

Article

Precise Crop Classification of Hyperspectral Images Using Multi-Branch Feature Fusion and Dilation-Based MLP

Haibin Wu ¹, Huaming Zhou ¹, Aili Wang ^{1,*}  and Yuji Iwahori ² 

¹ Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China; woo@hrbust.edu.cn (H.W.); 1705030430@stu.hrbust.edu.cn (H.Z.)

² Department of Computer Science, Chubu University, Kasugai 487-8501, Japan; iwahori@isc.chubu.ac.jp

* Correspondence: aili925@hrbust.edu.cn

Abstract: The precise classification of crop types using hyperspectral remote sensing imaging is an essential application in the field of agriculture, and is of significance for crop yield estimation and growth monitoring. Among the deep learning methods, Convolutional Neural Networks (CNNs) are the premier model for hyperspectral image (HSI) classification for their outstanding locally contextual modeling capability, which facilitates spatial and spectral feature extraction. Nevertheless, the existing CNNs have a fixed shape and are limited to observing restricted receptive fields, constituting a simulation difficulty for modeling long-range dependencies. To tackle this challenge, this paper proposed two novel classification frameworks which are both built from multilayer perceptrons (MLPs). Firstly, we put forward a dilation-based MLP (DMLP) model, in which the dilated convolutional layer replaced the ordinary convolution of MLP, enlarging the receptive field without losing resolution and keeping the relative spatial position of pixels unchanged. Secondly, the paper proposes multi-branch residual blocks and DMLP concerning performance feature fusion after principal component analysis (PCA), called DMLPFFN, which makes full use of the multi-level feature information of the HSI. The proposed approaches are carried out on two widely used hyperspectral datasets: Salinas and KSC; and two practical crop hyperspectral datasets: WHU-Hi-LongKou and WHU-Hi-HanChuan. Experimental results show that the proposed methods outshine several state-of-the-art methods, outperforming CNN by 6.81%, 12.45%, 4.38% and 8.84%, and outperforming ResNet by 4.48%, 7.74%, 3.53% and 6.39% on the Salinas, KSC, WHU-Hi-LongKou and WHU-Hi-HanChuan datasets, respectively. As a result of this study, it was confirmed that the proposed methods offer remarkable performances for hyperspectral precise crop classification.



Citation: Wu, H.; Zhou, H.; Wang, A.; Iwahori, Y. Precise Crop Classification of Hyperspectral Images Using Multi-Branch Feature Fusion and Dilation-Based MLP. *Remote Sens.* **2022**, *14*, 2713. <https://doi.org/10.3390/rs14112713>

Academic Editors: Liguo Wang,
Yanfeng Gu and Peng Wang

Received: 23 April 2022

Accepted: 4 June 2022

Published: 5 June 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 imaging instruments can capture rich spectral signatures and intricate spatial information of observed scenes [1]. Plentiful spectral signatures and spatial information of hyperspectral images (HSIs) offer great potentials for fine crop classification [2,3] and detection [4,5]. Therefore, hyperspectral remote sensing can obtain the spectral characteristics and their differences more comprehensively and meticulously than panchromatic remote sensing [6]. Therefore, this paper uses hyperspectral techniques to finely classify crops and to promote the development of specific applications of hyperspectral techniques in agricultural remote sensing, such as monitoring the development of agriculture and optimizing the management of the agricultural industry.

Many methods have been applied to hyperspectral image classification in recent years. Early-stage classification methods are support vector machine (SVM) [7], random forest (RF) [8], multiple logistic regression [9] and decision tree [10], which can provide promising

classification results. What is more, these classification methods can only extract the shallow feature information of hyperspectral images, which have limited ability to handle the highly nonlinear HSI data and limit the further improvement of their classification accuracy.

Recently, deep learning-based models have also been extended to HSI classification. In [11], Chen et al. used a deep stacked auto-encoder (SAE) to extract features from the spectral domain for HSI classification tasks. In [12], Tao et al. introduced a modified auto-encoder model, called multiscale sparse SAE, to construct two variants of feature-learning procedures for sparse spectral feature learning and multiscale spatial feature learning from unlabeled data in HSIs. Sun et al. proposed a hybrid classification method combined deep belief network (DBN) with principal component analysis (PCA) to improve the HSI classification performance [13]. Chen et al. introduced the Convolutional Neural Network (CNN) into HSI classification in [14]. In this paper, a regularized deep feature extraction (FE) method is presented for HSI classification using CNN. Zhang [15] proposed a novel spatial residual block combined parallel network, which extracted rich spatial context information to improve hyperspectral classification accuracy. Kanthi et al. [16] proposed a new 3D deep feature extraction CNN model for HSI classification in which the HSI data are divided into 3D patches and fed into the proposed model for deep feature extractions. Zhang et al. [17] proposed a multi-scale dense network for HSIs, which can extract more refined features and make full use of multi-scale features. Zhu [18] proposed a self-supervised contrastive efficient asymmetric dilated network for HSI classification, which designed a lightweight feature extraction network EADNet in the contrastive learning framework.

While CNN-based models have yielded positive scores in HSI classification, the complexities intrinsic to remote-sensing hyperspectral images retain limitations concerning the performance of numerous CNN models. First, the parametrization of CNNs multiplies as convolution layers multiply exponentially and becomes larger as the computational capabilities rise. Additionally, the computational has turns into a bottleneck for practical implementations due to the long runtime of multiplication and summation. Lastly, the translation invariance and local connectivity of CNNs can interfere with the effectiveness of HSI classification.

MLP, as a neural network with fewer constraints, can eliminate the adverse effects of local connectivity and focus on spatial structure and information. It has been proved to be a promising machine-learning technology. Ilya [19] proposed an architecture based on MLP, including channel-mixing MLPs and token-mixing MLPs. However, the MLP-Mixer can achieve performance comparable to CNN. Yu [20] proposed a novel pure MLP architecture, which only contains channel-mixing MLPs. It devised a spatial-shift operation for achieving the communication between patches and attained higher recognition accuracy than the MLP-Mixer. Lian [21] proposed an Axial Shifted MLP architecture that pays more attention to local feature interaction. Yu [22] improved the Spatial-Shift MLP Architecture (S2-MLP) for vision backbone to adopt smaller-scale patches and use a pyramid structure to boost image recognition accuracy. Chen [23] presented a simple MLP-like architecture, Cycle-MLP, a versatile backbone for visual recognition and dense prediction, which maintained computational complexity and expanded the receptive field to some extent.

MLP solves translation invariance and local connectivity problems, and residual blocks preserve original information to prevent model degradation and facilitate rapid model effects [24]. Multi-branch feature fusion can make full use of different levels of features. Therefore, we proposed two MLP-based classification frameworks: a Dilation-based MLP (DMLP) model, and DMLP combined with feature fusion network (DMLPFFN) to improve the different level feature representation capability of the model. As a summary, the following are the main contributions of this study.

1. MLP, as a less constrained network, can eliminate the negative effects of translation invariance and local connectivity. Therefore, this paper modified MLP combined with dilated convolution to fully obtain spectral–spatial features of each sample and improve HSI remote sensing scene classification performance, called DMLP. The dilated convolutional layer replaced the ordinary convolution of MLP, which can

enlarge the receptive field without losing resolution and keep the relative spatial position of pixels unchanged.

2. This paper composes multi-branch residual blocks and DMLP to form a multi-level feature fusion network, called DMLPFFN. Firstly, the residual structure can retain the original characteristics of the HSI data, and avoid the problems of gradient explosion and gradient disappearance in the training process. In addition, DMLP can improve the feature extraction capability of the residual blocks and strengthen the model with essential features while retaining the original features of the hyperspectral data. In DMLPFFN, three branches of features are fused to obtain a feature map with more comprehensive information, which integrates the spectral information, spatial context information, spatial feature information and spatial location information of HSI to improve classification accuracy.
3. Comprehensive experiments are designed and executed to prove the effectiveness of DMLPFFN by different hyperspectral datasets. DMLPFFN achieved better classification performance and generalization ability for fine crop classification.

The rest of this article is organized as follows. Section 2 describes our proposed classification approach in detail. Section 3 reports the experimental results and evaluates the performance of the proposed method. The application of the model to fine crop classification is given in Section 4. Section 5 analyzes how to choose experimental parameters in DMLPFFN and Section 6 gives the conclusion.

2. The Proposed MLP-Based Methods for HSI Classification

Figure 1 shows the overall framework of our proposed DMLPFFN for HSI classification, which takes the WHU-Hi-LongKou dataset as an example. First, principal component analysis (PCA) is applied to the original HSI to reduce its spectral dimension to weaken the Hughes phenomenon and decrease the burden of model training. At first, the DMLP is structured by altering the normal convolution with the dilated convolution in the local perceptron module of the MLP, thus facilitating the aggregation of contextual information without conceiving a loss of feature map resolution so as to upgrade the classification performance of hyperspectral features.

In addition, DMLPFFN combines residual blocks of different sizes and DMLP to obtain three feature extraction branches, which can fuse three different levels of features and achieve feature maps with more comprehensive information. In the DMLPFFN, multiscale features of the HSI are extracted by a hierarchical different-scale feature extraction branches at different stages of the network. The low-level feature extraction branch of DMLPFFN extracts texture feature information such as color and the edge of ground objects, the middle-level branch extracts regional information and high level is used to extract semantic information with DMLP. The feature fusion is then performed by element summation of the results of the three branches, which can achieve feature maps with more comprehensive information. Then, the global average pooling transforms the feature maps into feature vectors and subsequently obtains the classification results by the softmax function.

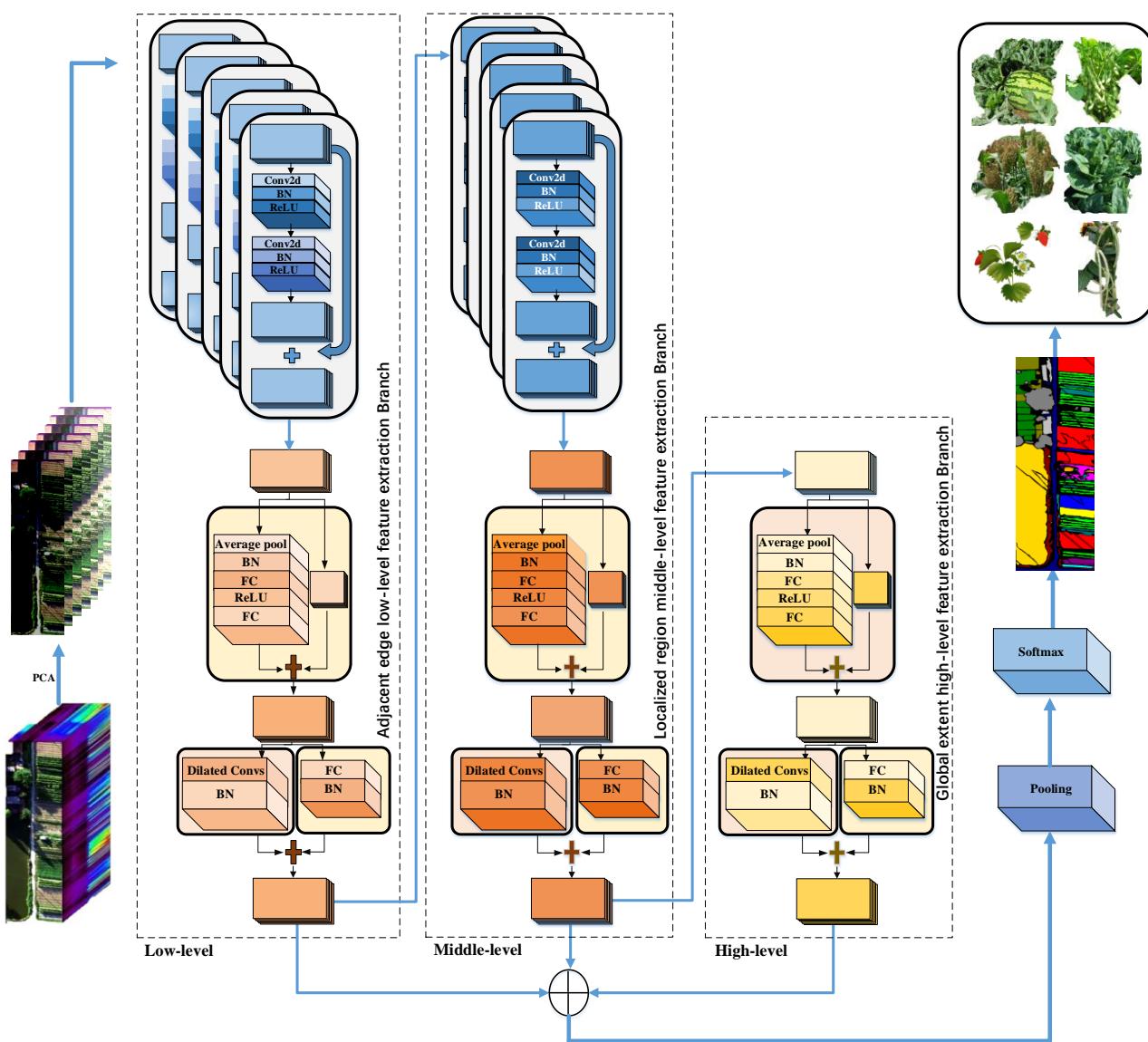


Figure 1. Framework of the proposed DMLPFFN for HSI classification.

2.1. The Proposed Dilation-Based MLP (DMLP) for HSI Classification

Figure 2 shows the overall architecture of the proposed DMLP for HSI classification. The network consists of the global perceptron module, the partition perceptron module and the local perceptron module. Since the MLP has a more powerful representation than convolution, we propose DMLP to accurately represent the feature location information, and retain spatial resolution without loss of detail information.

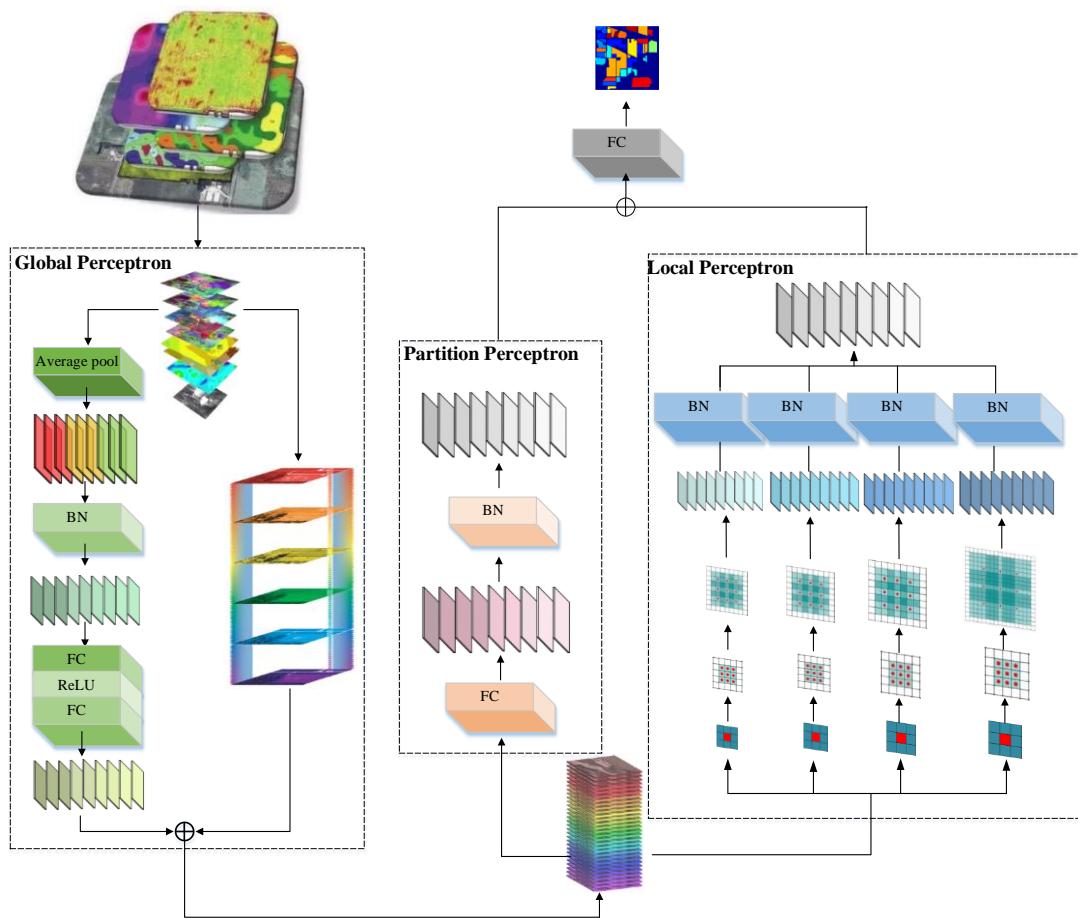


Figure 2. The structure of DMLP for HSI classification.

2.1.1. The Global Perceptron Module Block

It is assumed that the HSI dataset is size $H \times W \times nBand$, where H and W represent spatial height and width, and $nBand$ is the number of bands. First, each pixel of the hyperspectral image is processed with a fixed window size $y \times x$, and a single sample with a shape of $y \times x \times nBand$ is generated. The global perceptron uses shared parameters for different partitions, diminishing the parameters taken for computation and increasing the connection and correlation between the partitions. The global perceptron module block consists of two branches. The first branch splits up the input hyperspectral feature image. The hyperspectral feature map changes from (H_1, W_1, C_1) to (h_1, w_1, O) . H_1, W_1, C_1 indicate the height, width and number of input channels of the input hyperspectral feature map. h_1, w_1, O , respectively, represent the height, width and number of output channels of the split hyperspectral feature image.

In the second branch, the original feature map (H_1, W_1, C_1) is average pooled, and the size of the hyperspectral feature map becomes (h, w, O) as follows:

$$h_1 = \frac{H_1}{h}, w_1 = \frac{W_1}{w} \quad (1)$$

where h and w indicate the height and width of the hyperspectral feature image after average pooling; the second branch uses h and w to obtain a pixel for each hyperspectral feature image, and then feeds them through batch normalization (BN) and a two-layer MLP. The hyperspectral feature map (h, w, O) is sent to a BN layer and two fully connected layers. The Rectified Linear Unit (ReLU) function is introduced between the two fully connected layers to effectively avoid gradient explosion and gradient disappearance. For the fully

connected layer, $X^{(in)}$ and $X^{(out)}$ represent input and output; the kernel $W \in R^{Q \times P}$ is the matrix multiplication (MMUL) defined as follows:

$$X^{(out)} = MMUL\left(X^{(in)}, W\right) = X^{(in)} \cdot W^T \quad (2)$$

The hyperspectral vector was transformed into $(1, 1, C_1)$ by the BN layer and two fully connected layers. Then, the hyperspectral feature images were obtained after all branches were added. Next, we directly fed the input hyperspectral feature into partition perceptron and local perceptron without splitting.

2.1.2. The Partition Perceptron Module Block

The partition perceptron module block contains a BN layer and a group convolution. The input of the partition perceptron is (h, w, O) . After the BN layer and group convolution processing, (h, w, O) becomes the original hyperspectral feature input (H_1, W_1, C_1) . $Y^{(out)} \in R^{C_1 \times H_1 \times W_1}$ indicates the output hyperspectral feature and can be obtained as follows:

$$Y^{(out)} = g\left(Y^{(in)}, F, g, p\right), F \in R^{C_1/g \times K \times K} \quad (3)$$

where p is the number of pixels filled. $F \in R^{C_1/g \times K \times K}$ is the convolution kernel and g indicates the number of convolution groups.

2.1.3. The Local Perceptron Module Block

To enhance the extraction of high-level semantic information from hyperspectral feature maps without multiplying the calculation parameters, the local perceptron module introduces a dilated convolutional layer [25] and BN layer. First, the local perceptron module simultaneously sends the segmented hyperspectral feature image (h, w, O) to the dilated convolution layer. Then, the feature graph is fed into the BN layer. Finally, the output of all convolution branches and the partition perceptron is summarized as the final result.

Specifically, the dilated convolutional layer uses the odd–even mixed dilation rates to stack in each chain, resulting in the expanded receptive field. In addition, under the premise of the same receptive field, the dilated convolution with increased dilation rate consumes fewer training parameters than the extended receptive field with a large convolution kernel. The calculation of the size of the dilated convolution kernel and the receptive field is shown in Formulas (4) and (5), respectively:

$$f_n = f_k + (f_k - 1) * (D_r - 1) \quad (4)$$

$$l_m = l_{m-1} + \left[(f_n - 1) * \prod_{i=1}^{m-1} S_i \right] \quad (5)$$

f_k represents the size of the original convolution kernel; f_n represents the size of the dilated convolution kernel; D_r represents expansion rate; l_{m-1} represents the receptive field size of the $(m-1)$ layer; l_m is the size of the m layer receptive field after the convolution; S_i represents the step size of layer i .

The equivalently fully connected layers (FC) kernel of a Dilated Conv kernel is the result of convolution on an identity matrix with proper reshaping. Formula (6) shows exactly how to build $W^{(F,p)}$ from F and p .

$$W^{(F,p)} = DilatedCONV(Y, F, p), (Chw, Ohw)^T \quad (6)$$

The convolution after multiple superimpositions of the expansion may lead to a grid effect, as shown in Figure 3. This will cause some dilution between pixels, which causes some pixels to be omitted, resulting in the loss of local information and undermining the continuity of information.

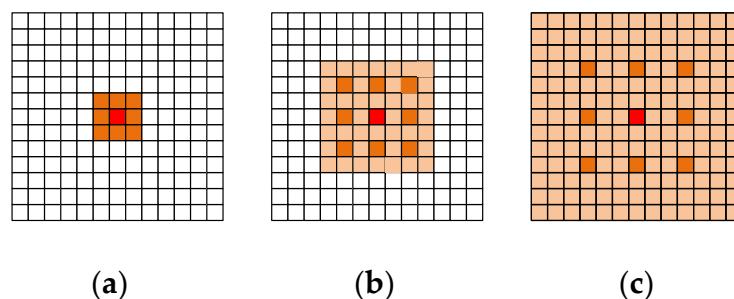


Figure 3. Dilated convolution with different dilation rates. (a) $r = 1$; (b) $r = 2$; (c) $r = 3$.

Considering the grid effect, the design of the expansion rate in the DMLP model proposed in this paper follows Equation (7).

$$M_i = \max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i] \quad (7)$$

where r_i is the expansion rate of layer i and M_i is the maximum expansion rate of layer i . Mixed dilated convolution requires that the expansion rate of superposition convolution cannot have a common divisor greater than 1. As shown in Figure 4, in this paper, the method of mixed parity expansion rate is used to expand the convolution kernel, and the expansion rate is set to the cyclic structure [1,2,5], which can cover every pixel on the image to avoid information loss.

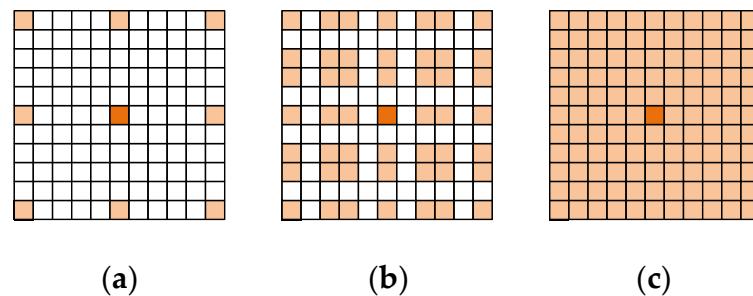


Figure 4. Odd–even mixed dilation rates. (a) $r = 5$; (b) $r = 2$; (c) $r = 1$.

2.2. The Proposed DMLPFFN Model for HSI Classification

Features at various levels contain diverse information distribution. The lower-level features contain rich spatial structure information, but their high resolution leads to weak global background information. The higher-level features have rich semantic information and can effectively classify hyperspectral images, but their poor resolution lacks spatial details for the hyperspectral images [26]. For this reason, the fusion of these different levels of feature information can significantly strengthen the classification accuracy of hyperspectral images. The paper proposes DMLPFFN, which can extract sufficiently different level features by fusing three feature extraction branches, as shown in Figure 1.

2.2.1. Fusion of Multi-Branch Features

As the layer of the network deepens, the feature information obtained during the feature extraction of the convolutional network will be different for each branch. Figure 5 shows the structure of residual blocks with DMLP, called the adjacent edge low-level feature extraction branch (the left branch in Figure 1), which is used to obtain texture characteristic information such as the color and border of the ground target.

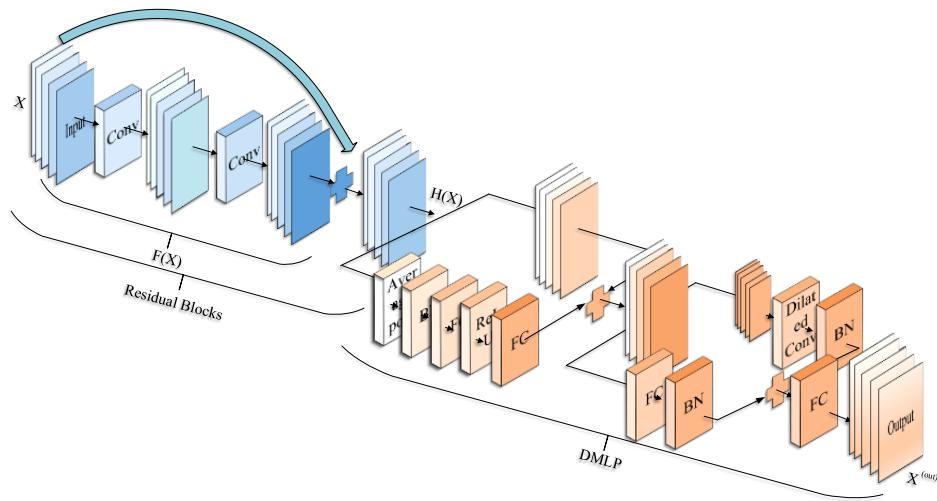


Figure 5. The structure of residual blocks with DMLP.

The residual block is introduced to connect each layer to other layers in a feed-forward fashion. According to the structure of the residual unit, X represents input, $H(x)$ represents output and $F(x)$ represents a residual unit. The residual unit carries out identity mapping of input at each layer from top to bottom, and the features of input are learned to form the residual function [27]. Then, the output of the residual unit becomes $H(x) = F(x) + x$. Therefore, the residual function can deal with more advanced abstract features when the number of network layers increases, and is easier to optimize. The calculation process of the residual element is shown in Formula (8):

$$F(x) = W_2 \sigma(W_1 x) \quad (8)$$

where σ stands for nonlinear function ReLU and W_1 and W_2 are the weights of layer 1 and layer 2, respectively. Then, the residual unit goes through a shortcut and a second ReLU layer to obtain the output $H(x)$:

$$H(x) = F(x, \{W_i\}) + x \quad (9)$$

When the dimension size of the input and the output needs to be changed, a linear transformation can be performed in a shortcut operation, as shown in Formula (10):

$$H(x) = F(x, \{W_i\}) + W_s x \quad (10)$$

By stacking multiple residual blocks, the extracted features become increasingly discriminative. Then, we connect the output of the residual block to the input of the DMLP. $H(x)^{(in)}$ and $X^{(out)}$ represent input and output and the kernel $W \in R^{Q \times P}$ is the matrix multiplication (MMUL) defined as follows:

$$X^{(out)} = MMUL(H(x)^{(in)}, W) \quad (11)$$

This structure extracts more abstract features and discards redundant information through the DMLP module. The introduction of DMLP brings fewer parameters and higher operational efficiency and speed compared to simply increasing the depth of the residual network. In addition, it improves the global feature learning capability and the nonlinearity for the model, resulting in a better abstract representation of the model.

The middle-level branch focuses on extracting regional information with a similar structure of a low-level feature extraction branch. Middle-level features focus more on regional features than lower-level features, which is of great significance to the extraction of spatial structure features of HSIs. The high-level branch uses DMLP to extract global

features, which keeps the relative spatial position of pixels unchanged and obtains the context information of the HSIs.

In detail, assume that O_1 , O_2 and O_3 refer to the outputs of the low-, middle- and high-level feature extraction branch, which has 16, 32 and 64 feature maps, respectively. Then, the resultant maps of the three branches are convolved with 64 kernels of size 1×1 in this paper. By means of such convolution operations, the number of feature maps of O_1 , O_2 and O_3 all become 64. Eventually, feature fusion can be conveniently performed by element summation as follows:

$$T = \sum_{i=1}^3 \& \sum_{j=1}^3 \text{Pooling}(f_i(O_j)) \quad (12)$$

where T represents the fused features, f_1 , f_2 and f_3 are the dimension matching function and Pooling is the global averaging function.

The proposed DMLPFFN model enhances the resemblance between the same hyperspectral feature objects and the variability between the exotic objects to accomplish high-precision classification of crop species.

2.2.2. Feature Output Visualization and Analysis

In order to better analyze the characteristics of feature extraction of DMLPFFN, this paper visualizes the feature maps of different branches, as shown in Figure 6.

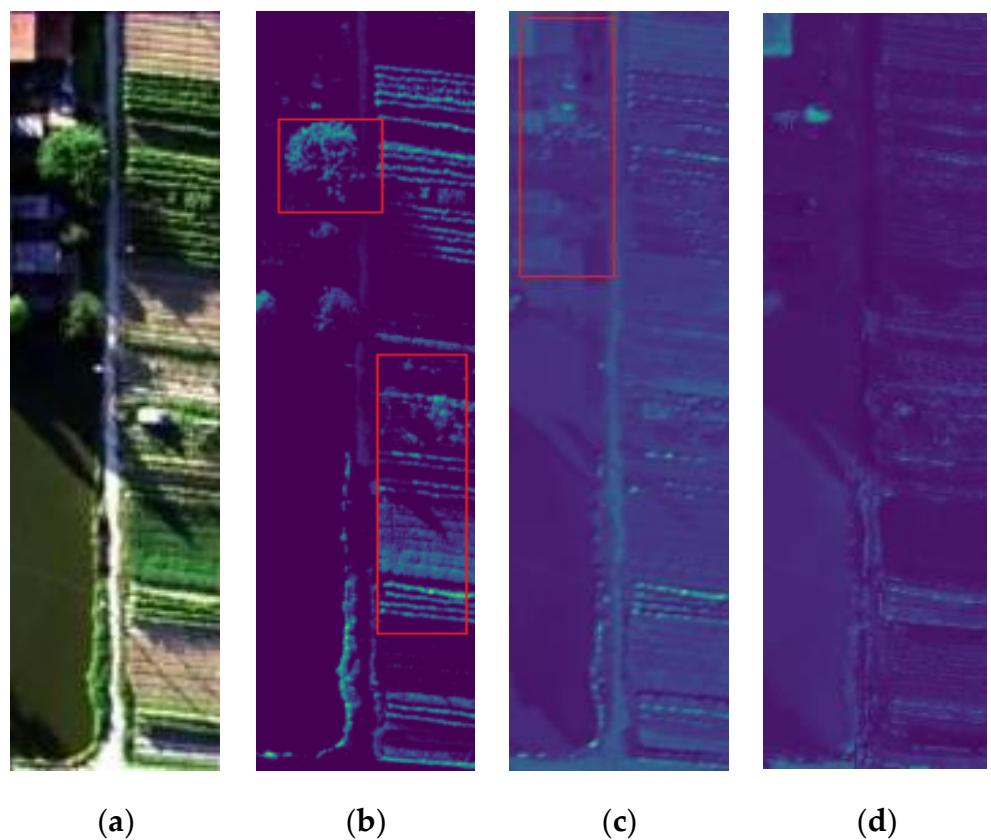


Figure 6. Feature output visualization. (a) HSI; (b) low level; (c) middle level; (d) high level.

Figure 6b–d shows the feature output plots for the adjacent edge low-level feature extraction branch, localized region middle-level feature extraction branch and global extent high-level feature extraction branch, respectively. As shown in the red frame in Figure 6b, detailed features as edges and textures of trees and farmland are highlighted. Figure 6c shows the crop regionality is enhanced and this branch extracts the regional information of the image. In Figure 6d, the global and abstract nature of the extracted features of the

image is made more apparent. In summary, Figure 6 shows the difference of the extracted features in each branch and it is necessary to fuse multi-branch features to fully dig spatial and spectral features of HSI.

3. Experimental Results

3.1. Public HSI Dataset Description

In order to verify the effectiveness of the proposed method, classification experiments were performed on two standard hyperspectral datasets (Salinas and KSC) [28,29]. The details of each dataset are as follows. The Salinas dataset was acquired by an Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) sensor at the Salinas Valley in California and consists of 512×217 pixels and 224 spectral reflectance bands. The number of bands was reduced to 204 by removing the bands covering the water-absorbing area (108–112, 154–167, 224). Ground Truth contains 16 types of land cover. The KSC dataset was picked up by AVIRIS sensors flying over the Kennedy Space in Florida. The number of spectral bands is 176, and the size is 512×614 pixels with 13 categories.

Tables 1 and 2 report the detailed number of pixels available in each class for the two datasets, respectively, and show the false-color composite image and ground truth map.

Table 1. Salinas Dataset Labeled Sample Counts.

No	Name	Color	Number	False-Color Map	Ground-Truth Map
1	Brocoli_green_weeds_1		1997		
2	Brocoli_green_weeds_2		3726		
3	Fallow		1976		
4	Fallow_rough_pow		1394		
5	Fallow_smooth		2678		
6	Stubble		3979		
7	Celery		3579		
8	Grapes_untrained		11,213		
9	soil_vinyard_develop		6197		
10	Corn_snesced_green_weeds		3249		
11	Lettuce_romaine_4wk		1058		
12	Lettuce_romaine_5wk		1908		
13	Lettuce_romaine_6wk		909		
14	Lettuce_romaine_7wk		1061		
15	Vinyard_untrained		7164		
16	Vinyard_vertical_trellis		1737		
Total Numbers			53,785		

Table 2. KSC Dataset Labeled Sample Counts.

No	Name	Color	Number	False-Color Map	Ground-Truth Map
1	Scrub		1997		
2	Willow		3726		
3	Palm		1976		
4	Pine		1394		
5	Broadleaf		2678		
6	Hardwood		3979		
7	Swap		3579		
8	Graminoid		11,213		
9	Spartina		6197		
10	Cattail		3249		
11	Salt		1058		
12	Mud		1908		
13	Water		909		
Total Numbers			5211		

3.2. Experimental Parameter Setting

All experiments were performed on Intel(R) Xeon(R) 4208 CPU @ 2.10GHz processor and Nvidia GeForce RTX 2080Ti graphics card. In order to reduce experimental errors, the model randomly selected a limited number of samples from the training set for training. The epoch was set to 200. All experimental results were averaged from 10 experiments. Overall accuracy (OA), average accuracy (AA) and Kappa coefficient (K) were used as evaluation indexes to measure the performance of each method. The initial learning rate of this method was 0.1 and was then divided by 10 when the error plateaued. The networks are trained for 2×10^4 iterations and the training minibatch has a size of 100. We use a weight decay of 0.0001 and a momentum of 0.9.

3.3. Comparison of the Proposed Methods with the State-of-the-Art Methods

The experiment mainly compares the proposed algorithm DMLP and DMLPFFN with the Radial Basis Function (RBF) Support Vector Machine algorithm (RBF-SVM) [30] and Extended Morphological Profile (EMP) Support Vector Machine Methods (EMP-SVM) [31], Convolutional Neural Network (CNN) [32], Residual Network (ResNet) [33], MLP-Mixer, RepMLP [34] and Deep Feature Fusion Network (DFFN) [35] classification performance for the hyperspectral dataset. Ten percent of the total sample number was used as the training sample number for hyperspectral classification as shown in Tables 3–5. Compared with other methods, the DMLPFFN method proposed in this paper has the highest classification accuracy for two datasets.

Table 3. Classification results on the Salinas dataset by different classification methods.

Method	RBF-SVM	EMP-SVM	CNN	ResNet	MLP-Mixer	RepMLP	DFFN	DMLP	DMLPFFN
1	85.13 ± 0.76	93.59 ± 0.26	94.57 ± 2.05	95.35 ± 1.05	96.90 ± 1.87	96.95 ± 0.02	97.20 ± 2.57	98.15 ± 0.79	99.26 ± 3.24
	91.27 ± 1.95	96.37 ± 0.15	94.59 ± 1.14	96.35 ± 1.29	96.95 ± 0.31	95.09 ± 0.78	97.41 ± 0.67	97.88 ± 2.13	98.13 ± 3.59
2	89.59 ± 2.68	81.65 ± 0.78	79.38 ± 2.21	94.51 ± 0.48	95.03 ± 1.28	96.21 ± 0.02	95.02 ± 1.12	96.39 ± 2.47	97.08 ± 1.52
	94.05 ± 3.61	95.34 ± 2.03	96.07 ± 1.08	96.49 ± 1.85	97.24 ± 2.05	97.39 ± 1.38	98.26 ± 0.81	98.04 ± 2.76	98.86 ± 3.03

Table 3. Cont.

Method	RBF-SVM	EMP-SVM	CNN	ResNet	MLP-Mixer	RepMLP	DFFN	DMLP	DMLPFFN
5	86.52 ± 2.64	92.24 ± 0.35	96.48 ± 1.32	97.25 ± 2.34	97.61 ± 0.54	97.78 ± 0.24	98.53 ± 2.07	98.65 ± 3.51	98.87 ± 1.45
6	93.14 ± 2.71	95.57 ± 0.29	96.86 ± 1.55	96.29 ± 0.17	98.76 ± 0.62	97.98 ± 2.01	97.49 ± 3.54	98.32 ± 0.94	98.69 ± 2.71
7	93.68 ± 0.53	95.21 ± 1.65	94.09 ± 2.29	95.38 ± 1.16	96.96 ± 0.35	97.39 ± 1.43	96.08 ± 3.49	97.12 ± 2.54	97.78 ± 3.66
8	85.21 ± 2.49	86.52 ± 0.46	91.39 ± 1.23	93.25 ± 0.74	94.54 ± 1.82	95.61 ± 1.29	96.25 ± 3.76	97.32 ± 1.54	98.15 ± 2.85
9	91.25 ± 0.83	92.74 ± 1.26	94.25 ± 0.46	94.47 ± 0.56	95.65 ± 1.47	96.97 ± 0.02	97.18 ± 5.51	97.46 ± 2.34	98.06 ± 0.67
10	81.21 ± 2.64	90.57 ± 1.37	92.52 ± 1.15	93.70 ± 0.59	94.13 ± 1.56	95.72 ± 0.15	95.68 ± 1.34	96.07 ± 0.51	97.92 ± 3.28
11	86.41 ± 2.09	91.37 ± 1.23	92.36 ± 2.68	94.71 ± 2.52	95.28 ± 0.92	96.23 ± 1.09	96.98 ± 4.06	97.24 ± 3.49	98.64 ± 0.28
12	92.91 ± 1.48	93.97 ± 0.15	94.57 ± 0.19	95.19 ± 0.45	95.60 ± 1.01	96.34 ± 2.45	97.73 ± 1.52	98.52 ± 0.67	98.97 ± 2.02
13	97.45 ± 2.37	98.22 ± 2.65	94.07 ± 1.09	96.96 ± 0.54	97.06 ± 0.37	96.63 ± 0.28	96.87 ± 4.26	97.16 ± 3.69	98.21 ± 3.69
14	87.04 ± 1.68	94.35 ± 2.04	95.13 ± 0.76	96.58 ± 1.45	96.41 ± 0.24	97.33 ± 0.37	96.61 ± 1.37	96.82 ± 2.58	97.22 ± 4.56
15	68.87 ± 2.54	66.19 ± 4.23	91.57 ± 0.49	92.34 ± 0.67	93.79 ± 3.17	94.58 ± 0.89	94.91 ± 0.32	95.46 ± 1.49	96.08 ± 2.64
16	83.14 ± 0.65	80.78 ± 1.32	94.53 ± 2.73	95.53 ± 1.86	96.68 ± 2.33	96.92 ± 0.02	96.24 ± 3.65	97.68 ± 0.34	98.34 ± 6.19
OA(%)	86.04 ± 1.67	88.89 ± 0.34	92.24 ± 0.67	94.57 ± 0.28	95.78 ± 0.38	96.45 ± 0.13	96.98 ± 3.59	98.12 ± 2.03	99.05 ± 3.29
AA(%)	87.36 ± 0.54	90.35 ± 2.17	92.91 ± 0.56	93.73 ± 1.84	94.93 ± 1.92	95.50 ± 0.40	96.16 ± 1.49	97.24 ± 0.91	98.83 ± 2.48
100 K	88.54 ± 1.79	89.26 ± 4.05	92.35 ± 3.67	94.84 ± 1.13	95.79 ± 2.04	96.17 ± 0.23	96.95 ± 1.46	97.79 ± 2.55	99.26 ± 2.86

Table 4. Classification results on the Salinas dataset with the water absorption bands.

Method	RBF-SVM	EMP-SVM	CNN	ResNet	MLP-Mixer	RepMLP	DFFN	DMLP	DMLPFFN
1	81.54 ± 0.63	89.58 ± 0.61	90.68 ± 2.05	95.74 ± 1.28	93.28 ± 0.42	96.35 ± 0.21	95.32 ± 2.64	96.25 ± 0.35	99.20 ± 2.84
2	87.49 ± 1.36	92.36 ± 0.23	90.35 ± 1.14	93.68 ± 1.35	92.11 ± 1.93	95.17 ± 0.16	95.43 ± 0.31	96.58 ± 2.86	98.63 ± 1.30
3	85.02 ± 2.07	74.13 ± 0.35	75.27 ± 2.21	93.87 ± 0.49	92.03 ± 0.57	93.89 ± 0.82	93.17 ± 1.45	94.79 ± 1.43	96.09 ± 0.15
4	90.16 ± 3.36	90.04 ± 2.02	91.64 ± 1.08	93.91 ± 1.23	94.32 ± 1.04	95.76 ± 0.68	96.34 ± 0.25	96.24 ± 0.68	99.31 ± 1.46
5	82.65 ± 1.94	88.75 ± 0.57	91.96 ± 1.32	89.66 ± 1.65	94.40 ± 0.63	95.62 ± 1.27	95.22 ± 2.93	97.35 ± 2.30	98.82 ± 1.61
6	89.31 ± 2.02	91.34 ± 0.64	91.85 ± 1.55	92.29 ± 0.67	95.32 ± 1.51	95.33 ± 0.51	95.17 ± 3.32	96.82 ± 0.35	98.38 ± 2.83
7	89.15 ± 0.41	90.47 ± 1.33	90.54 ± 2.29	91.38 ± 1.34	93.49 ± 0.37	95.29 ± 1.48	94.72 ± 3.41	95.62 ± 1.22	96.10 ± 0.59
8	81.22 ± 2.35	82.73 ± 0.41	87.32 ± 1.23	92.45 ± 0.47	91.97 ± 0.50	92.12 ± 0.75	94.81 ± 3.37	95.34 ± 1.34	98.51 ± 2.61
9	87.25 ± 0.82	87.39 ± 1.21	90.63 ± 0.46	90.84 ± 0.61	91.86 ± 0.65	93.94 ± 1.53	95.57 ± 5.07	95.48 ± 0.62	97.76 ± 1.62
10	77.95 ± 1.15	86.52 ± 1.30	88.47 ± 1.15	87.91 ± 0.83	92.46 ± 1.79	92.48 ± 0.62	92.69 ± 1.02	95.23 ± 0.10	97.37 ± 1.10

Table 4. Cont.

Method	RBF-SVM	EMP-SVM	CNN	ResNet	MLP-Mixer	RepMLP	DFFN	DMLP	DMLPFFN
11	81.19 ± 1.61	87.97 ± 1.27	88.65 ± 2.68	85.68 ± 1.26	91.75 ± 1.82	93.83 ± 0.14	94.05 ± 4.25	95.64 ± 2.31	98.82 ± 2.23
	1.53	0.45	0.19	3.25	0.04	93.35 ± 1.16	95.92 ± 1.67	97.31 ± 0.26	98.43 ± 1.68
13	92.57 ± 2.08	92.34 ± 2.24	90.01 ± 1.09	86.56 ± 1.39	94.57 ± 0.57	93.69 ± 0.28	94.38 ± 4.39	95.65 ± 3.25	97.35 ± 2.46
	87.08 ± 1.69	90.07 ± 2.05	90.65 ± 0.76	89.33 ± 2.75	93.16 ± 1.32	95.73 ± 1.20	94.25 ± 1.83	95.64 ± 1.62	97.68 ± 3.65
15	64.38 ± 2.32	61.35 ± 2.78	87.35 ± 0.49	90.93 ± 0.86	90.35 ± 2.53	91.66 ± 0.82	92.73 ± 0.15	93.34 ± 1.51	98.12 ± 2.83
	78.67 ± 0.94	76.64 ± 1.46	90.16 ± 2.73	89.68 ± 0.52	93.24 ± 1.82	93.41 ± 0.97	94.21 ± 3.46	96.23 ± 0.54	97.05 ± 1.25
OA(%)	81.21 ± 1.42	83.90 ± 0.62	88.49 ± 0.56	91.65 ± 0.32	92.47 ± 0.27	93.22 ± 0.53	94.30 ± 3.34	96.32 ± 1.64	98.10 ± 1.41
	82.13 ± 0.57	86.23 ± 2.13	88.46 ± 0.94	92.18 ± 0.96	91.24 ± 1.83	92.91 ± 0.31	94.08 ± 1.31	95.15 ± 0.48	97.23 ± 2.06
100 K	84.02 ± 1.62	85.46 ± 2.84	88.65 ± 2.48	90.68 ± 0.84	92.82 ± 1.24	93.87 ± 0.48	94.82 ± 1.02	96.34 ± 2.23	98.63 ± 1.37

Table 5. Classification results on the KSC dataset by different classification methods.

Method	RBF-SVM	EMP-SVM	CNN	ResNet	MLP-Mixer	RepMLP	DFFN	DMLP	DMLPFFN
1	89.59 ± 3.05	90.24 ± 1.68	91.75 ± 0.21	93.68 ± 3.74	93.24 ± 0.43	94.63 ± 0.65	94.98 ± 1.45	95.82 ± 3.06	97.25 ± 4.20
	80.25 ± 1.52	82.66 ± 0.85	86.69 ± 1.28	91.24 ± 2.80	94.24 ± 0.54	94.68 ± 0.49	95.81 ± 1.65	96.53 ± 3.55	96.99 ± 0.42
3	84.73 ± 0.64	85.91 ± 1.21	83.52 ± 0.98	87.67 ± 2.91	88.62 ± 3.72	89.98 ± 0.76	87.71 ± 1.24	91.36 ± 0.41	92.75 ± 3.28
	61.82 ± 3.44	63.75 ± 0.56	72.22 ± 0.52	81.08 ± 2.64	84.01 ± 1.91	86.02 ± 1.64	86.71 ± 0.68	89.52 ± 2.06	91.02 ± 1.56
5	61.56 ± 0.34	63.42 ± 4.57	71.09 ± 2.90	78.50 ± 1.63	82.55 ± 2.67	84.15 ± 1.53	85.53 ± 0.16	87.57 ± 0.46	89.59 ± 3.46
	66.38 ± 0.54	69.65 ± 3.10	70.24 ± 1.24	77.43 ± 0.93	85.15 ± 2.24	90.80 ± 2.35	89.62 ± 3.58	92.47 ± 3.05	94.56 ± 1.48
7	62.29 ± 0.66	66.56 ± 3.36	69.95 ± 4.02	83.88 ± 1.90	84.70 ± 0.23	85.68 ± 1.85	86.06 ± 2.89	88.40 ± 3.93	90.77 ± 4.51
	70.25 ± 1.48	74.82 ± 0.98	79.60 ± 4.22	92.10 ± 0.76	95.17 ± 0.93	96.52 ± 0.19	95.25 ± 0.35	97.88 ± 4.03	98.67 ± 3.51
9	82.64 ± 1.43	86.32 ± 2.36	89.94 ± 0.48	93.93 ± 1.30	94.78 ± 0.94	95.82 ± 4.24	95.94 ± 3.67	96.81 ± 0.79	97.69 ± 3.04
	88.78 ± 1.84	89.25 ± 1.22	91.52 ± 0.98	94.77 ± 1.34	96.30 ± 0.05	97.48 ± 0.38	96.24 ± 2.55	98.87 ± 1.29	99.04 ± 3.46
11	89.65 ± 0.46	91.38 ± 2.01	95.91 ± 3.55	96.51 ± 0.48	95.54 ± 3.06	96.98 ± 2.91	96.41 ± 1.68	97.03 ± 3.57	98.57 ± 2.11
	88.35 ± 2.19	91.01 ± 0.58	93.39 ± 2.20	95.09 ± 3.95	96.30 ± 1.47	94.84 ± 0.91	95.62 ± 0.85	96.87 ± 0.24	97.88 ± 4.62
13	92.26 ± 0.24	93.31 ± 0.32	95.84 ± 0.04	96.65 ± 0.05	96.28 ± 0.18	97.85 ± 0.33	96.81 ± 2.76	98.63 ± 3.28	99.35 ± 2.16
	81.65 ± 2.08	83.97 ± 0.27	86.04 ± 1.62	90.75 ± 3.54	93.41 ± 1.08	94.93 ± 3.83	95.82 ± 0.14	96.76 ± 1.73	98.49 ± 2.64
AA(%)	79.91 ± 1.63	82.57 ± 3.21	86.05 ± 2.56	89.11 ± 4.06	92.35 ± 2.16	93.18 ± 1.74	94.21 ± 2.03	95.24 ± 3.25	97.65 ± 4.26
	78.39 ± 2.46	80.98 ± 1.31	84.67 ± 5.78	88.86 ± 0.96	93.16 ± 2.04	94.35 ± 1.98	94.05 ± 3.72	96.22 ± 1.28	97.83 ± 3.29

Taking Salinas dataset as an example, compared with RBF-SVM, OA, AA and Kappa coefficients of DMLPFFN increased by 13.01%, 11.47% and 10.72%, and improved by 2.07%, 2.67% and 2.31% compared with DFFN, respectively. Taking the KSC dataset as an example, OA reached 98.49%, compared with RBF-SVM, EMP-SVM, CNN, ResNet, MLP-Mixer, RepMLP and DFFN, increased by 16.84%, 14.52%, 12.45%, 7.74%, 5.08%, 3.56%, 2.67% and 1.73% respectively. AA reached 97.65%, compared with DFFN, RepMLP, MLP-Mixer, ResNet, CNN, EMP-SVM and RBF-SVM, increased by 2.41%, 3.44%, 4.47%, 5.30%, 8.54%, 11.60%, 15.08% and 17.74%. All the experimental results show that the proposed DMLPFFN is superior to other methods.

In order to fully analyze the effect of the water absorption band on the experimental results, we downloaded the Salinas dataset with the water absorption band from the official website and conducted experimental analysis on it. As shown in Table 4, compared with RBF-SVM, OA, AA and Kappa coefficients of DMLPFFN increased by 16.89%, 15.10% and 14.61%, and improved by 3.80%, 3.15%, and 3.81% compared with DFFN, respectively. All the experimental results show that the proposed DMLPFFN is superior to other methods on the Salinas dataset with the water absorption bands.

Besides the quantitative classification results reported, we simultaneously visualized the classification maps of different methods discussed above, as shown in Figures 7 and 8.

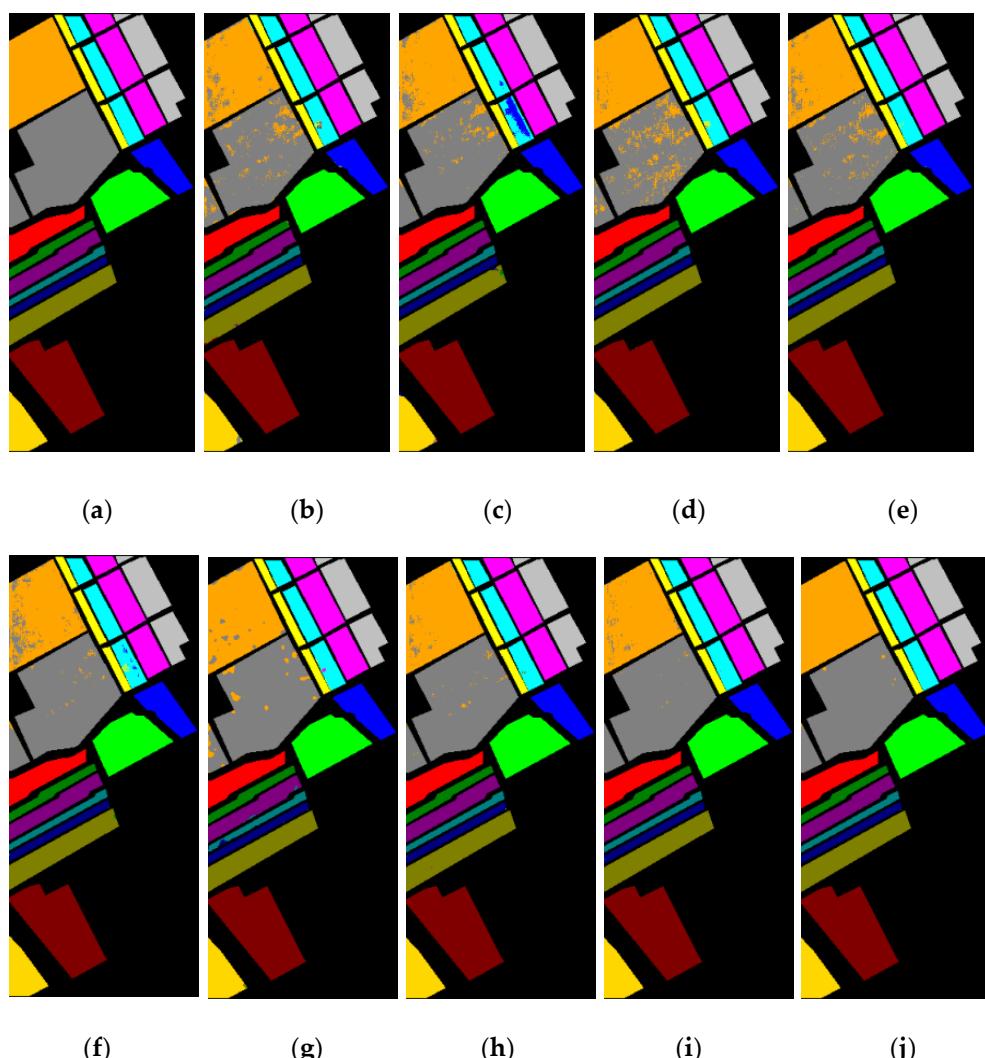


Figure 7. The classification results of Salinas dataset. (a) Ground Truth; (b) RBF-SVM; (c) EMP-SVM; (d) CNN; (e) ResNet; (f) MLP-Mixer; (g) RepMLP; (h) DFFN; (i) DMLP; (j) DMLPFFN.

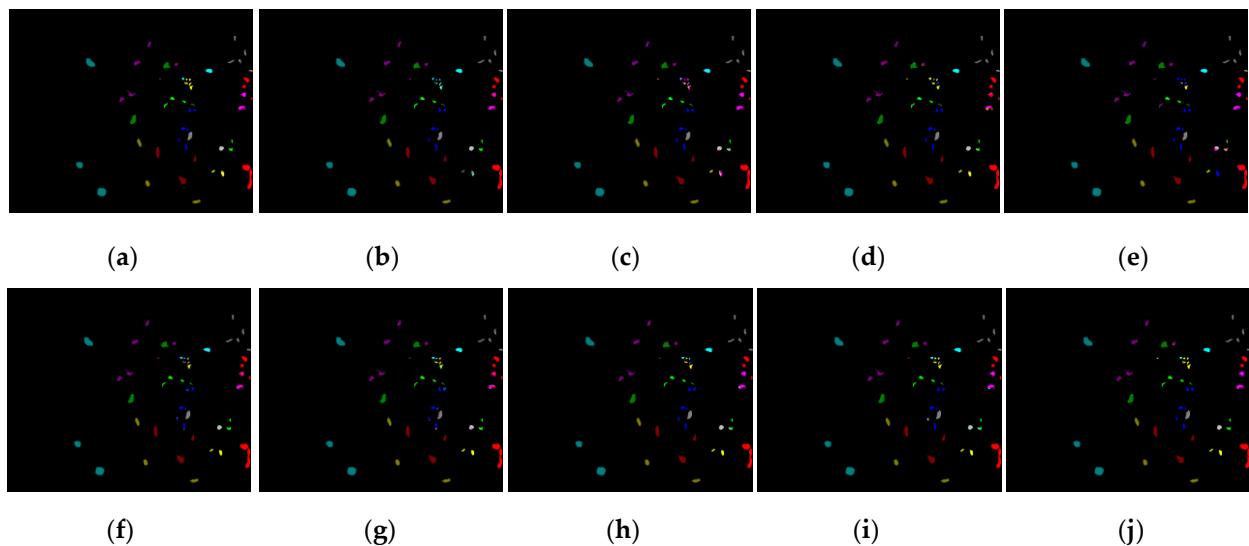


Figure 8. The classification results of KSC dataset. (a) Ground Truth; (b) RBF-SVM; (c) EMP-SVM; (d) CNN; (e) ResNet; (f) MLP-Mixer; (g) RepMLP; (h) DFFN; (i) DMLP; (j) DMLPFFN.

Obviously, it can be seen that RBF-SVM results have the most misclassified pixels in all classification maps, with many pretzel noises throughout, and each part is unavoidable for the classification confusion. Taking the dataset of Salinas as an example, in Figure 7b–d, a large amount of noise is generated in the upper left corner. Part of the area Vinyard untrained was misclassified as grapes_untrained. The classification confusion of Grapes_untrained, Grapes_untrained and Fallow_rough_pow in the middle part is serious. Compared with SVM, CNN and ResNet classification methods, the classification effect of MLP-Mixer, RepMLP and DFFN is improved, but there are still some misclassification phenomena. In addition, Figure 7i,j are the classification renderings of our algorithm; an obvious observation is that the classification map of the proposed method is the closest to the reference ground truth, which produces less internal noise and a cleaner boundary. Experiments show that the proposed method can effectively extract more refined features from two kinds of datasets, and cross-dimensional information interaction focuses on more important features, thus improving the classification accuracy.

4. Application in Fine Classification of Crops

In order to verify the classification performance and generalization ability of the DMLP and DMLPFFN, the WHU-Hi-LongKou and WHU-Hi-HanChuan hyperspectral datasets were selected in this paper for fine crop classification [36,37].

The WHU-Hi-LongKou dataset is located in a simple agricultural area and was captured by an 8 mm focal length steeple-wall Headwall Nano-HyperSpec sensor equipped with a receiver Matrix 600 Pro UAV platform with six kinds of crops. The image size was 550×400 pixels, with 270 bands between 400 and 1000 nm. The WHU-Hi-HanChuan dataset was collected in HanChuan, Hubei Province, using a 17 mm focal length Headwall Nano-HyperSpec sensor installed on the Leica Aibot X6 UAV V1 platform. The trial area has seven kinds crops and size 1217×303 pixels with 274 bands ranging from 400 to 1000 nm. Tables 6 and 7 report the detailed number of pixels available in each class for the two datasets, respectively, and show the false-color composite image and Ground Truth map.

Table 6. WHU-Hi-LongKou Dataset Labeled Sample Counts.

No	Name	Color	Number	False-Color Map	Ground-Truth Map
1	Corn		34,511		
2	Cotton		8374		
3	Sesame		3031		
4	Broad-leaf soybean		63,212		
5	Narrow-leaf soybean		4151		
6	Rice		11,854		
7	Water		67,056		
8	Roads and houses		7124		
9	Mixed weed		5229		
Total Numbers			204,542		

Table 7. WHU-Hi- HanChuan Dataset Labeled Sample Counts.

No	Name	Color	Number	False-Color Map	Ground-Truth Map
1	Strawberry		44,735		
2	Cowpea		22,753		
3	Soybean		10,287		
4	Sorghum		5353		
5	Water spinach		1200		
6	Watermelon		4533		
7	Greens		5903		
8	Trees		17,978		
9	Grass		9469		
10	Red roof		10,516		
11	Gray roof		16,911		
12	Plastic		3679		
13	Bare soil		9116		
Total Numbers			257,530		

In LongKou dataset, soybean occupies a prominent position, and its plot is continuous and extensive. Sesame and cotton are interlaced around the corn planting field. As shown in Table 8, the OA of DMLPFFN obtained 99.16%. Among all classification groups, the classification method DMLPFFN proposed in this paper has the highest OA, AA and Kappa coefficients, reaching 99.16%, 98.59% and 96.88, respectively. Compared with RBF-SVM, EMP-SVM, CNN, ResNet, MLP-Mixer, RepMLP, DFFN and DMLP, OA increased by 10.00%, 6.95%, 4.38%, 3.53%, 2.84%, 1.58%, 1.19% and 0.91%, respectively.

As shown in Table 9, in the HanChuan dataset, there are only a small number of soybean samples with 1335 pixels, thus affecting the classification effect of various algorithms. For soybean with poor classification performance in other algorithms, the accuracy of the two algorithms proposed in this paper can reach 93.37% and 94.16%, indicating that DMLP and DMLPFFN algorithms are suitable for separating similar ground objects. The methods proposed in this paper effectively solve the problem of spectral variation and heterogeneity within the same object.

Table 8. Classification results on the LongKou dataset by different classification methods.

Table 9. Classification results on the HanChuan dataset by different classification methods.

Method	RBF-SVM	EMP-SVM	CNN	ResNet	MLP-Mixer	RepMLP	DFFN	DMLP	DMLPFFN
1	80.25 ± 0.12	82.62 ± 3.61	88.34 ± 3.49	90.91 ± 1.96	91.67 ± 1.51	93.89 ± 0.57	92.05 ± 2.06	94.25 ± 1.87	96.33 ± 1.28
	64.22 ± 2.02	70.86 ± 4.09	76.15 ± 2.36	83.04 ± 1.49	85.26 ± 6.24	86.12 ± 2.48	88.38 ± 0.74	90.02 ± 1.23	92.97 ± 2.29
3	73.27 ± 1.28	78.15 ± 2.81	85.37 ± 3.63	91.65 ± 1.91	90.95 ± 1.25	92.39 ± 2.36	92.65 ± 3.28	93.37 ± 3.68	94.16 ± 1.61
	88.02 ± 0.88	89.34 ± 2.69	92.06 ± 1.67	94.65 ± 5.66	93.38 ± 3.97	94.93 ± 4.23	95.54 ± 1.08	95.37 ± 1.06	97.24 ± 0.98
5	78.22 ± 3.56	83.37 ± 1.63	89.38 ± 1.06	93.98 ± 3.26	94.33 ± 4.15	95.21 ± 1.89	95.48 ± 4.73	94.07 ± 1.36	95.19 ± 4.11
	70.52 ± 4.82	84.39 ± 2.57	86.38 ± 4.39	89.32 ± 2.32	90.72 ± 3.15	91.26 ± 1.37	92.22 ± 0.27	92.18 ± 3.03	93.51 ± 0.88
7	69.22 ± 2.67	72.38 ± 0.31	86.31 ± 0.98	90.70 ± 3.82	91.31 ± 10.97	92.64 ± 2.06	93.46 ± 4.79	93.37 ± 0.46	95.07 ± 1.54
	72.02 ± 5.92	74.20 ± 1.58	77.27 ± 3.18	82.08 ± 6.67	84.08 ± 4.37	86.10 ± 2.70	89.22 ± 3.17	90.56 ± 2.30	92.24 ± 1.85
9	82.20 ± 3.52	81.26 ± 4.54	88.09 ± 0.95	91.73 ± 5.95	89.90 ± 1.65	92.84 ± 1.19	91.34 ± 3.29	93.06 ± 3.75	94.51 ± 4.18
	85.22 ± 0.45	87.66 ± 3.10	89.09 ± 2.16	91.33 ± 5.14	92.98 ± 7.37	93.12 ± 0.58	94.05 ± 2.44	94.34 ± 2.88	95.46 ± 3.56
11	84.27 ± 3.74	86.57 ± 1.93	91.37 ± 1.06	94.36 ± 0.96	94.72 ± 2.61	93.71 ± 2.82	94.54 ± 1.28	94.09 ± 3.85	95.58 ± 2.60
	85.02 ± 4.31	87.34 ± 0.43	89.76 ± 0.41	90.17 ± 0.61	91.01 ± 0.61	92.70 ± 7.52	92.09 ± 5.06	93.45 ± 0.14	94.02 ± 1.03

Table 9. *Cont.*

Method	RBF-SVM	EMP-SVM	CNN	ResNet	MLP-Mixer	RepMLP	DFFN	DMLP	DMLPFFN
13	72.22 ± 2.59	79.61 ± 0.39	86.05 ± 3.28	88.39 ± 1.22	90.91 ± 2.54	91.89 ± 2.93	92.16 ± 3.08	93.07 ± 4.39	94.78 ± 2.37
	1.02	75.17 ± 2.09	84.36 ± 1.02	88.03 ± 2.36	88.55 ± 1.89	90.95 ± 1.70	92.81 ± 1.46	93.03 ± 2.69	94.36 ± 2.09
14	81.22 ± 3.05	84.20 ± 1.43	95.06 ± 2.47	94.84 ± 1.45	95.75 ± 3.26	94.65 ± 3.16	95.89 ± 2.04	94.90 ± 1.88	95.33 ± 2.76
	86.63 ± 0.98	88.05 ± 3.27	93.67 ± 4.09	93.87 ± 2.93	92.65 ± 2.79	94.35 ± 4.59	95.73 ± 2.17	95.37 ± 2.63	97.13 ± 1.58
OA(%)	81.05 ± 1.43	84.64 ± 0.47	89.21 ± 1.43	91.66 ± 0.60	93.61 ± 3.49	95.46 ± 3.91	95.95 ± 2.27	96.38 ± 4.67	98.05 ± 4.63
	77.17 ± 2.58	81.76 ± 2.14	83.65 ± 0.48	85.83 ± 3.37	88.76 ± 2.35	90.27 ± 0.12	91.96 ± 0.25	93.66 ± 2.23	95.24 ± 1.73
AA(%)	79.93 ± 3.86	82.59 ± 4.75	88.93 ± 1.28	89.34 ± 0.69	90.61 ± 1.89	91.78 ± 1.08	92.34 ± 4.87	93.92 ± 0.27	94.88 ± 1.83

Experimental results corresponding to different classification algorithms are shown in Figures 9 and 10 for HanChuan and LongKou datasets, respectively. As shown in Figure 9, it can be seen that there are still a large number of “salt and pepper” noises in RBF-SVM and EMP-SVM. The classification results of CNN and ResNet methods showed that the noise was greatly reduced after considering the contextual information. The classification results of ResNet, MLP-Mixer, RepMLP and DFFN methods showed that a large part of the samples of strawberry, cowpea, soybean and water oat were incorrectly classified into other categories in the middle region of the dataset. This is due to the fact that sowing at the edge of the field is even less compact than sowing in the center. The network misclassified crops as other categories at the margins of some plots owing to the sparse distribution of plants causing leakage of bare land area. Moreover, soybeans and cowpeas are crops of the same origin and exhibit highly similar spectral properties in a certain wavelength range, carrying a negative burden on the classification. Nevertheless, in our approaches, there is barely any misclassification of dense plants in the marginal areas and in the center of the plots, indicating that our method effectively discriminates between confusing crop classes due to spectral variation.

The color and edge features extracted from the low-level branch enable one to distinguish more conveniently between different types of crops and amplify the differences between different crops, whereas the regional features extracted in the middle-level branch lead to more apparent boundaries between crops in different places and perform better identification of crop areas and non-crop areas. The global features extracted at the high level minimize the clutter between sophisticated backdrops and crops to a certain extent and provide a better assessment of the overall crop area. Multi-level feature fusion can sufficiently extract and leverage the feature information of crops and fine classifications of them. Consequently, DMLPFFN is considered suitable for fine crop classification.

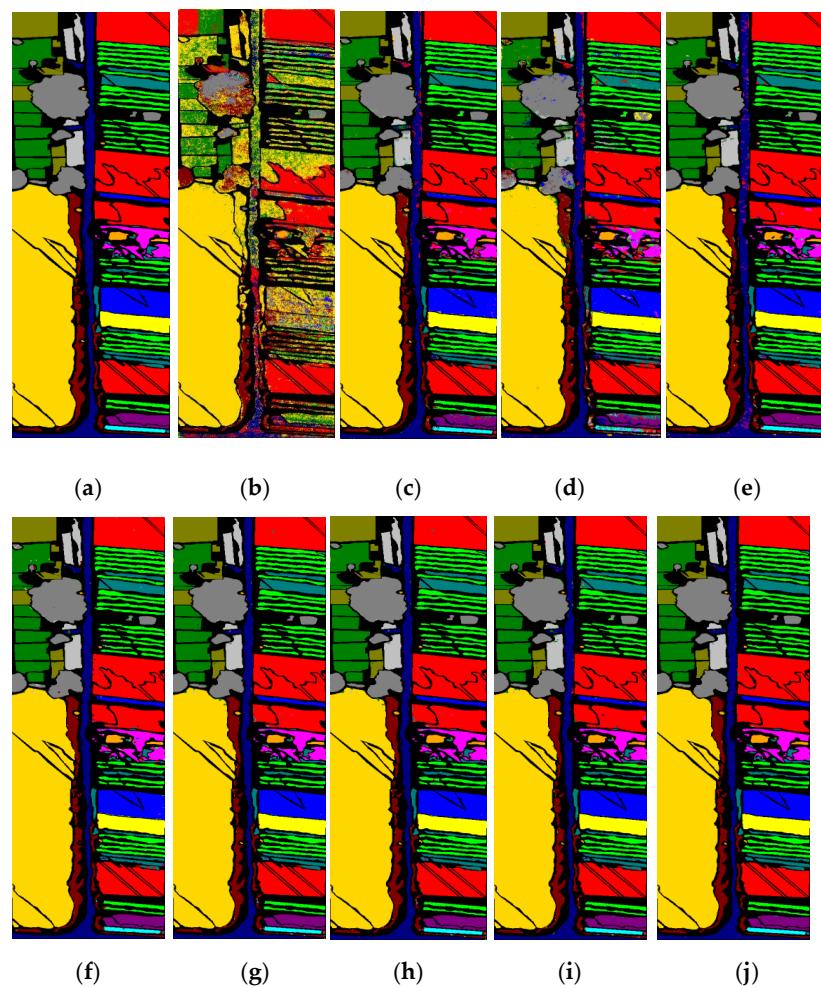


Figure 9. The classification results of HanChuan dataset. (a) Ground Truth; (b) RBF-SVM; (c) EMP-SVM; (d) CNN; (e) ResNet; (f) MLP-Mixer; (g) RepMLP; (h) DFFN; (i) DMLP; (j) DMLPFFN.

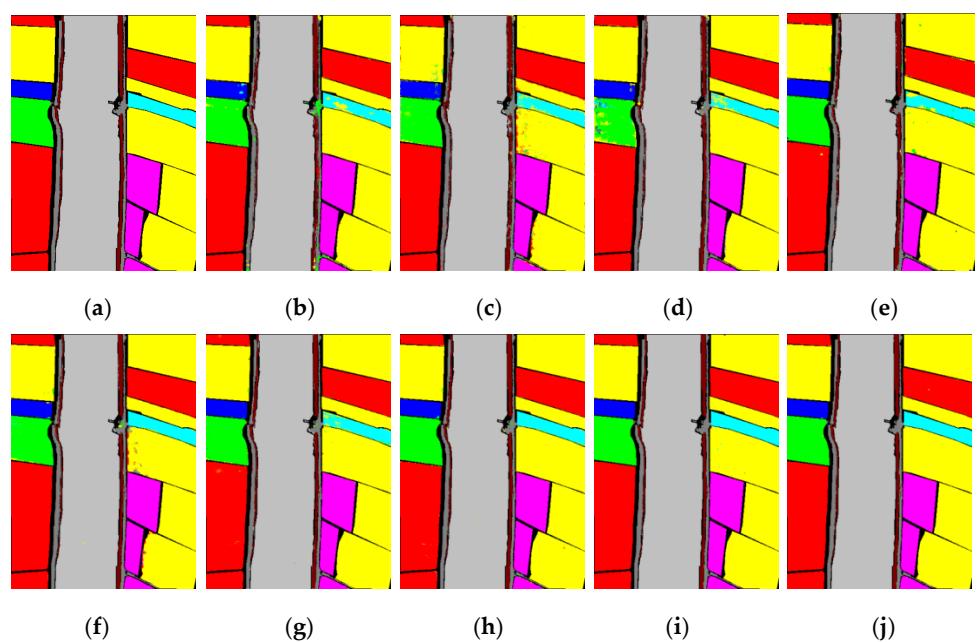


Figure 10. The classification results of LongKou dataset. (a) Ground Truth; (b) RBF-SVM; (c) EMP-SVM; (d) CNN; (e) ResNet; (f) MLP-Mixer; (g) RepMLP; (h) DFFN; (i) DMLP; (j) DMLPFFN.

5. Discussion

In order to find the optimal network structure, it is necessary to experiment with different parameters, which play a crucial role in the size of the model and the complexity of the proposed DMLPFFN. In this paper, the optimal parameter combination is determined by analyzing the influence of parameters on the accuracy of classification results, including the number of PCAs, the expansion rate of dilated convolution, the percentage of training samples and the number of branches in the feature fusion strategy.

5.1. The Number of Principal Components

The first parameter is the number of principal components selected for PCA on HSI, which is used to extract the main spectral components to improve the algorithm's efficiency and reduce noise interference. In the case of principal component number, the control variable method is used for all datasets in the experiment. That is to say, the value of training sample number, expansion rate and deep feature fusion strategy is fixed. As shown in Figure 11, OA increased and then tended to be stable with the increase of the principal component number for the four HSI datasets. Most of the information in the hyperspectral image exists in the first few principal components. However, it was concluded that using many principal components did not further improve performance.

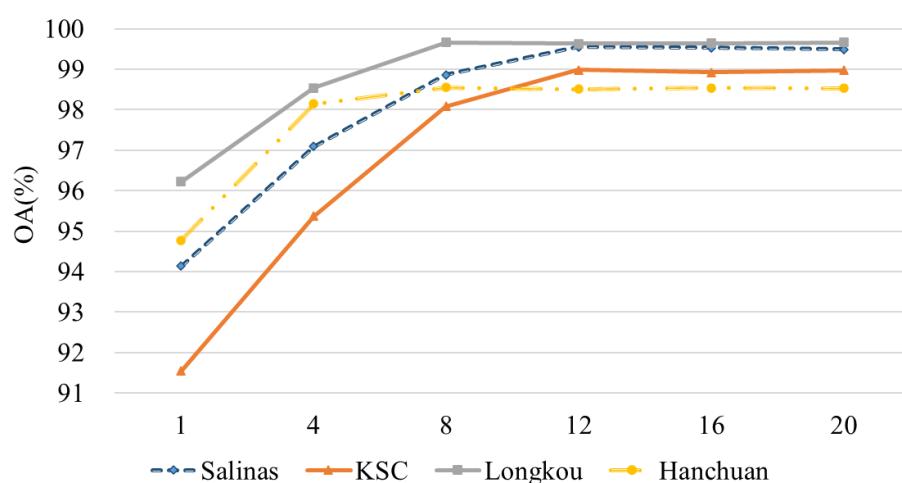


Figure 11. Results of DMLPFFN with different numbers of principal components.

5.2. The Expansion Rate of Dilated Convolution

The second parameter is the distribution of the expansion rate. In this experiment, seven circulation structures with expansion rate distributions of [1,1,2], [2,2,2], [1,2,2], [1,2,3], [1,2,4], [1,2,5] and [1,2,6] are selected for comparative analysis, as shown in Figure 12.

By comparing the experimental results, it can be found that the classification accuracy of the expansion rate distributions of [1,1,2] is lower than that of the expansion rate distributions of [1,2,2]. The receptive field size of [1,1,2] is 9×9 . In [2,2,2], although the range of the receptive field increases to 13×13 , the classification accuracy is lower than the average overall accuracy of the expansion rate distributions of [1,1,2]. This is because the superposition of three dilated convolutions will result in more feature information being omitted.

The latter five experiments used the combination of two dilated convolution layers and one ordinary convolution layer, and the receptive fields were 11×11 , 13×13 , 15×15 , 17×17 and 19×19 . Nonetheless, [1,2,5] has the largest range of receptive fields; with the increase of expansion rate, the data of input sampling become more and more sparse, resulting in local information loss and damage to information continuity. According to the experimental results in Figure 12, when the expansion rate distribution is [1,2,5], the four HSI datasets can obtain the optimal classification results.

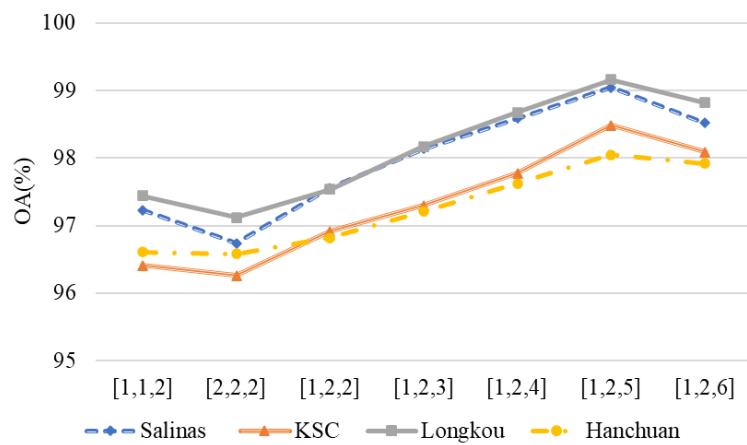


Figure 12. Results of DMLPFFN with different numbers of expansion rates.

5.3. The Percentage of Training Samples

The third parameter is the proportion of training samples to the total number of samples. We carried out experiments on the practical crop hyperspectral datasets LongKou and HanChuan, as shown in Figure 13. 0.4%, 0.6%, 0.8%, 1.0% and 1.2% of LongKou and HanChuan dataset training samples were selected for experiment, respectively. At the beginning, the classification accuracy increased with training samples. When the training sample of LongKou and HanChuan datasets was 1.2%, the OA value basically reached the highest point and then tended to be flat or even showed a downward trend. When the number of training samples reaches the required level, it can precisely illustrate the distribution of all pixels in the studied area; continuously increasing the number of training samples will not increase the classification accuracy. Therefore, 1.2% is chosen as the percentage of training samples, and the proposed DMLPFFN method always provides better performance than other comparison methods.

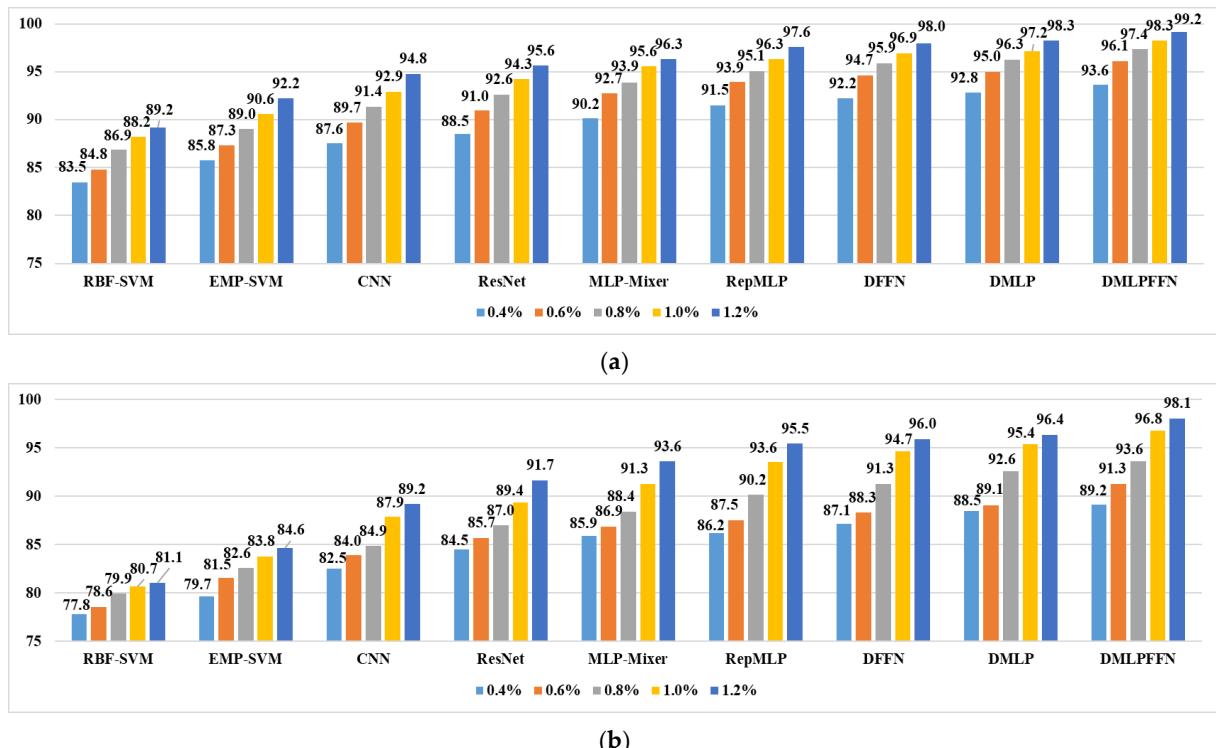


Figure 13. Comparison of the different number of training samples under different methods.
(a) LongKou; **(b)** HanChuan.

5.4. The Number of Branches in Feature Fusion Strategy

The fourth parameter is the number of branches in feature fusion strategy. This paper analyzes the correlation and complementarity of information in the deep network using multibranch feature fusion. DMLPFFN2, DMLPFFN3, DMLPFFN4 and DMLPFFN5 refer to methods that fuse two, three, four and five hierarchical branches. Among them, DMLPFFN2 represents the fusion of lower-level and higher-level sorts. It can be seen from Figure 14 that in different datasets, DMLPFFN3 obtained precision values that are superior to DMLPFFN2, DMLPFFN4 and DMLPFFN5. In addition, taking the LongKou dataset as an example, compared with the DMLPFFN2, the OA, AA and Kappa values of the DMLPFFN3 fusion strategy increased by 3.05%, 9.5%, and 1.97%, respectively. That is because the features extracted by DMLPFFN2 contain only details and global information, and regional feature information is dropped. To some extent, fusing multiple layers improves classification results. However, DMLPFFN5 has the lowest classification accuracy, which shows that too many fusion layers may bring redundant information and significantly reduce the performance. Specifically, middle-level information overlap can cause accuracy degradation. So the DMLPFFN method proposed in this paper used three branches for feature fusion and the structure is as shown in Figure 1.

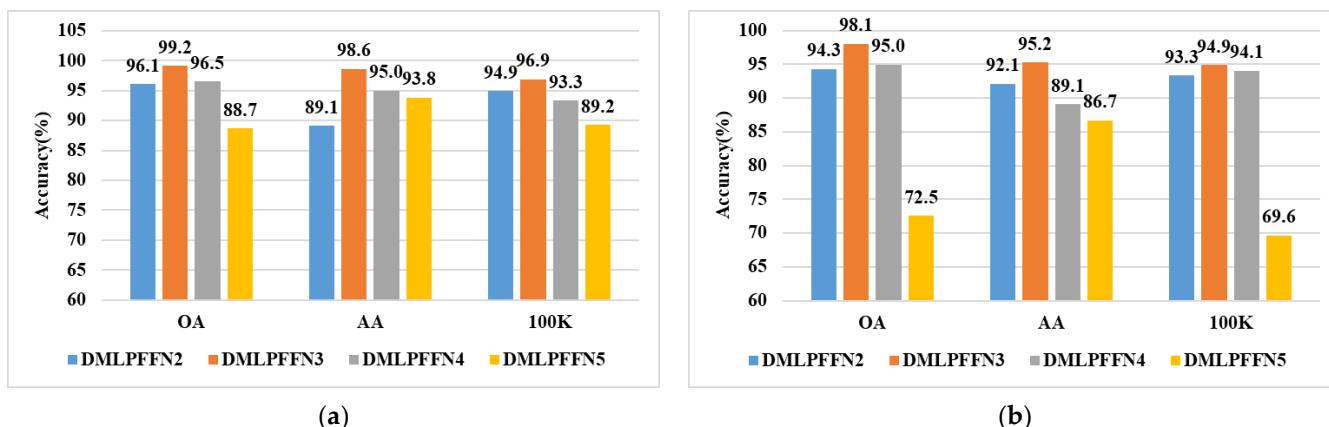


Figure 14. Comparison of the different branch combinations in feature fusion strategy. (a) LongKou; (b) HanChuan.

5.5. The Number of Classes for HSI Classification

We conducted experiments on the KSC dataset for different numbers of classes. The KSC dataset was picked up by AVIRIS sensors flying over the Kennedy Space in Florida. The number of spectral bands is 176, and the size is 512×614 pixels with 13 classes. Table 10 shows the OA, AA and Kappa values of the DMLPFFN method when the number of classes is 10, 11, 12 and 13. The results show that the accuracy of the result decreases when the number of classes reduces. The highest precision is achieved when the number of classes is 13, which is the original number of classes. This shows that if the number of classes in the experimental dataset is not utilized sufficiently, it will lead to a decrease in the accuracy of the experimental results.

Table 10. Classification results on the KSC dataset with different numbers of classes with the DMLPFFN method.

Number of Classes	10 Classes	11 Classes	12 Classes	13 Classes
OA(%)	92.87 ± 1.63	94.26 ± 1.35	95.37 ± 1.74	98.60 ± 2.26
AA(%)	92.13 ± 1.82	94.38 ± 1.61	95.52 ± 1.36	98.65 ± 1.57
100 K	90.47 ± 0.68	93.75 ± 1.87	94.94 ± 1.79	97.83 ± 1.65

5.6. Time Consumption and Computational Complexity

In order to comprehensively analyze the methods proposed in this paper and current research methods, this paper analyzes the average training time, average test time and total parameters of different methods. Table 11 reports the time consumption and computational complexity of different methods.

Table 11. Comparison of time consumption and computational complexity of different classification methods.

Datasets	Methods	Training Time (s)	Test Time (s)	Parameters (M)	OA(%)
Long Kou	CNN	56.37	3.82	3.29	94.78
	ResNet	1150.61	215.12	22.12	95.63
	MLP-Mixer	421.29	61.63	5.81	96.32
	RepMLP	418.29	66.42	7.84	97.58
	DFFN	111.7	7.95	8.55	97.97
	DMLP	459.71	71.24	6.36	98.25
	DMLPFFN	83.35	5.94	9.86	99.16
Han Chuan	CNN	71.09	2.86	3.48	89.21
	ResNet	1233.39	484.61	22.15	91.66
	MLP-Mixer	586.49	97.87	5.14	93.61
	RepMLP	471.27	75.79	6.83	95.46
	DFFN	201.76	15.78	7.96	95.95
	DMLP	497.61	51.16	5.26	96.38
	DMLPFFN	112.26	9.51	8.31	98.05

In terms of running time, taking the HanChuan dataset as an example, although DMLP has a larger receptive field to extract more delicate features and consumes more training time than RepMLP, the total parameters are reduced by 22.98%. Moreover, compared with ResNet and the MLP-Mixer, the training times of DMLP are reduced by 59.65% and 15.15%, and DMLP has better classification accuracy. The results show that, compared with ResNet, DMLP and DMLPFFN have fewer parameters on all datasets. Compared with CNN and the MLP-Mixer, the proposed method has a few more parameters because of its greater depth and width, but the accuracy of the proposed method is the highest. Moreover, compared with DFFN, DMLPFFN has a shorter training time on four datasets because DMLPFFN improves the training performance of the model by combining the fusion strategy with MLP. Taking the LongKou dataset as an example, DMLPFFN training time and test time are reduced by 25.20% and 25.28%, respectively, compared with DFFN. In addition, DMLPFFN has the lowest training time and test time next to CNN among all deep learning methods and achieves better OA than other classification algorithms.

6. Conclusions

In this paper, two classification frameworks based on MLP are proposed: DMLP and DMLPFFN. Firstly, in order to expand the perceptual field and aggregate multi-branch contextual information and avoid losing the feature map resolution, we introduced a dilated convolution layer instead of ordinary convolution. Secondly, for the purpose of fully utilizing the features of HSI to improve the classification efficiency, we use fusing residual blocks and the DMLP mechanism to extract deeper features and obtain the state-of-the-art performance. Finally, we designed comprehensive experiments and executed them to prove the effectiveness of DMLPFFN by different hyperspectral datasets and to prove that it has better classification performance and generalization ability for agricultural classification.

The proposed DMLP and DMLPFFN were tested on two public datasets (Salinas and KSC) and two real HSI datasets (LongKou and HanChuan). Compared with the classical methods (RBF-SVM and EMP-SVM) and deep learning-based methods (CNN, ResNet, MLP-Mixer, RepMLP and DFFN), the experiments show that the proposed DMLP algorithm

and DMLPFFN algorithm are meaningful and can obtain better classification results. We also validate the classification performance and generalization ability of DMLPFFN in fine crop classification, which contributes to promoting the specific application of hyperspectral remote sensing technology in agricultural development.

However, in the task of hyperspectral image classification, the available marker samples are usually very limited. When analyzing the classification effect of the number of training samples, take the KSC dataset as an example; the DMLPFFN proposed in this paper reveals that an effectiveness of 10% of the sample size is superior to the others. As a future step, we are conducting further experiments to probe the prevalence suitability of DMLPFFN in small sample cases.

Author Contributions: Conceptualization, A.W., H.Z. and H.W.; methodology, software, validation, H.Z.; writing—review and editing, H.W. and A.W.; supervision, Y.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China under Grant NSFC-61671190.

Data Availability Statement: The data are available at <https://www.ehu.eus/ccwintco/index.php?%20title=Hyperspectral-Remote-Sensing-Scenes> (accessed on 22 April 2022); http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm (accessed on 22 April 2022).

Acknowledgments: We thank for Kaiyuan Jiang for his valuable comments and discussion. Iwahori's research is supported by JSPS Grant-in-Aid for Scientific Research (C) (20K11873) and Chubu University Grant.

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

References

1. Czaja, W.; Kavalerov, I.; Li, W. Exploring the High Dimensional Geometry of HSI Features. In Proceedings of the 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 March 2021; pp. 1–5.
2. Zhang, Y.; Wang, D.; Zhou, Q. Advances in crop fine classification based on Hyperspectral Remote Sensing. In Proceedings of the 2019 8th International Conference on Agro-Geoinformatics, Istanbul, Turkey, 16–19 July 2019; pp. 1–6.
3. Kim, Y.; Kim, Y. Hyperspectral Image Classification Based on Spectral Mixture Analysis for Crop Type Determination. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 23–27 July 2018; pp. 5304–5307.
4. Spiller, D.; Ansalone, L.; Carotenuto, F.; Mathieu, P.P. Crop Type Mapping Using Prisma Hyperspectral Images and One-Dimensional Convolutional Neural Network. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 8166–8169.
5. Pignatti, S.; Casa, R.; Harfouche, A.; Huang, W.; Palombo, A.; Pascucci, S. Maize Crop and Weeds Species Detection by Using Uav Vnir Hyperpectral Data. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 7235–7238.
6. Kefauver, S.C.; Romero, A.G.; Buchaillot, M.L.; Vergara-Díaz, O.; Fernandez-Gallego, J.A.; El-Haddad, G.; Akl, A.; Araus, J.L. Open-Source Software for Crop Physiological Assessments Using High Resolution RGB Images. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 4359–4362.
7. Liu, C.; Li, M.; Liu, Y.; Chen, J.; Shen, C. Application of Adaboost based ensemble SVM on IKONOS image Classification. In Proceedings of the 2010 18th International Conference on Geoinformatics, Beijing, China, 18–20 June 2010; pp. 1–5.
8. Cuozzo, G.; D'Elia, C.; Puzzolo, V. A method based on tree-structured Markov random field for forest area classification. In Proceedings of the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; Volume 4, pp. 2352–2354.
9. Li, Z.; Li, X.; Chen, E.; Li, S. A method integrating GF-1 multi-spectral and modis multitemporal NDVI data for forest land cover classification. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 3742–3745.
10. Delalieux, S.; Somers, B.; Haest, B.; Spanhove, T.; Borre, J.V.; Mücher, C.A. Heathland conservation status mapping through integration of hyperspectral mixture analysis and decision tree classifiers. *Remote Sens. Environ.* **2012**, *126*, 222–231. [[CrossRef](#)]
11. 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](#)]

12. Tao, C.; Pan, H.; Li, Y.; Zou, Z. Unsupervised Spectral–Spatial Feature Learning with Stacked Sparse Autoencoder for Hyperspectral Imagery Classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2438–2442.
13. Sun, Q.; Liu, X.; Fu, M. Classification of hyperspectral image based on principal component analysis and deep learning. In Proceedings of the 2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC), Shenzhen, China, 21–23 July 2017; pp. 356–359.
14. 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](#)]
15. Zhang, B.; Qing, C.; Xu, X.; Ren, J. Spatial Residual Blocks Combined Parallel Network for Hyperspectral Image Classification. *IEEE Access* **2020**, *8*, 74513–74524. [[CrossRef](#)]
16. Kanthi, M.; Sarma, T.H.; Bindu, C.S. A 3d-Deep CNN Based Feature Extraction and Hyperspectral Image Classification. In Proceedings of the 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), Virtual, 1–4 December 2020; pp. 229–232.
17. Zhang, H.; Yu, H.; Xu, Z.; Zheng, K.; Gao, L. A Novel Classification Framework for Hyperspectral Image Classification Based on Multi-Scale Dense Network. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 2238–2241.
18. Zhu, M.; Fan, J.; Yang, Q.; Chen, T. SC-EADNet: A Self-Supervised Contrastive Efficient Asymmetric Dilated Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 1–17. [[CrossRef](#)]
19. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. Mlp-mixer: An all-mlp architecture for vision. *arXiv* **2021**, arXiv:2105.01601.
20. Yu, T.; Li, X.; Cai, Y.; Sun, M.; Li, P. S2-MLP: Spatial-Shift MLP Architecture for Vision. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 4–8 January 2022; pp. 3615–3624.
21. Lian, D.; Yu, Z.; Sun, X.; Gao, S. AS-MLP: An Axial Shifted MLP Architecture for Vision. *arXiv* **2021**, arXiv:2107.08391.
22. Yu, T.; Li, X.; Cai, Y.; Sun, M.; Li, P. S²-MLPV2: Improved Spatial-Shift MLP Architecture for Vision. *arXiv* **2021**, arXiv:2108.01072.
23. Chen, S.; Xie, E.; Ge, C.; Liang, D.; Luo, P. CycLNMLP: A MLP-like Architecture for Dense Prediction. *arXiv* **2021**, arXiv:2107.10224.
24. Potghan, S.; Rajamenakshi, R.; Bhise, A. Multi-Layer Perceptron Based Lung Tumor Classification. In Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 29–31 March 2018; pp. 499–502.
25. Deng, F.; Bi, Y.; Liu, Y.; Yang, S. Deep-Learning-Based Remaining Useful Life Prediction Based on a Multi-Scale Dilated Convolution Network. *Mathematics* **2021**, *9*, 3035. [[CrossRef](#)]
26. Li, Z.; Wang, T.; Li, W.; Du, Q.; Wang, C.; Liu, C.; Shi, X. Deep Multilayer Fusion Dense Network for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1258–1270. [[CrossRef](#)]
27. Jiang, Y.; Li, Y.; Zou, S.; Zhang, H.; Bai, Y. Hyperspectral Image Classification with Spatial Consistence Using Fully Convolutional Spatial Propagation Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10425–10437. [[CrossRef](#)]
28. 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; pp. 464–469.
29. He, X.; Chen, Y. Transferring CNN Ensemble for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 876–880. [[CrossRef](#)]
30. Melgani, F.; Bruzzone, L. Support vector machines for classification of hyperspectral remote-sensing images. In *2002 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2002), Proceedings of the 24th Canadian Symposium on Remote Sensing, Toronto, ON, Canada, 24–28 June 2002*; IEEE: Piscataway Township, NJ, USA, 2002; Volume I.
31. Gu, Y.; Liu, T.; Jia, X.; Benediktsson, J.O.N.A.; Chanussot, J. Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification. *IEEE Trans. Geosci. Remote* **2016**, *54*, 3235–3247. [[CrossRef](#)]
32. Morchale, S.; Pauca, V.P.; Plemmons, R.J.; Torgersen, T.C. Classification of pixel-level fused hyperspectral and lidar data using deep convolutional neural networks. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016; pp. 1–5.
33. Liu, X.; Meng, Y.; Fu, M. Classification Research Based on Residual Network for Hyperspectral Image. In Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 19–21 July 2019; pp. 911–915.
34. Ding, X.; Xia, C.; Zhang, X.; Chu, X.; Han, J.; Ding, G. Repmlp: Reparameterizing convolutions into fully-connected layers for image recognition. *arXiv* **2021**, arXiv:2105.01883.
35. Song, W.; Li, S.; Fang, L.; Lu, T. Hyperspectral Image Classification With Deep Feature Fusion Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3173–3184. [[CrossRef](#)]
36. Zhong, Y.; Hu, X.; Luo, C.; Wang, X.; Zhao, J.; Zhang, L. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens. Environ.* **2020**, *250*, 112012. [[CrossRef](#)]
37. Zhong, Y.; Wang, X.; Xu, Y.; Wang, S.; Jia, T.; Hu, X.; Zhao, J.; Wei, L.; Zhang, L. Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications. *IEEE Geosci. Remote Sens. Mag.* **2018**, *6*, 46–62. [[CrossRef](#)]