Multiscale Densely Connected Attention Network for Hyperspectral Image Classification

Xin Wang and Yanguo Fan

***Abstract*—Hyperspectral image classification (HSIC) based on deep learning has always been a research hot spot in the field of remote sensing. However, most of the classification models extract relevant features based on fixed-scales convolution kernels, which ignores the complex features of hyperspectral images (HSIs) at different scales and impairs the classification accuracy. To solve this problem, a multiscale densely connected attention network (MSDAN) is proposed for HSIC. First, the model adopts three different scales modules with dense connection to enhance classification performance, strengthen feature reuse, prevent overfitting and gradient disappearance. Besides, in order to reduce the model parameters and strengthen the extraction of spatial–spectral features, the traditional three-dimensional convolution is replaced by three-dimensional spectral convolution block and three-dimensional spatial convolution block. Furthermore, the spectral–spatial–channel attention is embedded into the end of each scale to enhance the favorable features for classification and further extract the discriminant features of the corresponding scale. Finally, the key feature extraction module is developed to extract multiscale fusion features to further enhance the classification performance of the network. The experimental results carried out on real HSIs show that the proposed MSDAN architecture has significant advantages compared with other most advanced methods.**

***Index Terms*—3-D spectral convolution, hyperspectral image classification (HSIC), multiscale dense connection, spatial–spectral features, spectral–spatial–channel attention, three-dimensional (3-D) spatial convolution.**

I. Introduction

HYPERSPECTRAL images (HSIs) contain both hundreds of narrow spectral bands information and abundant spatial distribution information of land covers [1], which are widely used in agriculture, environmental monitoring, geosciences, surveying and mapping, and other fields [2], [3]. However, this feature also brings a series of challenges for hyperspectral image classification (HSIC), such as information redundancy caused by more spectral bands, low classification accuracy caused by less training samples, and single classification models, which cannot adapt to the complex data characteristics of HSIs. Therefore, it is of great significance to research how to classify HSIs accurately.

The early HSIC research focused on utilizing its spectral information to complete feature matching, such as spectral matching based methods [4], [5]. However, this kind of method cannot effectively distinguish the same kind of land covers with different reflection spectra and different kinds of land covers with the same reflection spectrum, and thus this kind of method tends to impair classification accuracy. Considering the high spectral resolution and spectral redundancy of HSIs, some basic feature extraction algorithms are also applied to HSIC, such as linear feature extraction based [6]–[8], nonlinear feature extraction based [9]–[11]. In addition, methods based on spectral features also include traditional machine learning methods, such as support vector machine (SVM) [12], extreme learning machine [13], sparse representation classification (SRC) [14], and so on. Compared with traditional methods based on spectral matching, these algorithms have better classification performance, but still depend on prior knowledge to set parameters. Besides, these methods rely on artificial feature extraction, when dealing with different HSI datasets, these features lack sufficient generalization ability and expression ability.

To address the above problems, some scholars have explored the spatial–spectral joint classification method. Zhang *et al.* [15] proposed nonlocal weighted joint SRC model, which adopted different weights for different adjacent pixels according to the structural similarity. Li *et al.* [16] proposed spatial–spectral kernel SVM, which extracted spectral–spatial features by principal component analysis and median filter, respectively, and then utilized SVM to classify spatial–spectral joint features. This method can effectively reduce the influence of noise and make full use of spatial–spectral features to improve the classification accuracy. However, it has been proved that the spectral and spatial domains of the original HSI data are highly correlated and redundant. Jia *et al.* [17] developed a new subspace-based multitask learning framework for HSIC. The original HSI data space was projected into several different subspaces, and SVM classification was carried out in each subspace. In order to make full use of the spatial information, Markov random field was applied to process the results of SVM classification, and finally, the classification results are determined through the decision fusion.

In recent years, deep learning (DL) technology has made a breakthrough. Its powerful feature extraction ability is far beyond the traditional classification methods, so researchers try to extend DL to HSIC. Typical DL classification models include stack autoencoder (SAE) [18], deep belief network (DBN) [19], recurrent neural network (RNN) [20], [21], and convolutional neural network (CNN) [22], [23]. Although SAE and DBN can extract deeper features, they need to transform the input data

Manuscript received August 27, 2021; revised December 5, 2021; accepted January 13, 2022. Date of publication January 25, 2022; date of current version February 11, 2022. *(Corresponding author: Xin Wang.)*

The authors are with the School of Oceanography and Spatial Information, China University of Petroleum (East China), Qingdao 266555, China (e-mail: 3166588225@qq.com; ygfan@upc.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3145917

into a one-dimensional (1-D) vector, which leads to the loss of spatial information. RNN uses a hidden layer or storage unit to learn state features, which has attracted extensive attention in sequence data analysis. By considering the spectral signature as a sequence, RNN has been successfully used to learn discriminative features from HSIs recently. Zhou *et al.* [21] used two long short-term memory networks to learn the features of spectral and spatial sequences of HSIs, and fused the learning results at the decision level However, decision level fusion depends on the previous results, which affects the classification results. In this situation, CNN has achieved excellent performance on computer vision tasks and attracts considerable attention. So scholars began to explore the use of CNN to extract the spatial–spectral information of HSI. Chen *et al.* [23] used two CNN frameworks to extract spectral features and spatial features, respectively. However, the method can lead to redundant computation by extracting spectral–spatial features separately. In order to further extract spatial–spectral joint features, researchers apply a 3-D convolutional neural network (3D-CNN) to HSIC [24], [25]. For example, Shi and Pun [24] first obtained the preliminary classification results by using super-pixel segmentation, and then further extracted the depth features by 3D-CNN. Liu *et al.* [25] transport the original features to 3D-CNN without any preprocessing. However, the calculation cost of 3D-CNN is expensive, and the phenomenon of overfitting is easy to occur.

Through previous studies, it is found that the simple overlay convolution layers cannot satisfy the requirements of HSIs, and increasing the depth of the network model requires more training samples, whereas the HSI training samples are often less. Therefore researchers began to focus on the limited training samples to further improve the accuracy of HSIC. Zhong *et al.* [26] employed a spatial–spectral residual 3-D convolutional neural network (SSRN) to extract the spatial–spectral joint features and spatial context discrimination features of HSIs, which effectively improved the classification accuracy. Wang *et al.* [27] proposed an end-to-end fast dense spectral–spatial convolution (FDSSC) framework for HSIC to reduce the training time and improve accuracy. Different from SSRN, FDSSC uses a dense connection structure instead of residual structure to construct the network model. Li and Shang [28] applied a spatial–spectral pseudo-three-dimensional dense connection network (SSP3DNet) to reduce the training parameters and the overfitting phenomenon in the process of model training. Li *et al.* [29] utilized a deep multilayer feature fusion dense connection network (MFDN) to extract spatial and spectral features simultaneously based on different input sizes, and high-level abstract features through 3-D dense blocks, which effectively alleviated the problem of vanishing gradient, enhances feature propagation, encourages feature reuse, and improves the accuracy of HSIC. Wang *et al.* [30] proposed a dual-branch spatial–spectral dense residual neural network (DRN), which adopted 1D-CNN to extract spectral features and 2D-CNN to extract spatial information, and each branch used dense residual structure to enhance the feature extraction and reduced the problem of gradient disappearance. Hang *et al.* [31] proposed a multitask generative adversarial network to alleviate the limited samples issue by taking advantage of the rich information from unlabeled samples. Through the confrontational learning method, the discrimination ability and generalization ability of classification tasks are indirectly improved. However, these methods can only extract feature information by fixed convolution kernel scale, which is not conducive to feature learning, ignores the complex features of HSIs at different scales, and damages the classification accuracy. Not only that the increase of model depth and special connection mode also bring too many training parameters, resulting in slow convergence of the model.

In order to further extract discriminative features, an attention mechanism is also applied to DL, which can consciously extract features beneficial to classification by simulating the characteristics of human eyes. Gao *et al.* [32] combined the attention module with dense connected network (DMSAN) to enhance the features that are more relevant to classification, and weaken the features that are less relevant. Yu *et al.* [33] employed the multiscale feedforward attention module to extract semantic features, which effectively improves the computational efficiency and recognition ability of feature representation. Hang *et al.* [34] proposed a spectral attention subnetwork and a spatial attention subnetwork for spectral and spatial classifications and combined classification results via adaptively weighted summation method to aid networks that focus on more discriminative channels or positions. Based on the above analysis, we propose a multiscale densely connected attention network (MSDAN) for HSIC. First, multiscale dense connection blocks are used to extract different scale features of HSI to avoid the defect of insufficient information. Second, 3-D convolution blocks in the network are replaced by 3-D spatial convolution blocks and 3-D spectral convolution blocks in series to reduce training parameters. Finally, the attention mechanism is embedded into the end of each scale to enhance the discriminative features that are beneficial to classification and improve the performance of the network.

The main contributions of this article are listed as follows.

1) This article adopts a multiscale dense connection model to synchronously extract the comprehensive features of different scales of HSI, so as to enhance the feature extraction ability of the network and make it suitable for the complex data characteristics of HSIs. Through dense connection, we can realize the close relationship between features of different layers, strengthen feature reuse, and avoid the phenomenon of overfitting and gradient disappearance.
2) In the dense connection model, 3-D spatial convolution block and 3-D spectral convolution block are employed in series instead of traditional 3-D convolution block to reduce training parameters and accelerate the model convergence.
3) An improved attention mechanism module is proposed, which extracts the weight information of spatial dimension, spectral dimension, and channel dimension, respectively, and fuses the weight information from the three dimensions by feature multiplication, and then embeds it into the end of each scale channel. The convolution kernel size is the same as that of the channel to strengthen

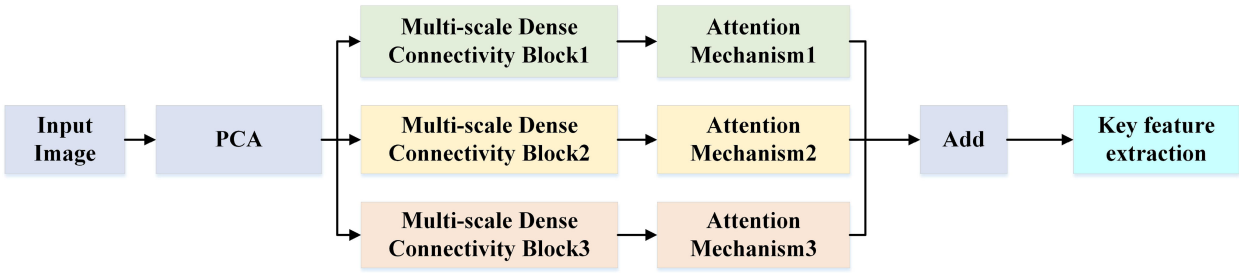

Fig. 1. Components of the proposed MSDAN.

the feature information of each channel, enhance features favorable for classification, and improve the feature extraction ability of the network.

The rest of this article is organized as follows. Section II introduces the detailed architecture of MSDAN. The experimental results and analysis are illustrated in Section III. Finally, Section IV concludes this article.

II. Methodology

The proposed network (MSDAN) is mainly composed of multiscale dense connection module, attention mechanism module, and key feature extraction module, as shown in Fig. 1. First, PCA transformation is adopted to retain the most important features and remove the interference of noise of the input data. Then, the features of the inputs are extracted synchronously by the multiscale dense connection module to adapt to the complex characteristics of HSIs. Next, the attention mechanism module composed of the corresponding convolution kernels is adopted to strengthen the relevant features that are beneficial to classification, and reduce the weight information of irrelevant features. Finally, the key feature extraction module is employed to extract discriminative features along the spectral, spatial, and channel dimensions.

A. Dense Connection Module

In order to avoid the problem of overfitting and gradient disappearing, researchers proposed DenseNet [35] to enhance feature transmission, encourage feature reuse and improve information flow in the network.

The DenseNet is connected layer by layer. The input features of each layer receive the output feature information from the previous layer and are superimposed on the channel dimension, which can be expressed as

$$x_l = H([x_0, x_1, \ldots, x_{l-1}]) \tag{1}$$

where x_l represents the output of the l layer, $[x_0, x_1, \ldots, x_{l-1}]$ represents the superposition of the characteristic maps from the input layer to the $l-1$ layer on the channel dimension, and $H(\bullet)$ represents the combination of nonlinear transformation functions including convolution, normalization, nonlinear activation, and other operations.

Assuming that the number of input characteristic graphs of the first layer in the dense connection layer is k_0 and the number of output characteristic graphs of each layer in the dense connection layer is k, the number of channels of the input characteristic graphs of the l layer can be formulated as

$$k_0 + (l-1)k. \tag{2}$$

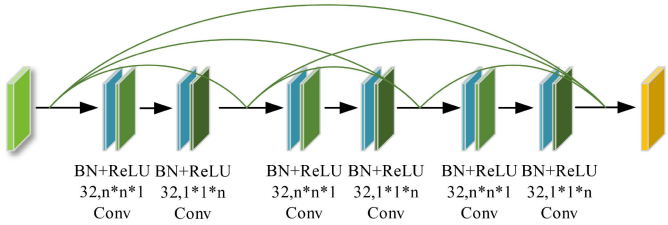

Fig. 2. Each scale dense connection structure. n represents the convolution kernel size of each channel.

Each layer in the dense connection layer receives the features from all previous layers, which encourages the features propagation, strengthens the feature reuse, and effectively suppresses the problems of gradient disappearance and overfitting.

B. Multiscale Dense Connection Module

Previous studies have found that the single-scale network model cannot extract the rich spatial–spectral features of HSIs, whereas the multiscale network model can extract more comprehensive information. In this article, we adopt multiscale dense connection to extract feature information of different scales synchronously. Each scale applies dense connection to enhance feature propagation and reuse, and employs 3-D spectral convolution and 3-D spatial convolution in series to replace 3-D convolution layer to reduce the training parameters caused by increased model complexity. The structure of each scale is shown in Fig. 2. The convolution kernels of the three scales are 3, 5, and 7, respectively, which are connected in $n \times n \times 1$ and $1 \times 1 \times n$ convolution order to form dense connection blocks. The $n \times n \times 1$ convolution is used to extract spatial information, and the $1 \times 1 \times n$ convolution is used to extract spectral information, where n represents the convolution kernel size of each scale. The number of output features of each dense connection layer is 32, and the number of dense connection layer is 3. Feature information of different scales is fused by feature addition, which can be expressed as

$$\begin{aligned} X = Add\,(&F_1([X_0, X_1, \ldots, X_{i-1}]),\\ &F_2([X_0, X_1, \ldots, X_{i-1}]),\\ &F_3([X_0, X_1, \ldots, X_{i-1}])) \end{aligned} \tag{3}$$

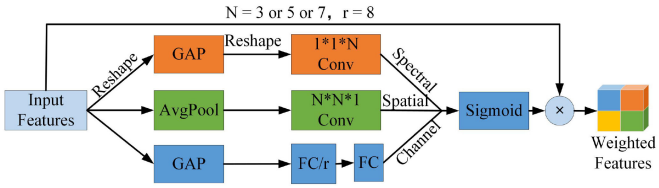

Fig. 3. Structure of the proposed attention.

where X represents the output features after fusion, $Add(\bullet)$ represents the additive function of the corresponding elements of the features, and $F_1(\bullet)$, $F_2(\bullet)$, $F_3(\bullet)$ represent the output features of the dense connected network of each scale.

C. Attention Mechanism Module

Inspired by the fact that human eyes consciously accept the characteristics of the region of interest, researchers put forward the attention mechanism and added it to the DL network model. It can consciously ignore the information irrelevant to the current task from many features and pay attention to the information related to the current task. In HSIC, the attention mechanism can improve the learning ability of the network by giving more weight to the features that are beneficial to classification. Common attention mechanism models include spatial attention mechanism and channel attention mechanism. The spatial attention mechanism is to redistribute the weight of the spatial information of the same feature graph to obtain feature graphs with different spatial weight information. Channel attention mechanism is to redistribute the weight of feature graph in channel dimension to obtain the weight feature graph of different channels.

Inspired by the above ideas, this article proposes an improved attention mechanism model. The model extracts the corresponding feature information in the spatial dimension, spectral dimension, and channel dimension, respectively, to strengthen the discriminative features favorable to classification and suppress the unimportant information, and fuses the weight information of the three features to obtain the weight feature graph. The structure of the proposed attention mechanism is shown in Fig. 3.

Suppose that the input feature map is expressed as $X \in R^{H\times W\times L\times C}$, where represents the spatial size of the feature map, L represents the spectral dimension of the feature map, and C represents the channel number of the feature map.

Spectral attention mechanism can give higher weight todiscriminant features in spectral dimension to obtain different spectral feature weight maps. First, we need to transform the dimension of the feature map $X \in R^{H\times W\times L\times C}$ to $X \in R^{H\times W\times C\times L}$, and then, through the global average pooling (GAP) layer, the dimension of the feature graph is transformed into $X \in R^{1\times 1\times 1\times L}$. Next, we need to transform the dimension $X \in R^{1\times 1\times 1\times H}$ to $X \in R^{1\times 1\times L\times 1}$, and extract the spectral features of the feature map by $1 \times 1 \times N$convolution. Finally, we map the eigenvalues to 0–1 through the sigmoid layer, and get the weighted feature map of the spectral dimension, where Ndenotes the convolution kernel size. It can be formulated as

$$F_{se}\left(Sigmoid\left(Conv\left(Re\left(GAP\left(Re\left(X\right)\right)\right)\right)\right)H\times W\right) \quad (4)$$

where $F_{se}(\bullet)$ represents the final spectral dimension weight feature map, $Sigmoid(\bullet)$ represents the sigmoid function, $Conv(\bullet)$ represents the convolution operation, $Re(\bullet)$ represents the dimension transformation, and $GAP(\bullet)$ represents the global average pooling.

Spatial attention mechanism can obtain spatial feature maps with different weights in spatial dimension. First, we need to use average pooling to transform the feature map from $X \in R^{H\times W\times L\times C}$ to $X \in R^{H\times W\times 1\times 1}$, and get the feature map that only contains spatial dimension. Then, we use $N \times N \times$ 1convolution to extract the spatial features of the feature map. Finally, we map the feature values to 0–1 through sigmoid layer and get the weighted feature map of spatial dimension, where N represents the convolution kernel size. It can be expressed as

$$F_{sa}\left(Sigmoid\left(Conv\left(Avepool\left(X\right)\right)\right)\right) \quad (5)$$

where $F_{sa}(\bullet)$ represents the final spatial dimension weight characteristic graph, $Sigmoid(\bullet)$ represents the sigmoid function, $Conv(\bullet)$ represents the convolution operation, and $Avepool(\bullet)$ represents the average pooling.

Channel attention mechanism can obtain feature maps with different weights in channel dimension. First, we need to use the GAP to transform the feature maps from $X \in R^{H\times W\times L\times C}$ to $X \in R^{1\times 1\times 1\times C}$, and get the feature map that only contains channel dimensions. Then, we use two full connection (FC) layers to extract the feature information of channel dimensions. Finally, we map the feature values to 0–1 through the sigmoid layer, and finally get the channel dimension weight feature map. It can be expressed as

$$F_{ca}\left(Sigmoid\left(FC\left(FC\left(GAP\left(X\right)\right)\right)\right)\right) \quad (6)$$

where $F_{ca}(\bullet)$ represents the final channel dimension weight characteristic graph, $Sigmoid(\bullet)$ represents the sigmoid function, $FC(\bullet)$ represents the full connection operation, and $GAP(\bullet)$ represents the global average pooling.

Finally, the obtained spectral, spatial, and channel weight feature maps are combined to obtain the 3-D weight feature map $F_w \in R^{H\times W\times L\times C}$, which is weighted by multiplying it with the input feature graph $X \in R^{H\times W\times L\times C}$, so as to improve the learning ability of the network. The process can be expressed as

$$F_w = F_{se} \otimes F_{sa} \otimes F_{ca} \quad (7)$$

$$F = X \otimes F_w \quad (8)$$

where $F(\bullet)$ represents the weighted input feature map, $F_w(\bullet)$ represents the 3-D weighted feature map, and $\otimes$ represents the multiplication of corresponding elements.

D. Key Feature Extraction Module

In order to further extract the discriminant features of multiscale fusion, this article sets up a key feature extraction module. The specific information of this module is shown in Fig. 4.

First, 3-D spectral convolution block and dimension transformation are used to extract the features of spectral dimension and channel dimension. Finally, the feature information of spatial dimension is extracted by maximum pooling and 3-D spatial convolution block. This operation can fully extract the feature

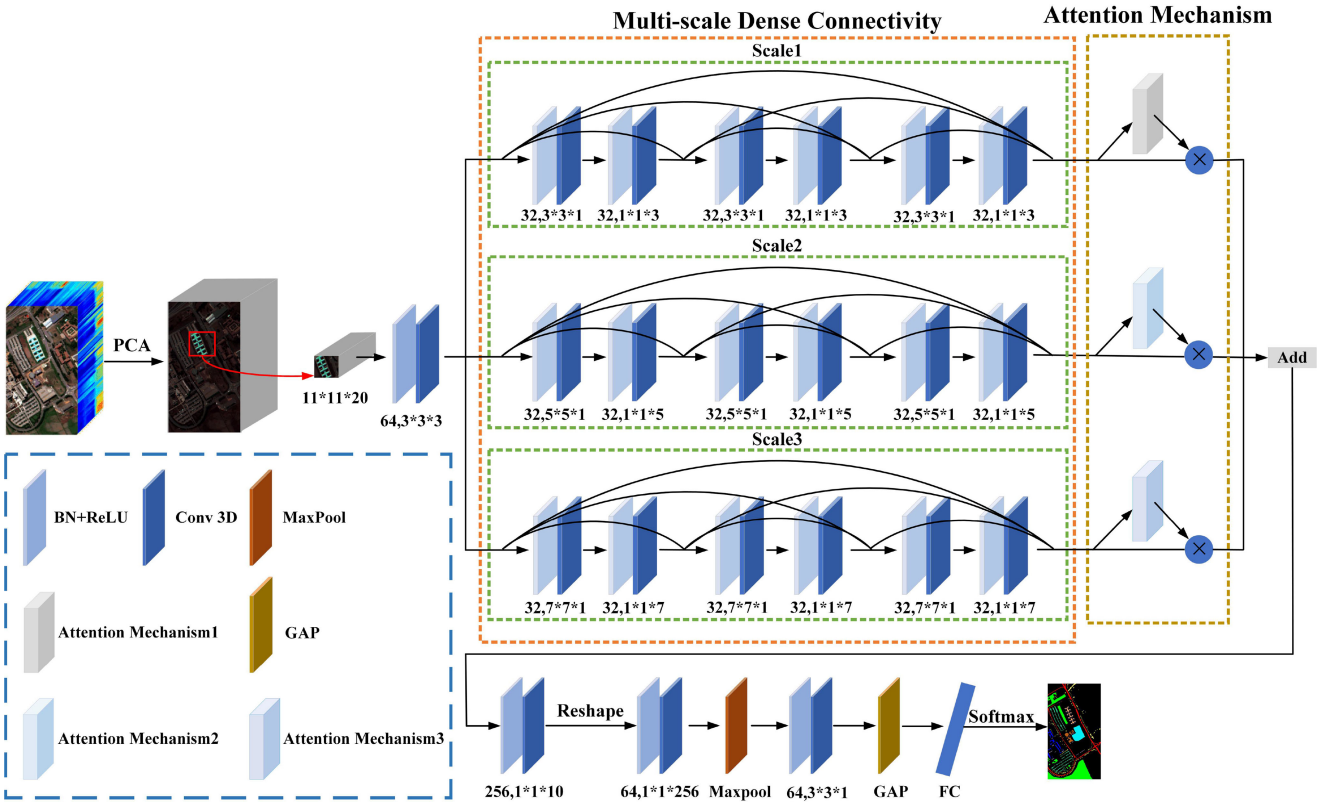

Fig. 4. Overall network model.

information from the multiscale dense connection layer and improve the classification performance of the network.

E. Overall Network Model

The proposed network model (MSDAN) is illustrated by the University of Pavia (UP) dataset, as shown in Fig. 4. The network is mainly composed of multiscale dense connection module, attention mechanism module, and key feature extraction module. First, the HSI has many spectral bands and high correlation between them, so it is necessary to reduce the feature dimension of the input data through PCA transformation, so as to retain the bands with large amount of information and remove the interference of noise and unnecessary information. The number of reserved bands is 20, and its spatial neighborhood is taken as the input of the network. The size of the spatial neighborhood is 11×11, and then through the initial convolution layer 64 $3 \times 3 \times 3$, whose strides are (2,1,1), and the dimension of the feature graph is transformed from $11 \times 11 \times 20$ to 64 $11 \times 11 \times 10$. Then, different features are extracted synchronously by the multiscale dense connection module to adapt to the complex characteristics of HSIs. Scale1 uses $3 \times 3 \times 1$ and $1 \times 1 \times 3$ by the series connection form to form the dense connection block to extract the spatial features and spectral features of the input data, respectively. Compared with the traditional 3D-CNN, it can reduce the training parameters. Second, the spatial–spectral–channel attention mechanism with a convolution kernel size of 3 is used to extract the important information of features, which further enhances the learning ability of the network. Scale2 uses $5 \times 5 \times 1$ and $1 \times 1 \times 5$ in series to form a dense connection block with three dense connection layers. A spatial–spectral–channel attention mechanism with convolution kernel size of 5 is embedded at the back end of the dense connection layer to extract discriminative features. Scale3 uses $7 \times 7 \times 1$ and $1 \times 1 \times 7$ to form dense connection blocks, and embeds the spatial–spectral–channel attention mechanism with convolution kernel size of 7 to enhance the relevant features that are beneficial to classification while reducing the weight information of irrelevant features. Finally, the multiscale fusion features are obtained by adding the corresponding elements, and the size of the feature map is transformed from 64 $11 \times 11 \times 10$ to 160 $11 \times 11 \times 10$. Then, the discriminative feature information of spectral and channel dimensions is extracted by 3-D spectral convolution block and dimension transformation, and the spatial features of the input features are further extracted by max pooling and 3-D spatial convolution block. Finally, through the FC layer and Softmax layer, the classification results of UP are obtained.

III. Experimental Results and Analysis

This part mainly introduces the datasets used in the validation experiment, a series of parameters of the model training, and the analysis of the experimental results.

A. Datasets Introduction

In order to verify the effectiveness of the proposed method, three HSI datasets are used for experimental verification.

Indian Pines (IN) dataset was acquired by airborne visible infrared imaging spectrometer (AVIRIS) from an Indian Pine

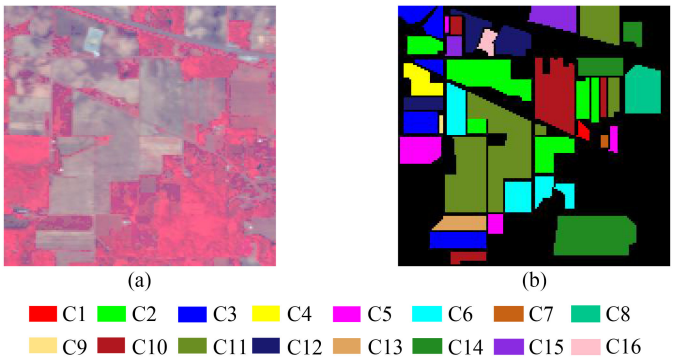

Fig. 5. IN dataset. (a) RGB composite image of three of the IN dataset. (b) Ground-truth map of the IN dataset.

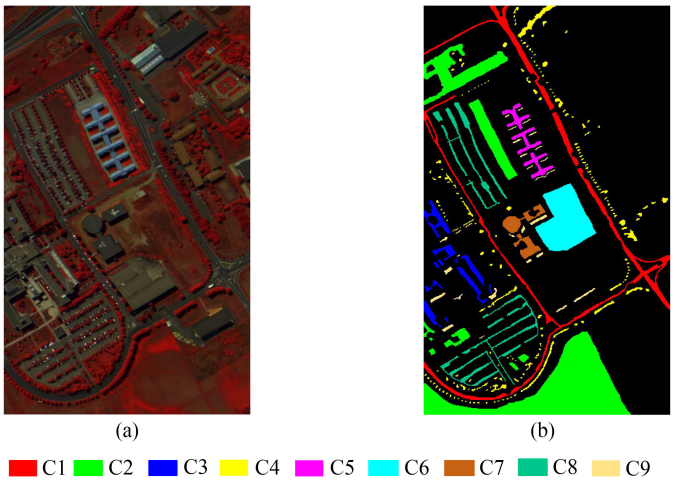

Fig. 6. UP dataset. (a) RGB composite image of three of the UP dataset. (b) Ground-truth map of the UP dataset.

tree in Indiana, USA. The image size is 145 × 145. AVIRIS imaging wavelength range is 0.4–2.5 μm. The spatial resolution is about 20 m. After eliminating 20 bands that cannot be reflected by water, the remaining bands are 200. The dataset contains 21 025 pixels, including 10 249 feature pixels, 10 776 background pixels, and 16 types of feature types, most of which are natural landscapes, and the distribution of samples is extremely uneven. The dataset is shown in Fig. 5.

Pavia University (UP) dataset was acquired by German reflective optics spectral imaging system (ROSIS-03) imaging Pavia city in Italy. The wavelength range of the image is 0.43–0.86 μm, and the spatial resolution is 1.3 m. After eliminating 12 bands affected by noise, 103 bands are left. The image size is 610 × 340, including 2 207 400 pixels, 42 776 feature pixels, and nine types of features, which are trees, asphalt roads, bricks, meadows, etc. The dataset is shown in Fig. 6.

University of Houston (HT) dataset was acquired by ITRES CASI-1500 sensor. The wavelength range of the image is 0.38–1.05 μm. The image is one of the multimodal optical remote sensing datasets released by the 2018 data fusion competition of the IEEE Geosciences and Remote Sensing Society [36], covering the HT campus and its surrounding urban areas. The original image size is 4172 × 1202, after clipping, selected part

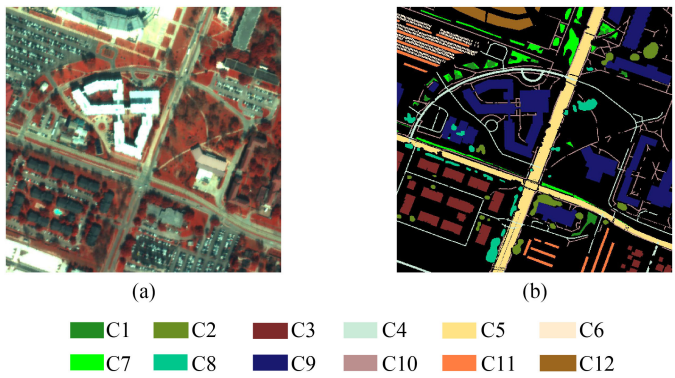

Fig. 7. HT dataset. (a) RGB composite image of three of the HT dataset. (b) Ground-truth map of the HT dataset.

TABLE I
NUMBER OF RANDOMLY SELECTED TRAINING SET, VERIFICATION SET, AND TEST SET OF IN DATASET

No.	Category	Train.	Val.	Test.
C1	Alfalfa	11	8	27
C2	Corn-notill	275	142	1011
C3	Corn-min	175	64	591
C4	Corn	44	26	167
C5	Grass/Pasture	104	44	335
C6	Grass/Trees	132	99	499
C7	Grass/Pasture-mowed	7	2	19
C8	Hay-windrowed	97	50	331
C9	Oats	4	2	14
C10	Soybeans-notill	199	98	675
C11	Soybeans-min	492	233	1730
C12	Soybeans-clean	118	61	414
C13	Wheat	37	20	148
C14	Woods	260	134	871
C15	Building-Grass-Trees-Drives	77	32	277
C16	Stone-steel Towers	17	10	66
	TOTAL	2049	1025	7175

of the image as the study area, the size of the clipped image is 541 × 710, including 48 bands and 12 categories. The dataset is shown in Fig. 7. For IN dataset, 20% of training samples are randomly selected, 10% of verification samples, and 70% test samples are selected. For UP dataset, 10% of training samples, 10% of validation samples, and 80% of test samples are selected. For HT dataset, considering its data characteristics, only 5% of training samples, 10% of validation samples, and 85% of test samples are selected. The number of three samples selected for each dataset is shown in Tables I–III.

B. Experimental Configuration

The hardware environment is Intel (R) Core (TM) i5-10400F CPU 2.90 GHz, the memory is 16.0 GB, and the software environment is Python 3.7.9, TensorFlow2.0.0, Keras2.2.4.

In the process of model training, the number of batch size is set to 16, the optimizer is RMSprop, the learning rate is 0.0002, and the number of training epochs is set to 100.

TABLE II
NUMBER OF RANDOMLY SELECTED TRAINING SET, VERIFICATION SET, AND TEST SET OF UP DATASET

No.	Category	Traini.	Val.	Test.
C1	Asphalt	1321	649	4661
C2	Meadows	3694	1839	13116
C3	Gravel	412	222	1465
C4	Trees	623	315	2126
C5	Sheets	255	126	964
C6	Baresoil	1037	490	3502
C7	Bitumen	267	147	916
C8	Bricks	755	390	2537
C9	Shadows	190	100	657
	TOTAL	8554	4278	29944

TABLE III
NUMBER OF RANDOMLY SELECTED TRAINING SET, VERIFICATION SET, AND TEST SET OF HT DATASET

No.	Category	Train.	Val.	Test.
C1	Healthy grass	75	165	1373
C2	Stressed grass	256	444	3986
C3	Evergreen trees	219	389	3161
C4	Deciduous trees	184	338	2951
C5	Residential buildings	436	851	7816
C6	Non-residential buildings	1358	2730	22903
C7	Roads	311	655	5234
C8	Sidewalks	431	912	7339
C9	Major thoroughfares	761	1515	13066
C10	Paved parking lots	197	430	3484
C11	Cars	137	284	2488
C12	Stadium seats	141	301	2816
	TOTAL	4506	9014	76617

In order to quantitatively evaluate the performance of the proposed method, this article uses three standard evaluation indexes: overall accuracy (OA), average accuracy (AA), and Kappa coefficient (κ) to evaluate the classification performance.

C. Parameter Analysis

This part mainly analyzes the influence of various parameters of the network on the classification results.

The classification performance of the model is not only related to the structure of the model but also related to the setting of various parameters in the model. This section mainly discusses the influence of spatial size, the number of convolution kernels in dense connection blocks and the number of dense connection layers in dense connection blocks on the classification accuracy of network. All experiments used the control variable method to analyze the influence of parameters.

First, the influence of the spatial size. Because the feature distribution of HSI is related to the spatial dimension, and the samples need to take a certain spatial size before inputting to the network. Different HSIs have different feature distribution, and different spatial sizes may produce different HSIC results. Therefore we fixed other factors and set spatial sizes to 7×7, 9×9, 11×11, 13×13, and 15×15 for the three HSI datasets to compare the overall classification accuracy. Fig. 8(a) shows the influence of different spatial sizes on the overall classification accuracy. It can be seen that the value of OA increases at first, and then decreases as the spatial size reaches 11×11 for all the three datasets. For IN and HT datasets, as the spatial size is 13×13, the OA value starts to increases. For UP dataset, as the spatial size is larger than 11×11, the OA value decreases all the time. Therefore, we choose the space size corresponding to the maximum accuracy, that is, 15×15 for IN dataset and 11×11 for UP and HT datasets.

Second, the influence of the number of convolution kernels in dense connection blocks. Dense connection blocks are composed of convolution layers, and the number of output characteristic graphs of each convolution layer affects the complexity of the network and the overall performance of classification. In order to analyze the number of optimal convolution kernels, the classification results of three HSI datasets are analyzed. The number of convolution kernels is set to 8, 16, 24, and 32, respectively. Fig. 8(b) shows the influence of the number of the convolution kernels on the OA. With the increase of the number of convolution kernels, the OA value of the three datasets increases at first, then decreases as the value is 16, and then increases as the number of convolution kernels is 24. Therefore, the number of the convolution kernels in the dense connection blocks of IN, UP, and HT datasets is set to 32.

Third, the influence of the number of dense connection layers in dense connection blocks. The number of dense connection layers determines the depth of dense connection blocks, and indirectly determines the depth and classification accuracy of network. Therefore, it is of great significance to explore the influence of the number of dense connection layers on the classification results. Fig. 8(c) shows the influence of the number of dense connection layers on the overall classification accuracy. The number of dense connection layers is set to 2, 3, and 4, respectively. For the three datasets, with the increase of the number of dense connection layers, the classification accuracy first increases and then decreases. When the number of dense connection layers is 3, the classification accuracy is the highest. Therefore, the number of dense connection layers of the three datasets is 3.

Finally, the influence of the percentage of training samples. In order to test the robustness and generality of the proposed MSDAN, 1%, 5%, 10%, 15%, and 20% are selected as the training samples for IN dataset; 1%, 3%, 5%, 7%, and 10% for UP dataset; and 0.1%, 1%, 5%, 10%, and 20% for HT dataset. Fig. 9 shows the impact of the proportion of different training samples of three datasets on classification performance. It can be seen that the MSDAN is more stable. With the increase of the proportion of training samples, the classification accuracy gradually improves and is the highest of other methods.

D. Analysis of Classification Results

In order to verify the effectiveness of the proposed method, the proposed MSDAN is compared with several classical network models. These classical networks are MFDN [29], FDSSC [27], SSP3DNet [28], DMSAN [32], SSRN [26], and DRN [30].

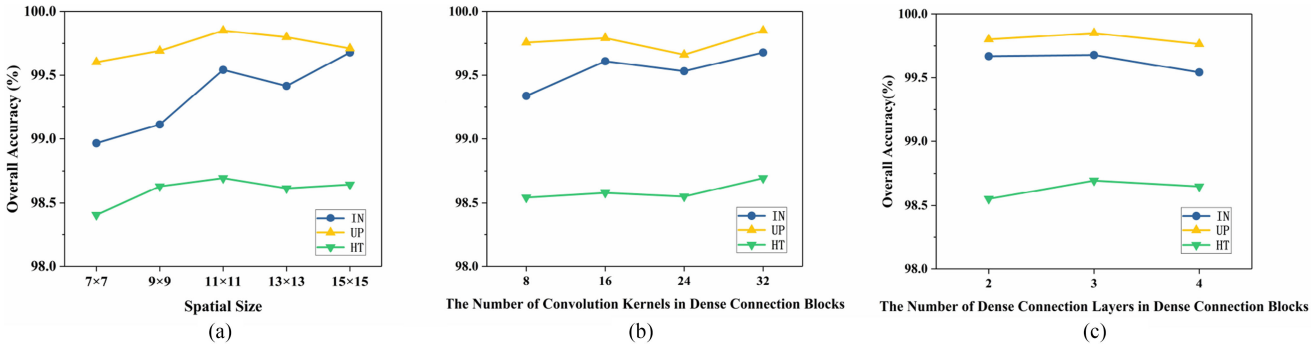

Fig. 8. Parameter analysis. (a) Influence of spatial sizes. (b) Influence of the number of convolution kernels in dense connection blocks. (c) Influence of the number of dense connection layers in dense connection blocks.

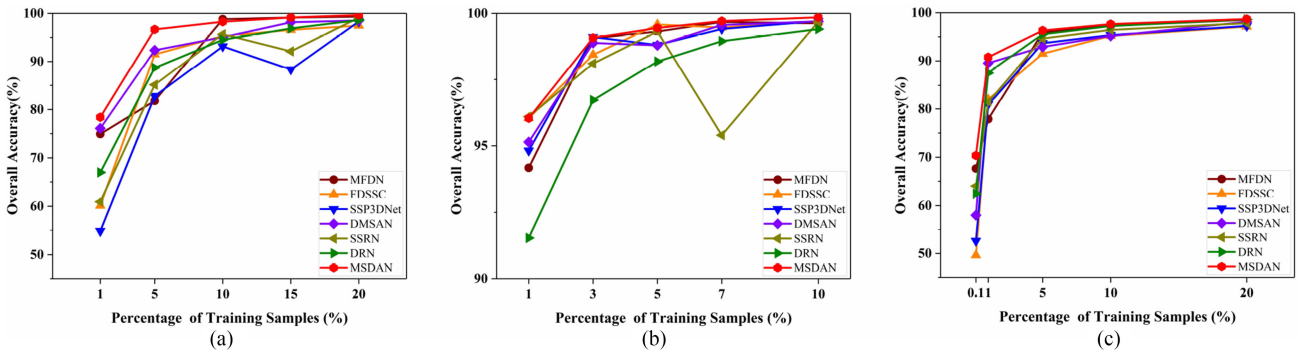

Fig. 9. Percentage of training samples. (a) IN dataset. (b) UP dataset. (c) HT dataset.

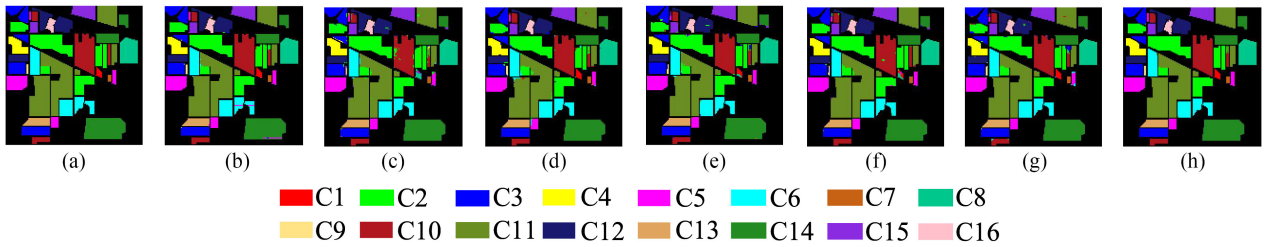

Fig. 10. IN dataset and classification results. (a) IN ground-truth map. (b) MFDN. (c) FDSSC. (d) SSP3DNet. (e) DMSAN. (f) SSRN. (g) DRN. (h) MSDAN.

Among them, SSRN and DRN use residual structure to extract spectral and spatial features, respectively. Compared with SSRN, FDSSC uses dense connection structure instead of residual structure to construct network model. MFDN utilized the deep multilayer feature fusion dense connection structure to extract spatial and spectral features simultaneously. SSP3DNet and DMSAN both adopt dense connection structure and pseudo-3D convolution structure. Besides, DMSAN adds attention mechanism and multiscale module. Different from the MSDAN proposed in this article, DMSAN is based on multiscale blocks to achieve dense connection between blocks.

In order to ensure the fairness of the experimental verification, all the methods are trained in the same environment, and the proportion of training samples is 20% for IN dataset, 10% for UP dataset, and 5% for HT dataset. In addition, other parameters of the comparison methods are the same as those of the original papers. All experiments were conducted for five times with randomly selected training samples and calculated the mean and standard deviation as the final main classification metrics.

Figs. 10–12 show the classification results of the three datasets. Compared with other methods, the classification result maps of the proposed MSDAN was most consistent with the ground-truth maps and it delivered the most accurate and smooth classification maps for all three HSIs.

Tables IV–VI present the accuracy evaluation results of the three datasets. It can be seen that our proposed MSDAN has the highest classification accuracy and lower standard deviation compared with other methods. In all three datasets, the classification results of FDSSC were worse than other methods and the standard deviations are large, showing an unstable trend. For IN dataset, from the classification accuracy of alfalfa and oats categories with few samples, FDSSC (7.97%, 23.33%), DMSAN (50%, 68.33%), and SSRN (52.17%, 51.67%) have worse classification effects on the two categories, whereas the

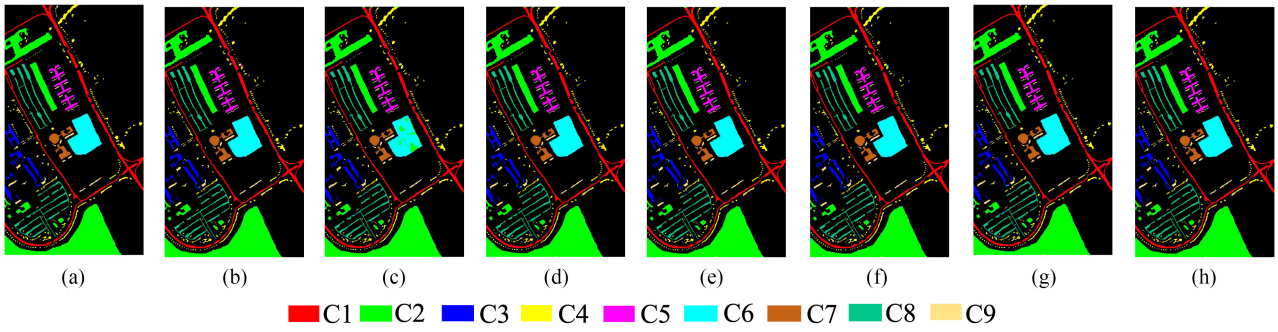

Fig. 11. UP dataset and classification results. (a) UP ground-truth map. (b) MFDN. (c) FDSSC. (d) SSP3DNet. (e) DMSAN. (f) SSRN. (g) DRN. (h) MSDAN.

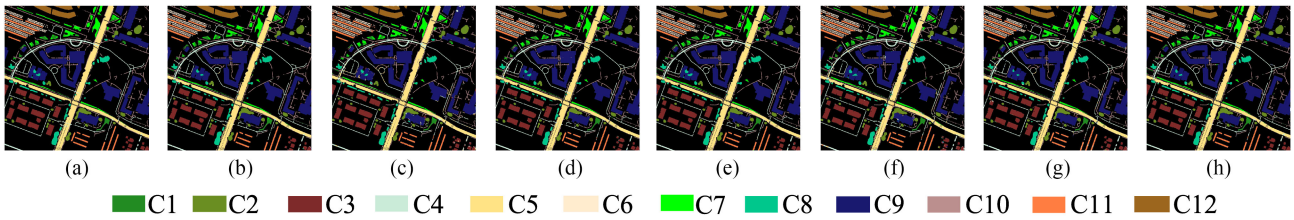

Fig. 12. HT dataset and classification results. (a) HT ground-truth map. (b) MFDN. (c) FDSSC. (d) SSP3DNet. (e) DMSAN. (f) SSRN. (g) DRN. (h) MSDAN.

TABLE IV
ACCURACY EVALUATION RESULTS OF THE IN DATASET

Classification Name	MFDN	FDSSC	SSP3DNet	DMSAN	SSRN	DRN	MSDAN
Alfalfa	92.03	7.97	97.10	50.00	52.17	89.86	**100.00**
Corn-notill	99.70	81.91	97.34	96.24	98.67	98.37	**99.88**
Corn-min	98.80	49.60	98.03	99.32	95.70	99.44	**99.82**
Corn	**100.00**	57.81	98.87	99.86	96.20	99.02	99.89
Grass/Pasture	99.31	96.69	96.20	96.55	96.83	96.55	**99.74**
Grass/Trees	99.22	96.85	99.59	99.45	97.67	99.77	**99.90**
Grass/Pasture-mowed	100.00	72.62	95.24	95.24	86.90	98.81	**100.00**
Hay-windrowed	100.00	100.00	99.72	99.93	100.00	100.00	**100.00**
Oats	96.67	23.33	96.67	68.33	51.67	95.00	**97.50**
Soybeans-notill	97.60	72.98	96.64	93.62	93.38	95.95	**98.66**
Soybeans-min	99.66	82.62	97.79	99.13	95.23	99.38	**99.70**
Soybeans-clean	**99.66**	90.89	94.66	98.99	99.49	98.48	99.11
Wheat	99.67	99.19	99.02	100.00	95.45	99.84	**100.00**
Woods	99.47	99.03	99.76	98.71	99.50	99.76	**99.94**
Building-Grass-Trees-Drives	99.83	99.05	98.19	98.10	98.36	99.83	**99.94**
Stone-steel Towers	98.21	94.62	96.77	93.91	**99.28**	94.27	98.66
OA (%)	99.31±0.17	83.93±2.27	97.89±0.34	97.76±0.73	96.70±1.77	98.75±0.04	**99.67±0.04**
AA (%)	98.74±0.79	76.57±0.70	97.60±0.37	92.96±1.72	91.03±4.05	97.77±0.36	**99.55±0.26**
κ×100	99.21±0.20	81.59±2.44	97.59±0.39	97.44±0.83	96.24±2.01	98.58±0.04	**99.62±0.04**

TABLE V
ACCURACY EVALUATION RESULTS OF THE UP DATASET

Classification Name	MFDN	FDSSC	SSP3DNet	DMSAN	SSRN	DRN	MSDAN
Asphalt	99.58	99.67	99.51	99.65	**99.74**	98.96	99.54
Meadows	**99.99**	99.98	99.97	99.94	99.93	99.97	99.98
Gravel	99.06	99.02	97.76	99.16	97.87	97.30	**99.59**
Trees	98.41	96.56	98.37	99.16	**99.41**	98.15	99.19
Sheets	100.00	100.00	100.00	100.00	100.00	99.68	**100.00**
Baresoil	99.90	88.03	99.80	99.98	99.46	99.65	**99.97**
Bitumen	100.00	99.92	98.92	99.50	99.70	98.32	**100.00**
Bricks	99.53	99.43	99.91	99.56	**99.98**	99.52	99.79
Shadows	99.75	99.86	99.47	99.37	**99.93**	99.08	99.65
OA (%)	99.71±0.12	98.19±1.18	99.61±0.14	99.75±0.04	99.71±0.01	99.40±0.04	**99.81±0.03**
AA (%)	99.58±0.20	98.05±1.21	99.30±0.22	99.59±0.09	99.56±0.08	98.96±0.06	**99.75±0.02**
κ×100	99.62±0.16	97.57±1.59	99.48±0.19	99.67±0.06	99.61±0.02	99.20±0.05	**99.75±0.04**

TABLE VI
ACCURACY EVALUATION RESULTS OF THE HT DATASET

Classification Name	MFDN	FDSSC	SSP3DNet	DMSAN	SSRN	DRN	MSDAN
Healthy grass	87.10	68.07	**88.68**	77.89	80.47	83.74	83.43
Stressed grass	92.04	94.65	90.65	92.17	**95.56**	92.92	92.37
Evergreen trees	98.32	98.29	94.57	97.05	98.24	97.61	**98.41**
Deciduous trees	95.61	93.69	94.81	95.81	**96.67**	91.74	96.21
Residential buildings	98.63	96.94	94.56	98.31	97.71	98.35	**99.54**
Non-residential buildings	**99.58**	97.42	98.45	98.85	98.61	98.81	99.34
Roads	92.45	84.00	84.89	87.20	85.73	86.16	**94.11**
Sidewalks	84.28	86.71	84.22	79.16	83.39	86.85	**90.30**
Major thoroughfares	**97.89**	83.76	92.82	96.11	93.41	97.55	96.01
Paved parking lots	98.41	97.42	95.40	97.35	**98.54**	96.76	96.47
Cars	**99.14**	90.70	97.82	95.68	94.67	97.18	99.12
Stadium seats	99.36	95.91	95.65	98.77	98.57	98.63	**99.84**
OA (%)	96.34±0.30	92.04±1.15	93.65±0.23	94.54±1.22	94.58±0.11	95.48±0.07	**96.64±0.22**
AA (%)	95.23±0.21	90.63±1.38	92.71±0.65	92.86±1.20	93.46±0.61	93.86±0.11	**95.43±0.17**
$\kappa\times100$	95.66±0.35	90.63±1.33	92.50±0.28	93.54±1.45	93.59±0.13	94.65±0.09	**96.03±0.26**

TABLE VII
TRAINING AND TESTING TIMES OF DIFFERENT MODELS FOR THREE HSI DATASETS

Dataset	Time (s)	MFDN	FDSSC	SSP3DNet	DMSAN	SSRN	DRN	MSDAN
IN	Train.	12.95	16.43	17.48	15.89	14.42	3.29	39.97
	Test.	12.34	15.63	16.76	18.18	9.44	3.95	35.64
UP	Train.	24.17	21.17	39.93	73.13	22.39	5.68	56.59
	Test.	48.13	36.51	73.73	112.78	31.52	10.09	104.92
HT	Train.	23.06	17.05	38.14	47.19	19.03	9.68	73.25
	Test.	80.51	48.71	55.62	160.18	45.83	54.36	222.85

TABLE VIII
TRAINABLE PARAMETERS OF DIFFERENT MODELS FOR UP DATASET

Model	MFDN	FDSSC	SSP3DNet	DMSAN	SSRN	DRN	MSDAN
Total Trainable Parameters	3, 887, 184	1, 350, 156	180, 395	187, 929	253, 931	6, 961, 288	1, 299, 793

classification accuracy of each category in MSDAN has reached more than 97%, indirectly indicating that the proposed MSDAN can be better applied to the classification of categories with unbalanced samples and has strong adaptability. For UP dataset, the average classification accuracy of every category for MSDAN is above 99%, which shows better stability. For HT dataset, all methods show poor performance in the classification accuracy of healthy grass, which may be related to the distribution of this category. Nevertheless, the OA of MSDAN still reaches 96.64%, about 0.3% higher than that of MFDN. These results validate the robustness of the MSDAN in the face of difficult conditions.

Besides, the training and testing times can provide a direct measure of computational efficiency for different models. Table VII records the training and testing times of each model. All records are the average results obtained by running five times in the same environment, in which the training times are the average times taken by each epoch. As shown in Table VII, the time of DRN is shorter than other models for all the three datasets, because DRN is simple without involving 3-D convolution. However, the MSDAN requires a larger amount of computational power than others. This is because MSDAN has multiscale modules, attention mechanism and dense connection structure, which increases the time of model training. On the contrary, it is the existence of these structures that improves the classification performance correspondingly. Therefore, without considering the computation time of the model, the MSDAN can be effectively applied to HSIC.

Table VIII presents the total trainable parameters of different models for UP dataset. As can be seen, DRN and MFDN have more trainable parameters than others, because they extract spatial and spectral information, respectively, through dual-channel structure, which causes the phenomenon of computational redundancy. Compared with SSRN, FDSSC has more trainable parameters, indicating that dense connection structure can increase training parameters. The number of parameters of our proposed MSDAN is less than that of FDSSC, DRN, and MFDN, but more than that of SSP3DNet, DMSAN, and SSRN. This is because the multiscale branches and dense connection structure lead to more training parameters.

E. Ablation Study

In order to further analyze the importance of attention mechanism, multiscale module, and dense connection structure in MSDAN, some comparative experiments are carried out. The classification results of the model without attention mechanism (MSDN), without multiscale module (DAN), or without dense connection structure (MSAN) are compared with MSDAN to verify the impact of each module on the classification results.

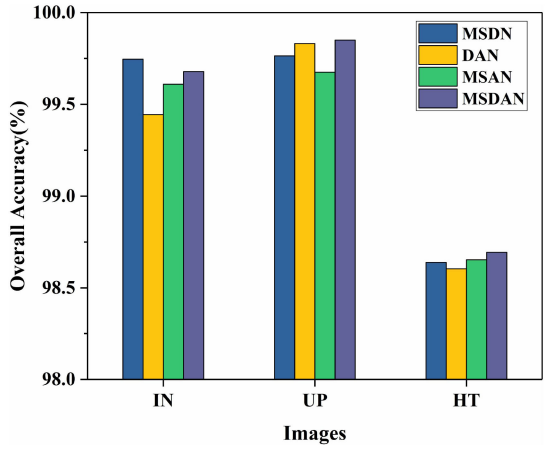

Fig. 13. Results of comparative experiments.

In Fig. 13, the addition of each module in MSDAN all promotes the model classification performance. First of all, for UP and HT datasets, the OA of MSDN is lower than that of MSDAN. For IN dataset, MSDN is 0.07% higher than that of MSDAN. Nevertheless, MSDAN has high classification accuracy for small samples, and therefore MSDAN is effective in the case of little loss of accuracy. Second, for IN, UP, and HT datasets, the OA of DAN is 0.24%, 0.02%, and 0.09% lower than that of MSDAN, respectively, which proves that multiscale module is beneficial for MSDAN to fully extract features. Similarly, compared with MSDAN, the OA of MSAN is reduced by 0.06%, 0.17%, and 0.04%, respectively, which indicates that the dense connection structure in MSDAN contributes to feature propagation. Therefore, to a certain extent, attention mechanism, multiscale structure and dense connection pattern in MSDAN are effective to improve the classification performance.

IV. Conclusion

In this article, we proposed an MSDAN model. The multiscale dense connection module integrates the feature information of different scales and layers to strengthen the feature extraction and feature propagation of the model. At the same time, the improved 3-D convolution blocks reduce the model parameters. Besides, the embedded spectral–spatial–channel attention module integrates the weight information of spectral, spatial, and channel dimensions. It not only fully extracts the discrimination features of the corresponding scale but also strengthens the weight information of different dimensions. At the end of the model, 3-D spectral convolution and 3-D spatial convolution are used to further extract the fusion features of different scales to enhance the learning ability of the network. The experimental results on three HSI datasets show that the proposed model has strong classification performance and adaptability. Although the proposed method has shown considerable results, further research should be developed to achieve higher classification accuracy with fewer samples in the future work.

Acknowledgment

The authors would like to thank NCALM and the Hyperspectral Image Analysis Laboratory at HT for providing the ITRES CASI-1500 HT dataset, and the Image Analysis and Data Fusion Technical Committee of the IEEE Geosciences and Remote Sensing Society for supporting the annual Data Fusion Contest.

References

[1] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.

[2] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, pp. 2659–2702, Aug. 2020.

[3] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[4] Y. Sohn and N. S. Rebello, "Supervised and unsupervised spectral angle classifiers," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 12, pp. 1271–1280, Dec. 2002.

[5] C. I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

[6] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 9, no. 3, pp. 447–451, May 2012.

[7] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.

[8] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[9] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2786–2803, Oct. 2006.

[10] B. Schlkopf and A. Smola, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1300–1319, Sep. 1998.

[11] D. Lee and H. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[12] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[13] Y. Bazi, N. Alajlan, F. Melgani, H. AlHichri, S. Malek, and R. R. Yager, "Differential evolution extreme learning machine for the classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 6, pp. 1066–1070, Jun. 2014.

[14] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[15] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.

[16] L. Li, H. Ge, and J. Gao, "A spectral-spatial kernel-based method for hyperspectral imagery classification," *Adv. Space Res.*, vol. 59, no. 4, pp. 954–967, Feb. 2017.

[17] S. Jia, B. Deng, J. Zhu, X. Jia, and Q. Li, "Superpixel-based multitask learning framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2575–2588, May 2017.

[18] J. Feng, L. Liu, X. Zhang, R. Wang, and H. Liu, "Hyperspectral image classification based on stacked marginal discriminative autoencoder," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Dec. 2017, pp. 3668–3671.

[19] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.

[20] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.
[21] F. Zhou, R. L. Hang, Q. S. Liu, and X. T. Yuan, "Hyperspectral image classification using spectral-spatial LSTMs," *Neurocomputing*, vol. 328, pp. 39–47, Feb. 2019.
[22] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, no. 2, Jul. 2015, Art. no. 258619.
[23] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
[24] C. Shi and C. M. Pun, "Superpixel-based 3D deep neural networks for hyperspectral image classification," *Pattern Recognit.*, vol. 74, pp. 600–616, Feb. 2018.
[25] B. Liu, X. Yu, P. Zhang, X. Tan, R. Wang, and L. Zhi, "Spectral-spatial classification of hyperspectral image using three-dimensional convolution network," *J. Appl. Remote Sens.*, vol. 12, no. 1, 2018, Art. no. 016005.
[26] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
[27] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral-spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, pp. 1068–1086, 2018.
[28] A. Li and Z. Shang, "A new spectral-spatial pseudo-3D dense network for hyperspectral image classification," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2019, pp. 1–7.
[29] Z. Li *et al.*, "Deep multilayer fusion dense network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1258–1270, Mar. 2020, doi: 10.1109/JSTARS.2020.2982614.
[30] Y. Wang, B. Liang, M. Ding, and J. Li, "Dual-branch dense residual network for hyperspectral imagery classification," *Int. J. Remote Sens.*, vol. 41, no. 7, pp. 2581–2602, Nov. 2019.
[31] R. Hang, F. Zhou, Q. Liu, and P. Ghamisi, "Classification of hyperspectral images via multitask generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1424–1436, Feb. 2021.
[32] H. Gao, Y. Miao, X. Cao, and C. Li, "Densely connected multiscale attention network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2563–2576, Feb. 2021.
[33] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "Feedback attention-based dense CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2021, Art. no. 5501916, doi: 10.1109/TGRS.2021.3058549.
[34] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, "Hyperspectral image classification with attention-aided CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2281–2293, Mar. 2021.
[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Nov. 2017, pp. 2261–2269.
[36] Y. Xu *et al.*, "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS Data Fusion Contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1709–1724, Jun. 2019.

Xin Wang received the B.Eng. degree in surveying and mapping engineering from the Xi'an University of Science and Technology, Xi'an, China, in 2019. She is currently working toward the M.Sc. degree in surveying and mapping with the China University of Petroleum (East China), Qingdao, China.

Her research interests include hyperspectral remote sensing, depth learning, and image classification.

Yanguo Fan received the B.S. and M.S. degrees in engineering survey from Wuhan University of Surveying and Mapping, Wuhan, China, in 1992 and 1998, respectively, and the Ph.D. degree in cartography and geographic information engineering from China University of Mining and Technology, Beijing, China, in 2007.

He is currently a Professor with the China University of Petroleum (East China), Qingdao, China, undertaking and completing more than 30 provincial, ministerial, and bureau level topics, and more than 40 papers have been published in Chinese core journals and international conferences. His research interests include hyperspectral remote sensing, intelligent remote sensing image processing, and the application of remote sensing in geoscience.