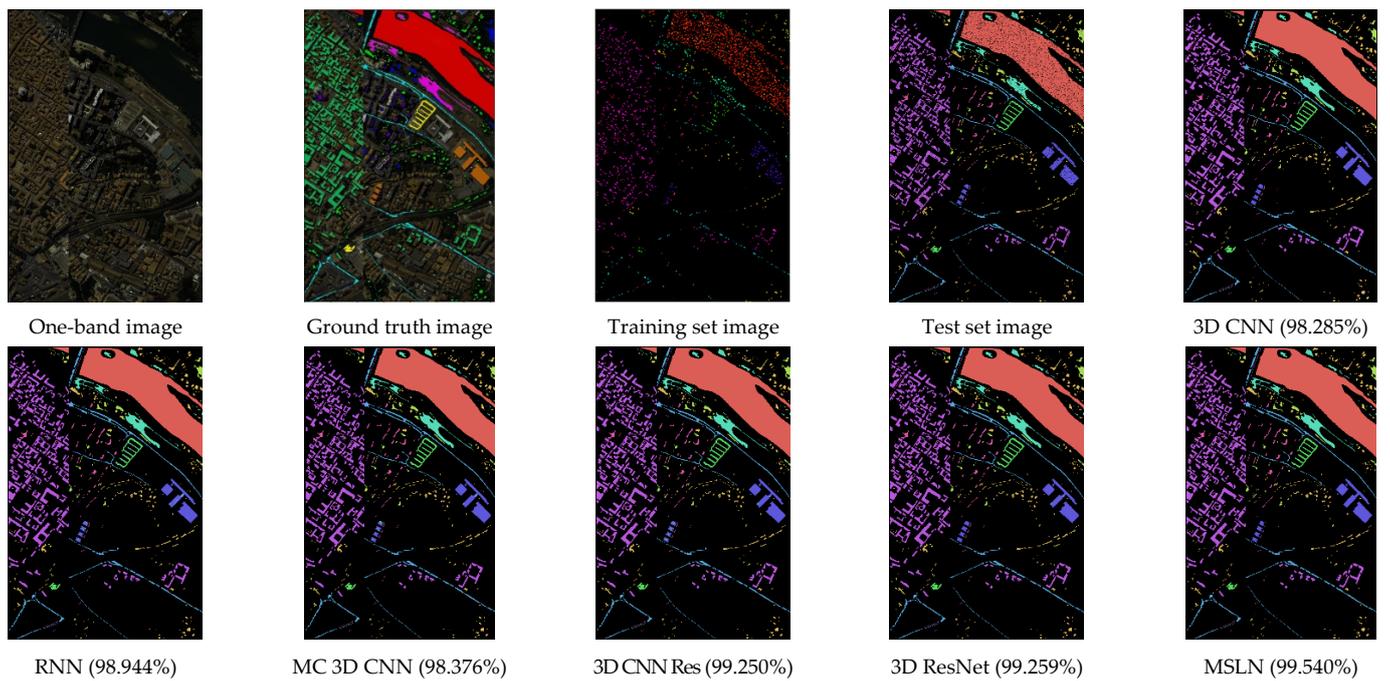


Figure 12. Classification maps and overall accuracy from using different methods on the Pavia Centre dataset.

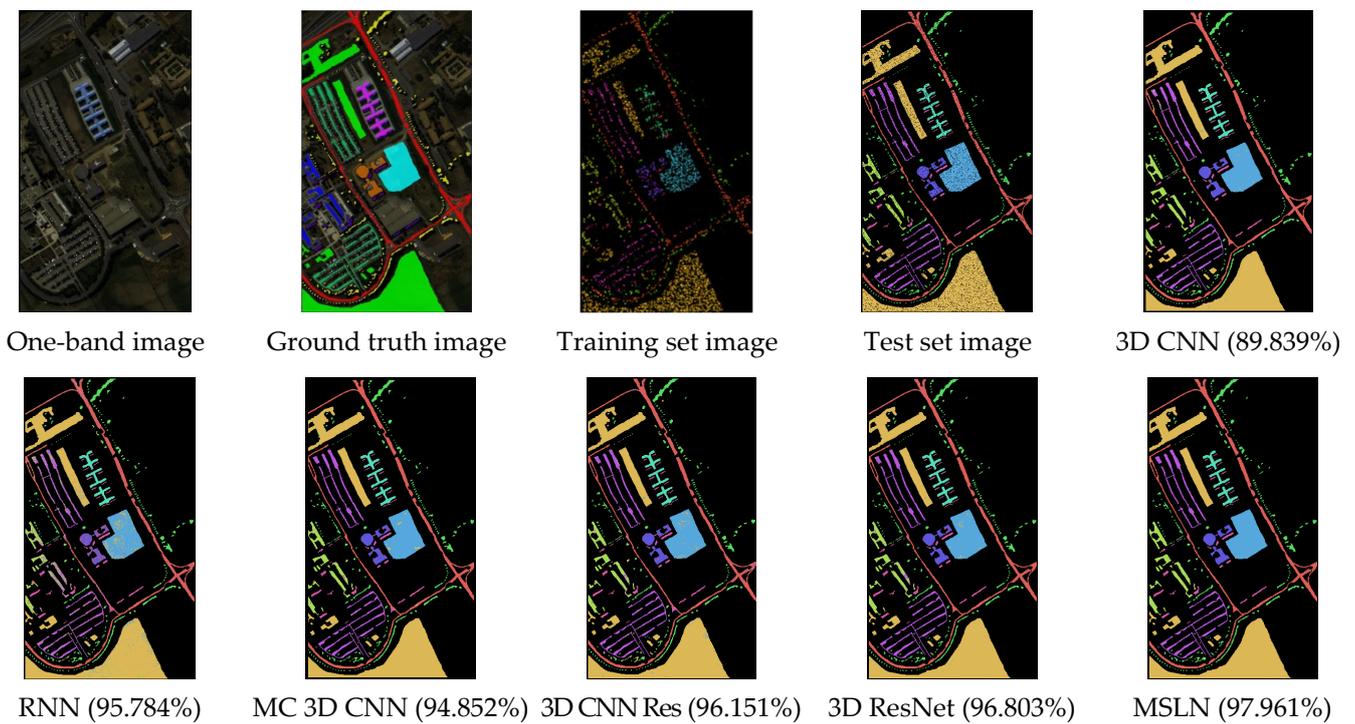


Figure 13. Classification maps and overall accuracy from using different methods on the Pavia University dataset.

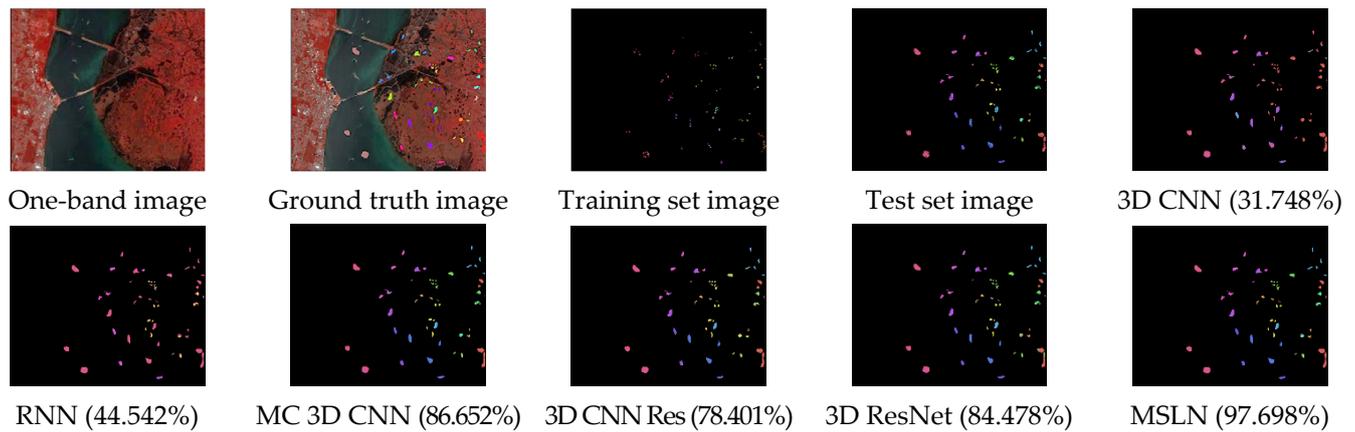
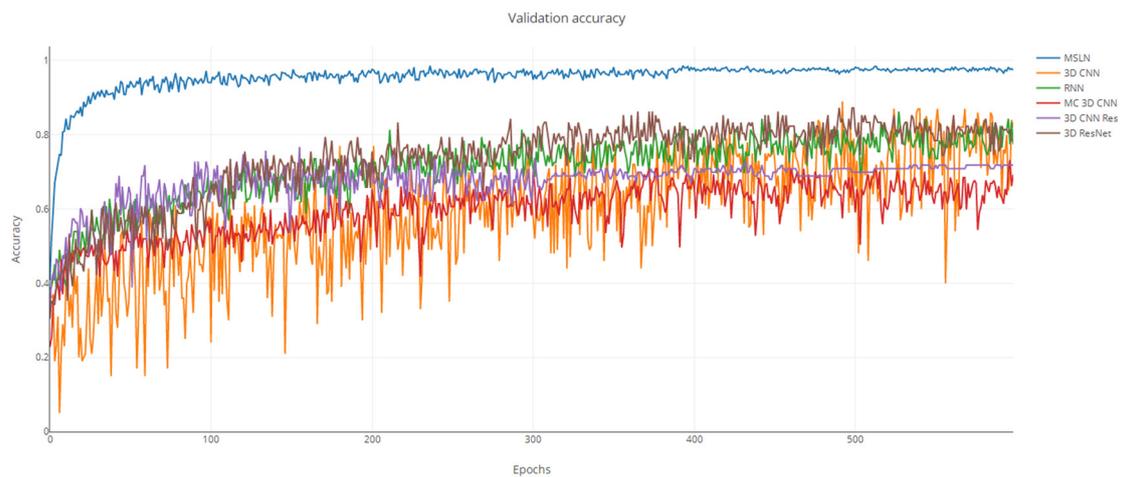


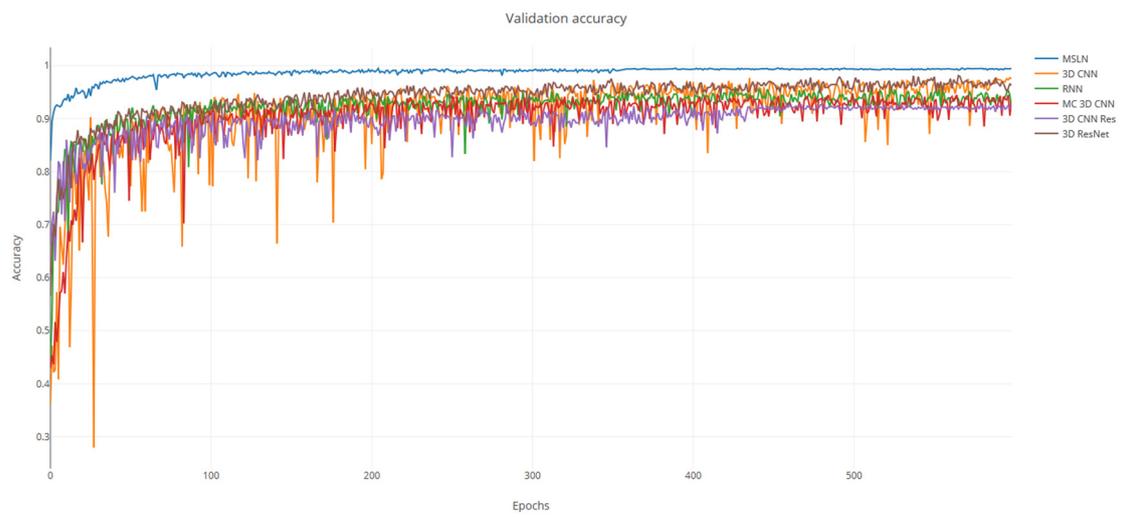
Figure 14. Classification maps and overall accuracy from using different methods on the KSC dataset.

The KSC dataset has 7 swamp types out of the 13 classes of terrain objects, which presented increased considerable difficulty regarding terrain classification, leading to all comparison algorithms being unable to effectively distinguish the discriminative features of different types of swamps during learning, mainly because of the “the same object with distinct spectra” circumstance being very obvious for swamps; therefore, the validation accuracies of all comparison algorithms shown in Figure 15e were highly unstable. According to the classification results, the fused features extracted via multi-shortcut-link networks by the algorithm produced in this study could express the abstract attributes of “the same object with distinct spectra” very well. The accuracy of classification results was 96% at minimum for the swamp type of terrain objects, and the classification results for salt marsh were all correct. However, the classification results remained not very high for cabbage palm and oak hammock, as well as oak and broadleaf hammock, which have extremely similar spectral signatures and inferior connectivity, in addition to more noise in the datasets such that the classification accuracies were only 88.6% and 89.4%, respectively.

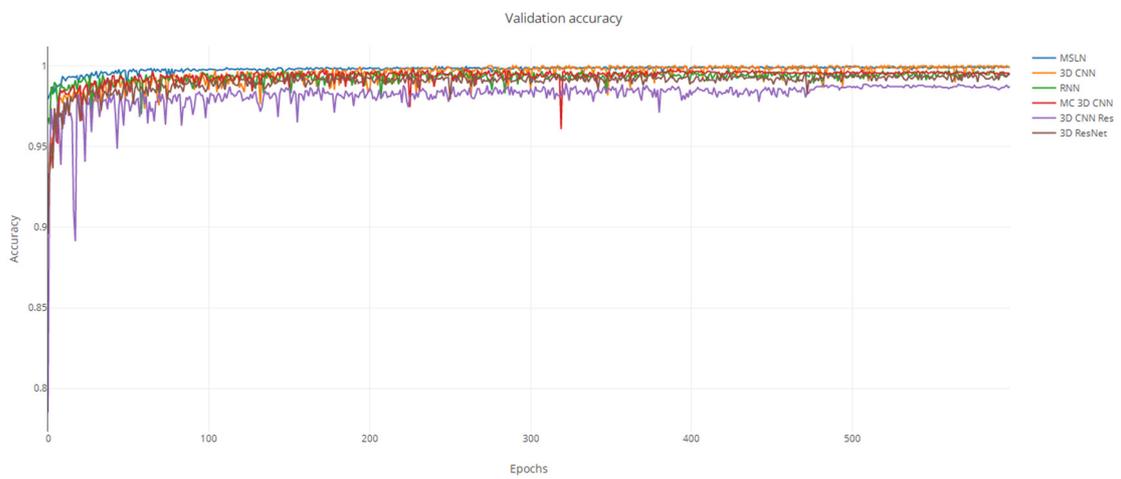


(a) Indian Pines

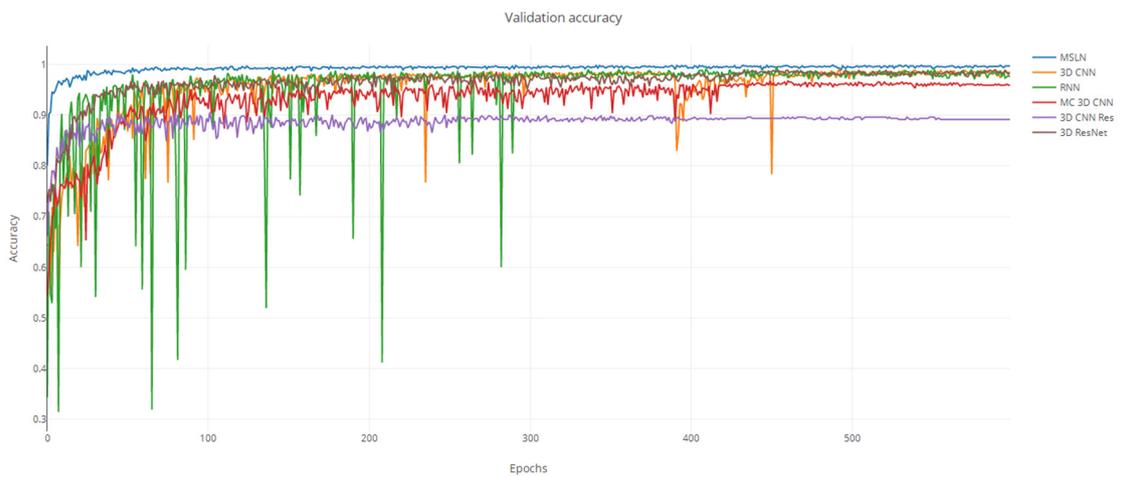
Figure 15. Cont.



(b) Salinas



(c) Pavia Centre



(d) Pavia University

Figure 15. Cont.

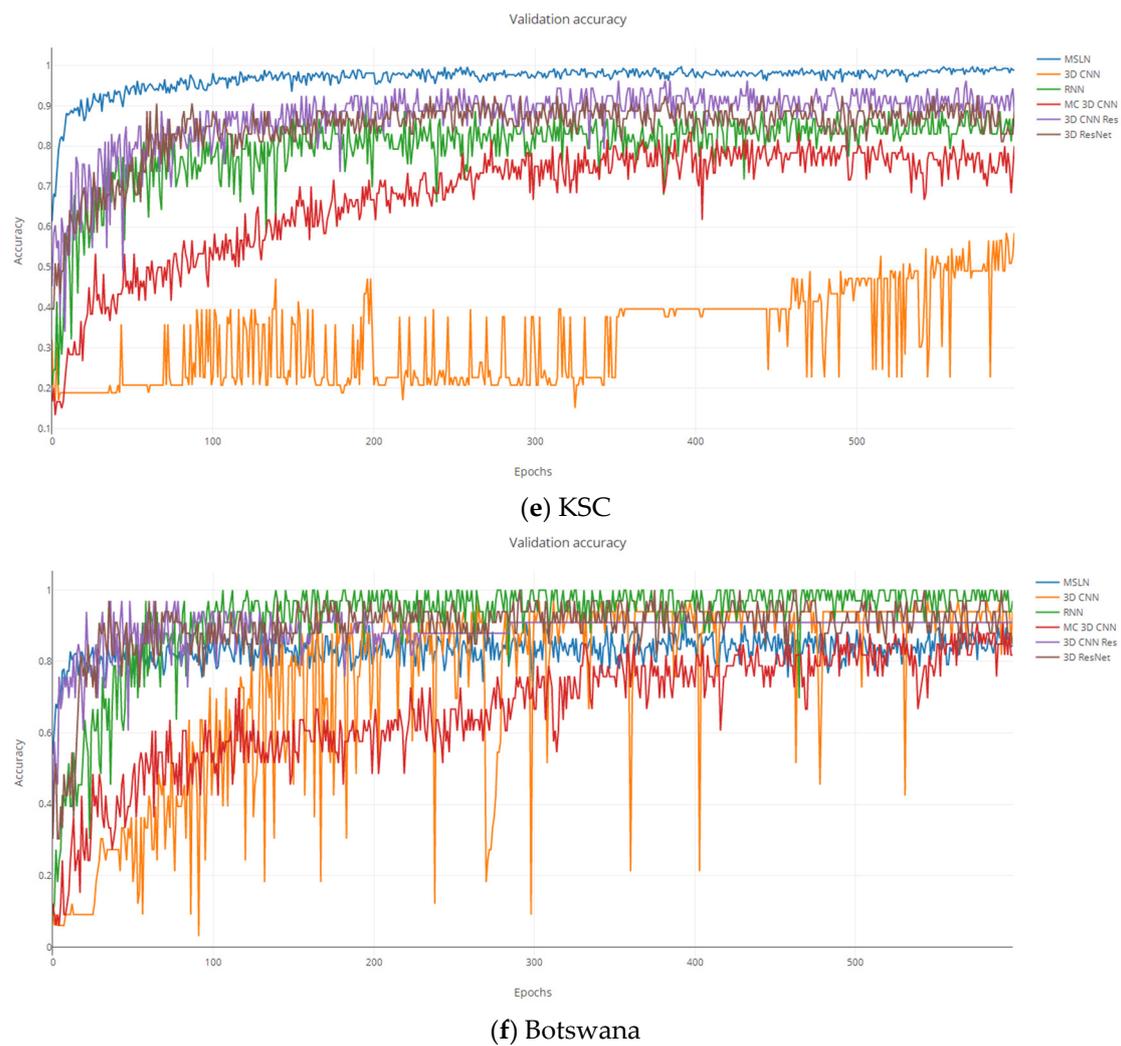


Figure 15. (a–f) Comparison of epoch validation accuracy and overall accuracy for different datasets.

To better indicate the parameter size of each network structure, the upper limit of the vertical coordinate in Figure 16 was set to 25 MB, mainly because of the great number of channels in the 3D ResNet algorithm and the large parameter size (130.20 ± 0.65 MB) when the six datasets were learned, which also coincided with the longest training time. It can be seen from Figure 16 that the parameter size of the MSLN network structure was 10.87 ± 0.27 MB, while Figure 17 indicates that the learning time of the algorithm proposed in this study failed to increase with the increase in parameter size. This fully illustrated that the multi-shortcut-link network structure proposed in this study not only improved the overall classification accuracy but also shortened the model's training and learning times while diminishing overfitting. Therefore, the algorithm proposed in this study took a shorter training time to achieve the highest accuracy compared with other models.

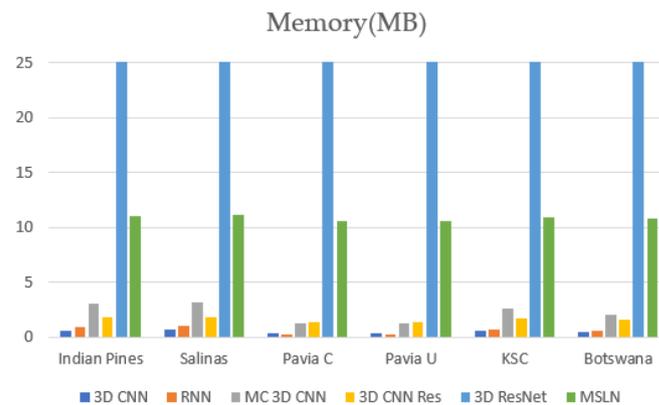


Figure 16. Memory usage of different datasets by different methods.

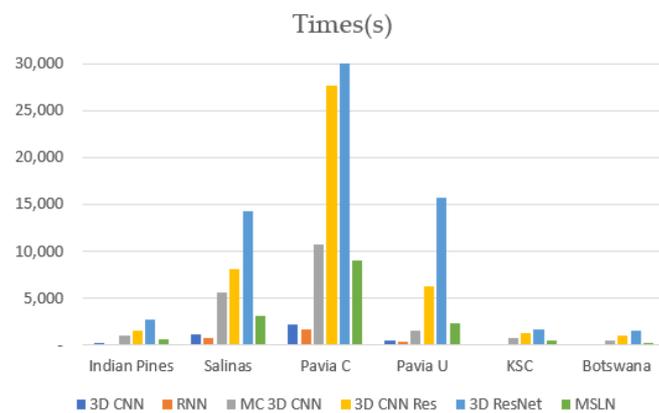


Figure 17. Time used on different datasets by different methods.

5. Conclusions

In view of the characteristics of hyperspectral images, such as few labeled samples, excess noise, and homogeneity with heterostructures, this study built a multi-shortcut-link network structure to extract the 3D spatial–spectral information of HSIs based on the properties of a 3D CNN and the shortcut link characteristics in a ResNet and tested six groups of HSI datasets by making full use of the shallow representational features and deep abstract features of HSIs. The results showed the following: (i) The MSLN could directly input the cube data of the HSIs and could then effectively extract the spatial–spectral information. The hybrid use of activation functions ensured the integrity of the nonlinear features of input data, which not only improved the use ratio of neurons but also increased the model’s rate of convergence. (ii) The multi-shortcut-link network’s structure fused shallow features and deep features, which reduced the gradient loss of deep features, solved the degeneration of the deep network satisfactorily, and enhanced the network’s generalization ability. The L-softmax loss function endowed the learned features with stronger discriminatory power, effectively addressing the issue of “the same object with distinct spectra, and distinct objects with the same spectrum” and achieving more significant classification results. Therefore, the MSLN proposed in this study could effectively improve the overall classification result.

Although the multi-shortcut-link network’s structure proposed in this study demonstrated extraordinary superiority regarding performance and classification accuracy, no discussion has ever covered the issues of information interaction and weight allocation between different channels. In future work, the attention mechanism will be introduced while the network is deepened, and the relevance between space and channels will be utilized to enhance the discriminatory power of the features of terrain objects with inferior classification accuracy to achieve higher classification accuracy.

Author Contributions: Conceptualization, M.S.; methodology, H.Z.; software, H.Z.; validation, Y.C., G.G. and X.G.; formal analysis, Y.J.; investigation, J.M.; resources, Y.C.; data curation, Y.C.; writing—original draft preparation, H.Z.; writing—review and editing, Y.J.; visualization, H.Z.; supervision, M.S.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

List of Acronyms:

HSI	Hyperspectral image
SVM	Support vector machine
ELM	Extreme learning machine
ANN	Artificial neural network
SAE	Stack autoencoding
DBN	Deep belief network
RNN	Recurrent neural network
CNN	Convolutional neural network
ResNet	Residual network
ResU	Residual unit
MSLN	Multi-shortcut-link network
ReLU	Rectified linear unit
SELU	Self-exponential linear unit
PReLU	Parametric rectified linear unit
MC 3D CNN	Multiscale 3D CNN
3D CNN Res	3D CNN residual
AA	Average accuracy
OA	Overall accuracy

References

- Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [\[CrossRef\]](#)
- Sun, L.; Wu, F.; He, C.; Zhan, T.; Liu, W.; Zhang, D. Weighted Collaborative Sparse and L1/2 Low-Rank Regularizations with Superpixel Segmentation for Hyperspectral Unmixing. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5.
- Maes, W.H.; Steppe, K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [\[CrossRef\]](#)
- Sun, H.; Zheng, X.; Lu, X.; Wu, S. Spectral-Spatial Attention Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3232–3245. [\[CrossRef\]](#)
- Shimoni, M.; Haelterman, R.; Perneel, C. Hypersectral Imaging for Military and Security Applications: Combining Myriad Processing and Sensing Techniques. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 101–117. [\[CrossRef\]](#)
- Yokoya, N.; Chan, J.C.; Segl, K. Potential of Resolution-Enhanced Hyperspectral Data for Mineral Mapping Using Simulated EnMAP and Sentinel-2 Images. *Remote Sens.* **2016**, *8*, 172. [\[CrossRef\]](#)
- Wu, H.; Lin, A.; Clarke, K.C.; Shi, W.; Cardenas-Tristan, A.; Tu, Z. A comprehensive quality assessment framework for linear features from Volunteered Geographic Information. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1826–1847. [\[CrossRef\]](#)
- Lin, A.; Sun, X.; Wu, H.; Luo, W.; Wang, D.; Zhong, D.; Wang, Z.; Zhao, L.; Zhu, J. Identifying urban building function by integrating remote sensing imagery and POI data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8864–8875. [\[CrossRef\]](#)
- Chehrehghan, A.; Ali Abbaspour, R. An evaluation of data completeness of VGI through geometric similarity assessment. *Int. J. Image Data Fusion* **2018**, *9*, 319–337. [\[CrossRef\]](#)
- Barrington-Leigh, C.; Millard-Ball, A. The world’s user-generated road map is more than 80% complete. *PLoS ONE* **2017**, *12*, e0180698.
- Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362. [\[CrossRef\]](#)
- Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [\[CrossRef\]](#)
- Huang, W.; Huang, Y.; Wang, H.; Liu, Y.; Shim, H.J. Local binary patterns and superpixel-based multiple kernels for hy-perspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4550–4563. [\[CrossRef\]](#)

14. Ye, Q.; Zhao, H.; Li, Z.; Yang, X.; Gao, S.; Yin, T.; Ye, N. L1-Norm Distance Minimization-Based Fast Robust Twin Support Vector k-Plane Clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 4494–4503. [[CrossRef](#)]
15. Ghamisi, P.; Benediktsson, J.A.; Ulfarsson, M.O. Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2565–2574. [[CrossRef](#)]
16. Liu, B.; Yu, X.; Zhang, P.; Tan, X.; Yu, A.; Xue, Z. semi-supervised convolutional neural network for hyperspectral image classification. *Remote Sens. Lett.* **2017**, *8*, 839–848. [[CrossRef](#)]
17. Qin, A.; Shang, Z.; Tian, J.; Wang, Y.; Zhang, T.; Tang, Y.Y. Spectral-spatial Graph Convolutional Networks for Semisupervised Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 241–245. [[CrossRef](#)]
18. Huang, G.B.; Ding, X.; Zhou, H. Optimization method based extreme learning machine for classification. *Neurocomputing* **2010**, *74*, 155–163. [[CrossRef](#)]
19. Hernández-Espinosa, C.; Fernández-Redondo, M.; Torres-Sospedra, J. Some experiments with ensembles of neural networks for classification of hyperspectral images. In *Proceedings of the International Symposium on Neural Networks, Dalian, China, 19–21 August 2004*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 912–917.
20. Yu, C.; Xue, B.; Song, M.; Wang, Y.; Li, S.; Chang, C.I. Iterative Target-Constrained Interference-Minimized Classifier for Hyperspectral Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1095–1117. [[CrossRef](#)]
21. Luo, F.; Du, B.; Zhang, L.; Zhang, L.; Tao, D. Feature Learning Using Spatial-Spectral Hypergraph Discriminant Analysis for Hyperspectral Image. *IEEE Trans. Cybern.* **2019**, *49*, 2406–2419. [[CrossRef](#)]
22. Yin, B.; Cui, B. Multi-feature extraction method based on Gaussian pyramid and weighted voting for hyperspectral image classification. In *Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Guangzhou, China, 15–17 January 2021; IEEE: New York, NY, USA, 2021; pp. 645–648.
23. Huang, W.; Li, G.; Chen, Q.; Ju, M.; Qu, J. CF2PN: A Cross-Scale Feature Fusion Pyramid Network Based Remote Sensing Target Detection. *Remote Sens.* **2021**, *13*, 847. [[CrossRef](#)]
24. Zhang, X.; Lu, W.; Li, F.; Peng, X.; Zhang, R. Deep Feature Fusion Model for Sentence Semantic Matching. *Comput. Mater. Contin.* **2019**, *61*, 601–616. [[CrossRef](#)]
25. Wu, H.; Liu, Q.; Liu, X. A Review on Deep Learning Approaches to Image Classification and Object Segmentation. *Comput. Mater. Contin.* **2019**, *60*, 575–597. [[CrossRef](#)]
26. Xu, Y.; Du, B.; Zhang, L. Beyond the Patchwise Classification: Spectral-spatial Fully Convolutional Networks for Hyperspectral Image Classification. *IEEE Trans. Big Data* **2020**, *6*, 492–506. [[CrossRef](#)]
27. Jiang, Y.; Li, Y.; Zou, S.; Zhang, H.; Bai, Y. Hyperspectral Image Classification with Spatial Consistency Using Fully Convolutional Spatial Propagation Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10425–10437. [[CrossRef](#)]
28. Xu, Q.; Xiao, Y.; Wang, D.; Luo, B. CSA-MSO3DCNN: Multiscale Octave 3D CNN with Channel and Spatial Attention for Hyperspectral Image Classification. *Remote Sens.* **2020**, *12*, 188. [[CrossRef](#)]
29. 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.* **2017**, *7*, 2094–2107.
30. Suk, H.-I.; Lee, S.-W.; Shen, D. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Struct. Funct.* **2015**, *220*, 841–859. [[CrossRef](#)]
31. Chen, Y.; Zhao, X.; Jia, X. Spectral-Spatial Classification of Hyperspectral Data Based on Deep Belief Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2392. [[CrossRef](#)]
32. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [[CrossRef](#)]
33. Seydgar, M.; Alizadeh Naeni, A.; Zhang, M.; Li, W.; Satari, M. 3-D convolution-recurrent networks for spectral-spatial classification of hyperspectral images. *Remote Sens.* **2019**, *11*, 883. [[CrossRef](#)]
34. He, C.; Sun, L.; Huang, W.; Zhang, J.; Zheng, Y.; Jeon, B. TSLRLN: Tensor subspace low-rank learning with non-local prior for hyperspectral image mixed denoising. *Signal Process.* **2021**, *184*, 108060. [[CrossRef](#)]
35. Sharma, V.; Diba, A.; Tuytelaars, T.; Van Gool, L. *Hyperspectral CNN for Image Classification & Band Selection, with Application to Face Recognition*; Technical Report: KUL/ESAT/PSI/1604; KU Leuven, ESAT: Leuven, Belgium, 2016.
36. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *J. Sens.* **2015**, *2015*, 258619. [[CrossRef](#)]
37. Hamida, A.B.; Benoit, A.; Lambert, P.; Amar, C.B. 3-D Deep Learning Approach for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4420–4434. [[CrossRef](#)]
38. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [[CrossRef](#)]
39. Ying, L.; Haokui, Z.; Qiang, S. Spectral-Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sens.* **2017**, *9*, 67.
40. Luo, Y.; Zou, J.; Yao, C.; Zhao, X.; Li, T.; Bai, G. HSI-CNN: A Novel Convolution Neural Network for Hyperspectral Image. In *Proceedings of the 2018 International Conference on Audio, Language and Image Processing (ICALIP)*, Shanghai, China, 16–17 July 2018; IEEE: New York, NY, USA, 2018.
41. Lee, H.; Kwon, H. Contextual Deep CNN Based Hyperspectral Classification. In *Proceedings of the Geoscience & Remote Sensing Symposium*, Beijing, China, 10–15 July 2016; IEEE: New York, NY, USA, 2016; pp. 3322–3325.

42. He, M.; Li, B.; Chen, H. Multi-scale 3D deep convolutional neural network for hyperspectral image classification. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; IEEE: New York, NY, USA, 2017; pp. 3904–3908.
43. Sun, L.; He, C.; Zheng, Y.; Tang, S. SLRL4D: Joint restoration of subspace low-rank learning and non-local 4-D transform filtering for hyperspectral image. *Remote Sens.* **2020**, *12*, 2979. [[CrossRef](#)]
44. Gao, H.; Chen, Z.; Li, C. Sandwich Convolutional Neural Network for Hyperspectral Image Classification Using Spectral Feature Enhancement. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3006–3015. [[CrossRef](#)]
45. Bishop, C.M. *Neural Networks for Pattern Recognition*; Oxford University Press: Oxford, UK, 1995.
46. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 2007.
47. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S-PLUS*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
48. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, R.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 17–20 June 2015; pp. 1–9.
49. Lee, C.Y.; Xie, S.; Gallagher, P.; Zhang, Z.; Tu, Z. Deeply-supervised nets. In Proceedings of the Artificial Intelligence and Statistics, San Diego, California, USA, 9–12 May 2015; PMLR: New York, NY, USA, 2015; pp. 562–570.
50. Raiko, T.; Valpola, H.; LeCun, Y. Deep learning made easier by linear transformations in perceptrons. In Proceedings of the Artificial intelligence and statistics, La Palma, Canary Islands, 21–23 April 2012; PMLR: New York, NY, USA, 2012; pp. 924–932.
51. Schraudolph, N. *Accelerated Gradient Descent by Factor-Centering Decomposition*; Technical report/IDSIA; IDSIA: Lugano, Switzerland, 1998; p. 98.
52. Schraudolph, N.N. Centering neural network gradient factors. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 207–226.
53. Vatanen, T.; Raiko, T.; Valpola, H.; LeCun, Y. Pushing stochastic gradient towards second-order methods—backpropagation learning with transformations in nonlinearities. In Proceedings of the International Conference on Neural Information Processing, Daegu, Korea, 3–7 November 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 442–449.
54. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; PMLR: New York, NY, USA, 2015; pp. 448–456.
55. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
56. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 630–645.
58. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the thirty-first AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
59. Lu, Y.S.; Li, Y.X.; Liu, B. Hyperspectral Data Haze Monitoring Based on Deep Residual Network. *Acta Opt. Sin.* **2017**, *37*, 1128001.
60. Liu, D.; Han, G.; Liu, P.; Yang, H.; Sun, X.; Li, Q.; Wu, J. A Novel 2D-3D CNN with Spectral-Spatial Multi-Scale Feature Fusion for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 4621. [[CrossRef](#)]
61. Meng, Z.; Jiao, L.; Liang, M.; Zhao, F. Hyperspectral Image Classification with Mixed Link Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2494–2507. [[CrossRef](#)]
62. 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](#)]
63. Cao, F.; Guo, W. Deep hybrid dilated residual networks for hyperspectral image classification. *Neurocomputing* **2020**, *384*, 170–181. [[CrossRef](#)]
64. Xu, Y.; Li, Z.; Li, W.; Du, Q.; Liu, C.; Fang, Z.; Zhai, L. Dual-channel residual network for hyperspectral image classification with noisy labels. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–11. [[CrossRef](#)]
65. Gao, H.; Yang, Y.; Li, C.; Gao, L.; Zhang, B. Multiscale residual network with mixed depthwise convolution for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3396–3408. [[CrossRef](#)]
66. Dang, L.; Pang, P.; Lee, J. Depth-Wise Separable Convolution Neural Network with Residual Connection for Hyperspectral Image Classification. *Remote Sens.* **2020**, *12*, 3408. [[CrossRef](#)]
67. Paoletti, M.E.; Haut, J.M.; Fernandez-Beltran, R.; Plaza, J.; Plaza, A.J.; Pla, F. Deep pyramidal residual networks for spectral–spatial hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 740–754. [[CrossRef](#)]
68. Wang, L.; Peng, J.; Sun, W. Spatial–spectral squeeze-and-excitation residual network for hyperspectral image classification. *Remote Sens.* **2019**, *11*, 884. [[CrossRef](#)]
69. Zhang, X.; Wang, Y.; Zhang, N.; Xu, D.; Luo, H.; Chen, B.; Ben, G. Spectral-spatial Fractal Residual Convolutional Neural Network with Data Balance Augmentation for Hyperspectral Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10473–10487. [[CrossRef](#)]

70. Xu, H.; Yao, W.; Cheng, L.; Li, B. Multiple Spectral Resolution 3D Convolutional Neural Network for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 1248. [[CrossRef](#)]
71. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
72. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
73. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-normalizing neural networks. In Proceedings of the 31st international conference on neural information processing systems, Long Beach, CA, USA, 4–9 December 2017; pp. 972–981.
74. Liu, W.; Wen, Y.; Yu, Z.; Yang, M. Large-Margin Softmax Loss for Convolutional Neural Networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 507–516.
75. Zhang, Y.Z.; Xu, M.M.; Wang, X.H.; Wang, K.Q. Hyperspectral image classification based on hierarchical fusion of residual networks. *Spectrosc. Spectr. Anal.* **2019**, *39*, 3501–3507.
76. Graña, M.; Veganzons, M.A.; Ayerdi, B. Hyperspectral Remote Sensing Scenes. Available online: http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 6 October 2021).