remote sensing

MDPI

Article

Wide and Deep Fourier Neural Network for Hyperspectral Remote Sensing Image Classification

Jiangbo Xi [1,2,3,4,†], Okan K. Ersoy [5,†], Ming Cong [1,2,*], Chaoying Zhao [1,2], Wei Qu [1,2] and Tianjun Wu [6]

[1] College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China; xijiangbo@chd.edu.cn (J.X.); cyzhao@chd.edu.cn (C.Z.); quwei@chd.edu.cn (W.Q.)
[2] Key Laboratory of Western China's Mineral Resources and Geological Engineering, Ministry of Education, Xi'an 710054, China
[3] Big Data Center for Geosciences and Satellites (BDCGS), Chang'an University, Xi'an 710054, China
[4] Key Laboratory of Ecological Geology and Disaster Prevention, Ministry of Natural Resources, Xi'an 710054, China
[5] School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA; ersoy@purdue.edu
[6] Department of Mathematics and Information Science, College of Science, Chang'an University, Xi'an 710064, China; tjwu@chd.edu.cn
* Correspondence: mingc@chd.edu.cn
† These authors contributed equally to this work.

Abstract: Hyperspectral remote sensing image (HSI) classification is very useful in different applications, and recently, deep learning has been applied for HSI classification successfully. However, the number of training samples is usually limited, causing difficulty in use of very deep learning models. We propose a wide and deep Fourier network to learn features efficiently by using pruned features extracted in the frequency domain. It is composed of multiple wide Fourier layers to extract hierarchical features layer-by-layer efficiently. Each wide Fourier layer includes a large number of Fourier transforms to extract features in the frequency domain from a local spatial area using sliding windows with given strides.These extracted features are pruned to retain important features and reduce computations. The weights in the final fully connected layers are computed using least squares. The transform amplitudes are used for nonlinear processing with pruned features. The proposed method was evaluated with HSI datasets including Pavia University, KSC, and Salinas datasets. The overall accuracies (OAs) of the proposed method can reach 99.77%, 99.97%, and 99.95%, respectively. The average accuracies (AAs) can achieve 99.55%, 99.95%, and 99.95%, respectively. The Kappa coefficients are as high as 99.69%, 99.96%, and 99.94%, respectively. The experimental results show that the proposed method achieved excellent performance among other compared methods. The proposed method can be used for applications including classification, and image segmentation tasks, and has the ability to be implemented with lightweight embedded computing platforms. The future work is to improve the method to make it available for use in applications including object detection, time serial data prediction, and fast implementation.

Keywords: hyperspectral image classification; deep learning; convolutional neural network; frequency domain learning; deep fourier neural network

Citation: Xi, J.; Ersoy, O.K.; Cong, M.; Zhao, C.; Qu, W.; Wu, T. Wide and Deep Fourier Neural Network for Hyperspectral Remote Sensing Image Classification. *Remote Sens.* **2022**, *14*, 2931. https://doi.org/10.3390/rs14122931

Academic Editor: Pedro Melo-Pinto

Received: 4 May 2022
Accepted: 16 June 2022
Published: 19 June 2022

1. Introduction

Hyperspectral remote sensing images can be used for land cover classification on different applications such as agriculture, environmental protection, and water resources analysis. In the past 30 years, many different machine learning methods such as decision trees [1], support vector machine (SVM) [2], multilayer perceptron (MLP) [3], and random forest (RF) [4] have been used for pixel classification. These methods mainly use the spectral information of HSI.

Including full use of spectral and spatial information, deep learning has been used successfully for HSI classification [5]. Zhang et al. [6] described in detail how to use deep learning with different inputs and applications. Diverse region-based CNN [7] was proposed to obtain good classification results. Researchers also used different features to learn HSI images, including fast dense spectral–spatial convolution (FDSSC) [8], which learn different features with different convolutional kernels, multiscale convolution (MS-CNNs) [9] to learn discriminative features, a CNN to learn various morphological profiles [10], and the method to extract hierarchical deep spatial features from CNN [11]. Recently, three dimensional features were used for HSI classification, such as 3D CNN [12], hybrid spectral CNN [13], and mixed CNN [14], which combined spectral spatial 3D features with 2D spatial features. An expansion convolution network (ECNet) was proposed to inject holes into convolution kernels to expand the receptive field; therefore, more context features can be extracted [15].

Researchers also combined emerging new methods with CNN to improve HSI classification performance. The fully convolutional network (FCN) was improved by using an efficient nonlocal block for HSI classification [16]. The model reduces parameters with less degrading of classification performance using lightweight and attentional block. The attentional mechanism was also combined with 3D octave convolutional neural network [17]. The HSI classification performance of CNN was improved by introducing active learning [18]. The CNNs were also trained with transfer learning [19,20] for HSI classification.

Different learning models were also proposed by researchers for HSI classification other than CNN-based deep learning methods. Originating with recurrent neural network, long short-term memory (LSTM), transformer, developed originally for natural language processing, deep recurrent neural network [21], cascaded recurrent neural network [22], bidirectional–convolutional LSTM [23], and attention-based bidirectional LSTM [24] were proposed for HSI classification. A two-branches multidirectional spectral and spatial long short-term memory (LSTM) attention network [25] was also proposed for HSI classification, which uses LSTMs to extract six directional spatial spectral features, and extract spectral spatial relationships along different directions. The transformers were used for HSI classification with improved performance using both spatial and spectral features [26,27]. Very recently, a spectral spatial self-mutual-attention network (S3MANet) was proposed to enhance both spectral and spatial features through self mutual attention method [28].

CNNs were also combined with Fourier Transform to improve their performance. A Fourier domain acceleration method was proposed for convolutional neural networks [29]. An efficient spectral-based CNN model [30] was proposed that uses only the lower frequency components. Spectral dropout method [31] was proposed to prevent overfitting by eliminating weak and noisy Fourier domain coefficients of the neural network activations. Other types of original machine learning methods were proposed for different goals such as using different learning kernels, learning efficiency, and unbalanced data. Researchers proposed different transform based networks, especially, Fourier transform is usually used to develop learning models [32,33]. A new neural operator [34] by parameterizing the integral kernel directly in Fourier space to generate an expressive and efficient learning architecture. The Global Filter Network (GFNet) [35] was proposed by replacing the self-attention layer in vision transformers with 2D discrete Fourier related operations. Harmonic networks [36] were proposed, and they were further developed by naive Gabor networks [37] for HSI classification to reduce the number of learning parameters. The deep support vector machine (DSVM) [38] was proposed through extending SVM in the deep direction. A generative adversarial minority oversampling [39] was proposed to deal with unbalanced instances in HSI. A spectral–spatial attention feature extraction method using generative adversarial network (GAN) [40] was proposed for HSI classification.

Because the training samples of HSI are pretty limited, it is usually difficult to train a learning model with a large number of parameters without overfitting. Incremental learning (IL) and scalable learning (SL) can be used to solve this problem by generating learning models with proper parameters according to the complexity of the learning tasks.

They were originally designed to overcome catastrophic forgetting [41], to and reduce computational load [42,43]. IL models have been proposed by different researchers, for example elastic weight consolidation (EWC) [44], which remembers old knowledge through reducing important weights selectively, and incremental moment matching (IMM) [45], which assumes that the moment of the posteriors can be matched incrementally by the posterior distribution of the parameters in Bayesian neural networks. SL models mainly include the earlier work called parallel, self-organizing, hierarchical neural networks (PSHNNs) [46] and parallel consensual neural networks (PCNNs) [47] which combine multistage neural networks with the sample rejection or statistical consensus methods. Scalable effort classifiers [42,43] were also generated with increasing complexity of learning models. Another way to generate a learning model for HSI classification is using wide learning [48,49] or both wide and deep [50] learning. It is demonstrated that the wide fully connected neural networks can generalize better [48,49]. Researchers have used incremental learning for remote sensing image scene classification, for example, an incremental learning with open set recognition (ILOSR) framework [51] was proposed by identifying unknown classes from a data stream and learning new classes incrementally. Parallel multistage wide neural network (PMWNN) [3] was proposed to learn unbalanced image data effectively and incrementally, and it also has good performance on HSI classification. Scalable wide neural network (SWNN) [52] was proposed for classification including HSIs, which can generate wide neural networks incrementally. Wide sliding and subsampling network (WSWS Net) [53] was proposed for HSI classification, and it can be extended in both wide and deep directions to learn both spatial and spectral features efficiently. Dynamic wide and deep neural network (DWDNN) [54] was proposed, which can generate wide and deep learning models dynamically to classify HSI effectively with proper computational load.

It is hard and to use limited HSI training samples to generate deep learning models with very deep architecture, and considering with other new proposed learning models, how to learn HSI features efficiently based on the limited training samples is still a still key issue to be solved. In this paper, a wide and deep Fourier network (WD-FNet) is proposed to learn hierarchical features efficiently in the frequency domain. It is composed of multiple wide Fourier layers in the deep direction to extract features layer-by-layer, and more abstract features can be extracted as the number of layers increases. Each wide Fourier layer includes a large number of Fourier transforms generated in the wide direction using sliding windows with given strides. These Fourier transforms extract features from the spatial domain in the corresponding sliding windows, and the extracted features are pruned to reduce the computational load and improve generalization. The fully connected layer follows wide Fourier layers, and its weights can be learned fast using the least squares method. The contributions of the proposed WD-FNet are as follows:

- Learning both spectral and spatial fine grained features efficiently using a large number of Fourier transforms in the wide direction, with computational load reduced by using pruning and retaining most effective features;
- Extracting hierarchical abstract features layer-by-layer in the deep direction with wide Fourier layers efficiently with limited training samples for HSI classification;
- Learning the weights in the fully connected layer by using least squares, which makes the training process very simple for HSI classification.

The motivation is that we would like to design a more representative and efficient feature learning model for HSI classification. The current convolutional neural network learns features in spatial domain with learnable kernels, which usually needs a large number of training samples and long training time. This is a disadvantage when there are only limited training samples such as for HSI classification, and it is hard to build and train a CNN with a very deep architecture. The WF model computes features in frequency domain, and the transforms can be computed with fast method (FFT). Only a few frequency components are used to represent the training data, and the number is further reduced by using subsampling method. Therefore, the training process are more efficient compared with CNN based methods for HSI classification. The training data is also learned with fine

grained using sliding windows with strides in wide direction; therefore, the learning model can be generated with proper complexity according to the number of training samples and the difficulty of learning tasks for HSI classification. This makes the proposed method with WF layers overcome overfitting more easily.

The rest of the paper is organized as follows: Section 2 presents the detailed description of the proposed WD-FNet. Section 3 gives the datasets and the experimental settings. Section 4 presents the experimental results with different HSI datasets. Sections 5 and 6 provide a discussion and conclusions.

2. Wide and Deep Fourier Neural Network for Hyperspectral Image Classification

Considering that CNN is not efficient for use with very deep architectures for HSI classification, and other newly proposed methods still need to be improved both in terms of their performance and efficiency, this paper proposed a wide and deep Fourier neural network for HSI classification. There are three advantages for the proposed method. Firstly, the wide Fourier layer (WF layer) is proposed to extract features in the frequency domain, and the transforms in WF layers can be computed with fast Fourier transform (FFT) fast. Then, the layers are stacked in deep direction to extract more abstract hierarchical features efficiently. Finally, only the weights in the fully connected layer need to be learned using least squares method, which is very simple and efficient for HSI classification. The limitation of the WD-FNet is that unlike DWDNN, which can learn features dynamically, the current learning model is a static model; therefore, it is hard and important to find proper hyperparameters for each WF layers. The capability of learning spatial features still needs to be improved. The architecture of the proposed model is shown in Figure 1. The deep Fourier neural network is composed of multiple wide discrete Fourier transform (DFT) layers to extract features in the frequency domain. Firstly, the HSI data is preprocessed including spectral band reduction by using PCA, and data splitting with overlapping to generate image patches. Then, these image patches are flattened into vectors, and a sliding window with a stride is used to generate input vectors. For each sub-vector in the sliding window, a discrete DFT is performed and, therefore, a large number of DFTs are generated in the wide direction, and the amplitudes of these DFTs are used as the outputs of the transforms. After that, the sorting and pruning operation are performed to reduce the number of outputs. Finally, the wide Fourier layer is added one-by-one in the deep direction to extract more abstract features, and a fully connected layer is given, and the weights are computed using least squares method. Because the proposed method is in the frequency domain, it can use fast Fourier transform to accelerate the computing process. Another advantage is that because only the important Fourier components are used, the number of outputs is highly reduced, and the WD-FNet only needs a small computational load.

2.1. Hyperspectral Remote Sensing Data Slipting

The HSI data $\mathbf{X} \in \mathbf{R}^{H \times W \times B}$ is first processed with principal component analysis (PCA) to reduce the spectral bands to B' to obtain $\mathbf{X}' \in \mathbf{R}^{H \times W \times B'}$, where (H, W) is the height and width of the HSI, and B is the number of bands. The image after PCA is split into fixed-size image patches $\mathbf{P}_m \in \mathbf{R}^{s_r \times s_c \times B'} (1 \leq m \leq M)$, and then these patches are flattened into vectors $\mathbf{P'}_m \in \mathbf{R}^{(s_r s_c) \times B'} (1 \leq m \leq M)$. These vectors are organized by subdivisions as training, validation, and testing sets to be used for learning and testing.

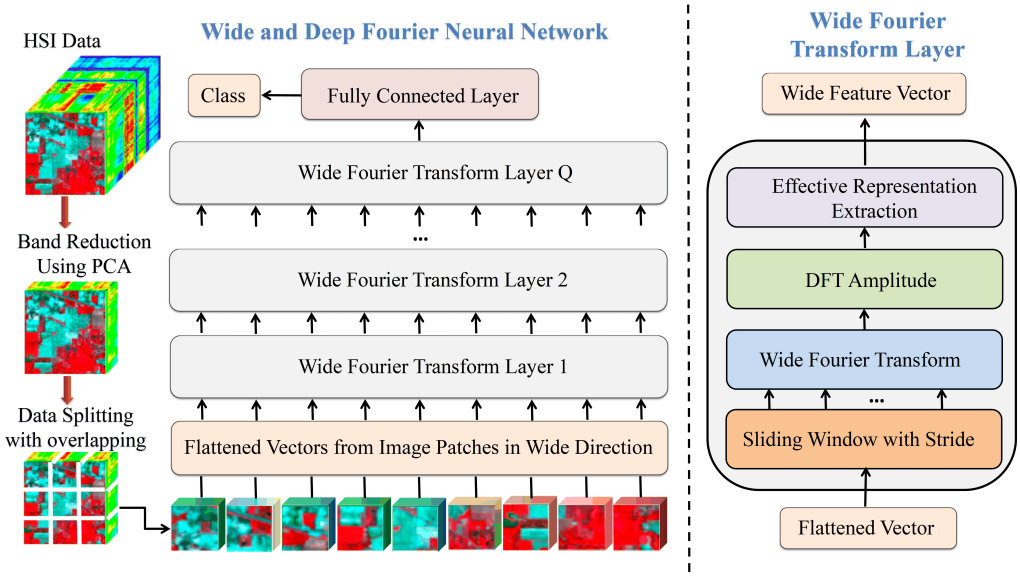

Figure 1. Architecture of the wide and deep Fourier neural network for hyperspcetral image classification.

2.2. The Wide Fourier Transform Layer

For each wide Fourier layer in Figure 2, the sliding window with size w and a sliding stride with the size s are selected. The sliding window is used to generate the input sub-vectors with a given stride. With the nth sliding window, the DFT is performed on part of the vector for the mth image patch (denoted as $\mathbf{v}_{mn}$ after zero padding according to the number of desired frequency components) as follows:

$$\tilde{\mathbf{u}}_{mn} = F(\mathbf{v}_{mn}) = \sum_{l=1}^{L-1} v_{mn}(l) e^{j\frac{2\pi}{L}lu}, 1 \le l \le L-1 \tag{1}$$

where n denotes the nth slice, L is the total number of frequency components, and l is the current component in frequency domain. $\tilde{\mathbf{u}}_{mn}$ is a complex vector, whose amplitude is denoted by

$$\mathbf{u}_{mn} = |\tilde{\mathbf{u}}_{mn}|^{\frac{1}{2}} \tag{2}$$

It is only the amplitude, and not the phase used for subsequent processing. For all N sliding windows with an image patch and writing v_{mn} as v_n, the outputs are written as

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N] \tag{3}$$

where $\mathbf{u}_n$ refers to the nth DFT amplitude vector.

In order to represent the features in the frequency domain more effectively, the DFT amplitudes in each slice are summed up with respect to the image patches as

$$\mathbf{u}_{\text{sum\_}n} = \sum_{m=1}^{M} \mathbf{u}_{mn} \tag{4}$$

Then, $\mathbf{u}_{\text{sum\_}n}$ is sorted from maximum to minimum according to

$$\mathbf{u}_{\text{sort\_}n} = \text{sort}(\mathbf{u}_{\text{sum\_}n}) \tag{5}$$

The indices of the sorted components with N_s elements determine the elements of u_n to be used, and the corresponding feature vectors is generated by

$$\mathbf{u}_{p\_n} = [u_n(1), u_n(2), \cdots, u_n(N_s)] \tag{6}$$

Finally, they are combined together as

$$\mathbf{U}_{FT} = \left[\mathbf{u}_{p\_1}, \mathbf{u}_{p\_2}, \ldots, \mathbf{u}_{p\_N}\right] \tag{7}$$

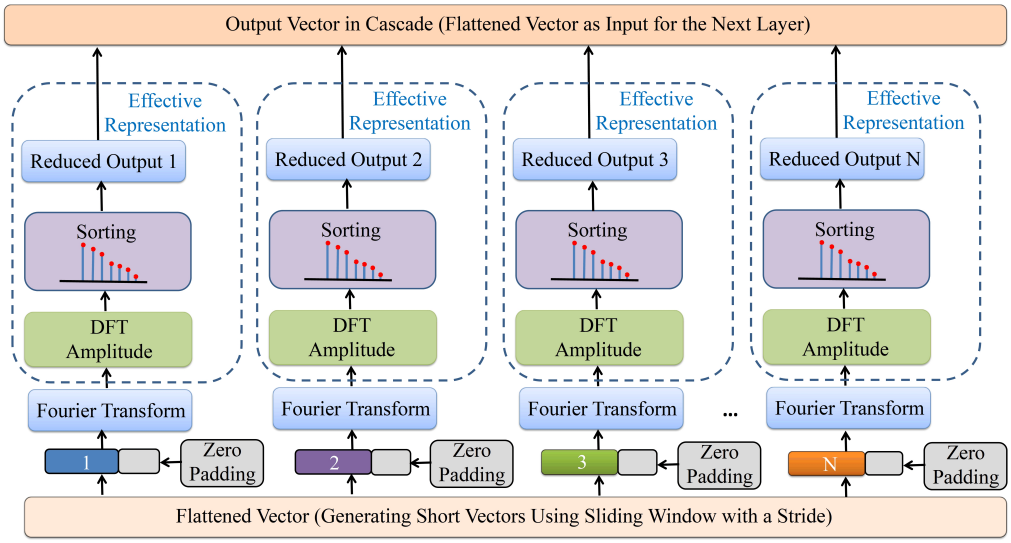

Figure 2. Architecture of the wide Fourier neural layer.

2.3. Wide and Deep Fourier Neural Network

The wide Fourier Layer is added layer-by-layer to obtain more abstract features. Suppose there are Q layers in the proposed network. The output in layer q, $(1 \leq q \leq Q)$ is denoted as $\mathbf{U}_{FT}^{(q)}$. A fully connected layer is added after the Fourier layers, and the output of the proposed network is

$$\mathbf{Y} = \mathbf{U}_{FT}^{(Q)}\mathbf{W} \tag{8}$$

Suppose the class ground truth of the instances is given by $\mathbf{D}$. The weights can be computed using least squares by computing

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \left\| \mathbf{U}_{\mathrm{FT}}^{(\mathrm{Q})}\mathbf{W} - \mathbf{D} \right\|^2 \tag{9}$$

That is because the final outputs are the linear combination of pruned features using weights in the fully connected layer. $\hat{\mathbf{W}}$ is finally computed with pseudoinverse of $\mathbf{U}_{FT}$ given by

$$\hat{\mathbf{W}} = \mathbf{U}_{FT}^{(Q)+}\mathbf{D} = \left(\mathbf{U}_{FT}^{(Q)T}\mathbf{U}_{FT}^{(Q)}\right)^{+}\mathbf{U}_{FT}^{(Q)T}\mathbf{D} \tag{10}$$

3. Datasets and Experimental Settings

3.1. Experimental Datasets

We used three datasets to evaluate the proposed method, and they are described in Tables 1 and 2.

Table 1. Description of Hyperspectral data.

Dataset	Description
Pavia University	It was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the Pavia University campus. It has nine classes of land cover. The image dimension is 610 × 610. After discarding the pixels without information, the dimension is 610 × 340 with 103 spectral bands.
KSC	The Kennedy Space Center (KSC, Merritt Island, FL, USA) dataset was acquired over the Kennedy Space Center in Florida. The image size is 512 × 614, and there are 13 classes of land cover. It has 176 bands after removing water absorption and low SNR bands.
Salinas	It was acquired in Salinas valley, California. The image size is 512 × 217 pixels, and the number of classes is 16. There are 204 bands after removing bands of water absorption.

Table 2. Hyperspectral Remote Sensing Datasets.

Class NO.	Pavia University		KSC		Salinas	
	Land Cover Class	NO.	Land Cover Class	NO.	Land Cover Class	NO.
1	Asphalt	6631	Scrub	347	Brocoli-green-weeds-1	2009
2	Meadows	18,649	Willow swamp	243	Brocoli-green-weeds-2	3726
3	Gravel	2099	CP hammock	256	Fallow	1976
4	Trees	3064	Slash pine	252	Fallow-rough-plow	1394
5	Painted metal sheets	1345	Oak/broadleaf	161	Fallow-smooth	2678
6	Bare soil	5029	Hardwood	229	Stubble	3959
7	Bitumen	1330	Swamp	105	Celery	3579
8	Self-blocking bricks	3682	Graminoid marsh	390	Grapes-untrained	11,271
9	Shadows	947	Spartina marsh	520	Soil-vinyard-develop	6203
10			Cattail marsh	404	Corn-senesced-green-weeds	3278
11			Salt marsh	419	Lettuce-romaine-4wk	1068
12			Mud flats	503	Lettuce-romaine-5wk	1927
13			Water	927	Lettuce-romaine-6wk	916
14					Lettuce-romaine-7wk	1070
15					Vinyard-untrained	7268
16					Vinyard-vertica-trellis	1807
Total		42,776		5211		54,129

3.2. Experimental Setup

The computer system used had Intel-i7-8700K CPU @ 3.7 GHz, 32G memory. The number of bands used was 15 after using PCA. The patch sizes used were with 15 × 15, 17 × 17, and 19 × 19 for Pavia Universiy, KSC, and Salinas datasets, respectively. The centered pixels in the patches were treated as instances for the learning process. The ratio of instances for training, validation was 0.2, and the remainder was for test. All the instances were chosen randomly for training, validaiton, and test. The overall accuracy (OA), average accuracy (AA) [53,54], and Kappa coefficient were used as evaluation indices in the experiments. OA is given by

$$OA = \frac{N_c}{N_t} \tag{11}$$

where N_c, and N_t are the correctly classified and the total number of testing instances, respectively.

AA is defined as

$$AA = \frac{1}{C}\sum_{i=1}^{C}\frac{N_{c\_i}}{N_{t\_i}} \tag{12}$$

where, N_{c_i} and N_{t_i} are the numbers of correctly classified and testing instances for class i, respectively.

The proposed WD-FNet in the experiments is composed of four WF layers. The hyper-parameters for each WF layer includes sliding window size, sliding stride, size of DFTs, and number of pruned outputs. For the Pavia University dataset, these parameters are 15,

0.9 of the length of the sliding window, 600, and 100 for the first layer; 0.35 of the input data length, 0.15 of the length of sliding window, 1000, and 100 for the second layer; 0.3 of the input data length, 0.15 of the length of sliding window, 1000, and 100 for the third layer; and 0.32 of the input data length, 0.15 of the length of sliding window, 3000, and 300 for the fourth layer. For the KSC dataset, these parameters are 20, 0.9 of the length of the sliding window, 600, and 100 for the first layer; they are the same for the second and third layers; and 0.37 of the input data length, 0.15 of the length of sliding window, 1000, and 50 for the fourth layer. For the Salinas dataset, these parameters are 15, 0.8 of the length of the sliding window, 600, and 100 for the first layer; they are the same for the second and third layers; and 0.37 of the input data length, 0.15 of the length of sliding window, 4000, and 400 for the fourth layer. The detailed settings are shown in Table 3.

We compared the proposed method with multilayer perceptron (MLP) [53], radial basis function (RBF) [53], CNN [53], RBF ensemble [53], CNN ensemble [53], 2D CNN [17,21], 3D CNN [17,21], SMSB [55], WSWS Net [53], and DWNNN [54].

Table 3. Hyperparameters of the WD-FNet for different datasets.

Dataset	Hyperparameters of the WD-FNet for Different Datasets							
Pavia University	Wide Fourier Layer 1				Wide Fourier Layer 2			
	Window 15	Stride 0.9	DFT Pts. 600	Pruned NO. 100	Window 0.35	Stride 0.15	DFT Pts. 1000	Pruned NO. 100
	Wide Fourier Layer 3				Wide Fourier Layer 4			
	Window 0.3	Stride 0.15	DFT Pts. 1000	Pruned NO. 100	Window 0.32	Stride 0.15	DFT Pts. 3000	Pruned NO. 300
KSC	Wide Fourier Layer 1				Wide Fourier Layer 2			
	Window 20	Stride 0.9	DFT Pts. 600	Pruned NO. 100	Window 0.35	Stride 0.15	DFT Pts. 1000	Pruned NO. 100
	Wide Fourier Layer 3				Wide Fourier Layer 4			
	Window 0.3	Stride 0.15	DFT Pts. 1000	Pruned NO. 100	Window 0.37	Stride 0.15	DFT Pts. 1000	Pruned NO. 50
Salinas	Wide Fourier Layer 1				Wide Fourier Layer 2			
	Window 15	Stride 0.8	DFT Pts. 600	Pruned NO. 100	Window 1.35	Stride 0.15	DFT Pts. 1000	Pruned NO. 100
	Wide Fourier Layer 3				Wide Fourier Layer 4			
	Window 0.5	Stride 0.15	DFT Pts. 1000	Pruned NO. 100	Window 0.37	Stride 0.15	DFT Pts. 4000	Pruned NO. 400

4. Experimental Results

Classification Performance on Different Datasets

The classification results on the Pavia University dataset are shown in Table 4 and Figure 3. It is seen from the results that the proposed WD-FNet achieved the best performance among the compared methods, and the OA, AA, and Kappa coefficients for WD-FNet are 99.77%, 99.55%, and 99.69%, respectively. DWDNN, WSWS Net, and SMSB also achieved excellent performance. It is also seen from the classification results in the figures that the proposed method has the most smooth predicted land cover results. For example, the class of bare soil at the lower right of Figure 3 is exactly predicted as compared with the original false color image.

Table 4. Classification results of different methods on Pavia University dataset (the unit of all the results is %).

Class NO.	MLP	CNN	2-D CNN	3-D CNN	SMSB	WSWS	DWDNN	WD-FNet
1	97.13	96.18	98.51	98.40	99.11	99.10	99.87	99.85
2	98.43	96.69	99.54	96.91	98.97	100.00	100.00	99.96
3	85.15	80.86	84.62	97.05	98.89	93.01	96.98	98.89
4	95.05	87.21	98.04	98.84	98.74	98.37	99.29	99.18
5	99.88	99.63	100.00	100.00	100.00	99.88	99.75	99.26
6	96.35	88.30	97.10	99.32	99.87	99.97	100.00	100.00
7	90.85	82.58	95.05	98.92	99.79	99.00	98.62	100.00
8	93.21	94.12	96.39	98.33	98.99	98.33	99.59	99.50
9	99.30	99.30	99.69	99.90	98.04	98.95	99.65	99.30
OA	96.47	93.66	97.84	96.52	99.11	99.19	99.69	99.77
AA	95.04	91.65	96.56	97.47	99.16	98.51	99.31	99.55
Kappa	95.36	91.72	97.19	95.50	98.79	98.93	99.59	99.69

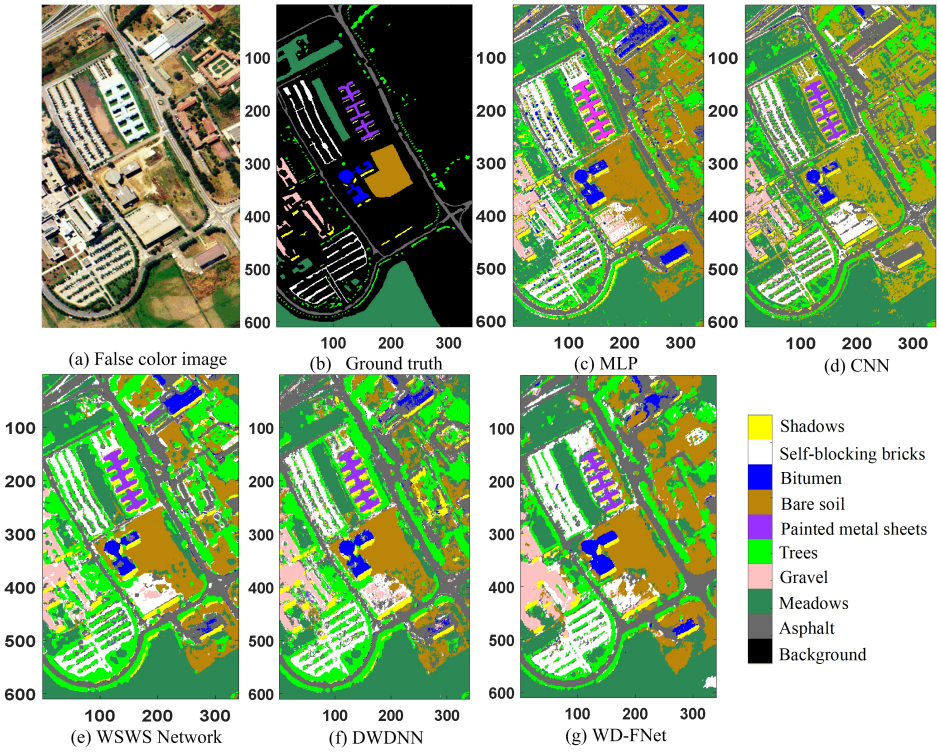

Figure 3. Classification results of Pavia University data (the unit for both horizontal and vertical axes is: pixel).

The classification results on KSC data are shown in Table 5 and Figure 4. It is observed that the proposed WD-FNet achieved the best performance among the compared methods. The OA, AA, and Kappa coefficients of the proposed method are 99.97%, 99.95%, and 99.96%, respectively. DWDNN, WSWS Net, and SMSB also achieved excellent performance. It is also seen that 12 of 13 classes achieved accuracies of 100% for both the proposed WD-FNet and DWNNN, while 11 of 13 classes achieved accuracies of 100% for WSWS Net. It is also seen from the classification results in the figures that the proposed method has smooth predicted land cover results with different classes.

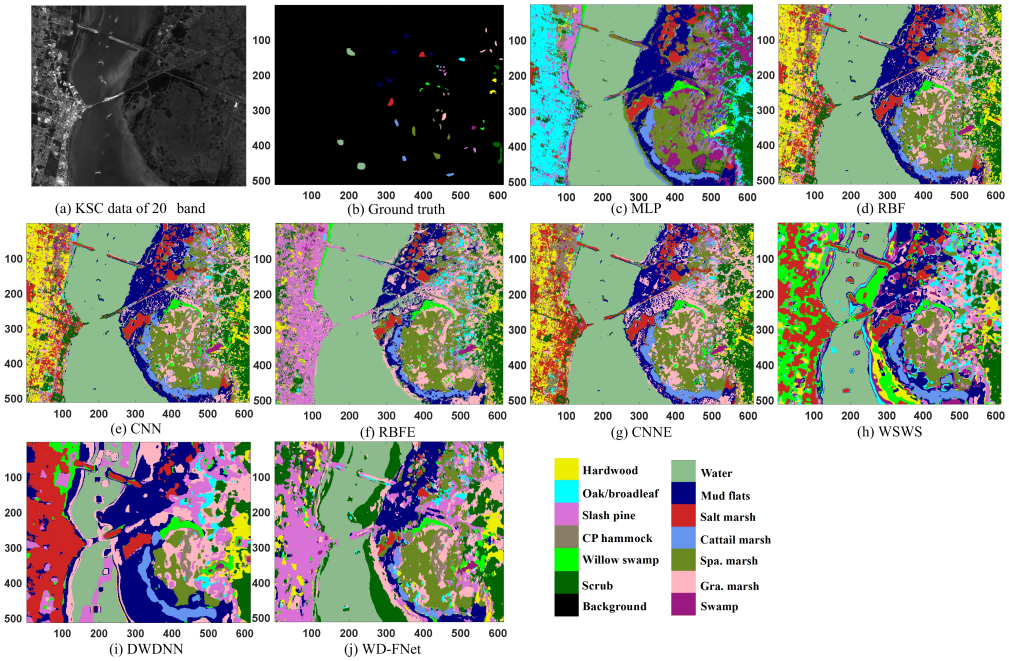

Figure 4. Classification results of KSC data (the unit for both horizontal and vertical axes is: pixel).

Table 5. Classification results of different methods for KSC (the unit of all the results is %).

Class NO.	MLP	RBF	CNN	RBFE	CNNE	WSWS	DWDNN	WD-FNet
1	99.78	98.47	97.37	96.94	97.81	100.00	100.00	100.00
2	99.31	88.28	94.48	92.41	94.48	100.00	100.00	100.00
3	92.86	96.75	95.45	96.10	98.70	99.35	100.00	100.00
4	79.61	64.74	76.97	71.71	70.39	100.00	98.68	99.34
5	87.63	90.72	72.16	92.78	69.07	96.91	100.00	100.00
6	99.27	88.32	83.21	83.21	86.13	100.00	100.00	100.00
7	100.00	96.83	100.00	95.24	90.48	100.00	100.00	100.00
8	100.00	98.07	96.53	94.98	99.61	100.00	100.00	100.00
9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
10	99.59	99.59	100.00	97.93	100.00	100.00	100.00	100.00
11	98.41	90.84	100.00	95.62	100.00	100.00	100.00	100.00
12	99.34	98.01	96.01	98.67	98.34	100.00	100.00	100.00
13	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
OA	97.95	95.36	95.75	95.52	95.97	99.87	99.94	99.97
AA	96.60	93.10	93.25	93.51	92.69	99.71	99.90	99.95
Kappa	97.72	94.85	95.28	95.03	95.52	99.86	99.93	99.96

The classification results for the Salinas dataset are shown in Table 6 and Figure 5. It is seen from the results that the proposed WD-FNet achieved the best performance among the compared methods. The OA, AA, and Kappa coefficients of the proposed method are 99.95%, 99.95%, and 99.94%, respectively. DWDNN also achieved excellent performance with OA, AA, and Kappa coefficients as high as 99.76%, 99.73%, and 99.73, respectively. It is also observed that the proposed method has the most smooth predicted results, which can be seen clearly at the bottom right of Figure 5.

Table 6. Classification results of different methods on Salinas dataset (The unit of all the results is %).

Class NO.	MLP	CNN	2-D CNN	3-D CNN	SMSB	WSWS	DWDNN	WD-FNet
1	100.00	98.51	100.00	98.41	99.78	100.00	100.00	100.00
2	100.00	99.82	99.96	100.00	99.97	99.87	99.91	100.00
3	99.41	99.66	99.63	99.23	99.94	98.82	99.24	100.00
4	99.52	98.68	99.28	99.90	99.28	97.73	98.80	100.00
5	97.70	99.38	99.20	99.43	99.54	99.38	99.88	99.88
6	100.00	99.96	100.00	99.55	99.97	99.96	100.00	100.00
7	0.00	99.95	100.00	99.72	99.88	99.91	99.95	100.00
8	90.64	74.24	93.62	89.75	98.87	99.72	99.81	99.91
9	100.00	100.00	100.00	99.81	99.91	99.76	99.70	99.92
10	99.08	93.44	98.82	98.36	98.85	99.64	99.95	100.00
11	99.53	96.72	99.73	98.12	99.79	100.00	99.84	100.00
12	100.00	99.74	100.00	98.96	99.94	99.91	99.74	100.00
13	99.64	98.91	100.00	98.93	99.03	99.82	99.82	100.00
14	99.84	100.00	99.86	98.60	98.86	100.00	99.69	100.00
15	85.53	88.65	91.52	79.31	97.63	99.52	99.56	100.00
16	99.91	98.53	99.92	94.51	99.92	100.00	99.72	99.45
OA	89.27	92.42	97.39	93.95	99.26	99.67	99.76	99.95
AA	91.92	96.64	98.85	97.02	99.45	99.63	99.73	99.95
Kappa	88.20	91.69	97.07	93.31	99.17	99.63	99.73	99.94

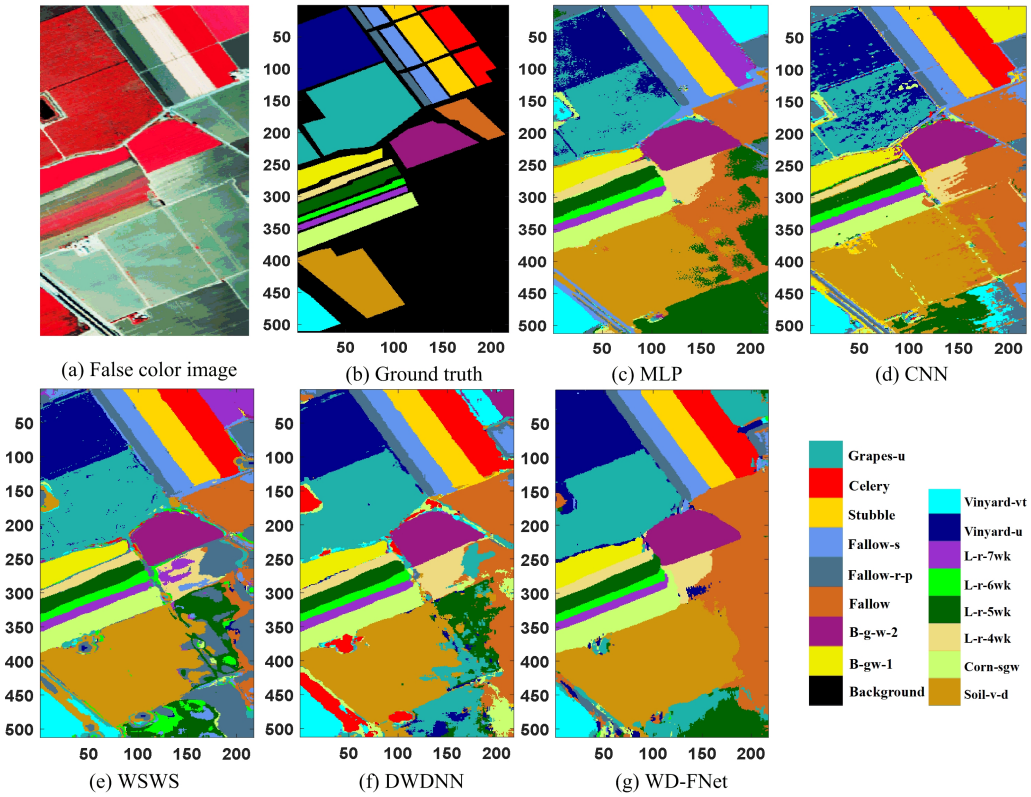

Figure 5. Classification results of Salinas data (the unit for both horizontal and vertical axes is: pixel).

5. Discussion

5.1. Effects of Different Sizes of Image Patches

The size of image patches is usually important for hyperspectral image classification, because it provides spatial features. We tested the proposed WD-FNet with different sizes of image patches on Pavia University, KSC, and Salinas datasets. The model settings are the same as the the previous part. The results are shown in Table 7. It can be seen that as the image patch increases, the test performance continue increasing, and then it decreases or stops increasing. For Pavia University, it is seen that the proposed method can reach the best OA, AA, and Kappa coefficients for the size of the image patches with 15×15, then these indices decrease. For the KSC dataset, it reaches the best OA, AA, and Kappa coefficients with the size of the image patches being as large as 17×17. For the Salinas dataset, the proposed method achieves the best test performance as the size of the image patches reach 21, and then the performance almost stops increasing. It also seen that the proposed method usually has good classification results with pretty large image patches, which demonstrates that the proposed method has good ability to learn spatial features.

5.2. Visualization of the Fourier Transform Layers of the Wd-Fnet

The learning process included the feed-forward process, and the computation of weights in the fully connected layer using least squares. The extracted features were computed layer-by-layer, and in each layer, the features were extracted from the amplitudes of a large number of Fourier transforms. These features were sorted, and the large feature values were retained as the output of each transform. Finally, these pruned features were concatenated as the feature vector of the current wide Fourier layer. The transformed outputs from the last sliding windows in different layers are shown in Figure 6. For the three datasets, 10% of the training samples in class 1 were chosen to visualize the Fourier transform layers. It is seen that the training vectors are similar in the same class, and the curves are different in different classes. The shapes of the whole curves of the transform amplitudes look a bit similar, but the values are quite different, representing the extracted features from different frequencies.

Table 7. Testing accuracies with different neighborhood sizes of Pavia Center, Pavia University, and KSC datasets.

Neighborhood Sizes	Pavia Univ. /(%)			KSC/(%)			Salinas/(%)		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
3×3	96.68	93.96	95.63	94.63	92.18	94.04	96.46	98.33	96.08
5×5	98.05	96.16	97.38	96.87	95.16	96.52	97.56	98.81	97.30
7×7	98.70	97.40	98.05	98.18	96.98	97.97	98.43	99.33	98.26
9×9	99.22	98.52	98.97	99.04	98.06	98.93	98.98	99.46	98.87
11×11	99.35	98.71	99.13	99.07	98.10	98.97	99.49	99.76	99.43
13×13	99.59	99.20	99.46	99.30	98.87	99.22	99.73	99.79	99.70
15×15	99.77	99.55	99.69	99.49	99.40	99.43	99.89	99.95	99.88
17×17	99.51	99.10	99.35	99.97	99.95	99.96	99.88	99.93	99.87
19×19	99.44	98.64	99.26	99.81	99.66	99.79	99.95	99.95	99.94
21×21	99.41	98.57	99.22	99.39	99.27	99.32	99.95	99.98	99.95

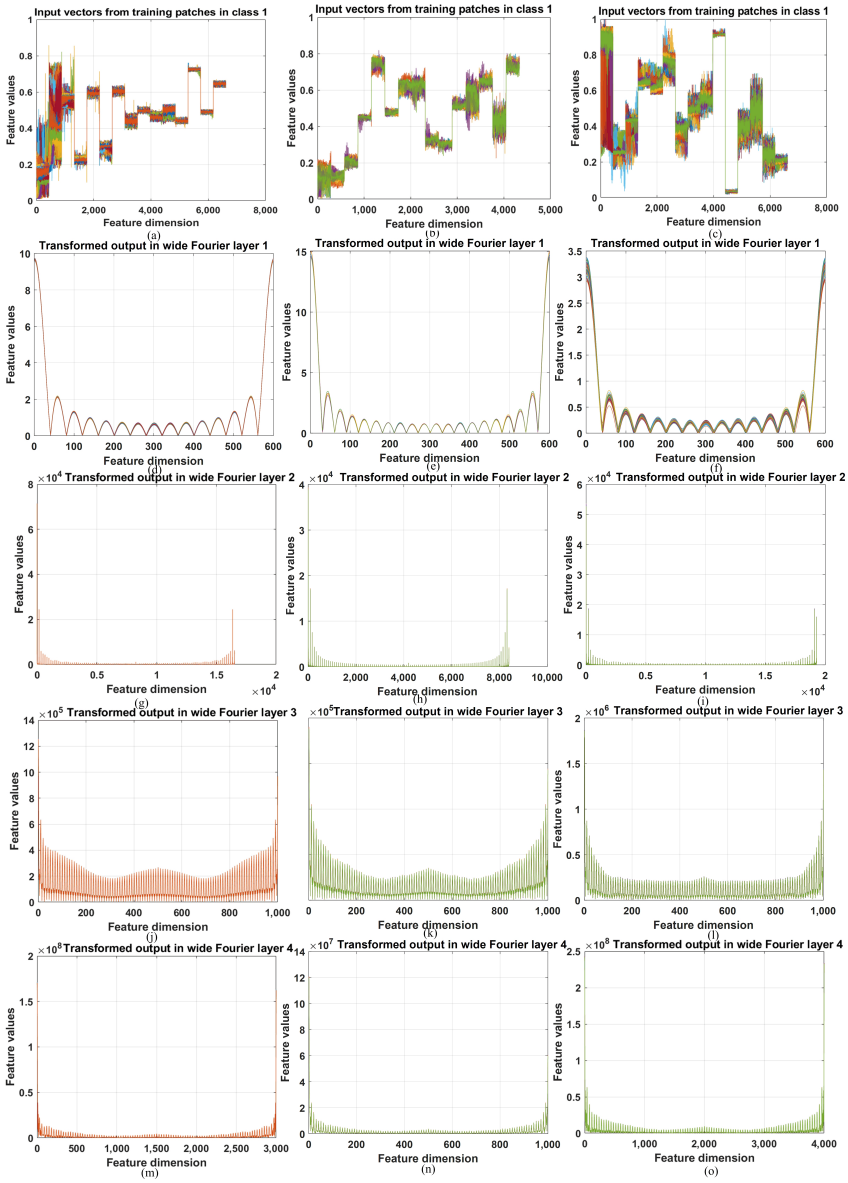

Figure 6. Visualization of the DFT layers of the WD-FNet. (**a**–**c**) Input training vectors of class 1 from Pavia University, KSC, and Salinas datasets, respectively. (**d**–**o**) are transformed outputs from the last sliding windows in layer 1, 2, 3, and 4 for Pavia University, KSC, and Salinas datasets, respectively.

5.3. Advantages and Limitations of the WD-Fnet Compared with Other Learning Models

For the HSI datasets in the experiments, SMSB, WSWS, DWDNN, and the proposed WD-FNet have comparable performance. It can be seen that the order of the performance of the models are the same on all the three datasets. The WD-FNet always has the best performance, and the DWDNN has the second best performance. DWDNN is composed of multiple base learners that can learn features dynamically; therefore, it is possible to find the learning model with proper complexity to overcome overfitting. However, the generating process will take some time until it is trained and stopped with proper architecture. WD-FNet has better performance on HSI classification than the WSWS Net and SMSB method. All these three models are static learning models, which means the proper hyperparameters need to be found. WD-FNet learns features in the frequency domain efficiently, while WSWS Net learns features with a group of Gassian kernels as basis functions. SMSB combines spectral and spatial data to reduce the computational complexity, but the performance can still be improved. For further validation of the WD-FNet, we also compared the results with methods in very recently published papers. For

the Pavia University data, the OA and Kappa coefficients of the deep high order tensor convolutional sparse coding model (DHTCSCNet) [56] are 99.11%, and 98.81%, respectively. The indices for spectral–spatial random patches network (SSRPNet) [57] are 99.05%, and 98.73%, respectively. The indices for deep multilayer fusion dense Network (MFDN) [58] are 98.89%, and 98.10%, respectively. For the KSC dataset, the OA and Kappa coefficients of the DHTCSCNet are 99.60%, and 99.56%, respectively. The indices for SSRPNet are 99.43%, and 99.37%, respectively. The indices for MFDN is 97.55%, and 97.27%, respectively. It can be seen that the proposed WD-FNet has very promising performance for HSI classification.

To summarize, the advantages of the proposed method are: (1) the WF layer extracts features in the frequency domain effectively, and can be computed fast with FFT. (2) The layers can extract more abstract hierarchical features by stacking WF layers in the deep direction. (3) The weights in the fully connected layer can be learned easily using least squares method. The limitations of the WD-FNet are that the current learning model is a static model, it is hard to find proper hyperparameters for each WF layer, and the performance of learning spatial features by using WD-FNet still needs to be improved.

6. Conclusions

It is important and useful to classify land covers with hyperspectral remote sensing images based on abundant spectral bands together with spatial features. The problem is that there are usually a very limited number of training samples to do the classification. In this paper, a wide and deep Fourier Network is proposed to do hyperspectral image classification efficiently with high performance. First, the hyperspectral image after PCA is split into image patches, and then stretched into vectors as inputs. Then, the wide Fourier layer is generated by combining a number of DFTs used with sliding windows to learn features in the frequency domain as DFT amplitudes with pruning to reduce the number of features. After that, the wide Fourier layers are stacked layer-by-layer to learn features hierarchically with the DFT amplitudes adopted as nonlinear activation functions. Finally, the fully connected layer is added and the weights are learned efficiently using least squares method. The experiments were performed with the Pavia University, KSC, and Salinas datasets, and the results show that the proposed method can learn features very efficiently, and has excellent classification performance. The OAs of the proposed method are 99.77%, 99.97%, and 99.95%, respectively. The AAs are 99.55%, 99.95%, and 99.95%, respectively. The Kappa coefficients are 99.69%, 99.96%, and 99.94%, respectively. The results show that the WD-FNet hast best performance among the compared methods. It can be used for image classification and segmentation, and can be potentially used on lightweight embedded computing platforms. In the future work, we will try to use it for applications such as object detection in images, and time serial data prediction.

Author Contributions: J.X. and O.K.E. contributed equally to this work. J.X., O.K.E., M.C., C.Z., W.Q. and T.W. made significant contributions to this work. All authors contributed to the methodology validation, results analysis, and reviewed the manuscript. Conceptualization, J.X. and O.K.E.; methodology, J.X. and M.C.; software and experiments, J.X. and M.C.; validation, W.Q. and T.W.; writing original draft preparation, J.X.; funding acquisition, C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grants 42171348, 41874005, 41929001, 41941019, 42174006, and 42090055; in part by the Fundamental Research Funds for the Central Universities, CHD under Grants 300102262202, 300102260301, 300102120201, and 300102262902; in part by Science Fund for Distinguished Young Scholars of Shaanxi Province (2022JC-18); in part by Key Research and Development Program of Shaanxi (Grant No. 2021NY-170); in part by Major Science and Technology Project of Inner Mongolia Autonomous Region under Grant 2021SZD0036; in part by Shaanxi Forestry Science and Technology Innovation Program, NO. SXLK2021-0225.

Data Availability Statement: The hyperspectral data in the experiments was downloaded from http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, accessed on 3 May 2022.

Acknowledgments: The authors are grateful to M. Graña, MA. Veganzons, and B. Ayerdi for collecting the hyperspectral datasets.

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

Abbreviations

The following abbreviations are used in this manuscript:

HSI	Hyperspectral Image
PCA	Principal component analysis
DFT	Discrete Fourier transform
FFT	Fast Fourier transform
PCA	Principal component analysis
MLP	Multilayer perceptron
CNN	Convolutional neural network
LSTM	Long short-term memory
FCN	Fully convolutional network
GAN	Generative adversarial network
EWC	Elastic weight consolidation
PSHNN	Parallel, self-organizing, hierarchical neural networks
PCNN	D-parallel consensual neural networks
WSWS	Wide sliding window and subsampling
SWNN	Scalable wide neural network
DWDNN	Dynamic Wide and Deep Neural Network
WF	Wide Fourier
WD-FNet	Wide and Deep Fourier Neural Network
OA	Overall Accuracy
AA	Average Accuracy

References

1. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [CrossRef]
2. Tarabalka, Y.; Fauvel, M.; Chanussot, J.; Benediktsson, J.A. SVM-and MRF-based method for accurate classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 736–740. [CrossRef]
3. Xi, J.; Ersoy, O.K.; Fang, J.; Wu, T.; Wei, X.; Zhao, C. Parallel Multistage Wide Neural Network. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–14. [CrossRef] [PubMed]
4. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geosci. Remote Sens. Lett.* **2012**, *10*, 318–322.
5. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef] [PubMed]
6. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
7. Zhang, M.; Li, W.; Du, Q. Diverse region-based CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2018**, *27*, 2623–2634. [CrossRef]
8. Wang, W.; Dou, S.; Jiang, Z.; Sun, L. A fast dense spectral–spatial convolution network framework for hyperspectral images classification. *Remote Sens.* **2018**, *10*, 1068. [CrossRef]
9. Gong, Z.; Zhong, P.; Yu, Y.; Hu, W.; Li, S. A CNN with Multiscale Convolution and Diversified Metric for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3599–3618. [CrossRef]
10. Gao, Q.; Lim, S.; Jia, X. Hyperspectral image classification using convolutional neural networks and multiple feature learning. *Remote Sens.* **2018**, *10*, 299. [CrossRef]
11. Cheng, G.; Li, Z.; Han, J.; Yao, X.; Guo, L. Exploring hierarchical convolutional features for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6712–6722. [CrossRef]
12. Paoletti, M.; Haut, J.; Plaza, J.; Plaza, A. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 120–147. [CrossRef]
13. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 277–281. [CrossRef]
14. Zheng, J.; Feng, Y.; Bai, C.; Zhang, J. Hyperspectral Image Classification Using Mixed Convolutions and Covariance Pooling. *IEEE Trans. Geosci. Remote Sens.* **2020**, 59, 522–534. [CrossRef]

15. Shi, C.; Liao, D.; Zhang, T.; Wang, L. Hyperspectral Image Classification Based on Expansion Convolution Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
16. Haut, J.M.; Paoletti, M.E.; Plaza, J.; Li, J.; Plaza, A. Active Learning With Convolutional Neural Networks for Hyperspectral Image Classification Using a New Bayesian Approach. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6440–6461. [CrossRef]
17. Tang, X.; Meng, F.; Zhang, X.; Cheung, Y.; Ma, J.; Liu, F.; Jiao, L. Hyperspectral Image Classification Based on 3-D Octave Convolution with Spatial-Spectral Attention Network. *IEEE Trans. Geosci. Remote Sens.* **2020**, 59, 2430–2447. [CrossRef]
18. Cao, X.; Yao, J.; Xu, Z.; Meng, D. Hyperspectral image classification with convolutional neural network and active learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, 58, 4604–4616. [CrossRef]
19. Xie, F.; Gao, Q.; Jin, C.; Zhao, F. Hyperspectral image classification based on superpixel pooling convolutional neural network with transfer learning. *Remote Sens.* **2021**, *13*, 930. [CrossRef]
20. Masarczyk, W.; Głomb, P.; Grabowski, B.; Ostaszewski, M. Effective Training of Deep Convolutional Neural Networks for Hyperspectral Image Classification through Artificial Labeling. *Remote Sens.* **2020**, *12*, 2653. [CrossRef]
21. 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]
22. Hang, R.; Liu, Q.; Hong, D.; Ghamisi, P. Cascaded recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5384–5394. [CrossRef]
23. Liu, Q.; Zhou, F.; Hang, R.; Yuan, X. Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1330. [CrossRef]
24. Mei, S.; Li, X.; Liu, X.; Cai, H.; Du, Q. Hyperspectral Image Classification Using Attention-Based Bidirectional Long Short-Term Memory Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–12. [CrossRef]
25. Song, T.; Wang, Y.; Gao, C.; Chen, H.; Li, J. MSLAN: A Two-Branch Multidirectional Spectral–Spatial LSTM Attention Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [CrossRef]
26. He, X.; Chen, Y.; Lin, Z. Spatial-Spectral Transformer for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 498. [CrossRef]
27. Qing, Y.; Liu, W.; Feng, L.; Gao, W. Improved Transformer Net for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 2216. [CrossRef]
28. Zhou, F.; Hang, R.; Li, J.; Zhang, X.; Xu, C. Spectral-Spatial Correlation Exploration for Hyperspectral Image Classification via Self-Mutual Attention Network. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
29. Lin, J.; Ma, L.; Yao, Y. A Fourier domain acceleration framework for convolutional neural networks. *Neurocomputing* **2019**, *364*, 254–268. [CrossRef]
30. Ayat, S.O.; Khalil-Hani, M.; Ab Rahman, A.A.H.; Abdellatef, H. Spectral-based convolutional neural network without multiple spatial-frequency domain switchings. *Neurocomputing* **2019**, *364*, 152–167. [CrossRef]
31. Khan, S.H.; Hayat, M.; Porikli, F. Regularization of deep neural networks with spectral dropout. *Neural Netw.* **2019**, *110*, 82–90. [CrossRef] [PubMed]
32. Uteuliyeva, M.; Zhumekenov, A.; Takhanov, R.; Assylbekov, Z.; Castro, A.J.; Kabdolov, O. Fourier neural networks: A comparative study. *Intell. Data Anal.* **2020**, *24*, 1107–1120. [CrossRef]
33. Silvescu, A. Fourier neural networks. In Proceedings of the IJCNN'99 International Joint Conference on Neural Networks, Washington, DC, USA, 10–16 July 1999; IEEE: Piscataway, NJ, USA, 1999; Volume 1, pp. 488–491.
34. Li, Z.; Kovachki, N.B.; Azizzadenesheli, K.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Fourier Neural Operator for Parametric Partial Differential Equations. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
35. Rao, Y.; Zhao, W.; Zhu, Z.; Lu, J.; Zhou, J. Global filter networks for image classification. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 980–993.
36. Worrall, D.E.; Garbin, S.J.; Turmukhambetov, D.; Brostow, G.J. Harmonic networks: Deep translation and rotation equivariance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5028–5037.
37. Liu, C.; Li, J.; He, L.; Plaza, A.; Li, S.; Li, B. Naive Gabor Networks for Hyperspectral Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 376–390. [CrossRef] [PubMed]
38. Okwuashi, O.; Ndehedehe, C.E. Deep support vector machine for hyperspectral image classification. *Pattern Recognit.* **2020**, *103*, 107298. [CrossRef]
39. Roy, S.K.; Haut, J.M.; Paoletti, M.E.; Dubey, S.R.; Plaza, A. Generative Adversarial Minority Oversampling for Spectral-Spatial Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]
40. Liang, H.; Bao, W.; Shen, X.; Zhang, X. Spectral-Spatial Attention Feature Extraction for Hyperspectral Image Classification Based on Generative Adversarial Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10017–10032. [CrossRef]
41. Parisi, G.I.; Kemker, R.; Part, J.L.; Kanan, C.; Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Netw.* **2019**, *113*, 54–71. [CrossRef]

42. Venkataramani, S.; Raghunathan, A.; Liu, J.; Shoaib, M. Scalable-effort classifiers for energy-efficient machine learning. In Proceedings of the 52nd Annual Design Automation Conference, San Francisco, CA, USA, 8–12 June 2015; ACM: New York, NY, USA, 2015; p. 67.
43. Panda, P.; Venkataramani, S.; Sengupta, A.; Raghunathan, A.; Roy, K. Energy-Efficient Object Detection Using Semantic Decomposition. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2017**, *25*, 2673–2677. [CrossRef]
44. Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A.A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3521–3526. [CrossRef]
45. Lee, S.W.; Kim, J.H.; Jun, J.; Ha, J.W.; Zhang, B.T. Overcoming catastrophic forgetting by incremental moment matching. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4652–4662.
46. Ersoy, O.K.; Hong, D. Parallel, self-organizing, hierarchical neural networks. *IEEE Trans. Neural Netw.* **1990**, *1*, 167–178. [CrossRef] [PubMed]
47. Benediktsson, J.A.; Sveinsson, J.R.; Ersoy, O.K.; Swain, P.H. Parallel consensual neural networks. *IEEE Trans. Neural Netw.* **1997**, *8*, 54–64. [CrossRef]
48. Neyshabur, B.; Li, Z.; Bhojanapalli, S.; LeCun, Y.; Srebro, N. The role of over-parametrization in generalization of neural networks. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
49. Lee, J.; Xiao, L.; Schoenholz, S.S.; Bahri, Y.; Sohl-Dickstein, J.; Pennington, J. Wide neural networks of any depth evolve as linear models under gradient descent. *arXiv* **2019**, arXiv:1902.06720.
50. Cheng, H.T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M. Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, Boston, MA, USA, 15 September 2016; ACM: New York, NY, USA, 2016; pp. 7–10.
51. Liu, W.; Nie, X.; Zhang, B.; Sun, X. Incremental Learning with Open-Set Recognition for Remote Sensing Image Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
52. Xi, J.; Ersoy, O.K.; Fang, J.; Cong, M.; Wei, X.; Wu, T. Scalable Wide Neural Network: A Parallel, Incremental Learning Model Using Splitting Iterative Least Squares. *IEEE Access* **2021**, *9*, 50767–50781. [CrossRef]
53. Xi, J.; Ersoy, O.K.; Fang, J.; Cong, M.; Wu, T.; Zhao, C.; Li, Z. Wide Sliding Window and Subsampling Network for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 1290. [CrossRef]
54. Xi, J.; Cong, M.; Ersoy, O.K.; Zou, W.; Zhao, C.; Li, Z.; Gu, J.; Wu, T. Dynamic Wide and Deep Neural Network for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 2575. [CrossRef]
55. Azar, S.G.; Meshgini, S.; Rezaii, T.Y.; Beheshti, S. Hyperspectral image classification based on sparse modeling of spectral blocks. *Neurocomputing* **2020**, *407*, 12–23. [CrossRef]
56. Cheng, C.; Li, H.; Peng, J.; Cui, W.; Zhang, L. Deep High Order Tensor Convolutional Sparse Coding for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]
57. Cheng, C.; Li, H.; Peng, J.; Cui, W.; Zhang, L. Hyperspectral Image Classification Via Spectral-Spatial Random Patches Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4753–4764. [CrossRef]
58. 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]