# Agricultural Hyperspectral Image Classification Based on Deep Separable Convolutional Neural Networks

Yangyang Liang[1], Yu Wu[2], Gengke Wang[2,3(✉)], and Lili Zhang[2]

[1] Henan University, Kaifeng 475004, China
[2] Chinese Academy of Sciences, Beijing 100094, China
wanggk@aircas.ac.cn
[3] University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract.** Due to the high computational complexity of traditional convolutional neural networks, the execution time is long and the computational cost is too high. In this paper, we propose a deep separable convolutional neural network with attention mechanism added to improve the classification accuracy and generalization ability of hyperspectral images. The network uses separable convolution combined with residual connections to construct residual units with fewer parameters and adds an attention mechanism layer at the end of the network, which helps to improve the overall performance of the model. So this model has stronger generalization ability now, shorter computation time, and stronger network performance. Finally, the overall accuracy of the model in this paper is 98.48%, 99.1% and 97.40% on the Salinas dataset and the more newly proposed Wuhan Longkou and Wuhan Hanchuan datasets, respectively. It proves that the model has better generalization ability and can complete the calculation in a shorter time. Improving the classification accuracy of hyperspectral images like the Wuhan Longkou dataset is important for agricultural development.

**Keywords:** Residual Network · Separable convolution · Convolutional neural network · Attention mechanism · Agricultural hyperspectrum

## 1 Introduction

Hyperspectral images are generally images with hundreds or even more spectral bands composed. Hyperspectral images are not only rich in spectral information, but also have more spatial information, and only reasonable and sufficient use of these two parts can maximize the classification accuracy of hyperspectral images. At the same time, the computational cost should be taken into account. In conclusion, the classification of hyperspectral images is beneficial to many industries such as environment, agriculture, and atmosphere.

From the development of convolutional neural network (CNN), since AlexNet [1] won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in


D. Jiang and H. Song (Eds.): SIMUtools 2021, LNICST 424, pp. 403–420, 2022.
https://doi.org/10.1007/978-3-030-97124-3_30

2012, the convolutional neural network has entered a rapid development stage. Firstly, the AlexNet network model deepens the previous neural network depth and integrates multiple transformation layers, which improves the classification accuracy and introduces regularization in CNNs again. This improvement directly reduced the ginger error rate from 25.8% to 16.4% compared to traditional ML techniques. Since then, CNNs have become more and more widely used in the field of computer vision (CV), and people have slowly started to try to improve the performance of CNNs by reducing the computational cost while maintaining the computational accuracy. Therefore, each new convolutional neural network will try to overcome the shortcomings of the previous networks. In 2013 and 2014, most of the efforts of experts working in this field were spent on the optimization of parameters, hoping to accelerate the computational speed of CNNs with increasing computational complexity. In 2013, Zeiler and Fergus [2] created a mechanism that is a filter that can be visualized for each layer of convolution. The method was designed to improve feature extraction by reducing the size of the convolutional kernel. Subsequently, a related group at Oxford University proposed the famous VGG network [3] in 2014, which was the runner-up in the ILSVRC competition that year. VGG reduces the perceptual field and expands the volume compared to the classical AlexNet, where the feature map after each layer of convolution in VGG gradually increases the number of channels. In the same year, GoogleNet [4], which won the ILSVRC competition, not only reduces the computational cost by improving the structure of the network, but also widens the channels of the feature map according to the depth of the network structure. The real major performance improvement of CNNs in the neural network domain is the residual network (ResNet) [5], both the very famous residual links or jump links, proposed by Kaiming He et al. in 2015. This design made it possible to design deep networks that greatly improved the classification accuracy of hyperspectral images. Zhang Lei et al. proposed a privacy protection scheme for data considering the security protection aspect of the computational process, again based on the verification tree as well as the signature mechanism to make security protection [6, 7].

In 2016 researchers in this field have explored mainly the width of the network with the hope of improving feature learning [8]. In addition there were no more prominent architectures proposed, almost always using a mixture of already emerged network structures and adding some newly researched mechanisms used to improve the overall network performance. This event gives the impression that for improving the performance of CNNs, in addition to proposing new network models, grid cells can also be properly assembled, and this practice can also be an important factor for improving the network performance. Therefore, Hu et al. in 2017 identified the role played by the grid representation in the whole training process of convolutional neural networks. They also introduced the idea of feature graph development, while pointing out that a small amount of information and domain-independent features may affect the performance of the network to a greater extent. Using this idea, a new network architecture called “Squeeze and Excite Network (SE-Network) [9]” was proposed. This network is designed to exploit the spectral information by designing a dedicated SE module that assigns weights to each feature map according to its role in class recognition. This idea has been further investigated by many researchers in the field who have shifted attention

to important regions by exploiting spatial and feature map (channel) information [10]. The model in this paper also incorporates the attention mechanism. In 2018, Khan et al. introduced a new idea of channel boosting [11]. The performance of CNN can be effectively improved by learning various features, and by using the already learned features through TL concept.

In fact, since 2012, many improvements on CNN network models have appeared one after another. Regarding the advancement of convolutional neural networks, the research in recent years has focused on designing new residual modules, performing deep convolutional operations, enhancing the network performance by means of feature maps and adding artificial channels. Zhang L, Huang Z and other researchers used the study of hyperspectral images for weather radar echo prediction [12, 13] and achieved better results.

Regarding the research on hyperspectral image classification, in the early days of hyperspectral feature extraction (HSIFE), the focus of classification was on the extraction of spectra, recognition of objects by spectra, and other spectral-based methods. The main methods include principal component analysis (PCA), independent component analysis (ICA), linear discriminant analysis, etc. [5, 14]. These methods mainly apply linear transformations to extract the features of the input data, but in the natural world natural objects and complex light scattering mechanisms, and hyperspectral data are inherently nonlinear [15], which makes these linear-based transformations not very suitable for analyzing hyperspectral data.

With the development of imaging technology, hyperspectral sensors are again hungry for higher spatial resolution, and the spatial information we obtain is becoming more and more detailed in hyperspectral data. In [16], a method that combines the use of morphological operators and support vector machines (SVM) was introduced, which unfortunately significantly improves the classification efficiency. However, traditional image classification methods including support vector machine support vector machine (SVM) [17], 3D wavelet transform [18], Gaussian mixing, etc. all use band selection and feature extraction methods, which can reduce the dimensionality of hyperspectral images, however, it destroys the overall think of the data, and the related deficiencies can lead to unsatisfactory classification accuracy of images. Wei Huang et al. contributed to image preprocessing using dense networks for reconstruction of hyperspectral compressed images [19], and then went on to classify hyperspectral images using local binary patterns and superpixel multicores [20].

In recent years, it is obvious that convolutional neural networks (CNNs) have an excellent performance in the field of image classification, and researchers related to the field of hyperspectral image classification have also used CNNs for hyperspectral image classification. hu et al. first used convolutional neural networks for hyperspectral image classification in [21]. In the paper he and his team used principal component analysis (PCA) to reduce the image dimensionality, followed by using two-dimensional convolution to extract the spatial features of the input data, and then one-dimensional convolution to extract the spectral features of each pixel, and then combining the results of both convolutions to obtain higher accuracy classification results [22]. It is not difficult to find that the method can destroy the continuity of the spectrum. With the development of convolutional neural networks, people began to extract the spatial spectral information

of HSI more fully, and due to the small data set, in order to achieve more accurate classification in a limited sample, related researchers proposed more godly and lighter complex network models, among which Lee et al. proposed an Inception-based deep CNN model (DC-CNN) [23], after Zhong et al. proposed a hyperspectral image classification by 3D convolution operation and proposed a spatial-spectral residual network (SSRN) [24], which can extract spatial and spectral features more completely. Later Wang et al. used $1 \times 1$ and $3 \times 3$ convolution kernels to extract spatial-spectral features of hyperspectral images by density linking, which can also be classified effectively [25]. Paoletti et al. [26] proposed a newer pyramidal residual network, mainly by stacking pyramidal bottleneck residual units [27] to construct a residual network (pResNet), which obtained a high classification accuracy.

However, it is important for agricultural hyperspectral classification, which can detect the nutrient and water content of each crop at any time, monitor the rise of crops, and further estimate the yield, so as to make corresponding countermeasures in advance and reduce unnecessary losses.

## 2 Analysis of the Model

### 2.1 Attention Mechanism

The whole attention mechanism acts to integrate global information (including the a priori information already obtained earlier) to extract important and useful information for the present features to adjust the corresponding weights. In this paper, we combine spatial attention and channel-wise attentiveness in a multilayer feature application. attentiveness is essentially the training of a weight that can then be used to select a channel or superimposed on each pixel of a feature map [28].

Spatial attention: each pixel of the current feature map is assigned a weight value for each pixel, which is a two-dimensional matrix; Channel-wise: a weight is assigned to each channel in terms of the feature map, so this weight is a vector.

The below figure is a diagram of one layer of SCA-CNN (see Fig. 1.), which takes the classical encoder-decoder structure and mainly contains two parts: CNN network (encoder) and LSTM network (decoder). The spatial attention can be understood as a unit of each pixel of the feature map, and each pixel of the feature map is assigned a weight value, so this weight value should be a matrix; the channel wise attention is a unit of the feature map, and each channel is assigned a weight value. The channel wise attention is assigned to each channel as a unit of feature map, so the weight value should be a vector.

Principles of attentional mechanisms:

Where $x \in R^{C \times N}$, A $1 \times 1$ convolution of x (the input to the previous layer) yields f, g, h. This changes the number of channels from C to C*.

$$f(x) = W_f x, g(x) = W_g x, h(x) = W_h x \tag{1}$$

These weights are called "attention maps" and essentially quantify the "importance" of pixel j relative to pixel i in the image. Since these weights ($\beta$) are computed over the

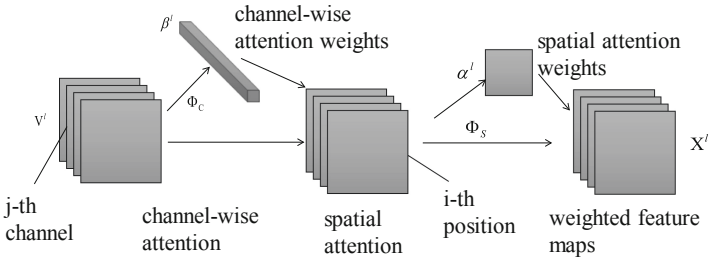


**Fig. 1.** A diagram of one layer of SCA-CNN.

entire height and width of the feature set, the receptive field is no longer limited to the size of the small kernel.

$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})}, s_{ij} = f(x_i)^T g(x_j) \tag{2}$$

The output of the self-attentive layer is calculated as:

$$o_j = v\left(\sum_{i=1}^{N} \beta_{j,i}(x_i)\right), v(x) = W_v x, W_v \in R^{C \times C^*} \tag{3}$$

Usually set $C^* = \frac{C}{8}$, As a final step, the input feature x is added to the output weighting ($\gamma$ is another scalar parameter that can be learned):

$$y_i = \gamma o_i + x_i \tag{4}$$

### 2.2 Deep Separable Convolution

There are two main types of separable convolution: depth separable convolution and spatial separable convolution.

Deeply separable convolution means that the input of $N \times H \times W \times C$ is divided into C groups, Then each group does the corresponding convolution operation so that the spatial features of each Channel are collected, Depth-wise features (see Fig. 2).

Spatially separable convolution means making k ordinary $1 \times 1 \times C$ convolutions of the input of $N \times H \times W \times C$, which is equivalent to collecting the features of each point features, Pointwise features. Usually the W and H of the convolution kernel are 1 (see Fig. 3).

The advantage of separable convolution is mainly that it reduces the number of parameters and thus the number of computations. The computational burden of the model is alleviated.

The number of parameters for the regular convolution operation is: $H \times W \times C \times k$;

The number of parameters of the separable convolution operation is: $H \times W \times C + 1 \times 1 \times C \times k$.

Deeply separable convolution changes the previous ordinary convolution operation to consider both channels and regions (convolution first considers only regions and then channels), and achieves the separation of channels and regions.

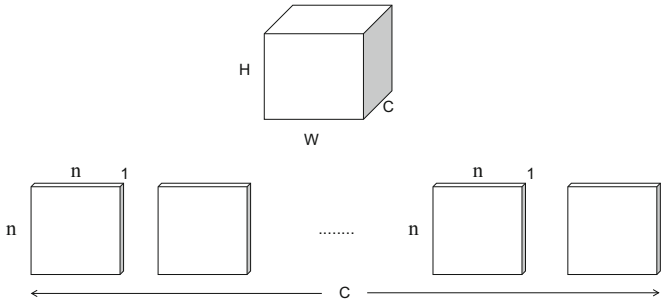


**Fig. 2.** Deeply separable convolution.

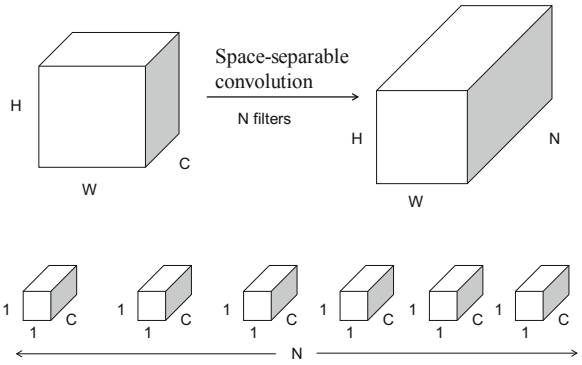


**Fig. 3.** Spatially separable convolution.

By separating the regions and channels, it is equivalent to compressing the computational effort of ordinary convolution is:

$$\frac{\text{depthwise} + \text{pointwise}}{conv} = \frac{H \times W \times C \times 3 \times 3 + H \times W \times C \times k}{H \times W \times C \times 3 \times 3} = \frac{1}{k} + \frac{1}{3 \times 3} \tag{5}$$

Depthwise: number of parameters for depth-separable convolution, pointwise: number of parameters for spatially separable convolution, conv: number of parameters for traditional convolution methods, k: number of convolution kernels.

### 2.3 Residual Unit

The residual unit in this paper consists of four BN (Batch Normalization) layers, three separable convolutions, and two ReLU activation layers (see Fig. 4).

The input hyperspectral image of $9 \times 9 \times 38$ is normalized and then subjected to separable convolution, which, as introduced in the previous section, reduces the number of parameters and thus the computation time. The separable convolution can effectively extract the spectral spatial information from the hyperspectral image, and after several such residual modules, the spectral spatial information of each region of the hyperspectral

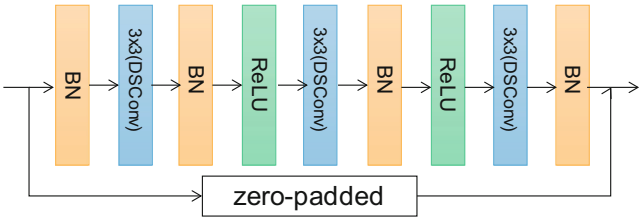


**Fig. 4.** Residual Unit is proposed.

image can be fully obtained. The final output after several residual modules is a matrix of shape $5 \times 5 \times 86$. This process uses a pyramidal type of residual units.

The pyramidal residual cell is a simple structure that increases the number of channels of the same feature map with each passing residual cell. This design approach can reduce the number of parameters and computational cost of the network model more significantly. Unlike the traditional pyramidal residual unit, the final ReLU layer is not used in this paper [29], and it is noteworthy that the data need to be normalized first when just entering the interior of the residual unit, i.e., first passing through the BN layer [30], and when passing through all residual units completely, the output will present a long strip-like feature map with smaller length and width and a larger number of channels. The most important thing is to replace the conventional convolution layer for each residual cell with a separable convolution. This can reduce the model parameters. In short, the core part of the network model proposed in this paper is composed of such units. The parameters do not increase significantly as the network is progressively deepened, resulting in the construction of a lightweight residual classification model.

### 2.4 HSI Classification Models

A depth-separable convolutional neural network model is built as shown below. The model first performs $1 \times 1$ convolutional dimensionality reduction on the HSI data cube after data pre-processing to extract the corresponding spectral information. Then, the output of the convolutional dimensionality reduction is put into the residual module as input to continuously extract the spatial contextual features and spectral features of the data cube, as shown in the figure below, each residual unit is composed of a BN layer, a ReLU layer and a $1 \times 1$ separable convolutional layer, and then, it enters a $1 \times 1$ convolutional layer with a global average pooling (GAP) layer, and finally, it passes through the attention mechanism module, which does not change the data structure, but modifies the weights of the corresponding pixel points to prepare for the next data processing, which can converge faster and complete the final classification.

In addition, taking Wuhan Hanchuan data set as an example, this paper proposes a new network model, as shown in the figure below (see Fig. 5).

First, the processed 3D hyperspectral data with the shape of $9 \times 9 \times 274$ (274 is the number of data channels) is input into this network model. In C1, the channel information of the input data is reorganized by 38 $1 \times 1$convolutional kernels, at which point the processed shape is $9 \times 9 \times 38$.

Than R1 block consisting of $3 \times 3$, stride $= 1$ convolutional kernels is used to extract the corresponding spatial features. After the R1block, the first layer of R2 is a $3 \times 3$ filter, stride $= 2$ for downsampling operation, and the second layer of kernel uses a step size of 1 to generate a $3 \times 3 \times 86$ feature cube with a smaller spatial size.

The final convolutional layer C2 of the model contains 16 $3 \times 3$ kernels for compressing the discriminative feature map, and the generated $5 \times 5 \times 16$ feature map is passed to the GAP layer and then to the Attention layer, where some of the weights are strengthened by back propagation of the model, and then the shape of the space is transformed into a $1 \times 16$ one-dimensional vector. The more detailed model flow are shown in Table 1.

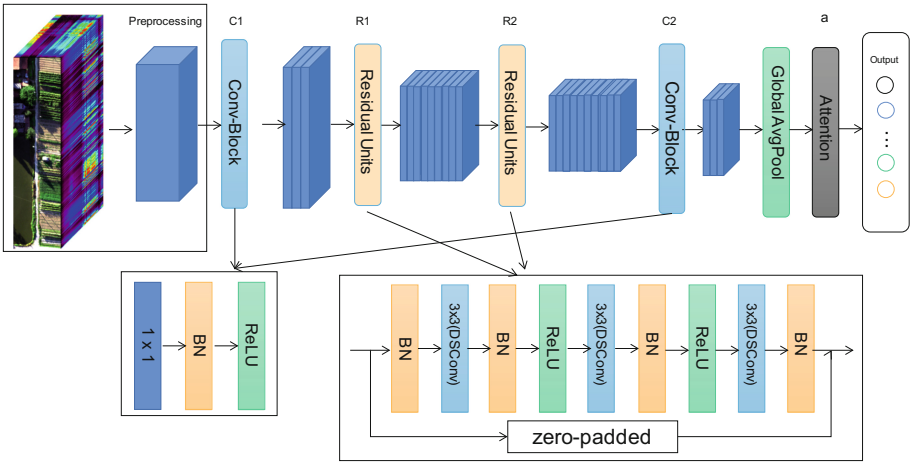


**Fig. 5.** Neural network model.

## 3 Detailed Design

The HSI has a continuum property and the data are relatively scattered in each waveform. To speed up the convergence and reduce the training time of the model, the input data cubes are first zero-averaged before being fed into the network. The standardized computational equation is defined as:

$$X^{a}_{\mathrm{m,n}} = \frac{X^{a}_{m,n} - \overline{X}^{a}}{\sigma^{a}} (1 \le m \le W, 1 \le n \le H, 1 \le a \le N) \tag{6}$$

$X^{a}_{\mathrm{m,n}}$ denotes the pixel value of the mth row and nth column in the ath band of the HSI, $\overline{X}$ is the mean value of all pixels in the ath band, and σ is the standard deviation of pixels in the ath band; W, H, and N denote the width, height, and total number of channels of the input HSI, respectively.

Considering the numerous spectral bands of HSI data, the Hughes phenomenon is easily generated [31]. That is, with hundreds of spectral bands and a small amount of

**Table 1.** Model flow details.

| Layer | Output size | Kernel size | Stride | Padding |
|---|---|---|---|---|
| Input | $9 \times 9 \times 274$ | | | |
| C1 | $9 \times 9 \times 38$ | $1 \times 1$ | 1 | 0 |
| R1 | $9 \times 9 \times 62$ | $3 \times 3$ | 1 | 1 |
| | $9 \times 9 \times 62$ | $3 \times 3$ | 1 | 1 |
| | $9 \times 9 \times 62$ | $3 \times 3$ | 1 | 1 |
| R2 | $5 \times 5 \times 86$ | $3 \times 3$ | 2 | 1 |
| | $5 \times 5 \times 86$ | $3 \times 3$ | 1 | 1 |
| | $5 \times 5 \times 86$ | $3 \times 3$ | 1 | 1 |
| C2 | $5 \times 5 \times 16$ | $1 \times 1$ | 1 | 1 |
| GAP | $1 \times 1 \times 16$ | | | |
| a | $1 \times 1 \times 16$ | | | |

data, this is prone to overfitting, so we use $1 \times 1$ convolution in the first layer of the model, which is used to reduce the number of channels of the hyperspectral data and does not change the spatial size. This approach not only ensures the spatial integrity of the data, but also effectively utilizes the multispectral information of the data, which in turn avoids the occurrence of overfitting phenomena. As shown in the above network structure, two residual network modules (R1, R2) can fully extract the spectral spatial information of the processed HSI data. Both modules use a $3 \times 3$ convolutional kernel. The value of stride in R1 is 1 and the value of padding is also 1 to ensure that the input and output data have the same size and the edge features of the data can be fully preserved. In R2, the value of stride is 2 and the value of padding is 1. This is designed to further reduce the size of the feature map and facilitate the final one-dimension visualization of the feature map, and then the second layer of convolution uses a design with both stride and padding of 1, which is designed to ensure that the input and output sizes are the same. Both residual blocks are connected with a jump with zero padding [32], so that the features of the data can be more fully utilized.

For the traditional convolution operation, to ensure that the output and input feature maps are of the same size, the Padding method is used to expand the edges of the feature maps, which undoubtedly increases the parameters and increases the computational cost of the computer. And it also increases the number of channels of the output feature map, while the pyramid residual structure used in this paper is an ordered small step to increase the number of channels of the output feature map, and this method reduces the number of parameters. The equations for the output channels of each residual module are as follows:

$$D_i = \mathrm{D_i} = \begin{cases} C; \mathrm{i} = 1 \\ \mathrm{D_{i-1}} + \frac{\alpha}{\mathrm{R}}; \mathrm{i} > 1 \end{cases} \tag{7}$$

Where $D_i$ is the number of output channels of the ith residual cell and C the initial number of channels of the first residual cell. In this paper that is the number of output channels after the C1 layer, R is the number of residual cells, and α is an integer greater than zero.

After the residual block, the data enters the 1 × 1 convolution and global average pooling layer, which uses fewer parameters than the traditional fully connected layer and provides mitigation of the overfitting problem and accelerates the convergence of the network [33]. Moreover, the attention mechanism is added at the end of the model, and the role of adding the attention mechanism layer is to adjust the relevant weights of the pixel points in the data as soon as possible to highlight the characteristics of the crops we are concerned about, such as variety, growth, and pests, which is more likely to meet the practical needs of our related work, and also plays the same role of accelerating the convergence of the network model.

## 4 Experimental Procedure

In this paper, we use publicly available datasets from Wuhan University: the Wuhan Longkou dataset and the Wuhan Hanchuan dataset [27, 34, 35], in addition to the international publicly available dataset Salinas. The following Table 2 is a basic introduction to these three types of datasets.

**Table 2.** Three kinds of data sets are used in this paper.

| | Salinas | WHHC | WHLK |
|---|---|---|---|
| Type of Sensor | AVIRIS | Aibot X6 | DJI matrix 600 Pro (DJI M600 Pro) |
| Spatial Size | 512 × 217 | 1217 × 303 | 550 × 400 |
| Spectral Range | 400–2500 nm | 400–1000 nm | 400–1000 nm |
| Spatial Resolution | 3.7 m | 0.109 m | 0.463 m |
| Bands | 204 | 274 | 270 |
| Num. Of Classes | 16 | 16 | 9 |

The experimental design of this paper is roughly as follows. Four experiments are set up in this paper, which are the model in this paper, the model in this paper without adding the attention mechanism, changing the separable convolution of the model in this paper to the ordinary two-dimensional convolution, and the last one changing the residual module in this paper to the classical residual module structure. The overall experimental results are measured by three parameters, which are overall classification accuracy (OA), average classification accuracy (AA) and Kappa coefficient (K). To ensure the accuracy of the experimental results, each experiment is done five times, and the final average is taken as the final experimental result.

## 5 Experimental Results

The results of the neural network model training and testing experiments in this paper is:

The below figure shows, from left to right, the Salinas dataset, the Wuhan Hanchuan dataset and the Wuhan Longkou dataset (see Fig. 6.), and the experimental results in the model experiments of this paper are shown in the following Table 3:

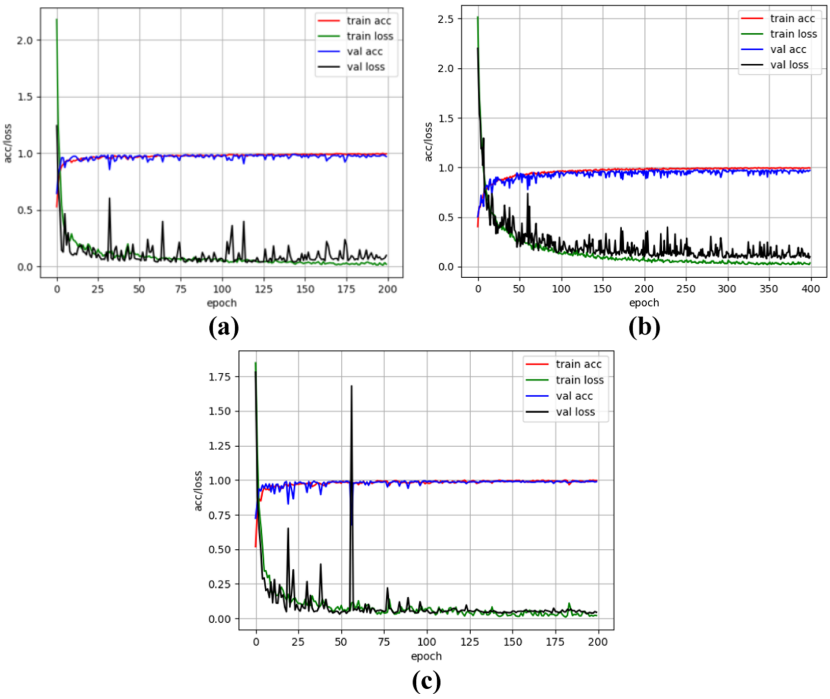


**Fig. 6.** Experimental results of the proposed model on Salinas dataset (a), Wuhan Hanchuan dataset (b), Wuhan Longkou dataset (c).

The model in this paper obviously has a better classification effect for Wuhan Longkou.

**Table 3.** Experimental results of the proposed model on three data sets are presented.

| | Salinas | WHHC | WHLK |
|---|---|---|---|
| Train time (ms) | 332.5548 | 452.1498 | 228.7295 |
| Test time (ms) | 12.5837 | 62.8690 | 66.1033 |
| Test loss (%) | 0.109 | 0.277 | 0.0307 |
| Test Accuracy (%) | 96.33 | 91.32 | 99.06 |
| AA (%) | 98.481220 | 97.399308 | 99.127253 |
| OA (%) | 96.332149 | 97.371132 | 99.067781 |
| Kappa (%) | 95.893613 | 96.921145 | 98.774291 |

### 5.1 Comparative Test

Without adding the attention mechanism:

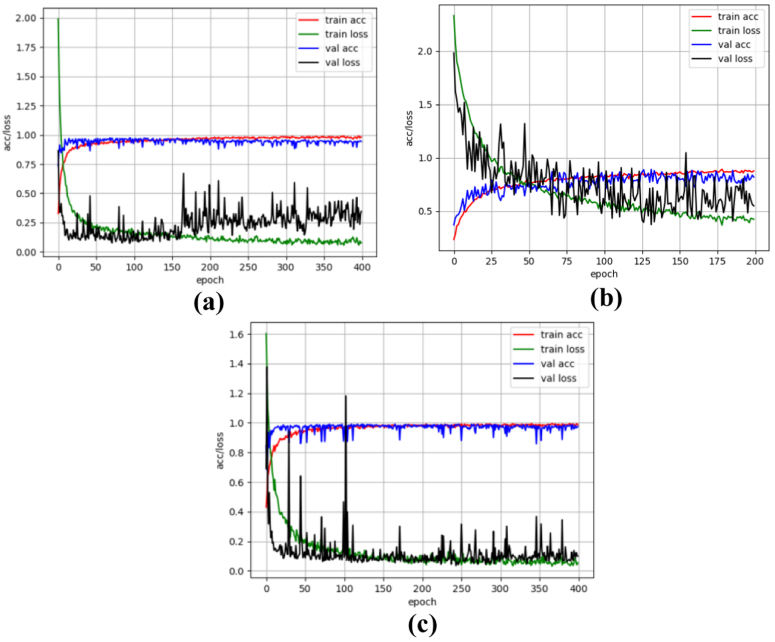


**Fig. 7.** Experimental results of the proposed model and delete the attention mechanism on Salinas dataset (a), Wuhan Hanchuan dataset (b), Wuhan Longkou dataset (c).

The above figure shows, from left to right, the Salinas dataset (a), the Wuhan Hanchuan dataset (b) and the Wuhan Longkou dataset (c) (see Fig. 7.), and their experimental results in the experiments without the inclusion of the attention mechanism are shown in the following Table 4:

**Table 4.** Experimental results of the proposed model and delete the attention mechanism on three datasets are presented.

| | Salinas | WHHC | WHLK |
|---|---|---|---|
| Train time (ms) | 2816.7585 | 1434.8909 | 862.5844 |
| Test time (ms) | 223.7624 | 159.9729 | 198.8493 |
| Test loss (%) | 0.206 | 0.277 | 0.061 |
| Test Accuracy (%) | 93.78 | 91.32 | 98.17 |
| AA (%) | 97.658851 | 88.729316 | 98.566262 |
| OA (%) | 93.787430 | 91.327016 | 98.179953 |
| Kappa (%) | 93.064872 | 89.846760 | 97.614027 |

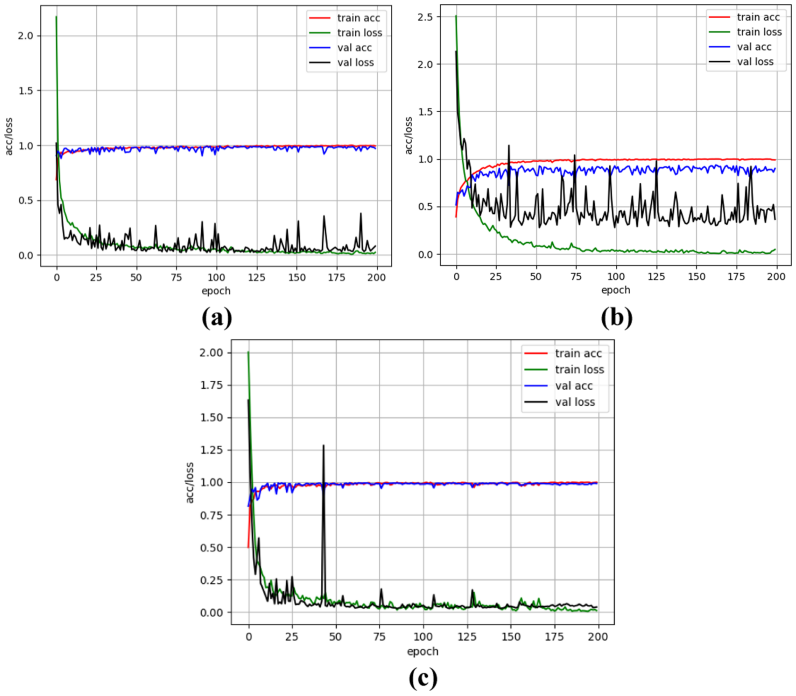


**Fig. 8.** Experimental results of the proposed model but use normal two-dimensional convolution on Salinas dataset (a), Wuhan Hanchuan dataset (b), Wuhan Longkou dataset (c)

Instead of separable convolution, a normal two-dimensional convolution is used:

The above figure shows, from left to right, the Salinas dataset (a), the Wuhan Hanchuan dataset (b) and the Wuhan Longkou dataset (c) (see Fig. 8.), and their experimental results in the experiments without the use of separable convolution are shown in the following Table 5:

**Table 5.** Experimental results of the proposed model but use normal two-dimensional convolution on three datasets are presented.

| | Salinas | WHHC | WHLK |
|---|---|---|---|
| Train time (ms) | 423.6212 | 471.5599 | 271.3779 |
| Test time (ms) | 10.2761 | 60.9518 | 52.6547 |
| Test loss (%) | 0.127 | 0.249 | 0.037 |
| Test Accuracy (%) | 95.47 | 93.34 | 99.15 |
| AA (%) | 98.253737 | 92.452533 | 98.961172 |
| OA (%) | 95.472128 | 93.348799 | 99.086524 |
| Kappa (%) | 94.934475 | 92.220816 | 98.798554 |

Experimental results using separable convolution, residual module using two separable convolution layers:

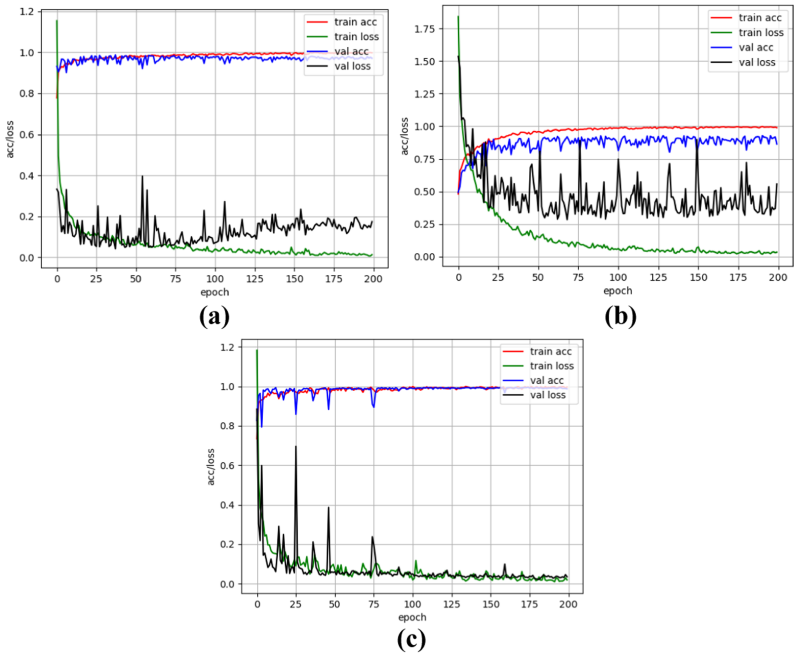


**Fig. 9.** The experimental result graph of each residual element of the proposed model using two separable convolution layers is presented on Salinas dataset (a), Wuhan Hanchuan dataset (b), Wuhan Longkou dataset (c)

The above figure shows, from left to right, the Salinas dataset (a), the Wuhan Hanchuan dataset (b) and the Wuhan Longkou dataset (c) (see Fig. 9.), and their experimental results are shown in the following Table 6:

**Table 6.** The experimental result graph of each residual element of the proposed model using two separable convolution layers is presented on three data sets are presented.

| | Salinas | WHHC | WHLK |
|---|---|---|---|
| Train time (ms) | 225.1533 | 255.6815 | 150.1188 |
| Test time (ms) | 9.2854 | 61.8522 | 39.6964 |
| Test loss (%) | 0.129 | 0.375 | 0.034 |
| Test Accuracy (%) | 96.11 | 91.62 | 98.95 |
| AA (%) | 98.485236 | 93.688709 | 99.192328 |
| OA (%) | 96.108308 | 93.916565 | 99.051504 |
| Kappa (%) | 95.645520 | 92.886840 | 98.753314 |

Through the above experimental comparisons, it is easy to conclude that the proposed deep separable convolutional neural network with added attention mechanism in this paper reduces the model parameters and shortens the training time while ensuring high classification accuracy for hyperspectral datasets.

## 6 Conclusion

In this paper, a new deeply separable convolutional neural network model for HSI classification is proposed, and an attention mechanism is added to the model, which is experimentally proven to have the advantages of high accuracy and fast convergence.

This lightweight model uses a $1 \times 1$ convolution kernel from the first layer to reorganize the channels of the hyperspectral data to reduce the number of spectral channels for further processing later. Then it enters the residual module to fully extract the spatial spectral features of the hyperspectral data through the residual unit. This is followed by a layer of $1 \times 1$ filters and a global average pooling to classify the data. The attention mechanism is added in the last layer to adjust the weights of the pixel points in the data without changing the execution results of the previous layer, and the weights of the pixel points where we are more concerned about the crop-related information are adjusted upward, which can effectively enhance our attention to the crop and also accelerate the convergence of the data.

The separable convolution used in the residual block of the model not only ensures the accuracy of the classification, but also further reduces the number of parameters of this model, decreases the load on the machine, and helps the model to classify more quickly. During the experimental comparison, we clearly see that the model used in this paper has the highest accuracy and the shortest operation time at the same time. This indicates that the model has a strong generalization ability and is fully capable of performing the classification task of HSI data excellently.

Therefore, in summary, the model proposed in this paper has high accuracy and fast speed, and is highly feasible for the task of classifying HSI data again. In the future research, I will devote myself to the classification of hyperspectral images, using 3D

convolution, or adding density connection and other methods, and continue to work deeply in this field. I will strive to build a better network model.

**Funding.** This research was funded by Civil space technology advance research project (D040401), Highly differentiated Earth surface system science research (E1K503010M), Key Technologies for Collaborative Processing and Joint Verification of Quantitative Remote Sensing Basic Common Products for "One Belt and One Road (E0BD030404), Evaluation of Space Integration Satellite Application Technology (Y7k00100kJ).

## References

1. Alom, M.Z., Taha, T.M., Yakopcic, C., et al.: The history began from AlexNet: a comprehensive survey on deep learning approaches, arXiv preprint arXiv:1803.01164 (2018)
2. Zeiler, M., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
3. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint, pp. 1409–1556 (2014)
4. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
5. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
6. Zhang, L., Huo, Y., Ge, Q., et al.: A privacy protection scheme for IoT big data based on time and frequency limitation. Wirel. Commun. Mobile Comput. Article ID 5545648, 10 p (2021)
7. Han, D., Chen, J., Zhang, L., et al.: A deletable and modifiable blockchain scheme based on record verification trees and the multisignature mechanism. CMES-Comput. Model. Eng. Sci. **128**(1), 223–245 (2021)
8. Cheng, H.T., Koc, L., Harmsen, J., et al.: Wide & deep learning for recommender systems. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. ACM (2016)
9. Jie, H., Li, S., Gang, S., et al.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 99 (2017)
10. Wang, X., Girshick, R., Gupta, A., et al.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803 (2018)
11. Khan, N., Afaq, F., Saleem, M., et al.: Targeting multiple signaling pathways by green tea polyphenol epigallocatechin-3-gallate. Can. Res. **66**(5), 2500–2505 (2006)
12. Zhang, L., Huang, Z., Liu, W., et al.: Weather radar echo prediction method based on convolution neural network and long short-term memory networks for sustainable e-agriculture. J. Clean. Prod. **298**, 126776 (2021)
13. Zhang, L., Xu, C., Gao, Y., et al.: Improved Dota2 lineup recommendation model based on a bidirectional LSTM. Tsinghua Sci. Technol. **25**(6), 712–720 (2020)
14. Licciardi, G., Marpu, P.R., Chanussot, J., et al.: Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. IEEE Geosci. Remote Sens. Lett. **9**(3), 447–451 (2012)
15. Mou, L., Bruzzone, L., Zhu, X.X.: Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. IEEE Trans. Geosci. Remote Sens. **57**, 1–12 (2018)

16. Chen, P.H., Lin, C.J., Schlkopf, B.: A tutorial on ν-support vector machines. Appl. Stoch. Model. Bus. Ind. **21**(2), 111–136 (2005)
17. Luts, J., Ojeda, F., Plas, R., et al.: A tutorial on support vector machine-based methods for classification problems in chemometrics. Anal. Chim. Acta **665**(2), 129–145 (2010)
18. Zhu, Z., Jia, S., He, S., et al.: Three-dimensional gabor feature extraction for hyperspectral imagery classification using a memetic framework. Inf. Sci. **298**, 274–287 (2015)
19. Huang, W., Xu, Y., Hu, X., et al.: Compressive hyperspectral image reconstruction based on spatial-spectral residual dense network. IEEE Geosci. Remote Sens. Lett. **17**(5), 884–888 (2020)
20. Wei, H., Yao, H., Hua, W., et al.: Local binary patterns and superpixel-based multiple kernels for hyperspectral image classification. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. **13**, 4550–4563 (2020)
21. Hu, W., Huang, Y., Wei, L., et al.: Deep convolutional neural networks for hyperspectral image classification. J. Sens. **2015** (2015)
22. Zhang, H., Li, Y., Zhang, Y., et al.: Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. Remote Sens. Lett. **8**(4–6), 438–447 (2017)
23. He, M., Li, B., Chen, H.: Multi-scale 3D deep convolutional neural network for hyperspectral image classification. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3904–3908. IEEE (2017)
24. Zhong, Z., Li, J., Luo, Z., et al.: Spectral–spatial residual network for hyperspectral image classification: a 3-D deep learning framework. IEEE Trans. Geosci. Remote Sens. **56**(2), 847–858 (2017)
25. Wen, J.W., Shu, G.D., Zhong, M.J., et al.: A fast dense spectral-spatial convolution network framework for hyperspectral images classification. Remote Sens. **10**(7), 1068 (2018)
26. Gao, H., Yang, Y., Li, C., Zhang, X., Zhao, J., Yao, D.: Convolutional neural network for spectral–spatial classification of hyperspectral images. Neural Comput. Appl. **31**(12), 8997–9012 (2019). https://doi.org/10.1007/s00521-019-04371-x
27. Zhong, Y., Hu, X., Luo, C., et al.: WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. Remote Sens. Environ. **250**, 11–20 (2020)
28. Chen, L., Zhang, H., Xiao, J., et al.: SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning. In: 2017 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5659–5667 (2017)
29. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML), Haifa, Israel, USA, 21–24 June 2010, pp. 807–814. Omnipress, Madison (2010)
30. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning, Paris, France, 6–11 July 2015, pp. 448–456 (2015)
31. Donoho, D.L.: High-dimensional data analysis: the curses and blessings of dimensionality. AMS Math Challenges Lect. 1–32 (2000)
32. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
33. Lin, M., Chen, Q., Yan, S.: Network in network. Comput. Sci. (2013)

34. Zhong, Y., Wang, X., Xu, Y., et al.: Mini-UAV-borne hyperspectral remote sensing: from observation and processing to applications. IEEE Geosci. Remote Sens. Mag **6**(4), 46–62 (2018)
35. Lv, L., Zheng, C., Zhang, L., et al.: Contract and Lyapunov optimization-based load scheduling and energy management for UAV charging stations. IEEE Trans. Green Commun. Network. **5**(3), 1381–1394 (2021)