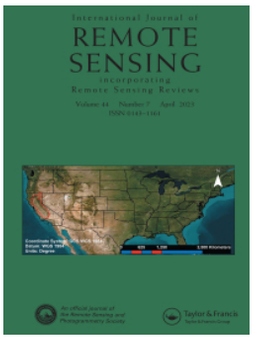


International Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

GDPA_LDSICS: graph and double pyramid attention network based on linear discrimination of spectral interclass slices for hyperspectral image classification

Haiyang Wu, Cuiping Shi & Liguo Wang

To cite this article: Haiyang Wu, Cuiping Shi & Liguo Wang (2023) GDPA_LDSICS: graph and double pyramid attention network based on linear discrimination of spectral interclass slices for hyperspectral image classification, International Journal of Remote Sensing, 44:17, 5283-5312, DOI: 10.1080/01431161.2023.2247523

To link to this article: https://doi.org/10.1080/01431161.2023.2247523

Published online: 01 Sep 2023.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2023, VOL. 44, NO. 17, 5283–5312
https://doi.org/10.1080/01431161.2023.2247523

Taylor & Francis
Taylor & Francis Group

GDPA_LDSICS: graph and double pyramid attention network based on linear discrimination of spectral interclass slices for hyperspectral image classification

Haiyang Wu[a], Cuiping Shi[a] and Liguo Wang[b]

[a]College of Electronic and Communication Engineering, Qiqihar University, Qiqihar, China; [b]College of Information and Communication Engineering, Dalian Nationalities University, Dalian, China

ABSTRACT

In recent years, convolution neural networks (CNNs) and graph convolution networks (GCNs) have been widely used in hyperspectral image classification (HSIC). CNNs can effectively extract the spatial spectral features of hyperspectral images (HSIs), while GCNs can quickly capture the structural features of HSIs, which makes the effective combination of the two is beneficial to improve classification performance of hyperspectral images. However, the high redundancy of feature information and the problem of small sample are still the major challenges of HSIC. In order to alleviate these problems, in this paper, a new graph and double pyramid attention network based on linear discrimination of spectral interclass slices (GDPA_LDSICS) is proposed. First, a linear discrimination of spectral inter class slices (LDSICS) module is designed. The LDSICS module can effectively eliminate a lot of redundancy in spectral dimension, which is conducive to subsequent feature extraction. Then, the spatial spectral deformation (SSD) module is constructed, which can effectively correlate the spatial spectral information closely. Finally, in order to alleviate the problem of small sample, a double branch structure of CNN and GCN is developed. On the CNN branch, a double pyramid attention (DPA) structure is designed to model context semantics to avoid information loss caused by long-distance feature extraction. On the GCN branch, an adaptive dynamic encoding (ADE) method is proposed, which can more effectively capture the topological structure of spatial spectral features. Experiments on four open datasets show that the GDPA_LDSICS can provide better classification performance and generalization performance than other most advanced methods.

ARTICLE HISTORY
Received 1 March 2023
Accepted 4 August 2023

KEYWORDS
Hyperspectral image classification; convolutional neural networks; graph convolutional networks

1. Introduction

Hyperspectral image (HSI) is a kind of three-dimensional data containing rich spatial information and spectral information, and each sample corresponds to a specific object (Malthus et al. 2003). This makes HSIs widely used in practical life and production. For example, military exploration, agricultural vegetation, pollution monitoring, medical

CONTACT Cuiping Shi shicuiping@qqhru.edu.cn College of Electronic and Communication Engineering, Qiqihar University, Qiqihar 161000, China

diagnosis, etc. (Yuen and Richardson 2010). This is all thanks to an important remote sensing technology, namely hyperspectral image classification (HSIC).

In the early days, HSIC techniques were mainly based on some traditional machine learning methods (Melgani and Bruzzone 2004). Representative methods include the support vector machine (SVM) proposed by (Pal 2008) and the sparse representation classification (SRC) proposed by Chen, Nasrabadi, and Tran (2010). These methods are relatively simple to model. However, only the spectral information of the HSI is concerned, and do not associate spatial information with spectral information, thus affecting the classification effect. In recent years, some deep learning models have been applied to HSIC, and gradually become the mainstream of research. Such as deep belief networks (DBNs) (Chen, Zhao, and Jia 2015), stacked autoencoder (SAE) networks (Chen et al. 2014), convolutional neural networks (CNNs) (Chen et al. 2016), graph neural networks (GNNs) (Kang et al. 2021), etc. DBNs and SAE are capable of processing HSI high-dimensional spectral information. However, there are strict requirements for the size of the input, which makes a part of the spatial information lose and affects the final classification result. CNNs and GCNs are different from them. CNN is a special sparse representation method, and the local connection method makes it have excellent feature extraction ability. Recently, some excellent CNNs methods have been proposed, such as one-dimensional convolutional neural networks (1-D CNNs)[23] and two-dimensional convolutional neural networks (2-D CNNs). However, they only focus on the feature information of a single spatial or spectral, and do not associate spatial features with spectral features. Therefore, an automatic design of convolutional neural network (AD-CNN) (Chen et al. 2019) was proposed by Chen et al. AD-CNN combines one-dimensional convolution with three-dimensional convolution (Ying, Zhang, and Shen 2017) to explore spatial spectral union features. And the best network framework is obtained by using the neural architecture search (NAS) (Nakai, Matsubara, and Uehara 2020). However, shallow CNNs cannot greatly improve the classification accuracy. Therefore, some deep network structures have been proposed. A spectral-spatial residual network (SSRN) (Zhong et al. 2018) was proposed to connect three-dimensional convolutional kernels using a residual structure that could deepen the network depth and extract deep features without overfitting the network. On this basis, a densely connected convolutional network (DenseNet) (Paoletti et al. 2018; Yang et al. 2018; Yu et al. 2021; Zhang et al. 2021; Zhao et al. 2021) has been proposed, which can repeatedly use the same feature to expand the depth of the network. After this, a unified multi-scale learning framework (UML) (Wang et al. 2020) was proposed. UML considers multiple ranges of features and uses a multi-scale strategy for channel shuffling to improve the robust performance of the network.

With the development of CNNs, attention-based CNNs (Sun et al. 2020, 2022) have also been applied to HSIC. The attention network can highlight important feature information, directly establish the dependency relationship between input and output, and process it with the CNN to reduce the information loss of long-distance feature extraction. For example, a residual spectral – spatial attention network (RSSAN) (Zhu et al. 2021) was proposed by Zhu et al. It uses a residual structure to unify spatial attention with spectral attention. And after attention, a large number of residual structure modules are used to extract the spatial spectral features, effectively

improving the classification accuracy. After this, the multi-attention fusion network (MAFN) (Li et al. 2022) was proposed by Li et al. MAFN integrates multiple attention and realizes the complementation of attention information. In addition, in order to improve the attention ability of attention, a dual-triple attention network (DTAN) (Cui et al. 2022) was proposed by Cui et al. and achieved significant classification effects. However, when using attention to explore spatial information, classification results are disturbed by the rotation of features. Thus, a rotation-invariant attention (RIA) (Zheng et al. 2022) was proposed by Zheng. RIA has invariance to spatial rotation and facilitates HSI classification.

However, CNNs' single fixed mesh processing method limits their feature extraction capabilities. GCNs, on the other hand, can act directly on graph structures and process graph data in non-Euclidean space (Liu et al. 2021). Therefore, some GCN-based HSIC have been proposed. The classic GCNs are capable of exploring complex non-Euclidean spatial graph data. However, GCNs need to input all pixels in the graph as vertices. Obviously, using this way to process large hyperspectral data sets will make computational costs higher. To this end, a multi-scale dynamic graph convolutional network (MDGCN) (Wan et al. 2020) was proposed, which first used hyperpixel segmentation to segment the entire HSI into multiple small pieces, each containing several pixels. These small blocks are then entered as vertex features to reduce computational costs. And MDGCN adopts a multi-scale segmentation method, so that GCN can fully extract spatial information. In order to reduce the influence of redundant information on the classification effect, a spatial pooling graph convolutional network (SPGCN) (Zhang et al. 2022) was proposed, SPGCN uses spatial pooling to retain basic characteristics, while removing some redundant information, so that the classification accuracy has been significantly improved.

While there are already many excellent CNN or GCN-based methods for HSIC, there are still some shortcomings. First, HSI is a high-dimensional stereo data, which contains a lot of redundant information. Dealing with it directly not only makes the calculation costly, but also affects the classification effect. Second, it is difficult to make full use of spatial and spectral features for classification using CNN or GCN alone. Finally, achieving high-precision classification in the case of small samples still requires continuous exploration of the feature extraction capabilities of the network.

To this end, this paper combines CNN with GCN, and a new graph and double pyramid attention network based on linear discrimination of spectral interclass slices (GDPA_LDSICS) is proposed. First of all, a linear discrimination of spectral inter class slices (LDSICS) module, which can effectively remove highly redundant information and facilitate subsequent feature extraction. And the high-dimensional spectral information is mapped to the low-dimensional linear space to reduce the computational cost. In addition, in order to further reduce the amount of parameters, GDPA_LDSICS uses patches input. Then, a spatial spectral deformation (SSD) module was designed to effectively correlate spatial and spectral information. Finally, the cross-fusion method of CNN and GCN is used to extract deep features. In particular, on the CNN branch, double pyramid attention (DPA) was proposed to model contextual semantics to avoid information loss caused by long-distance feature extraction. On the GCN branch, an adaptive dynamic encoder (ADE) is proposed to capture spatial spectral features more completely.

The main contributions to this article include the following four parts:

(1) A LDSICS module is proposed. The LDSICS can remove similar local bands between different objects in the spectral dimension by slicing. The remaining spectral information is projected to a low-dimensional space with the largest interclass distance and the smallest distance within the class. This module can not only remove redundant spectral information, but also significantly reduce the computing cost.
(2) A DPA module is designed. The double pyramid tight structure of DPA can extract more complete context information over a long distance, reduce the information loss during the process of feature extraction. In addition, the group convolution of the pyramid structure can extract richer feature information with fewer parameters.
(3) In order to better integrate CNN and GCN, an ADE module is constructed. The ADE module adaptively and dynamically encodes adjacency matrix through autocorrelation, which can dynamically capture topological structure of spatial spectral correlation features, and make CNN and GCN more compatible.
(4) In order to effectively extract the spatial spectral correlation features, an SSD module is designed. SSD uses spectral spatial convolution to extract spectral features and spatial features successively, and significantly improves the nonlinear representation ability of spectral features. And the structure of dense connections enables SSD to extract more abundant deep features.

The rest of this paper is organized as follows. Section 2 introduces the GDPA_LDSICS network proposed in this paper and the four modules proposed in detail. Section 3 presents the hyperspectral datasets covered in this paper. In addition, all experimental results and analysis are provided. Finally, the conclusions are given in Section 4.

2. Methodology

In this paper, a GDPA_LDSICS network is proposed. Through cross fusion of CNN and GCN, GDPA_LDSICS network enables better classification of HSIs with small samples training. GDPA_LDSICS consists of four modules, including LDSICS, SSD, ADE and DPA, which will be described in detail.

2.1. The overall structure of GDPA_LDSICS

The goal of HSIC is to accurately assign corresponding labels to the objects in the image. Represents the HSI data as $x \in R^{h\times w\times b}$ and the label data as $y \in R^{h\times w}$. Where h, w, and b represent the height, width, and number of bands of HSI, respectively. First, GDPA_LDSICS preprocesses x through LDSICS. Second, a small number of pixel samples are randomly selected from x, and the pixels centred on them will be split into different patches as input. GDPA_LDSICS can make predictions for all pixels after training. The overall GDPA_LDSICS framework is shown in Figure 1. Specifically, the step 1 of GDPA_LDSICS first uses LDSICS to remove spectral bands that are highly similar between classes in x and project the remaining bands into low-dimensional space. It avoids the interference of redundant information on training and greatly reduces the amount of parameters for

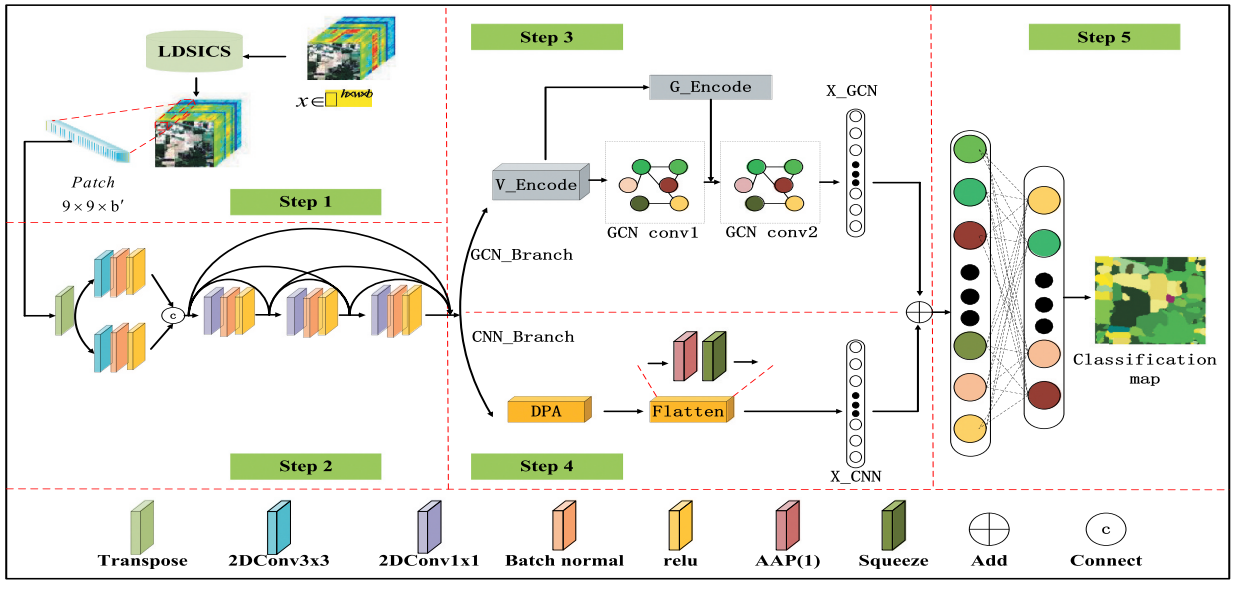


Figure 1. The overall structure of GDPA_LDSICS (the processing of GDPA_LDSICS is divided into 5 steps, and they are divided by red dotted lines in the figure).

network training. Then, in the step 2, the SSD module is used to extract spectral features and spatial features successively, and correlate spectral features with spatial features to achieve joint extraction of spatial spectral features. Subsequently, in the step 3 and step 4, the features are further extracted by two branches. In particular, on the GCN branch, the ADE module is proposed to dynamically extract the spatial topological features fused with spectral features and make the GCN and CNN structures more consistent. On the CNN branch, the DPA module is designed. The DPA uses multi-scale modelling to make the context semantic more complete, and the tight structure of the double pyramid avoids the loss of information for long-distance feature extraction. In addition, group convolution has been introduced in the DPA, enabling richer features to be obtained at low computational cost. Finally, in the step 5, the fully connected layer is used to obtain classification results.

2.2. LDSICS module

As shown in Figure 2, the raw data of HSI has interclass similarity problems in spectral bands, which seriously interferes with the feature extraction of the network. At the same time, the high-dimensional HSI data also increases the difficulty of network processing. In addition, directly using the entire HSI as input will also bring a huge amount of parameters, making it difficult for the network to converge. Inspired by the linear discriminant analysis (LDA) method, a LDSICS module was proposed. LDSICS can preprocess HSI to remove similar spectral bands between classes and map high-dimensional HSI data to low-dimensional space.

Specifically, first transform HSI data $x \in R^{h \times w \times b}$ into a two-dimensional matrix $x'$with size $s \times b$, and the process can be described as

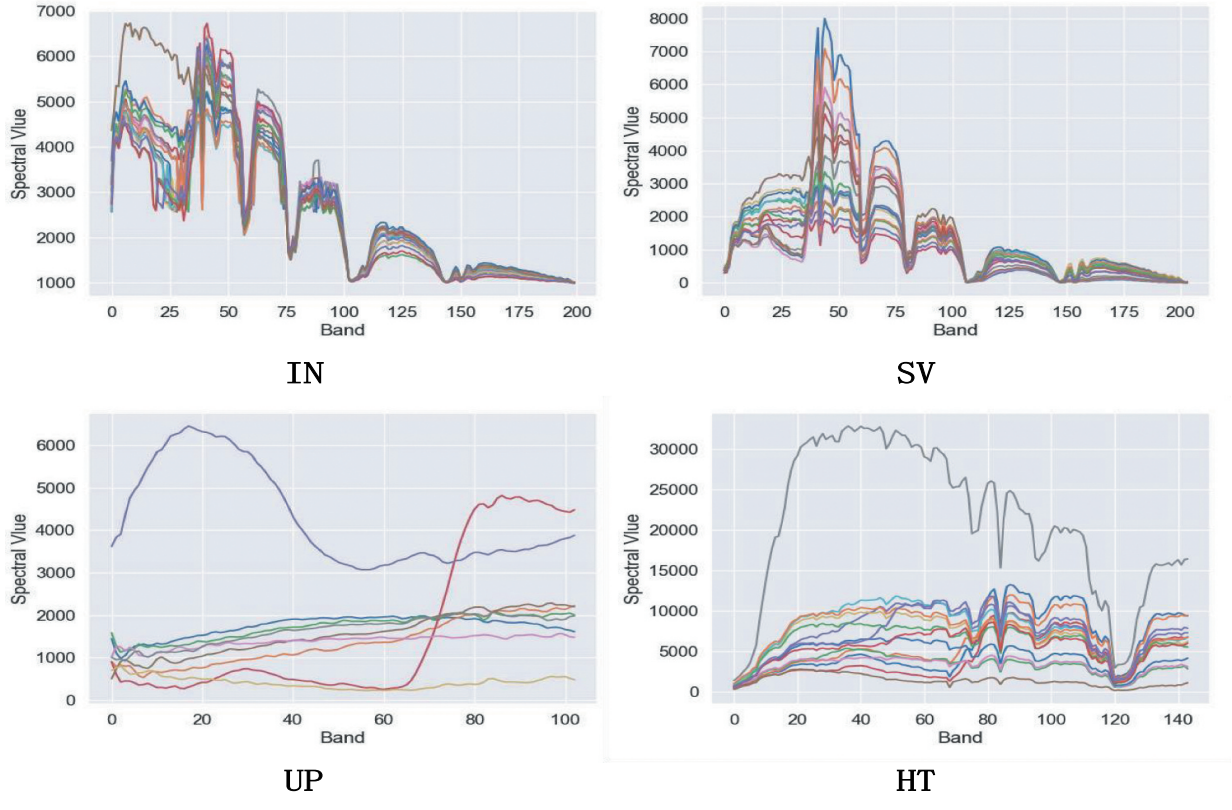


Figure 2. Local spectral of different data sets (different color curves correspond to different ground object).

$$x' = re(x) \quad x \in R^{h\times w\times b} x' \in R^{s\times b} \tag{1}$$

Where, $s = h \times w$, $re(\cdot)$ represents the reshaping operation. Slicing x' column by column yields the set θ, and normalizes each element in θ to get θ'. The details are as follows

$$\theta^{'} = BN(\theta) = BN(\phi_{\_1}(x')) \quad x' \in R^{s\times b} \tag{2}$$

Where $BN(\cdot)$ and $\phi_{\_1}(\cdot)$ represent the normalization function and the operation of slicing column-by-column, respectively. θ' is a slice set containing b elements. Then, the coefficient of variation is used to determine the similarity of bands between classes, and the bands are sorted according to the similarity. Next, n inter class similar bands are removed by slicing, and the basic characteristics of HSI are retained to obtain X. The related process can be represented as

$$\xi = \left\{ \sigma | \sigma = \frac{\sqrt{\frac{1}{s}\sum_{i=1}^{s}\left(\theta'_i - \overline{\theta'_i}\right)^2}}{\overline{\theta'_i}} \right\} (s = h \times w, 1 \le \mathrm{i} \le \mathrm{b}) \tag{3}$$

$$X = \{\partial | \partial = \phi_{-1}(H(\xi))\} \tag{4}$$

In Equation (3), θ'_i represents the i-th element in set θ', and $\bar{\theta}'_i$ is the mean value of the i-th element in θ'. Similar to θ', ξ is also a set containing b elements. Furthermore, the elements in ξ are coefficients of variation of the elements in θ'. In Equation (4), $H(\cdot)$ is the sort operation. The coefficients of variation are sorted to facilitate the removal of bands with smaller coefficients of variation, that is, the removal of similar bands between classes. In particular, these processes lay a solid foundation for the subsequent search for the optimal projection spatial. After removing the interclass similarity bands, proceed to calculate the between-class scatter matrix and the within-class scatter matrix of X. And use them to construct the objective function $\ell(\omega)$. In particular, in order to make the processed HSI data more conducive to classification tasks, the between-class scatter matrix should be as large as possible and the within-class scatter matrix should be as small as possible, that is, to project X in the direction of $\ell(\omega)$ maximum. In this way, the spectral features of the same class in HSI are more concentrated, while the spectral features of different classes are more dispersed, which is more conducive to the classification of HSI. Thus, using the Lagrange multiplier method to optimize the objective function $\ell(\omega)$ yields the projection spatial Ω, which can be represented as

$$BS_w = \sum_{i=1}^{c} BS_{w_i} = \sum_{i=1}^{c} \sum_{a \in X_i} (a - \mu_i)(a - \mu_i)^{\mathrm{T}} \tag{5}$$

$$BS_\tau = \sum_{i=1}^{c} m_i(\mu_i - \mu)(\mu_i - \mu)^{\mathrm{T}} \tag{6}$$

$$\Omega = \arg\max_W(I(W)) = \arg\max_W \frac{tr\left(W^{\mathrm{T}} BS_\tau W\right)}{tr(W^{\mathrm{T}} BS_w W)} \tag{7}$$

In Equation (5), c is the number of categories of samples, μ_i is the average value of the sample vector of the i-th category, and BS_w represents the within-class scatter matrix. In (6), m_i is the total number of samples in the i-th category, μ is the average value vector of the entire sample set, and BS_τ is between-class scatter matrix. In Equation (7), $tr(\cdot)$ represents the trace of the calculated matrix, $\arg\max(\cdot)$ is the optimization function, and Ω is the resulting projection spatial.

The proposed LDSICS is a linear, supervised pretreatment approach. The processing process of LDSICS is shown in algorithm 1. In the LDSICS, the redundant spectral information is removed first, and then the linear dimensionality reduction is carried out, which can maximize the retention of the characteristic information conducive to HSIC. In addition, the projection spatial obtained by LDSICS makes the inter-class features after dimensionality reduction more dispersed and the intra-class features more concentrated. The HSIC accuracy is significantly improved.

Algorithm 1 LDSICS module

Input: HSI raw data $x \in R^{h\times w\times b}$.

1. The spatial dimension of the input three-dimensional data x is flattened by Equation (1) to obtain the two-dimensional data $x' \in R^{s\times b}$. ($s = h \times w$)
2. Using Equation (2), slice $x^{'}$ column by column to get the set θ, and then normalize each element in θ to get θ'.
3. Calculate the coefficient of variation for each element in $\theta^{'}$. The coefficients of variation are then sorted from largest to smallest and the corresponding indexes are extracted.
4. Use slices for the original input data $x \in R^{h\times w\times b}$ and remove the bands corresponding to the first n indexes to get data X that removes similar bands between classes.
5. Calculating the between-class scatter matrix BS_τ and the within-class scatter matrix BS_w of X.
6. The objective function $\ell(\omega)$ is constructed by BS_τ and BS_w.

Output: Optimize the objective function $\ell(\omega)$ using the Lagrange multiplier method to get the projection spatial $\Omega \in R^{(b-n)\times(c-1)}$.

2.3. SSD module

HSI is a three-dimensional stereoscopic data containing a wealth of spatial and spectral information. To effectively extract the joint features of spectral and spatial is one of the important means to improve the performance of HSIC. A module capable of extracting spatial and spectral features in turn is designed by using the structure of spectral- spatial convolution (SSC) and dense connection. First, the SSD processes the spectral information as a channel for the image, and then extracts the spatial and spectral features sequentially using SSC. Moreover, when extracting spectral features, point-wise convolution is used for feature extraction, which enhances the nonlinear characterization ability of spectral features. Specifically, firstly, the spectral dimension is transposed to the channel dimension by SSD, and spectral information is processed as the channel. Then, two different branches are used to extract spatial features with different scales. Subsequently, multiple point-wise convolutional layers are used to extract spectral features. In particular, when extracting spectral features, a dense connection structure is embedded to prevent the network from over-fitting. This process can be expressed as

$$S\left(W_{(x,y)}\right) = relu\left(BN\left(cat\left(\sum_{W_{(x,y)}\in X} f\left(W_{(x,y)}\right) + b, \sum_{W_n\in X} f'\left(W_{(x,y)}\right) + b'\right)\right)\right) \tag{8}$$

$$S_L = \psi(S_0, S_1, \cdots, S_{L-2}, S_{L-1})\ L \in (0, 3], S_0 = S\left(W_{(x,y)}\right) \tag{9}$$

In Equation (8), X is the input mesh data. $W_{(x,y)}$ enumerates each position of X. $cat(\cdot)$ is a connection operation. $BN(\cdot)$ represents a normalized function. $relu(\cdot)$ is the activation function. $f(\cdot)$ and $f'(\cdot)$ represent pointwise convolution kernels with different numbers of channels, respectively. b and b' correspond to the biases of the convolutional kernels of $f(\cdot)$ and $f'(\cdot)$, respectively. $S\left(W_{(x,y)}\right)$ is the extracted spectral features. In (9), we define $\psi(\cdot)$ as a nonlinear composite function, including normalization, activation functions, and two-dimensional convolution operations at different scales. S_L represents the spatial features of the L-th layer.

2.4. ADE module

GCN mainly captures the spatial topology of the graph by modelling the vertex features and the adjacency matrix. Therefore, the performance of GCN on HSI classification depends largely on the construction of vertex features and adjacency matrices. In addition, in this article, the input features of GCN include spatial features with invariance as well as spectral features without variability. Therefore, if the encoding of the neighbour matrix continues to remain fixed in the traditional GCN way, it will affect the performance of classification. To this end, we propose an ADE module. Using autocorrelation, the adjacency matrix is dynamically encoded. The structure of the ADE is shown in Figure 3.

In general, vertex features are encoded by vertex encoders. Then, the vertex feature adjacency relationship is found by self-correlation, and the adjacency matrix is constructed for GCN modelling graph structural relationship. Specifically, ADE first linearly processes the vertex features and constructs vectors Q and vector E. Q and E then self-correlate and establish adjacencies and are weighted by Softmax, highlighting important structural features. This process can be described as

$$V = V\_Encode(X_S, w_v)\ V \in R^{s\times b} \tag{10}$$

$$Adj = A\_Encode(V, W_a) = V \cdot DN(V \cdot Soft\max(I(V) \cdot I(V))) \tag{11}$$

Where, X_S is the output feature of the SSD, $V\_Encode(\cdot)$ represents the vertex feature encoder, and w_v is the learnable weight of the encoder. In (11), $I(\cdot)$ represents the point convolutional kernel. $Soft\max(\cdot)$ is the activation function. $DN(\cdot)$ is denoted as double normalization. $A\_Encode(\cdot)$ is defined as a neighbour matrix encoder. Adj is the resulting neighbour matrix.

The ADE module is designed to construct a dynamic adjacency matrix and enable GCN to dynamically update vertex features through the adjacency matrix and training parameters. This not only enables GCN to maximize the mining of the spatial structure features of the image, but also makes GCN framework more compatible with CNN framework.

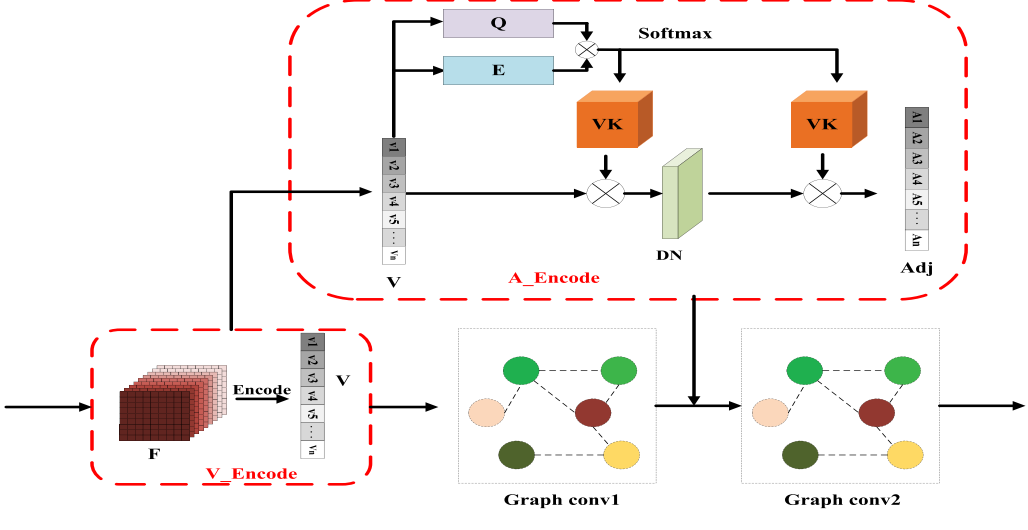


Figure 3. The structure of ADE module.

2.5. DPA module

In order to improve the HSI classification performance of the network, the method of increasing the network depth is usually used. However, long-distance feature extraction often causes a loss of a part of the information and weakens the interdependence of the context. With the introduction of attention mechanisms, this situation has been alleviated to some extent. However, the pooling of traditional attention networks will still cause some information loss. Therefore, the DPA module is proposed in this article.

The DPA has a dense structure of double pyramids, enabling the network to extract rich features. And reduce the loss of information during feature extraction. Its structure is shown in Figure 4. Specifically, first, the channels of the input feature map are split into four parts by DPA. Then, the features are extracted using multi-scale pyramid group convolution. Let the input feature be $X_S \in R^{c\times h\times w}$, and the process can be represented as

$$[X_0, X_1, X_2, X_3] = Split(X_S) \tag{12}$$

$$X'_i = F_i(X_i, w_i)\ \ i \in [0, 3] \tag{13}$$

In (12), $Split(\cdot)$ is the split function. X_0, X_1, X_2 and X_3 represent the four features after the split, respectively. In formula (13), $F_i(\cdot)$ is the group convolution. It is worth noting that this paper divides the convolution into four groups. The size of each group convolution is different. The size of the convolution corresponds to the i-value. The values of i are 0, 1, 2, and 3, corresponding to the size of the convolution of 3×3, 5×5, 7×7, and 9×9, respectively. In particular, the use of group convolution can reduce the computational cost of DPA. Specifically, the parameter quantities of ordinary convolution and the parameter quantities of group convolution can be expressed as such

$$P_O = k \times k \times c_{in} \times c_{out} \tag{14}$$

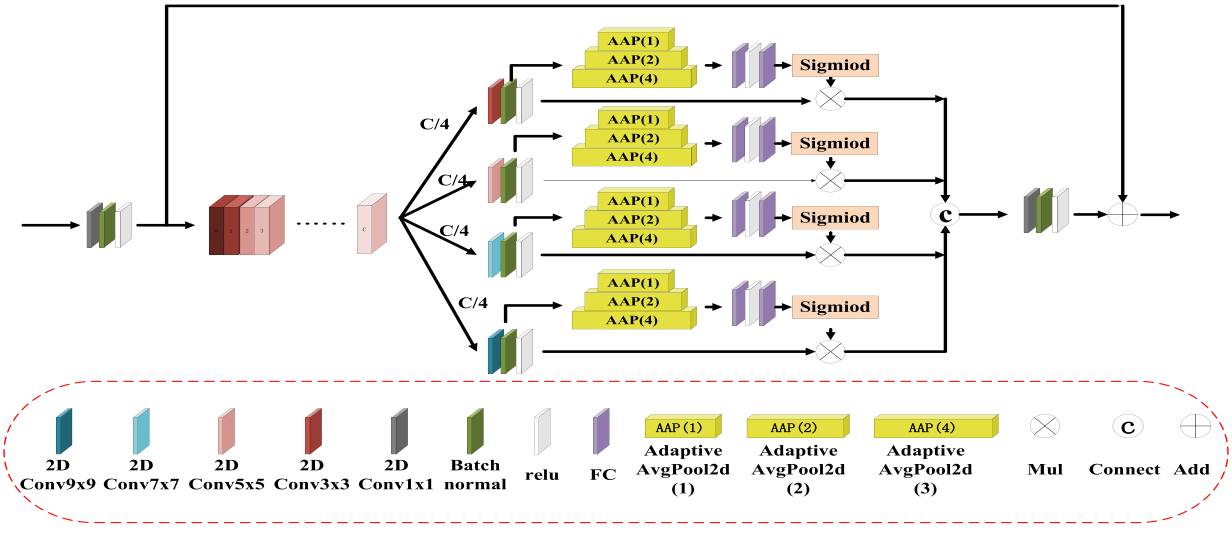


Figure 4. The structure of DPA module.

$$P_G = k \times k \times \frac{c_{in}}{G} \times \frac{c_{out}}{G} \times G \tag{15}$$

Where k is the size of the convolution. c_{in} and c_{out} are the input and output channels of the convolution, respectively. P_O and P_G are the parameter quantities of ordinary convolution and the parameter quantities of group convolution, respectively. In (15), G represents the number of groups for group convolution. Obviously, in the case of the same size of the convolution, the amount of parameters for the group convolution is much smaller than the amount of parameters for the ordinary convolution. And, as the number of groupings for group convolution increases, this advantage becomes more obvious.

For traditional attention methods, the model establishes using a single global average pooling (GAP) will lead to the loss of some feature information. Inspired by Wang et al. (2018), we first build the pyramid pooling structure using three adaptive average pooling (AAP) with different sizes. It is then optimized through two fully connected layers. An activation function has been added between the two fully connected layers to enhance the nonlinear representation of features. Next, the attention mask is obtained by normalization, and the attention mask is multiplied and connected together with the features extracted by the multi-scale pyramid group convolution. Finally, in order to prevent overfitting of network training, residual connections are also added to the DPA. This process can be described as

$$O_i = Sigmiod(FC(PP(X'_i))) \ i \in [0, 3] \tag{16}$$

$$O = relu(X_S + F(cat(O_i))) \ i \in [0, 3] \tag{17}$$

In (16), X'_i is the output feature of the multi-scale pyramid group convolution. $PP(\cdot)$ is the pyramid pooling function, and $PP(\cdot)$ is composed of three AAPs of different sizes 2×2, 4×4, and 8×8. $FC(\cdot)$ is a composite function consisting of two fully connected layers and an activation function. $Sigmiod(\cdot)$ is the activation function. O_i is the attention mask. In (17), $F(\cdot)$ is a composite function that contains point-size convolution, batch normalization, and activation function. O is the output of DPA.

3. Experimentation and analysis

In this part, the data set selected for the experiment and the parameter settings of the experiment are described. In addition, in order to verify the effectiveness of GDPA_LDSICS, a large number of experiments were conducted.

3.1. HSI datasets

To verify the network performance of GDPA_LDSICS, four classic datasets are selected in this paper. As shown in Figures 5–8, they are Indian Pines (IN), Salinas Valley (SV), Pavia University (UP), and Houston2013 (HT). In order to accurately evaluate the network, three standards of overall accuracy (OA), mean accuracy (AA) and Kappa coefficient (Kappa) are uniformly selected. Specifically, IN contains 16 kinds of vegetation. Most of them are agricultural vegetation. IN has total of 21,025 pixels and has 200 spectral bands. However, only 10,249 pixels have real object. SV contains 16 kinds of agricultural vegetation,

No.	Legend	Name	Sample
1		Alfafa	46
2		Corn–n	1428
3		Corn–m	830
4		Corn	237
5		Grass–p	483
6		Grass–t	730
7		Grass–p–m	28
8		Hay-w	478
9		Oats	20
10		Soybean-n	972
11		Soybean-m	2455
12		Soybean-c	593
13		Wheat	205
14		Woods	1265
15		Buildings-G-T	386
16		Stone-S-T	93

Figure 5. Categories and number of ground objects in IN.

including 204 bands. A total of 111,104 pixels, the object covers 54,129 pixels. The UP dataset has a resolution of 610 $\times$ 340 and contains 103 bands. Compared with the first two data sets, not only the types of ground objects have changed greatly, but also the number of classes and bands of ground objects have decreased. UP has only 9 kinds of ground objects, covering 42,776 pixels. Finally, the HT dataset is an image obtained from the spectral range of 346 nm to 1049 nm by ITRES CASI-1500. The HT consists of 144 bands in total and includes 15 categories. Its ground objects cover 15,029 pixels.

3.2. Experimental environment and parameter settings

The GDPA_LDSICS is implemented on the Pycharm software of Pytorch 1.10.0 and Python 3.7.4. For fair comparison, the hyperparameters of the comparison algorithm are the same as those of the original network. And, all experiments use a unified hardware platform. The CPU we use is AMD Ryzen 75,800 H with Radeon Graphics. The GPU uses the NVIDIA GeForce RTX 3070 with 8GB of memory. To improve the computational speed, the GPU is used for network training. Therefore, the operating environment of CUDA11.2 was chosen. First, GDPA_LDSICS uses the Adam optimizer for training optimization and sets the learning rate to 0.001. The number of epochs is 300 times. Then, for different HSI datasets, different proportions of training samples were randomly selected. In IN, 3% of the sample is used as the training set. In SV and UP, 0.5% of the sample is used as the training set. In HT, a 5%

Ground truth

No.	Legend	Name	Sample
1		weeds_1	2009
2		weeds_2	3726
3		Fallow	1976
4		Fallow-r-p	1394
5		Fallow-s	2678
6		Stubble	3959
7		Celery	3579
8		Grapes-u	11271
9		Soil-v-y-d	6203
10		C-s-g-weeds	3278
11		L-r-4wk	1068
12		L-r-5wk	1927
13		L-r-6wk	916
14		L-r-7wk	1070
15		VIN-yard-u	7268
16		VIN-yard-v-t	1807

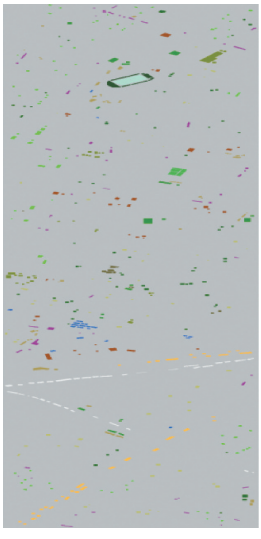
Figure 6. Categories and number of ground objects in SV.

Ground truth

No.	Legend	Name	Sample
1		Asphalt	6631
2		Meadows	18649
3		Gravel	2099
4		Trees	3064
5		Painted m s	1345
6		B Soil	5029
7		Bitumen	1330
8		S-B Bricks	3682
9		Shadows	947

Figure 7. Categories and number of ground objects in UP.

sample is used as the training set. In addition, through a large number of experimental comparisons, the n value of the optimal input patch size and LDSICS was determined. Specifically, the input patch size and n are set to 5 different values, and the OA changes are observed experimentally.



Ground truth

No.	Legend	Name	Sample
1		Healthy grass	1251
2		Stressed grass	1254
3		Synthetic grass	697
4		Trees	1244
5		Soil	1242
6		Water	325
7		Residential	1268
8		Commercial	1244
9		Road	1252
10		Highway	1227
11		Railway	1235
12		Parking Lot 1	1233
13		Parking Lot 2	469
14		Tennis Court	428
15		Running Track	660

Figure 8. Categories and number of ground objects in HT.

As can be seen from Figure 9, the higher OA values are all concentrated near the patch with input 9×9 and behave similarly on four datasets. Therefore, this article sets the patch size of the network to 9×9. In addition, on the four datasets, when the size of the input patch of the network is determined, the classification accuracy of the network first increases and then decreases as n increases. Firstly, the classification accuracy of the network increases with the increase of n. This is because the redundant information in HSIs is effectively removed by LDSICS, thereby improving the classification performance of the network. When the classification performance of the network reaches its peak, the classification accuracy of the network will decrease if the n value continues to increase. This is because some spectral information that is conducive to classification in HSIs has also been removed. Therefore, in order to achieve the best performance of the network, different n is used on different datasets. Specifically, set the n values of IN, SV, HT and UP to 20, 15, 10 and 5 in turn. Finally, to ensure the credibility of the experimental results, all results are the average of 10 replicated experiments, and the optimal experimental data in all tables will be marked in bold.

3.3. Performance analysis of proposed modules

In this section, in order to verify the performance of the proposed modules, ablation experiments are performed on different modules. And the experimental results were analysed.

3.3.1. Performance analysis of the LDSICS

The LDSICS module is proposed to remove the influence of highly redundant spectral information on the classification effect, and reduce the dimension of spectral information

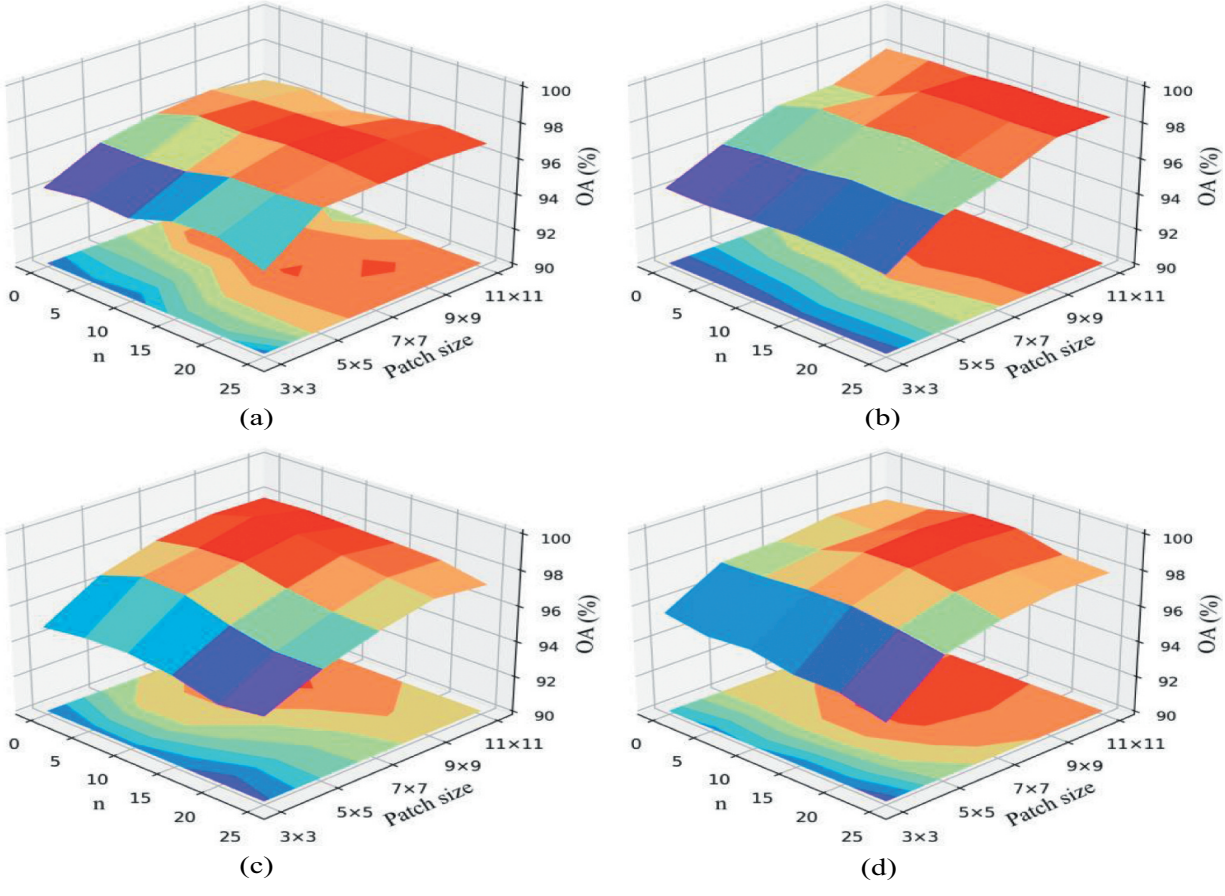


Figure 9. When inputting different patch sizes, the relationship between the *n* value of LDSICS and the OA. (a)-(d) IN、SV、UP、HT.

to facilitate network feature extraction. To verify the performance of the LDSICS, a series of ablation experiments were conducted. Specifically, keeping other experimental conditions unchanged, only GDPA_LDSICS with the LDSICS modules removed will be compared experimentally with GDPA_LDSICS with the LDSICS modules. The result is shown in Figure 10. Obviously, the GDPA_LDSICS using the LDSICS has significant improvements in classification performance on different datasets compared to unused the LDSICS. Specifically, on the IN and SV, the improvement of classification accuracy by the LDSICS is more obvious. This is due to the fact that IN and SV contain a larger number of ground objects and that most of these ground objects are similar crops. Therefore, their spectral information is relatively similar. Therefore, when using the LDSICS to remove these indistinguishable inter-class similar bands, the classification effect is naturally significantly improved. This also fully proves that the LDSICS can effectively improve the classification performance of the network.

In order to more intuitively demonstrate the effectiveness of LDSCIS, Scatter plots before and after processing using LDSICS on the IN dataset are compared in this section. As shown in Figure 11(a) the spectral values of different classes in the original spectral scatter plots of the IN dataset almost completely overlap in some bands, and these overlapping spectral

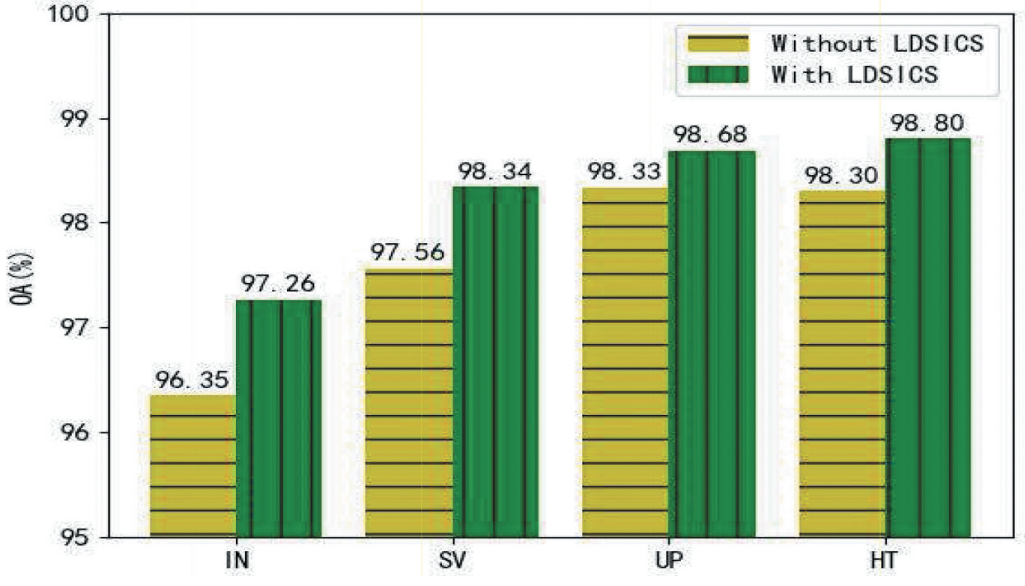


Figure 10. Effect of the LDSICS on classification results.

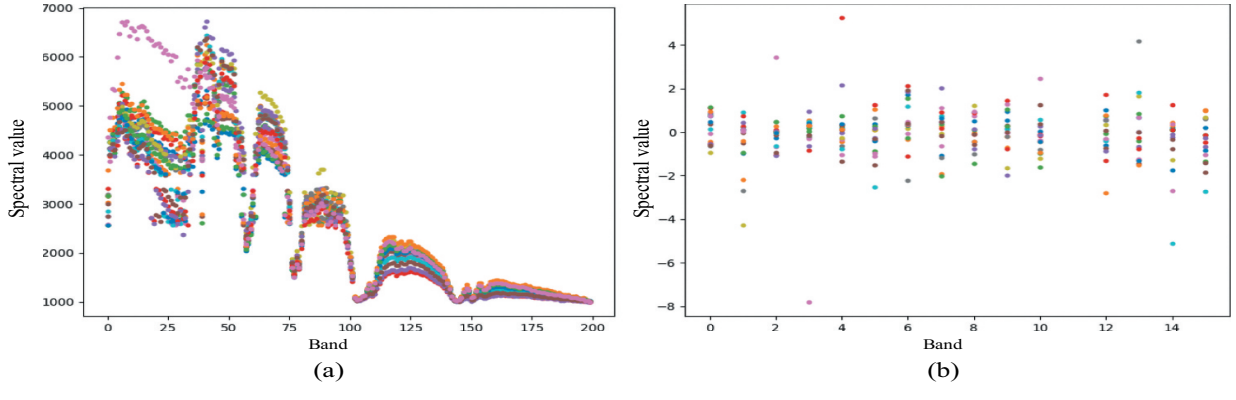


Figure 11. Local spectral scatter plots of the in dataset. (a) spectral scatter plots of the original in dataset; (b) spectral scatter plots of the in dataset processed by LDSICS. (different colored dots correspond to spectral values of different classes).

information will seriously affect the final classification results. As shown in Figure 11(b), the IN dataset shows significant distinguish-ability of spectral values between different classes after the LDSICS processing proposed in this article. It is obvious that the data processed by LDSICS is more conducive to the classification of HSI.

In addition, when the GDPA_LDSICS uses the LDSICS, the cost of computing is significantly reduced. As shown in Table 1, the amount of parameters in the GDPA_LDSICS is almost halved after using the LDSICS. It is explained that the LDSICS can reduce the calculation difficulty of the network, save computing resources, and improve the computing speed of the network. And the LDSICS as a plug-and-play module, it provides great convenience for the study of lightweight networks of HSICs.

Table 1. The LDSICS effect on the amount of parameters.

METHOD	IN	SV	UP	HT
Without LDSICS	680K	531K	686K	594K
With LDSICS	**396K**	**385K**	**396K**	**394K**

3.3.2. Performance analysis of the SSD

In this section, the classification performance of the GDPA_LDSICS that only removes SSDs is compared with the classification performance of the complete GDPA_LDSICS. The experimental results are shown in Figure 12. Obviously, with the SSD module, the classification accuracy of GDPA_LDSICS on all four data sets has been significantly improved. This shows that SSD first extracts spectral features and spatial features, and then realizes the joint extraction of spatial spectral features, which can more effectively extract the spatial spectral features of HSI and use them for classification.

3.3.3. Performance analysis of the ADE

For the performance analysis of ADE, the same experimental method is used. The experimental results are shown in Figure 13. Similarly, from Figure 13, it can be concluded that the ADE is beneficial to improve the classification performance of GDPA_LDSICS. This is due to the ability of ADE to dynamically code adjacency matrices to update vertex features that enable the GDPA_LDSICS to mine rich spatial topology features.

3.3.4. Performance analysis of the DPA

DPA has a tight structure of double pyramids that can extract rich deep features. And the attention modelling is carried out by using the way of pyramid pooling, which reduces the loss of feature information. To verify the performance of DPA, some experiments were

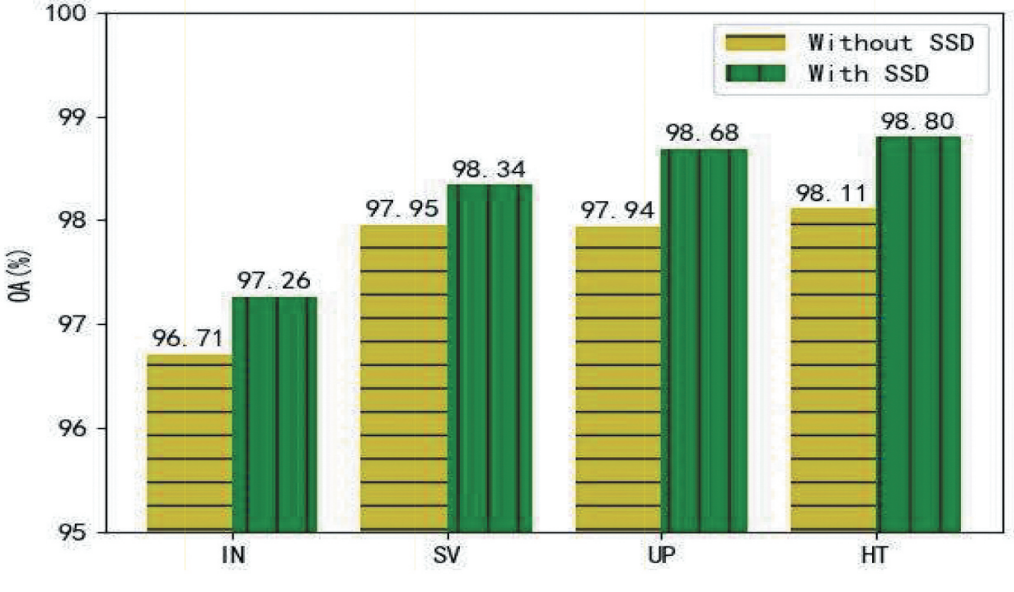


Figure 12. Effect of the SSD on classification results.

performed and the results were shown in Figure 14. As can be seen from Figure 14, the GDPA_LDSICS using DPA shows a steady improvement in classification accuracy on all four data. This shows that the DPA has strong generalization and can adapt to different scenarios.

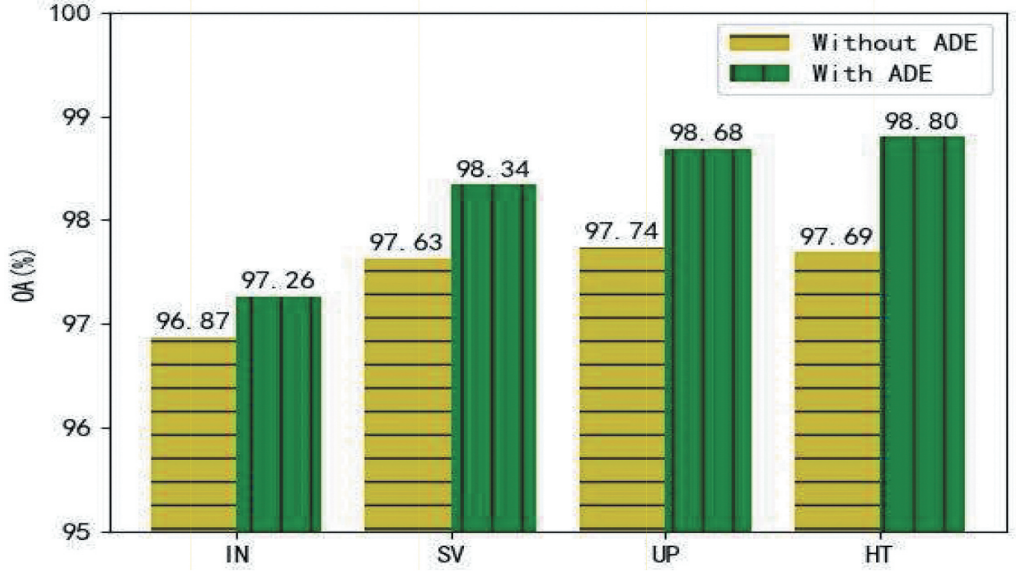


Figure 13. Effect of the ADE on classification results.

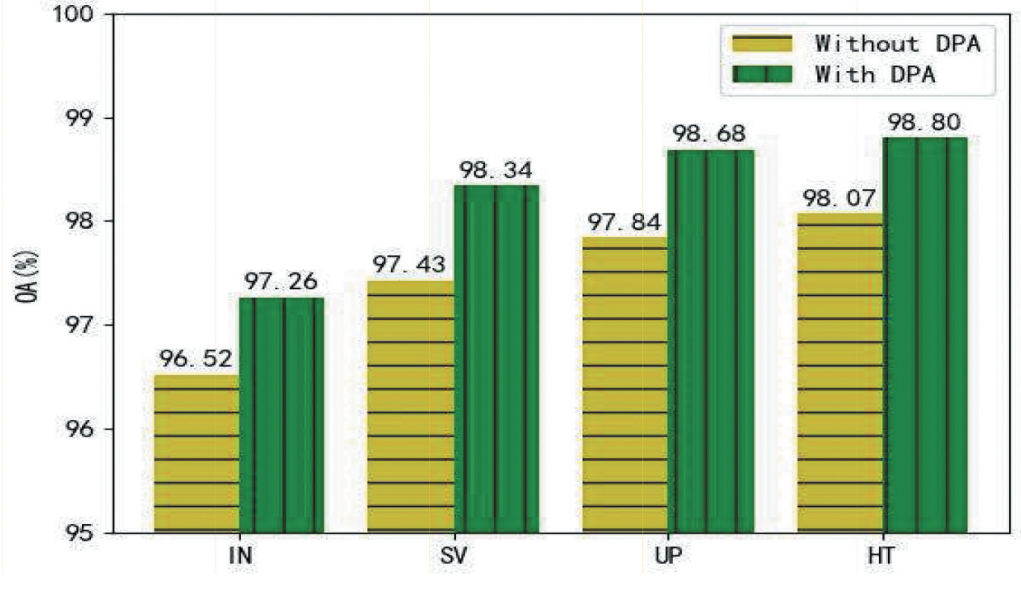


Figure 14. Effect of the DPA on classification results.

3.4. Analysis of feature fusion

In the method proposed in this paper, CNN and GCN can work collaboratively. The classification performance of the network can be improved effectively by fusing features extracted from CNN and GCN. In order to intuitively analyse the impact of feature fusion on the network, the analysis of feature fusion on the IN dataset is given in this section.

As shown in Figures 15(a) and 11(b), the same classes in the GCN branch are more clustered compared to the CNN branch. This is because GCN can capture the internal structural relationships between ground objects. However, there is more mixing between different categories in the GCN branch compared to the CNN branch, which is due to the stronger ability of CNN to extract fine features and local features compared to GCN. Overall, although both GCN and CNN working alone can achieve good classification, there are also some shortcomings. To this end, in this paper, CNN and GCN are integrated to effectively improve network classification performance. As shown in Figure 15(c), compared to using CNN or GCN alone, not only are the same classes more clustered, but there is also less mixing between different classes.

3.5. Comparison of different methods

In this section, GDPA_LDSICS proposed in this paper is compared with seven methods. These methods include CDCNN (Lee and Kwon 2017), SVM, DBMA (Ma et al. 2019), SSRN, FDSSC (Wang et al. 2018), DBDA (Li et al. 2020), FECNet (Shi et al. 2022), A2S2KResNet (Roy et al. 2021) and FDGC (Liu et al. 2022). Among them, the CDCNN is a simple and effective CNN classification method. It combines multi-branch convolution with maximum pooling and uses a residual structure. The SVM is a classic machine learning and is a pixel-by-pixel classification method based on radial basis functions (RBF) kernels. The SSRN and FDSSC optimize the network using residual and dense connections, respectively, to deepen network depth. Both the DBDA and DBMA are classification methods based on attention. The difference is that the DBDA uses double-branched and dual-attention, while the DBMA uses double-branched multi-attention. In FECNet, dilation convolution is used to reduce the amount of network parameters and a feedback mechanism is used to fuse the shallow and deep features extracted by the network. An improved 3D-ResNet is adopted

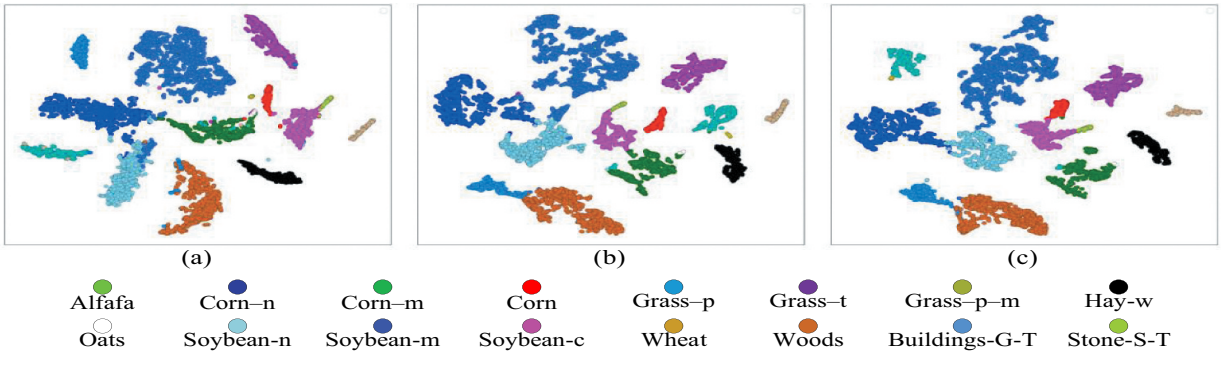


Figure 15. Analysis of feature fusion on in datasets. (a) visualization of GCN branch; (b) visualization of CNN branch; (c) visualization of GDPA_LDSICS.

by A2S2ResNet to extract image features and this network can improve classification accuracy through effective feature recalibration (EFR). The FDGC uses three branches to extract features separately and is a classification method that combines CNN and GCN.

3.5.1. Comparison of classification results

Table 2–5 reports the classification results of GDPA_LDSICS and other methods on four datasets. As can be seen from Table 2, on the IN, GDPA_LDSICS's OA, AA, KAPPA three indicators are higher than other methods. For the classification accuracy of each category on the IN, GDPA_LDSICS has 12 categories with higher classification accuracy than other methods. On the other 3 datasets, similar conclusions were drawn. In particular, Table 4 reports the classification results of all methods on the UP. As can be seen from Table 4, GDPA_LDSICS has achieved greater advantages in all evaluation indicators than other methods. This is due to the fact that the different categories of spectral information in the UP are more dispersed than other datasets, which is more conducive to the work of the GDPA_LDSICS preprocessing phase. Looking at the classification results of all the methods in 4 datasets, GDPA_LDSICS always showed the best classification effect. This fully proves the excellent classification performance and generalization of GDPA_LDSICS. In addition, in order to explore the network computing complexity of the proposed method, the running times of different methods on different datasets, including testing time and training time, are also shown in Tables 3 to 6. Specifically, on different datasets, the runtime of the proposed method is only slightly higher than that of FDGC, but much lower than that of other methods. This indicates that the proposed method has a lower network computational complexity. This is thanks to the proposed network reducing spectral dimensions through LDSICS. On the other hand, group convolution in DPA can significantly reduce the computational complexity of the network.

Table 2. Comparison between GDPA_LDSICS and other methods on IN.

Class	CDCNN	SVM	DBMA	SSRN	FDSSC	DBDA	FECNet	A2S2KResNet	FDGC	GDPA_LDSICS
1	24.67	36.62	77.33	81.03	85.72	99.06	99.06	94.32	49.46	**100.0**
2	49.93	55.48	78.68	87.48	93.35	92.74	96.63	92.99	88.34	**98.06**
3	31.93	62.33	76.12	76.35	89.51	92.28	93.25	94.51	91.24	**98.05**
4	33.52	42.53	75.11	73.88	93.68	94.46	95.61	94.43	98.98	**99.06**
5	71.47	85.05	94.91	84.28	92.61	**99.10**	98.55	97.62	81.20	98.62
6	73.52	83.31	93.99	92.68	98.31	98.28	97.51	97.68	96.76	**99.12**
7	36.29	59.86	44.33	79.06	82.46	66.57	73.14	76.17	68.49	**100.0**
8	78.08	89.67	97.49	96.85	97.54	99.24	99.84	**100.0**	99.13	**100.0**
9	42.14	39.27	45.66	73.57	71.07	93.64	79.71	61.96	97.44	**100.0**
10	41.71	62.31	77.27	84.45	89.30	**93.77**	90.09	91.23	91.99	92.10
11	55.67	64.72	83.89	86.95	93.97	93.83	96.11	94.71	95.67	**98.11**
12	27.68	50.54	77.65	83.31	88.25	90.65	93.20	92.44	85.46	**97.71**
13	67.88	86.73	96.74	98.83	99.53	97.48	98.34	96.62	91.59	**99.46**
14	76.39	88.67	93.51	95.13	95.82	97.69	97.28	97.68	**97.77**	94.47
15	47.30	61.81	76.83	88.58	92.48	93.79	95.11	92.37	91.57	**98.71**
16	65.67	**98.66**	92.33	96.52	98.22	93.67	97.26	91.71	89.9	97.77
OA(%)	59.19	68.76	82.38	86.68	92.37	92.26	95.47	94.57	92.55	**97.26**
AA(%)	51.48	66.72	80.11	86.18	91.36	92.90	93.79	91.65	87.97	**98.20**
KAPPA × 100	78.1	63.98	79.86	84.76	91.30	91.24	94.83	93.80	91.51	**96.88**
Training time (s)	57.16	—	423.03	934.14	2124.94	425.72	256.58	959.01	**24.55**	57.16
Testing time (s)	1.91	—	16.34	4.69	9.98	16.38	16.02	6.18	**1.05**	1.91

Table 3. Comparison between GDPA_LDSICS and other methods on SV.

Class	CDCNN	SVM	DBMA	SSRN	FDSSC	DBDA	FECNet	A2S2KResNet	FDGC	GDPA_LDSICS
1	68.15	99.42	99.98	99.53	99.96	99.68	**100.0**	**100.0**	99.68	**100.0**
2	73.63	98.79	99.20	99.59	98.90	98.88	99.95	99.92	**100.0**	**100.0**
3	75.56	87.98	97.65	94.22	96.91	97.94	98.41	99.18	99.90	**100.0**
4	92.79	**97.54**	92.64	96.98	94.54	94.89	95.85	96.76	93.52	95.64
5	92.90	95.09	98.79	98.84	99.16	98.40	99.64	99.48	95.85	**99.73**
6	96.25	99.89	98.49	99.87	99.82	99.92	99.89	99.86	99.90	**100.0**
7	93.76	95.59	98.41	98.21	98.16	98.46	99.40	99.20	99.85	**100.0**
8	74.03	71.66	90.72	86.20	91.72	90.87	94.79	94.18	96.37	**98.97**
9	94.72	98.08	99.61	99.14	99.53	99.24	99.67	99.45	**99.99**	99.64
10	76.65	85.39	92.23	97.99	97.61	97.62	98.01	**99.29**	98.97	98.99
11	69.32	86.97	93.10	94.48	95.70	94.92	97.18	96.39	92.40	**99.90**
12	80.43	94.20	99.24	98.53	98.16	99.54	99.55	99.15	98.06	**99.94**
13	69.55	93.53	98.56	98.28	98.35	99.76	99.94	98.51	97.23	**100.0**
14	87.22	92.03	96.56	96.05	96.21	**96.66**	96.51	96.44	95.59	94.29
15	63.71	71.02	88.27	81.04	89.07	90.48	93.68	90.41	**94.50**	93.06
16	98.38	97.81	99.68	99.48	99.77	99.83	**100.0**	99.82	97.81	**100.0**
OA(%)	80.67	86.97	94.77	92.51	95.91	95.74	97.49	96.81	97.59	**98.34**
AA(%)	81.69	91.55	96.45	96.32	97.49	97.41	98.01	98.00	97.47	**98.49**
KAPPA × 100	78.35	85.45	94.18	91.67	95.45	95.26	97.21	96.45	97.31	**98.15**
Training time (s)	51.65	—	375.41	805.15	1819.6	380.11	268.62	733.87	**23.54**	44.17
Testing time (s)	10.56	—	92.57	28.37	56.23	92.78	124.55	20.72	**5.1**	8.5

Table 4. Comparison between GDPA_LDSICS and other methods on UP.

Class	CDCNN	SVM	DBMA	SSRN	FDSSC	DBDA	FECNet	A2S2KResNet	FDGC	GDPA_LDSICS
1	79.88	81.26	89.41	93.22	84.30	90.77	96.77	92.66	89.44	**98.20**
2	86.66	84.52	94.34	93.75	95.47	98.07	99.20	97.65	98.19	**99.27**
3	32.74	56.56	84.99	64.97	80.58	90.04	96.56	84.15	84.95	**99.79**
4	84.88	94.34	96.58	94.69	98.13	97.96	97.87	98.58	82.76	**98.54**
5	94.77	95.38	98.06	97.69	99.21	98.67	97.38	98.18	90.04	**99.85**
6	71.54	80.66	94.60	93.14	92.36	98.85	97.72	96.69	95.06	**99.96**
7	31.77	49.13	93.18	73.06	69.35	96.95	96.50	97.67	97.08	**100.0**
8	65.54	73.15	76.61	79.98	73.36	87.65	87.34	83.51	78.64	**94.19**
9	73.50	97.93	91.45	98.71	97.39	97.81	98.98	98.91	74.80	**100.0**
OA	80.69	82.06	90.98	89.12	90.71	95.32	96.96	94.83	92.19	**98.68**
AA	69.02	79.21	91.02	87.69	87.91	95.20	96.48	94.24	87.88	**98.76**
KAPPA × 100	73.55	75.43	87.87	85.44	87.38	93.77	95.97	93.12	89.62	**98.26**
Training time (s)	42.83	—	130.39	606.40	1406.32	130.68	181.27	619.50	**18.81**	29.73
Testing time (s)	8.13	—	39.7	12.34	25.19	40.37	69.60	17.34	**3.95**	7.1

3.5.2. Comparison of classification results

Figures 16–19 show the classification maps on the four datasets. In which, IN is similar to SV, including more the classes of ground objects in the four data sets, and the classes of ground objects are also similar. In addition, in the 4 datasets, the number of spectral bands contained in the IN and the SV is also higher. Therefore, it is also more difficult to classify them. As can be seen from Figures 16 and 17, the classification maps obtained by other methods show a lot of noise and without clear category boundaries. This is due to the large number of inter-class similar bands between IN and SV datasets, which interferes with the feature extraction of the network. However, the GDPA_LDSICS still exhibits a high classification effect compared to other methods. This is because the GDPA_LDSICS can remove the similar bands between different categories through pre-processing, aggregate the feature information of the same classes, and disperse the feature

Table 5. Comparison between GDPA_LDSICS and other methods on HT.

Class	CDCNN	SVM	DBMA	SSRN	FDSSC	DBDA	FECNet	A2S2KResNet	FDGC	GDPA_LDSICS
1	87.59	95.62	98.11	99	98.48	86.76	97.42	96.09	92.19	**99.21**
2	93.83	98.37	99.81	99.02	99.82	99.9	99.39	99.69	94.91	**100.0**
3	91.23	98.5	100	97.49	99.52	**100.0**	99.92	100.0	97.55	**100.0**
4	97.15	96.35	97.88	98.25	97.99	**99.43**	99.02	97.53	96.00	97.90
5	97.69	95.17	99.19	97.02	99.9	98.92	99.63	**100.0**	**100.0**	**100.0**
6	98.73	99.32	99.3	99.61	**100**	98.47	99.65	99.73	99.04	97.56
7	88.24	92.56	98.7	91.53	89.97	93.36	96.69	96.34	97.17	**99.10**
8	95.3	83.08	96.45	99.5	97.95	93.91	99.61	**99.92**	97.15	99.90
9	78.46	82.32	96.53	85.63	92.44	88.83	94.84	96.75	**98.83**	95.50
10	70.99	88.78	98.05	88.13	96.88	83.29	95.20	96.62	99.19	**99.63**
11	75.51	92.52	97.42	91.59	99.13	91.24	97.86	98.42	99.49	**99.82**
12	80.23	81.91	93.77	98.54	96.26	80.28	98.72	98.35	98.75	**98.90**
13	94.76	77.77	92.05	86.76	83.54	97.45	92.31	95.81	95.24	**100.0**
14	98.17	95.97	**100.0**	99.48	**100.0**	**100.**	99.10	97.93	**100.0**	**100.0**
15	91.95	99.67	99.33	98.51	98.67	95.06	98.97	99.43	**99.52**	98.57
OA(%)	87.47	91.40	97.69	94.94	96.67	92.32	97.79	98.05	97.44	**98.80**
AA(%)	89.32	91.86	97.77	95.38	96.70	93.79	97.89	98.17	97.43	**98.87**
KAPPA × 100	86.46	90.69	97.51	94.53	96.39	91.70	97.61	97.89	97.23	**98.70**
Training time (s)	127.63	—	666.47	2177.71	5023.52	678.17	811.18	2333.97	**80.56**	102.18
Testing time (s)	2.79	—	16.73	5.07	10.47	16.83	27.98	6.75	9.82	**2.48**

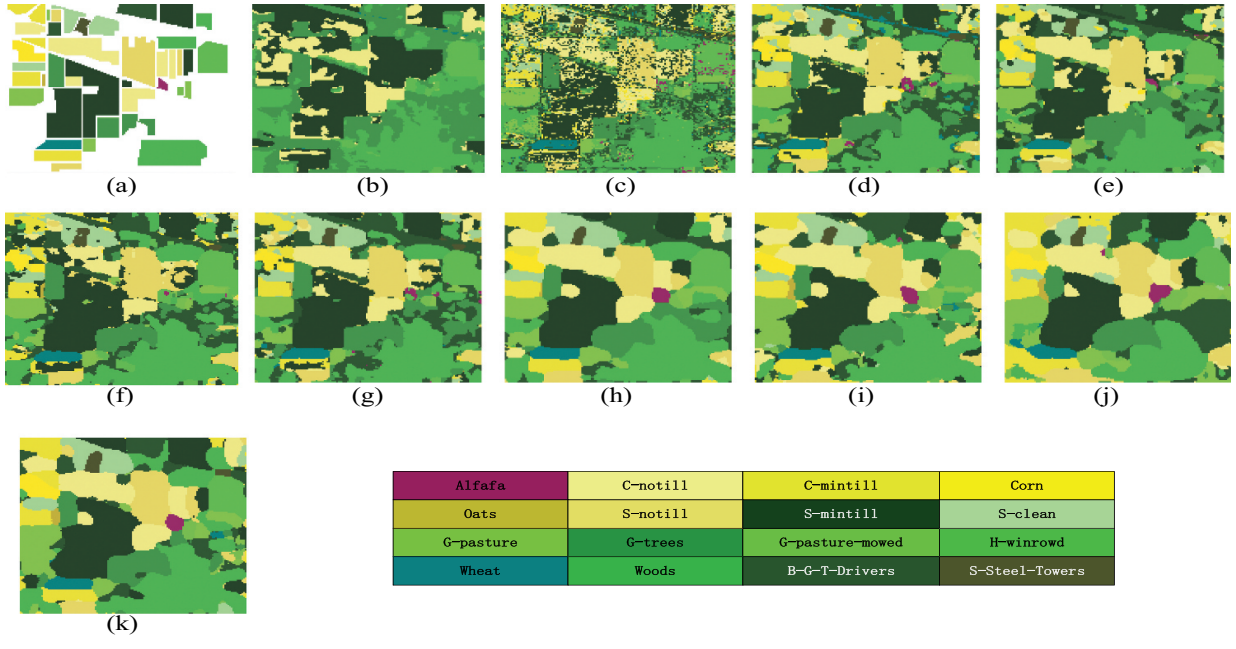


Figure 16. Comparison of classification maps on in (the different colors in the figure correspond to different ground objects). (a) ground truth, (b–k) CDCNN, SVM, DBMA, SSRN, FDSSC, DBDA, FECNet, A2S2KResNet, FDGC, GDPA_LDSICS).

information of different classes. At the same time, the spectral dimension is reduced, reducing the computational cost of the network. In addition, compared with other methods, GDPA_LDSICS has stronger feature extraction ability. In the

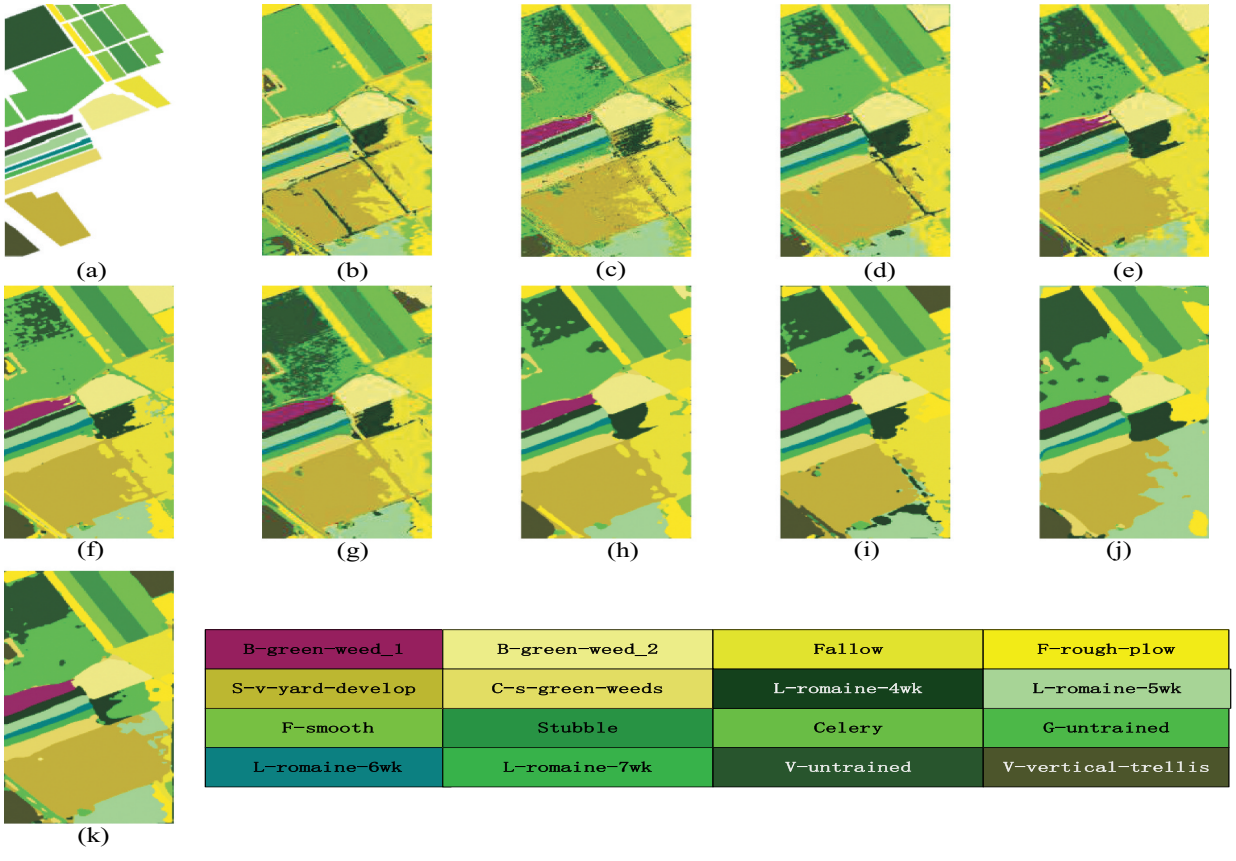


Figure 17. Comparison of classification maps on SV (the different colors in the figure correspond to different ground objects). (a) ground truth, (b–k) CDCNN, SVM, DBMA, SSRN, FDSSC, DBDA, FECNet, A2S2KResNet, FDGC, GDPA_LDSICS).

process of feature extraction, the context semantic dependency is enhanced as much as possible, and the information loss is reduced.

Figures 18–19 show the classification map of UP and HT respectively. Compared with the other two datasets, the UP and HT datasets have fewer classes, and the spectral curves of different classes differ greatly, which is more conducive to HSIC. Nevertheless, the classification performance of other comparison methods is unsatisfactory. For example, Figure 18(d) is a classification map of DBMA on UP. It can be seen from Figure 18(d) that different classes interact with each other, and the boundaries between classes are ambiguous. This is due to the fact that DBMA adopts a single GAP to model attention, resulting in a loss of feature information. However, on UP, the classification map of GDPA_LDSICS proposed in this article has a clearer class boundary, and there are very few misclassifications, which is more in line with the real scene. This is due to the pyramid pooling structure of DPA, which reduces the information loss during attention modelling.

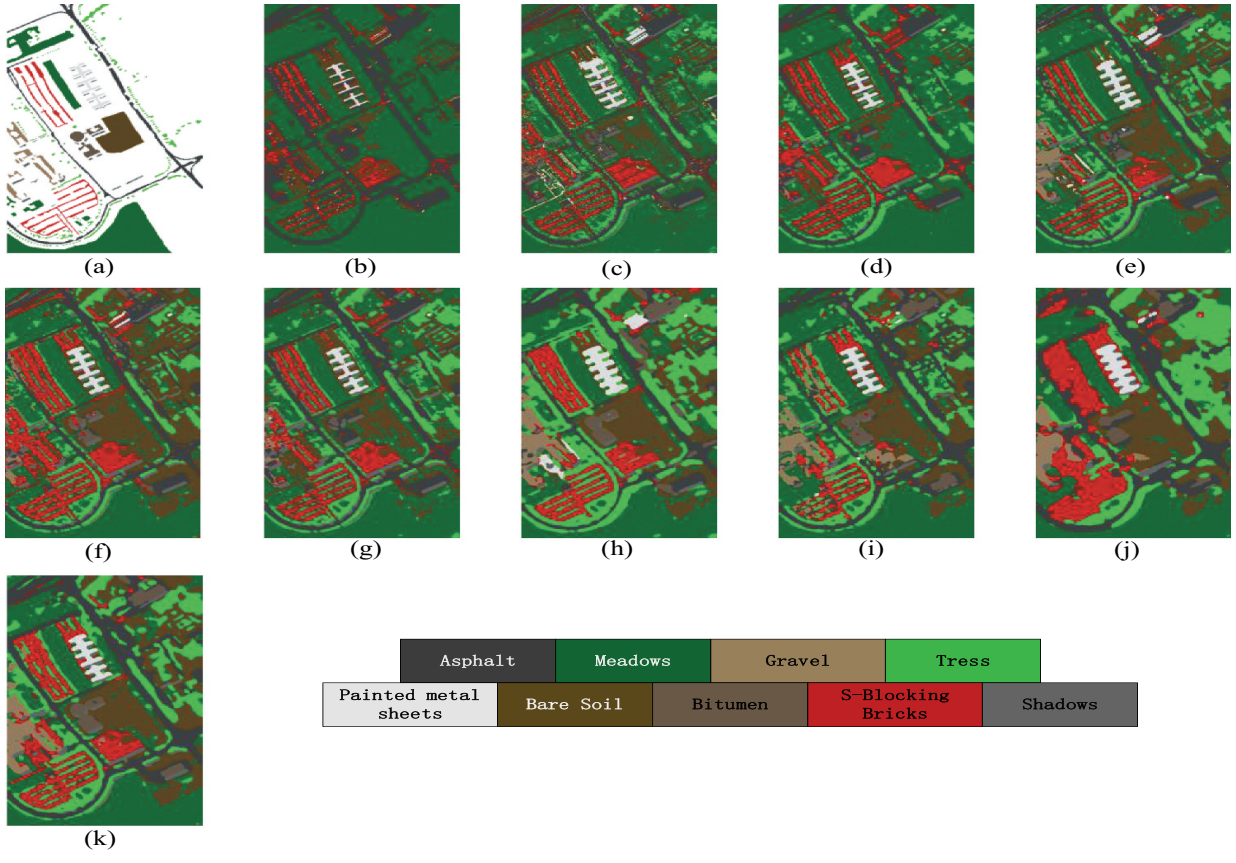


Figure 18. Comparison of classification maps on UP (the different colours in the figure correspond to different ground objects). (a) ground truth, (b–k) CDCNN, SVM, DBMA, SSRN, FDSSC, DBDA, FECNet, A2S2KResNet, FDGC, GDPA_LDSICS).

3.5.3. Comparison of classification results with different training sample proportions

In practical projects, marking the ground objects of data sets will consume huge resources. A high-performance network must be able to achieve accurate classification results with a small number of training samples. In order to prove that GDPA_LDSICS has efficient classification performance in small sample scenarios, this paper compares GDPA_LDSICS with different methods in different proportions of samples. As shown in Figure 20, the experiment sets the training sample proportions to 1%, 5%, 10%, and 20%, respectively. Experiments were conducted on 4 datasets and OA was used as an evaluation indicator. As can be seen from Figure 20, on different training sets, GDPA_LDSICS shows an upward trend in classification accuracy as the proportion of training samples increases. And GDPA_LDSICS has always shown optimal classification performance. However, other methods have a situation where the number of training samples increases but the classification accuracy decreases. It is proved that GDPA_LDSICS has better classification in the context of small samples. At the same time, it is also shown that GDPA_LDSICS has strong robustness and can adapt to different sample proportions.

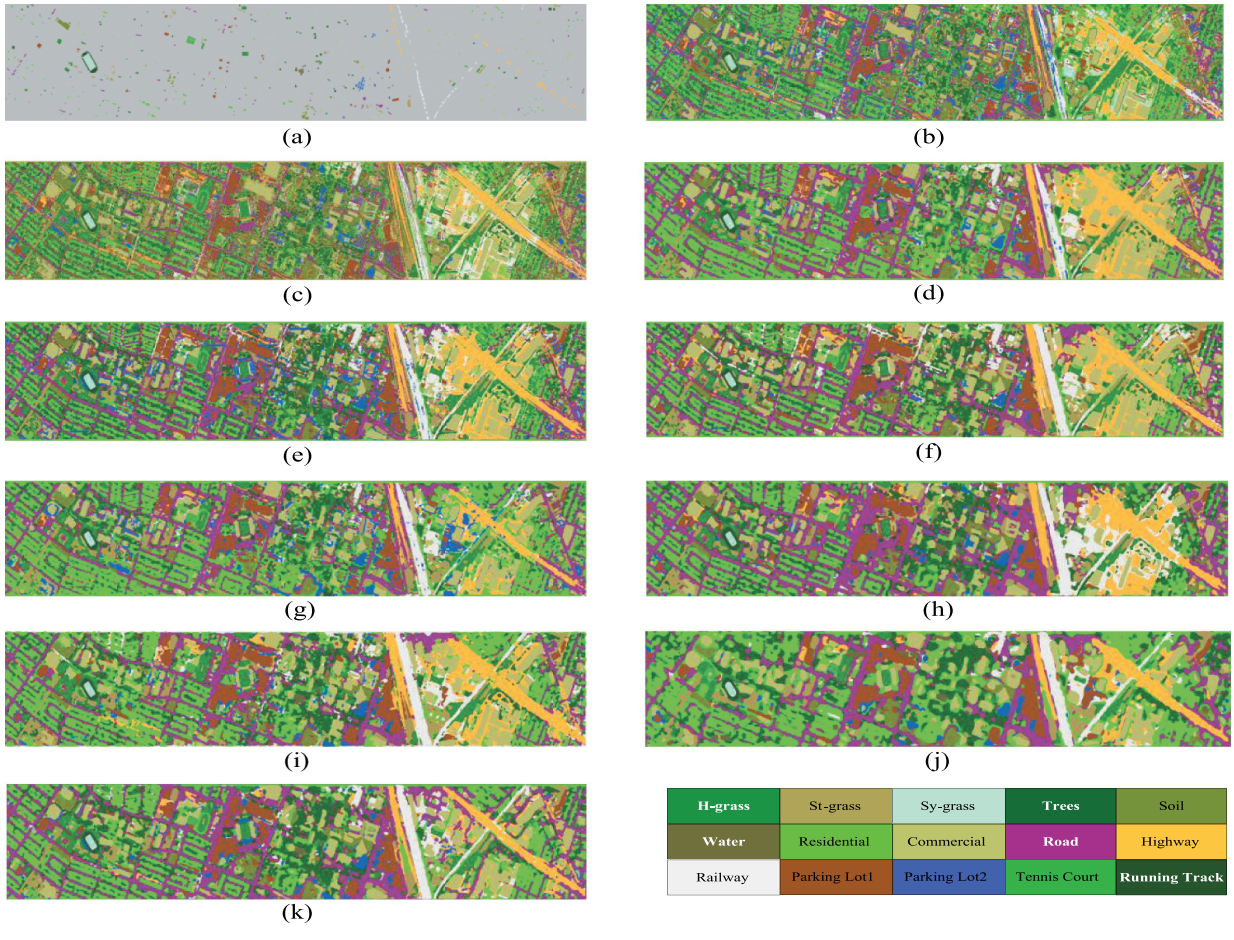


Figure 19. Comparison of classification maps on HT (the different colours in the figure correspond to different ground objects). (a) ground truth, (b–k) CDCNN, SVM, DBMA, SSRN, FDSSC, DBDA, FECNet, A2S2KResNet, FDGC, GDPA_LDSICS).

3.5.4. Comparison of T-SNE visualization

For the classification of hyperspectral images, different categories often interfere with each other. To demonstrate that GDPA_LDSICS can effectively distinguish different categories, this paper uses T-SNE (Vander Maaten and Hinton 2008) to visualize the classification results. As shown in Figures 21 and 22, the proposed method is compared with four methods with good performance on the more challenging SV and UP datasets. As can be seen from the results of Figures 21 and 22, all methods can obtain good clustering results. However, there are still cases of sample misclassification in some categories, which are better alleviated by the method proposed in this paper. For example, some other methods in Figure 21 perform poorly in the classification of 'Grapes-n' and 'VIN-yard-u', while the method proposed in this paper can better distinguish these two categories. This is because on the one hand, other methods do not solve the problem of interference of redundant information to classification. On the other hand, because of GDPA_LDSICS has better feature extraction ability than other methods. This also fully proves the effectiveness of the proposed method.

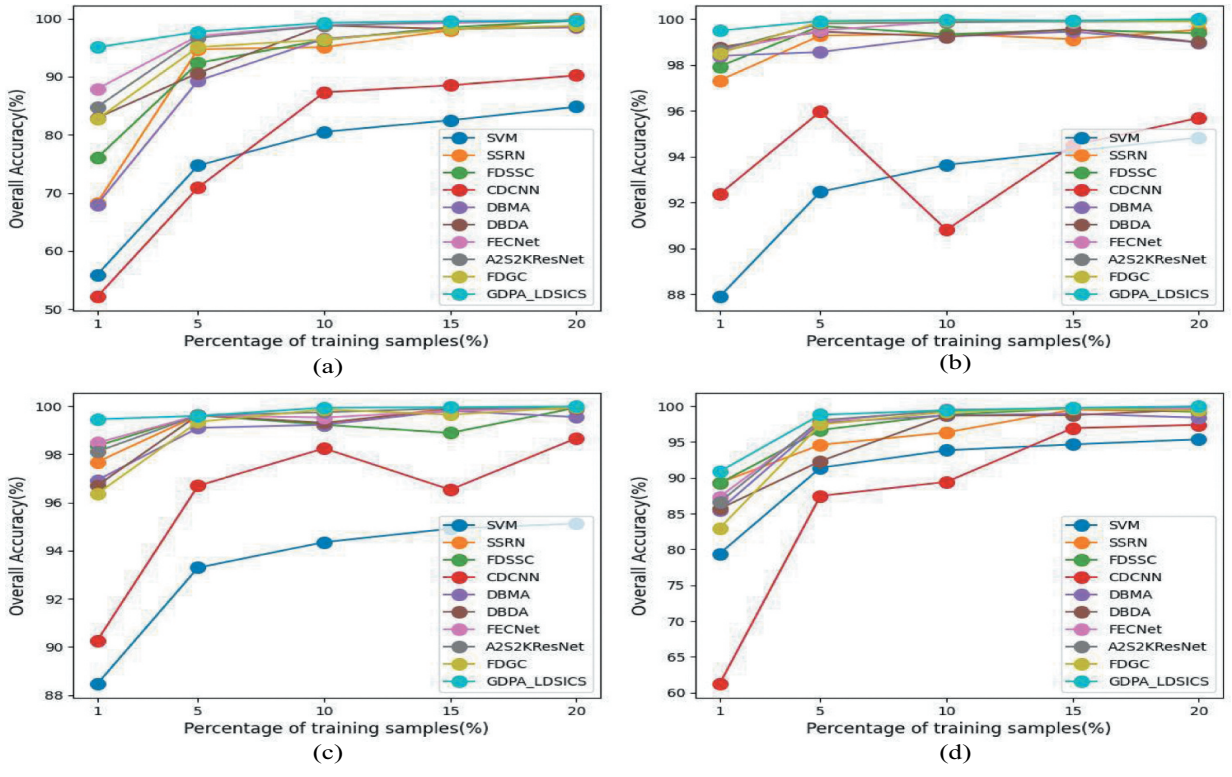


Figure 20. Classification performance of different methods with different training sample proportions on four data sets. (a–d) IN, SV, UP, HT.

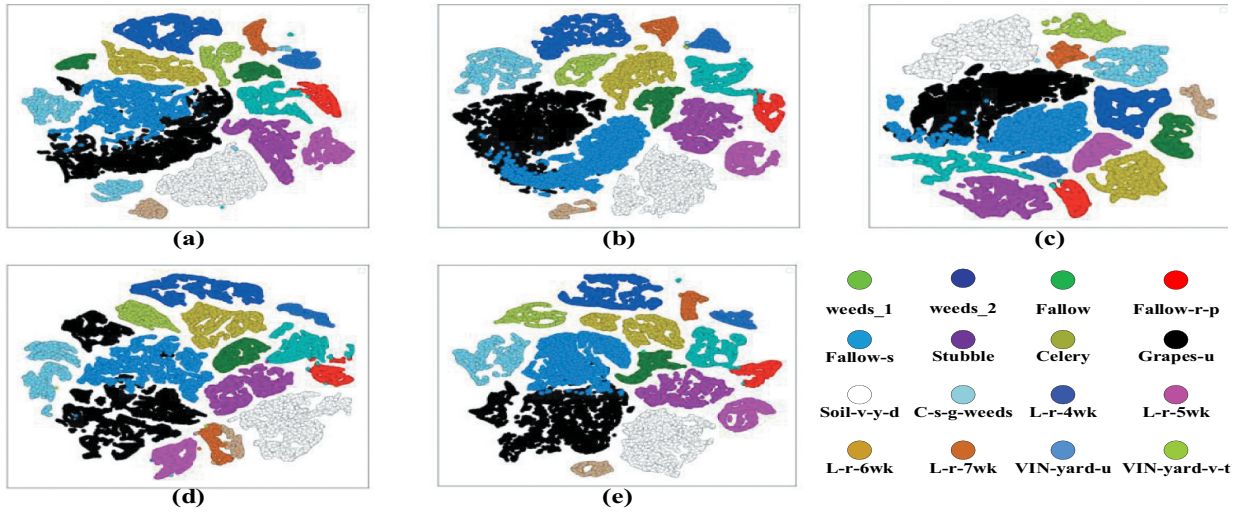


Figure 21. T-SNE visualization of classification results on SV dataset. (a)DBDA; (b)FECNet; (c) A2S2KResNet; (d)FDGC; (e)GDPA_LDSICS.

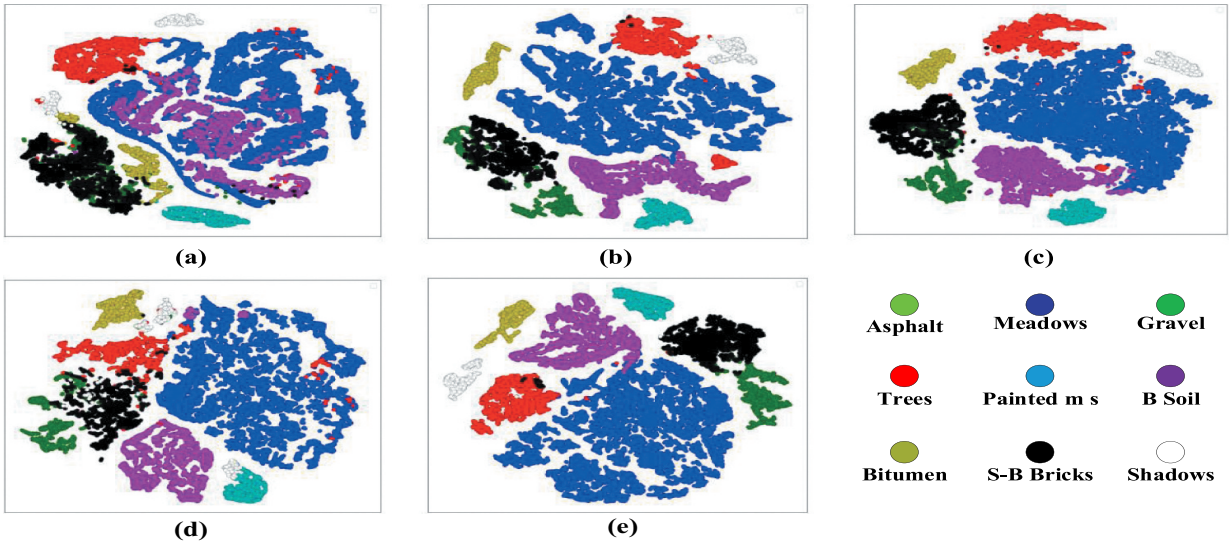


Figure 22. T-SNE visualization of classification results on UP dataset. (a)DBDA; (b)FECNet; (c)A2S2KResNet; (d)FDGC; (e)GDPA_LDSICS.

4. Conclusions

In this paper, in order to avoid the interference of redundant information on classification, and make full use of the spatial spectral joint features to achieve high performance hyperspectral image classification under the training set of small samples, a new GDPA_LDSICS is proposed. The GDPA_LDSICS mainly includes four parts. First, a LDSICS strategy is proposed to effectively eliminate highly redundant information, and map high-dimensional spectral information to low dimensional linear spatial to reduce computing costs. Then, a SSD module is proposed, which can effectively associate the spatial spectral information closely and realize the spatial spectral joint feature extraction. In addition, on CNN branch and GCN branch, DPA module and ADE module are proposed respectively. DPA can extract more complete context semantics and reduce the information loss during feature extraction. ADE module dynamically constructs adjacency matrix through self-correlation, which can enable GCN to more fully mine graph structure features. Finally, the cross fusion results of CNN and GCN are classified. A large number of experiments prove that the GDPA_LDSICS method is superior to some state-of-the-art methods in the case of small samples, both in terms of classification performance and robustness.

Acknowledgements

The authors would like to thank the editors and the reviewers for their help and suggestion.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded in part by the National Natural Science Foundation of China [42271409, 62071084], in part by the Heilongjiang Science Foundation Project of China under Grant LH2021D022, in part by the Leading Talents Project of the State Ethnic Affairs Commission, and in part by the Fundamental Research Funds in Heilongjiang Provincial Universities of China under Grant [145209150].

Data availability statement

The IN, SV and UP datasets are available online at http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. The HT dataset is available online at https://hyperspectral.ee.uh.edu/?page_id=459.

References

Chen, Y., H. Jiang, C. Li, X. Jia, P. Ghamisi. 2016. "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks." *IEEE Transactions on Geoscience & Remote Sensing* 54 (10): 6232–6251. https://doi.org/10.1109/TGRS.2016.2584107

Chen, Y., Z. Lin, X. Zhao, G. Wang, and Y. Gu. 2014. "Deep Learning- Based Classification of Hyperspectral Data." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 7 (6): 2094–2107. https://doi.org/10.1109/JSTARS.2014.2329330

Chen, Y., N. M. Nasrabadi, and T. D. Tran 2010. "Classification for Hyperspectral Imagery Based on Sparse Representation." In *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. Reykjavik. https://doi.org/10.1109/WHISPERS.2010.5594882

Chen, Y., X. Zhao, and X. Jia. 2015. "Spectral- Spatial Classification of Hyperspectral Data Based on Deep Belief Network." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 8 (6): 2381–2392. https://doi.org/10.1109/JSTARS.2015.2388577.

Chen, Y., K. Zhu, L. Zhu, X. He, P. Ghamisi, and J. A. Benediktsson. 2019. "Automatic Design of Convolutional Neural Network for Hyperspectral Image Classification." *IEEE Transactions on Geoscience & Remote Sensing* 57 (9): 7048–7066. https://doi.org/10.1109/TGRS.2019.2910603.

Cui, Y., Z. Yu, J. Han, S. Gao, and L. Wang. 2022. "Dual-Triple Attention Network for Hyperspectral Image Classification Using Limited Training Samples." *IEEE Geoscience and Remote Sensing Letters* 19 (5504705): 1–5. https://doi.org/10.1109/LGRS.2021.3067348.

Kang, J., R. Fernandez-Beltran, D. Hong, J. Chanussot, and A. Plaza. 2021. "Graph Relation Network: Modeling Relations Between Scenes for Multilabel Remote-Sensing Image Classification and Retrieval." *IEEE Transactions on Geoscience and Remote Sensing* 59 (5): 4355–4369. https://doi.org/10.1109/TGRS.2020.3016020.

Lee, H., and H. Kwon. 2017. "Going Deeper with Contextual CNN for Hyperspectral Image Classification." *IEEE Transactions on Image Processing* 26 (10): 4843–4855. https://doi.org/10.1109/TIP.2017.2725580 .

Liu, Q., Y. Dong, Y. Zhang, and H. Luo. 2022. "A Fast Dynamic Graph Convolutional Network and CNN Parallel Network for Hyperspectral Image Classification." *IEEE Transactions on Geoscience & Remote Sensing* 60:1–15. https://doi.org/10.1109/TGRS.2022.3179419.

Liu, Q., L. Xiao, J. Yang, and Z. Wei. 2021. "CNN-Enhanced Graph Convolutional Network with Pixel- and Superpixel-Level Feature Fusion for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 59 (10): 8657–8671. https://doi.org/10.1109/TGRS.2020.3037361.

Li, Z., X. Zhao, Y. Xu, W. Li, L. Zhai, Z. Fang, and X. Shi. 2022. "Hyperspectral Image Classification with Multiattention Fusion Network." *IEEE Geoscience & Remote Sensing Letters* 19 (5503305): 1–5. https://doi.org/10.1109/LGRS.2021.3052346

Li, R., S. Zheng, C. Duan, Y. Yang, and X. Wang. 2020. "Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network." *Remote Sensing* 12 (3): 582. https://doi.org/10.3390/rs12030582 .

Malthus, T. J., and P. J. Mumby. 2003. "Remote sensing of the coastal zone: an overview and priorities for future research." *International journal of remote sensing* 24 (13): 2805–2815.

Ma, W., Q. Yang, Y. Wu, W. Zhao, and X. Zhang. 2019. "Double-Branch Multi-Attention Mechanism Network for Hyperspectral Image Classification." *Remote Sensing* 11 (11): 1307. https://doi.org/10.3390/rs11111307.

Melgani, F., and L. Bruzzone. 2004. "Classification of Hyperspectral Remote Sensing Images with Support Vector Machines." *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society* 42 (8): 1778–1790. https://doi.org/10.1109/TGRS.2004.831865 .

Nakai, K., T. Matsubara, and K. Uehara, 2020. "Att-DARTS: Differentiable Neural Architecture Search for Attention." In *International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN48605.2020.9207447.

Pal, M. 2008. "Ensemble of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 29 (10): 3043–3049. https://doi.org/10.1080/01431160802007624.

Paoletti, M. E., J. M. Haut, J. Plaza, and A. Plaza. 2018. "Deep&dense Convolutional Neural Network for Hyperspectral Image Classification." *Remote Sensing* 10 (9): 1454. https://doi.org/10.3390/rs10091454 .

Roy, S. K., S. Manna, T. Song, and L. Bruzzone. 2021. "Attention-Based Adaptive Spectral–Spatial Kernel ResNet for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 59 (9): 7831–7843. https://doi.org/10.1109/TGRS.2020.3043267.

Shi, C., D. Liao, T. Zhang, and L. Wang. 2022. "Hyperspectral Image Classification Based on Expansion Convolution Network." *IEEE Transactions on Geoscience and Remote Sensing* 60 (5528316): 1–16. https://doi.org/10.1109/TGRS.2022.3174015.

Sun, H., X. Zheng, X. Lu, and S. Wu. 2020. "Spectral–Spatial Attention Network for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society* 58 (5): 3232–3245. https://doi.org/10.1109/TGRS.2019.2951160.

Vander Maaten, L., and G. Hinton. 2008. "Visualizing data using t-SNE." *Journal of Machine Learning Research* 9 (86): 2579–2605. http://www.jmlr.org/papers/v9/vandermaaten08a.html.

Wang, W., S. Dou, Z. Jiang, and L. Sun. 2018. "A Fast Dense Spectral–Spatial Convolution Network Framework for Hyperspectral Images Classification." *Remote Sensing* 10 (7): 1068. https://doi.org/10.3390/rs10071068 .

Wan, S., C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang. 2020. "Multiscale Dynamic Graph Convolutional Network for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 58 (5): 3162–3177. https://doi.org/10.1109/TGRS.2019.2949180.

Wang, T., G. Wang, K. E. Tan, and D. Tan. 2020. "Spectral Pyramid Graph Attention Network for Hyperspectral Image Classification." arXiv. https://doi.org/10.48550/arXiv.2001.07108.

Yang, G., U. B. Gewali, E. Ientilucci, M. Gartley, and S. T. Monteiro. 2018. "Dual-Channel Dense Net for Hyperspectral Image Classification." In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2595–2598. https://doi.org/10.1109/IGARSS.2018.8517520.

Ying, L., H. Zhang, and Q. Shen. 2017. "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network." *Remote Sensing* 9 (1): 67. https://doi.org/10.3390/rs9010067.

Yuen, P. W., and M. Richardson. 2010. "An Introduction to Hyperspectral Imaging and Its Application for Security, Surveillance and Target Acquisition." *Imaging Science Journal* 58 (5): 241–253. https://doi.org/10.1179/174313110X12771950995716.

Yu, C., R. Han, M. Song, C. Liu, and C. I. Chang. 2021. "Feedback Attention-Based Dense CNN for Hyperspectral Image Classification." *IEEE Transactions on Geoscience & Remote Sensing* 60 (5501916): 1–16. https://doi.org/10.1109/TGRS.2020.3040273.

Zhang, X., S. Chen, P. Zhu, X. Tang, J. Feng, and L. Jiao. 2022. "Spatial Pooling Graph Convolutional Network for Hyperspectral Image Classification." *IEEE Transactions on Geoscience & Remote Sensing* 60 (5521315): 1–15. https://doi.org/10.1109/TGRS.2022.3140353.

Zhang, H., H. Yu, Z. Xu, K. Zheng, and L. Gao. 2021. "A Novel Classification Framework for Hyperspectral Image Classification Based on Multi-Scale Dense Network." In 2021 IEEE *International Geoscience and Remote Sensing Symposium IGARSS*, 2238–2241. https://doi.org/10.1109/IGARSS47720.2021.9555010.

Zhao, C., B. Qin, T. Li, S. Feng, and Y. Yan. 2021. "Hyperspectral Image Classification Based on Dense Convolution and Conditional Random Field." In *2021 IEEE International Geoscience and Remote SensINg Symposium IGARSS*, 3653–3656. https://doi.org/10.1109/IGARSS47720.2021.9554062.

Zheng, X., B. Wang, X. Du, X. Lu, and B. Luo. 2022. "Mutual Attention Inception Network for Remote Sensing Visual Question Answering." *IEEE Transactions on Geoscience & Remote Sensing* 60 (5606514): 1–14. https://doi.org/10.1109/TGRS.2021.3079918.

Zhong, Z., J. Li, Z. Luo, and M. Chapman. 2018. "Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework." *IEEE Transactions on Geoscience and Remote Sensing* 56 (2): 847–858. https://doi.org/10.1109/TGRS.2017.2755542.

Zhu, M., L. Jiao, F. Liu, S. Yang, and J. Wang. 2021. "Residual Spectral–Spatial Attention Network for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 59 (1): 449–462. https://doi.org/10.1109/TGRS.2020.2994057.

报告编号：**L24Y2023-0423**

检　索　报　告

一、检索要求

1. 委 托 人：石翠萍 Shi, CP (Shi, Cuiping)
2. 委托单位：齐齐哈尔大学
3. 检索目的：论文被 SCI-E 收录情况

二、检索范围

Science Citation Index Expanded (SCI-EXPANDED)	1990-present	网络版
JCR-(Journal Citation Reports)	2022	网络版
中国科学院文献情报中心期刊分区表(升级版)	2022	网络版

三、检索结果

委托人提供的**1**篇论文被SCI-E收录，论文被收录、所在期刊的JCR影响因子、中科院期刊分区(升级版)情况见附件一。

特此证明！

检索报告人：杨茗惠

东北师范大学科技查新咨询中心

教育部科技查新工作站(L24)

2023 年 **9** 月 **12** 日

附件一：SCI-E收录情况

标题: GDPA_LDSICS: graph and double pyramid attention network based on linear discrimination of spectral interclass slices for hyperspectral image classification

作者: Wu, HY (Wu, Haiyang); **Shi, CP (Shi, Cuiping)**; Wang, LG (Wang, Liguo)

来源出版物: INTERNATIONAL JOURNAL OF REMOTE SENSING 卷: 44 期: 17 页: 5283-5312 DOI: 10.1080/01431161.2023.2247523 出版年: SEP 2 2023

Web of Science 核心合集中的 "被引频次": 0

被引频次合计: 0

使用次数 (最近 180 天): 0

使用次数 (2013 年至今): 0

引用的参考文献数: 35

摘要: In recent years, convolution neural networks (CNNs) and graph convolution networks (GCNs) have been widely used in hyperspectral image classification (HSIC). CNNs can effectively extract the spatial spectral features of hyperspectral images (HSIs), while GCNs can quickly capture the structural features of HSIs, which makes the effective combination of the two is beneficial to improve classification performance of hyperspectral images. However, the high redundancy of feature information and the problem of small sample are still the major challenges of HSIC. In order to alleviate these problems, in this paper, a new graph and double pyramid attention network based on linear discrimination of spectral interclass slices (GDPA_LDSICS) is proposed. First, a linear discrimination of spectral inter class slices (LDSICS) module is designed. The LDSICS module can effectively eliminate a lot of redundancy in spectral dimension, which is conducive to subsequent feature extraction. Then, the spatial spectral deformation (SSD) module is constructed, which can effectively correlate the spatial spectral information closely. Finally, in order to alleviate the problem of small sample, a double branch structure of CNN and GCN is developed. On the CNN branch, a double pyramid attention (DPA) structure is designed to model context semantics to avoid information loss caused by long-distance feature extraction. On the GCN branch, an adaptive dynamic encoding (ADE) method is proposed, which can more effectively capture the topological structure of spatial spectral features. Experiments on four open datasets show that the GDPA_LDSICS can provide better classification performance and generalization performance than other most advanced methods.

入藏号: WOS:001056161400001

语言: English

文献类型: Article

作者关键词: Hyperspectral image classification; convolutional neural networks; graph convolutional networks

地址: [Wu, Haiyang; Shi, Cuiping] Qiqihar Univ, Coll Elect & Commun Engn, Qiqihar 161000, Peoples R China.

[Wang, Liguo] Dalian Nationalities Univ, Coll Informat & Commun Engn, Dalian, Peoples R China.

通讯作者地址: Shi, CP (通讯作者)，Qiqihar Univ, Coll Elect & Commun Engn, Qiqihar 161000, Peoples R China.

电子邮件地址: shicuiping@qqhru.edu.cn

Affiliations: Qiqihar University; Dalian Minzu University

出版商: TAYLOR & FRANCIS LTD

出版商地址: 2-4 PARK SQUARE, MILTON PARK, ABINGDON OR14 4RN, OXON, ENGLAND

Web of Science Index: Science Citation Index Expanded (SCI-EXPANDED)

Web of Science 类别: Remote Sensing; Imaging Science & Photographic Technology

研究方向: Remote Sensing; Imaging Science & Photographic Technology

IDS 号: Q2WD8

ISSN: 0143-1161

eISSN: 1366-5901

29 字符的来源出版物名称缩写: INT J REMOTE SENS

ISO 来源出版物缩写: Int. J. Remote Sens.

来源出版物页码计数: 30

基金资助致谢:
基金资助机构 授权号
The authors would like to thank the editors and the reviewers for their help and suggestion.
The authors would like to thank the editors and the reviewers for their help and suggestion.

输出日期: 2023-09-12

期刊影响因子 ™ 2022：3.4

中国科学院文献情报中心期刊分区(升级版, 2022)截图如下:

INTERNATIONAL JOURNAL OF REMOTE SENSING

刊名	INTERNATIONAL JOURNAL OF REMOTE SENSING
年份	2022
ISSN	0143-1161
Review	否
Open Access	否
Web of Science	SCIE

	学科	分区	Top期刊
大类	工程技术	3	否
小类	IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY 成像科学与照相技术	3	
	REMOTE SENSING 遥感	3	

The End