





# Dual Graph U-Nets for Hyperspectral Image Classification

Fangming Guo , Zhongwei Li , Ziqi Xin, Xue Zhu, Leiquan Wang , *Member, IEEE*, and Jie Zhang 

**Abstract**—Graph convolutional neural networks (GCNs) have been widely used in hyperspectral images (HSIs) classification for their superiority in processing non-Euclidean structure data. The performance of GCNs relies on the initial graph structure. Most GCN models only utilize spectral information to construct a graph, which is inaccurate because they fail to take the relationship between adjacent nodes into consideration. In addition, due to the over-smooth phenomenon, most GCN models are shallow and unable to extract effective features. To address these issues, a dual graph u-nets is proposed by integrating spatial graph and spectral graph for HSIs classification, denoted by DGU-HSI. To integration the spectral and spatial information, two graphs are constructed for feature extraction. The spectral graph is created by spectral similarity among all pixels where multirange spectral information is retained, and the spatial graph is constructed by exploiting the neighborhood relationship of the center pixel, which describes spatial information. Then, a dual GCN is utilized to extract spatial and spectral graph features simultaneously. To relieve the over-smooth phenomenon, the graph u-nets structure is applied on constructed spectral and spatial graph to extract effective features. Finally, the extracted spectral and spatial features are fused for classification. Experiments conducted on the public datasets demonstrate the effectiveness of the proposed method on HSIs classification.

**Index Terms**—Graph convolutional networks(GCN), hyperspectral image (HSI) classification, spectral-spatial fusion, spectral-spatial graph.

## I. INTRODUCTION

**H**YPERSPECTRAL images (HSIs) contain hundreds of continuous and narrow spectral channels with rich spatial and spectral information. Emerging as one of the most rapid developments in earth science and remote sensing field, HSIs classification aims to assign each pixel to a set of land-use/land-over classes with a classifier. In the past decades, HSIs classification has attracted extensive attention in atmospheric environment research [1], ocean remote sensing [2], environmental

monitoring [3], [4], urban planning [5], [6], and so on. However, HSIs classification still faces numerous challenges, particularly in dealing with the influences of high spectral similarity and spectral variability among materials.

In the early stage of HSIs classification, several conventional pattern recognition methods have been studied, such as k-nearest neighbor classifier and linear classifier [7]. Meanwhile, better classifiers from machine learning, such as e support vector machine (SVM) [8] and random forest [9], have also been applied in HSIs classification to obtain satisfactory classification performance. Notably, these methods employ spectral information of a single pixel to determine its class label and have the advantage of conceptual simplicity and easy implementation. Nevertheless, given the phenomenon that the same objects may have different spectral reflectance and different objects share the same spectra characteristics, it is difficult to get satisfactory classification performance from spectral data alone. Besides, the single pixel spectral data have limited feature representation ability for lacking effective data fitting ability. Subsequently, because spatially neighboring pixels usually carry correlated information, the spatial context is regarded as naturally helpful information for HSIs classification. For example, Markov random field (MRF)-based models [10] and morphological profile-based methods [11] have been employed to integrate spectral and spatial features.

The aforementioned methods adopt a series of handcrafted spectral-spatial features that rely heavily on professional expertise and parameter setting. However, they fail to fully represent spectral and spatial information. Recently, the deep learning methods [12]–[16] have displayed satisfactory performance in HSIs classification for their ability to extract higher quality features automatically. The stack auto-encoder obtained the high-level feature, the first attempt to classify HSIs with deep learning methods [17]. With the development of computer vision, the convolutional neural network (CNN) has been regarded as one of the most effective methods to deal with image processing problems. Therefore, more attention has been paid to HSIs classification methods based on CNN [18]–[21]. Hu *et al.* [22] employed CNN to extract features for HSIs classification, whose performance is better than that of SVM. Li *et al.* [23] developed a 1-D CNN with the pixel-pair features strategy to learn spectral features, which boasted a robust classification performance. In addition, CNN-based methods have shown great potential for integrating spectral and spatial features. For instance, Lee *et al.* [24] established a 2-D CNN similar to AlexNet to extract the spectral and spatial features using a multiscale windows sensing field. Pan *et al.* [25] proposed a multigrained network to

Manuscript received June 1, 2021; revised July 27, 2021; accepted August 7, 2021. Date of publication August 10, 2021; date of current version August 30, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62071491, in part by the Joint Funds of the National Natural Science Foundation of China under Grant U1906217, and in part by the Fundamental Research Funds for the Central Universities under Grant 19CX05003A-3. (*Corresponding author: Zhongwei Li.*)

Fangming Guo, Zhongwei Li, and Jie Zhang are with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China (e-mail: 492478392@qq.com; lizhongwei@upc.edu.cn; zhangjie@fio.org.cn).

Ziqi Xin, Xue Zhu, and Leiquan Wang are with the College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China (e-mail: 17806281107@163.com; s19070024@s.upc.edu.cn; richiewlq@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2021.3103744

extract the joint spectral-spatial information and combine different grains spectral and spatial relationship for HSIs classification.

Although the CNN-based methods have demonstrated great capability in HSIs classification [26]–[28], there are still some drawbacks. A common feature of these methods is that CNN can only be conducted on data with a grid-like structure. Consequently, they cannot naturally catch the geometric characteristics variations of different regions in HSIs. Besides, due to the spectral similarity and spectral variability among pixels, the relevant characters of spectral cannot be extracted adequately with a fixed convolutional kernel. Therefore, the convolutional kernel with fixed size and weights is not suitable for all the regions in HSIs. To address this issue, graph convolutional network (GCN) has been applied to HSIs classification [29]–[31], which can aggregate and transform feature information from the neighbors of each graph node [32]–[34]. Qin *et al.* [35] developed a semisupervised GCN-based network that explicitly utilized the adjacency nodes in the graph to approximate the convolution. Nevertheless, owing to the well-known over-smooth phenomenon of GCN-based methods, the nodes may converge to the same state with the increase of network layers, which limits the depth of GCN to extract discriminative features. Therefore, the increase of the depth of the GCN network to improve the feature extracting ability remains challenging. Furthermore, most GCN methods perform semisupervised node classification on graphs. With the increase of nodes, the required huge computing resources will limit the application of these methods. Therefore, scaling down computing resources is another challenge for GCN in the HSIs classification.

To tackle the problems mentioned above, a dual graph u-nets to integrate spatial graph information and spectral graph feature was built for HSIs classification, denoted by DGU-HSI. Images divided by original HSIs were applied in HSIs classification to reduce the computation cost. Then, the spectral graph was constructed based on the spectral similarity among all pixels, and the spatial graph was designed based on the neighbor pixel similarity of the central pixel. For feature extraction, the dual graph structure was deployed to acquire spectral and spatial features separately, and a u-nets structure was applied to extract effective features to improve the network depth. Finally, the spectral and spatial features were fused and sent to full connection for HSIs classification. To summarize, the main contributions of this article are as follows.

- 1) To solve the problem that node-based classification methods require a high computational cost, a graph-based method is proposed for HSIs classification.
- 2) Spectral graph and spatial graph are constructed to resolve the inaccuracy of spectral graph. The spectral graph retains the spectral similarity among pixels, while the spatial graph contains the neighboring information of the central pixel.
- 3) A graph u-net is employed to reduce the influence of over-smooth, which utilizes the pooling and unpooling layers and attention mechanism to acquire effective feature extraction ability.

- 4) Experimental results conducted on the public datasets demonstrate the effectiveness of the proposed DGU-HSI, which achieves competitive performances based on existing methods.

The rest of this article is organized as follows. In Section II, the related work of proposed DGU-HSI is expounded. Section III describes the details of proposed method DGU-HSI. The experimental results and analysis are shown in Section IV. Finally, Section V concludes this article.

## II. RELATED WORK

### A. Graph Convolutional Network

A graph is a non-Euclidean structure Data that is applied to introduce the nonlinear relationship of data nodes. In this study, the relations of spectral and spatial information can be represented as an undirected graph. Consider an undirected graph as  $G = (V, E)$ , in which  $V$  and  $E$  denote the vertex and edges sets of the graph. Besides, the adjacency matrix is denoted as  $A \in \mathbb{R}^{N \times N}$  ( $N$  is the number of vertexes), which contains the relationships between vertexes and can also be regarded as the matrix representation of edge. The corresponding Laplacian matrix  $L$  of a graph  $G$  can be defined as

$$L = D - A \quad (1)$$

where  $D$  is the degree matrix of  $A$ ,  $D$  is a diagonal matrix and can be calculated by  $D_{ii} = \sum_{j=1}^N A_{ij}$ .

Generally, the above mentioned Laplacian matrix is called Combinatorial Laplacian. To improve the generalization ability of the graph, the Symmetric normalized Laplacian is proposed as follows:

$$\begin{aligned} L^{sym} &= D^{-1/2} L D^{-1/2} \\ &= D^{-1/2} (D - A) D^{-1/2} \\ &= I - D^{-1/2} A D^{-1/2} \end{aligned} \quad (2)$$

where  $I$  is an identity matrix.

Given a graph signal  $f \in R^l$  (the feature representation of vertexes) and a filter  $g_\theta = \text{diag}(\theta)$  parameterized by  $\theta \in R^l$ , the convolution between  $f$  and  $g$  on a graph can be formulated as

$$g_\theta \star f = U g_\theta U^T f \quad (3)$$

where  $\star$  is the operation of graph convolution,  $U$  is the matrix of eigenvectors of  $L^{sym}$  and can be obtained by eigen-decomposition, i.e.,  $L^{sym} = U \Lambda U^T$ .  $\Lambda$  is the diagonal matrix of eigenvalues of  $L^{sym}$ .  $g_\theta$  is the convolutional kernel of graph convolution and can be treated as the function of the eigenvalues of, i.e.,  $g_\theta(\Lambda)$ .

However, (3) has a high computational complexity because it involves eigen decomposition of Laplacian matrix and a large number of matrix calculations. Therefore, the Chebyshev polynomials are deployed to approximate the filter  $g_\theta$ . Hammond *et al.* [36] suggested that the function of the eigenvalues  $g_\theta(\Lambda)$

can be fitted by Chebyshev polynomials  $T_k(x)$  up to  $K$ th order

$$g_{\theta} \approx \sum_{k=0}^K \theta' T_k(\tilde{\Lambda}) \quad (4)$$

where  $\theta' \in \mathbb{R}^K$  is the coefficients of Chebyshev polynomials,  $\tilde{\Lambda} = \frac{2\lambda}{\lambda_{max}} - I$ ,  $\lambda_{max}$  is the largest eigenvalue of  $L^{sym}$ ,  $\tilde{\Lambda}$  is the rescaled diagonal matrix of eigenvalues of  $L^{sym}$ . The recursive Chebyshev polynomials are defined as  $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ ,  $T_0(x) = 1$ , and  $T_1(x) = x$ . Therefore, the graph convolution can be defined as follows:

$$\begin{aligned} g_{\theta} \star f &\approx U \sum_{k=0}^K \theta' T_k(\tilde{\Lambda}) U^T f \\ &= \sum_{k=0}^K \theta' T_k(U \tilde{\Lambda} U^T) f \\ &= \sum_{k=0}^K \theta' T_k(\tilde{L}) f \end{aligned} \quad (5)$$

where  $\tilde{L} = \frac{2L^{sym}}{\lambda_{max}} - I$  denotes the normalized  $L^{sym}$ .

Kipf and Welling [37] limited  $K = 1$  and assigned the largest eigenvalue  $\lambda_{max}$  of  $\tilde{L}$  to 2. By doing so, (5) can be rewritten as

$$g_{\theta} \star f \approx \theta(I + D^{-1/2} A D^{-1/2}) f \quad (6)$$

Using (6), the following propagation rule for GCN:

$$H^{l+1} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{1/2} H^l W^l) \quad (7)$$

where  $\tilde{A} = A + I$  and,  $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$  are the renormalization terms of  $A$  and  $D$ .  $H^l$  is the feature of GCN in  $l$ th layer and  $W^l$  represents the weights of  $l$ th layer.  $\sigma(\cdot)$  is the activation function (i.e., ReLU).

### B. Graph Pooling and Unpooling

Graph u-nets contains two important parts: graph pooling layer and graph unpooling layer [38]. In graph pooling layer, the information all nodes retaining is measured as follows:

$$y = X^l P^l / \|P^l\| \quad (8)$$

where  $P \in \mathbb{R}^l$  is a trainable projection vector,  $X$  is the feature matrix of all nodes,  $\|\cdot\|$  is employed to get Euclidean norm of  $P$ .

Using  $y$ , the process of graph pooling can be defined as

$$idx = \text{rank}(y, k) \quad (9)$$

$$\tilde{y} = \text{sigmoid}(y(idx)) \quad (10)$$

$$\tilde{X}^l = X^l(idx, :) \quad (11)$$

$$A^{l+1} = A^l(idx, idx) \quad (12)$$

$$X^{l+1} = \tilde{X}^l \odot (\tilde{y} 1_C^T) \quad (13)$$

where  $\text{rank}(\cdot)$  is the operation of ranking, and  $\text{rank}(y, k)$  will returns indices of the  $k$ -largest values in  $y$ .  $idx$  is the indices of nodes in the current graph for the new graph. Equations (11)–(13), is utilized to select the new feature matrix and the new

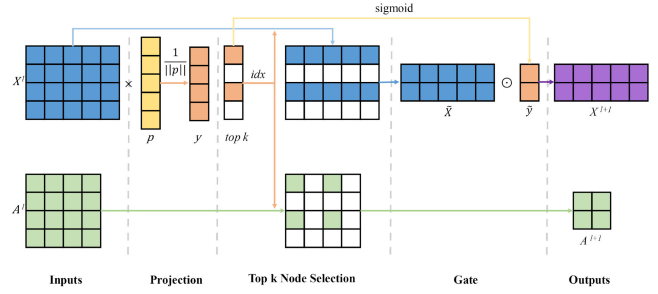


Fig. 1. Illustration of the operation of graph pooling.

adjacency matrix.  $1_C \in \mathbb{R}^c$  is a vector of size  $C$  with all components being 1, and  $\odot$  is the element-wise matrix multiplication. Fig. 1 illustrates the operation of graph pooling.

In graph unpooling layer, the layer-wise propagation rule of feature matrix is proposed as follows:

$$X^{l+1} = \text{distribute}(0_{N \times C}, X^l, idx) \quad (14)$$

where  $0_{N \times C}$  is the initially empty feature matrix for the new graph, whose size is the same as that of  $X^{l+1}$ .  $\text{distribute}(\cdot)$  uses  $X^l$  to update the initial feature matrix  $0_{N \times C}$  by the row index of  $idx$ , and other row vectors remain zero.

In addition, because the u-nets structure needs to concatenate two same size graphs, the adjacency matrix of unpooling layer is the same size as the adjacency matrix of the corresponding pooling layer.

## III. METHOD

In this section, the proposed DGU-HIS is elaborated in details that contains three blocks. The overall framework of DGU-HIS is illustrated in Fig. 2. The first block is applied to build the graph structure data of spectrum and spatial, the second one is implemented to extract the spectral and spatial feature, and the third is applied to integrate the spectral and spatial features for HSIs classification.

### A. Graph Representation for Hyperspectral Images

The spectral and spatial information contained in HSIs play an important role in HSIs classification. The spectral and spatial graph will be designed by the graph representation method. Assume  $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{1 \times 1 \times D}$ , where  $D$  is the spectral dimensions of original HSIs and  $n$  is the number of pixels that have a ground truth label, namely  $Y = \{y_1, y_2, \dots, y_n\} \in \mathbb{R}^1$ . For the set of  $X$ , an operation of dimensionality reduction is performed to reduce the redundant of spectral data. Therefore, the set  $X$  can be represented as  $X = \{x'_1, x'_2, \dots, x'_n\} \in \mathbb{R}^{1 \times 1 \times d}$ , where  $d$  is the new dimension of each pixel and  $d$  is far less than  $D$ . For each pixel  $x'_i$ , the image patch  $x_i^*$  with a size of  $w \times w \times d$  ( $w$  is the window size) are extracted, where  $x'_i$  is its center pixel, and the ground truth of the image patch is  $y_i$ . For each image patch  $x_i^*$ , two undirected graphs are designed to represent spatial and spectral graph respectively. A spatial graph is defined as  $G_{spa} = (V_{spa}, E_{spa})$ , in which  $V_{spa}$  is composed of the pixels in the image patch  $x_i^*$  and  $E_{spa}$  is the spatial

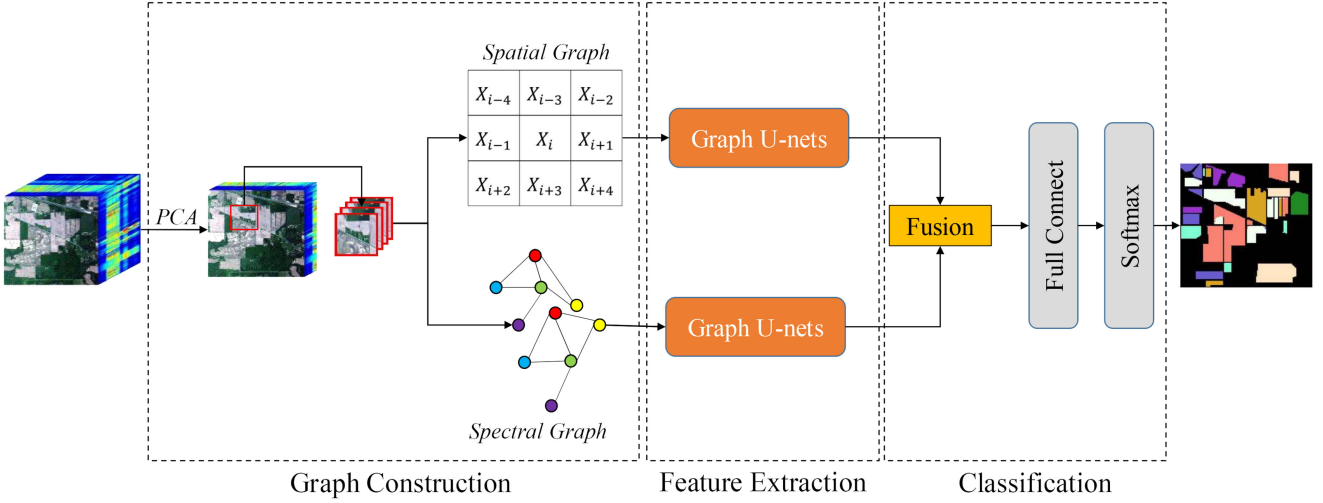


Fig. 2. Overall frameworks of DGU-HSI.

similarity distance between each pixel and its neighbors. The spatial similarity distance can be calculated as

$$E_{ij}^{spa} = e^{-\lambda \|x'_i - x'_j\|^2}, \text{ if } x'_j \in \text{Nei}(x'_i) \text{ or } x'_i \in \text{Nei}(x'_j) \quad (15)$$

where  $\text{Nei}(x)$  is the neighbors of  $x$ , the parameter  $\lambda$  is empirically set to 0.2 in the experiments.

The notation  $A^{spa}$  denotes the adjacency matrix of  $G_{spa}$ , which can be calculated as

$$A_{ij} = \begin{cases} E_{ij}^{spa}, & \text{if } x'_j \in \text{Nei}(x'_i) \text{ or } x'_i \in \text{Nei}(x'_j) \\ 1, & i = j \\ 0, & \text{others.} \end{cases} \quad (16)$$

A spectral graph is defined as  $G_{spe} = (V_{spe}, E_{spe})$ , where  $V_{spe}$  is the pixels in the image patch  $x'_i$ ,  $E_{spe}$  is the spectral similarity distance between each pixel and other pixels, which can be calculate by (17)

$$E_{ij}^{spe} = e^{-\lambda \|x'_i - x'_j\|^2}. \quad (17)$$

The notation  $A^{spe}$  denotes the adjacency matrix of  $G_{spe}$ , which can be calculated as

$$A_{ij} = \begin{cases} E_{ij}^{spe}, & \text{rank}(E_{ij}^{spe}) \geq k \\ 1, & i = j \\ 0, & \text{others} \end{cases} \quad (18)$$

where  $\text{rank}(E_{ij}^{spe})$  is the descending order of the spectral similarity distance of  $x'_i$ ,  $\text{rank}(E_{ij}^{spe}) \geq k$  represents the most relevant  $k$  pixels of  $x'_i$ .

Different from the traditional CNN, the spatial graph measures the similarity between the central pixel and its contiguous pixels, which can provide fine spatial relation information. Compared with the spectral data of original HSIs, the spectral graph contains multirange spectral information. The pixel close to the calculated pixel is called a short-range pixel, and the pixel far from the calculated pixel is defined as a long-range pixel. By calculating different range spectral information of calculated pixel, the multi-range spectral information can be extracted and

relieve the defect of fixed CNN kernel receptive field in feature extraction.

### B. Dual Graph U-Nets

The dual graph u-nets is composed of a feature extraction model and an attention mechanism model for acquiring the spectral and spatial graph features. The spectral and spatial graph u-nets are applied by the same network structure. The detailed information of a single graph u-nets in dual graph u-nets is illustrated in Fig. 3. In the feature extraction model, the input of spectral and spatial feature extraction model is spectral and spatial graph data, respectively. The graph convolutional layer is utilized to extract spectral and spatial features. The traditional GCN obtains information by transmitting the adjacent nodes feature, thereby acquiring effective shallow feature by one or two layers. However, HSIs contain complex spectral information, and shallow features are not enough for HSIs classification. To get abundant information, a deeper GCN is required. However, this transfer characteristic of over smooth of deeper GCN limits the development of GCN. Therefore, the graph pooling layers are adopted to reduce the number of parameters in the network and improve the generalization ability of networks. Meanwhile, the graph pooling layers can retain the nodes with abundant key information and remove the invalid nodes.

The input graph is defined as  $G_{ip} = (V_{ip}, E_{ip})$ , and the outputs of a graph convolution layer are convolved features of each node. The rule of graph convolution in DGU-HIS is as follows:

$$X^{l+1} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{1/2} X^l W^l) \quad (19)$$

where  $\tilde{A} = A + I$ ,  $A$  is the adjacent matrix of  $G_{ip}$ ,  $D$  is the degree matrix of  $\tilde{A}$ ,  $X^l$  is the feature matrix of layer  $l$ , and  $W$  is the weights to-be-learned.

After a series of graph convolution and graph pooling layers, graph features  $\{G_i = (V_i, E_i), i = 1, 2, 3\}$  is captured from different graph pooling layers. In the GCN model, early convolutional layers with high spatial resolutions capture local details

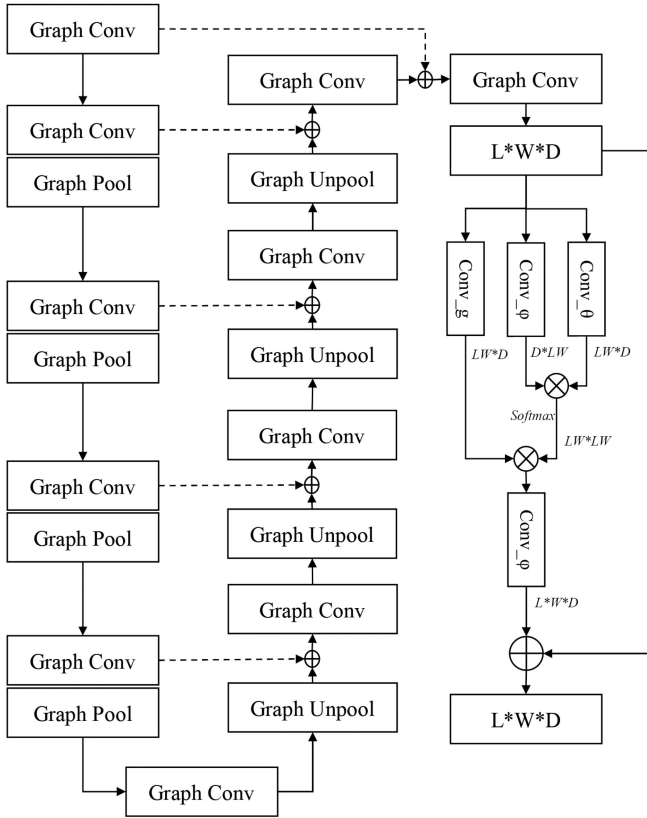


Fig. 3. Detailed information of a single graph u-nets in dual graph u-nets.

while the later convolutional layers with low spatial resolutions extract high-level semantics, the same as the CNN model. To fuse the advantages of the different stages, the DGU-HSI adopts graph unpooling and skip connection to combine different scale graph features. It is inadequate to join the two pieces of information simply for HSIs classification. Hence, the graph convolution layers are also implemented to fuse local information and high-level features.

Then, the attention mechanism model is adopted to find the feature weight of each node. The nonlocal network is introduced to calculate the similarities in latent feature space, which can be defined as

$$s(x_i, x_j) = \frac{1}{\sum_{\forall j} e^{((W_\theta x_i)^T (W_\varphi x_j))}} e^{((W_\theta x_i)^T (W_\varphi x_j))} W_g x_j \quad (20)$$

where  $W_\theta, W_\varphi, W_g$  represent  $Conv_\theta, Conv_\varphi, Conv_g$  respectively, using  $1 \times 1$  convolution kernel.

Finally, a ResNet structure is utilized to add an original feature for getting the spectral or spatial feature with attention mechanism constraint. The attention mechanism finds more important ones in the current feature map and distributes them with higher weights. This operation makes the feature more effective in classification.

Spectral and spatial graphs have different graph structures and relationship features. Thus, single graph u-nets cannot extract

effective features from spectral and spatial graphs simultaneously. To tackle this issue, a dual graph u-nets is adopted to acquire more discriminative spectral and spatial features for HSIs classification.

### C. Fusion Schemes

Spectral and spatial features are capable of presenting different characteristics of HSIs. For instance, spectral features can perform difference in different materials and spatial features are able to reflect the influence of position. Thus, only spectral or spatial features may not provide optimal results in terms of classification performance. In this section, two fusion strategies are proposed to enhance feature discrimination ability. Additive(A), Concatenation(C) are considered, and the two fusion strategies (A, C) can be respectively formulated as follows:

$$F_{Fu-A}^{l+1} = F_{spe}^l \oplus F_{spa}^l \quad (21)$$

$$F_{Fu-C}^{l+1} = [F_{spe}^l, F_{spa}^l] \quad (22)$$

where the operations  $\oplus, [\cdot, \cdot]$  denote the elementwise addition and concatenation, respectively. The  $F_{spe}^l, F_{spa}^l$  is the  $l$ th layer feature acquired from the spectral and spatial feature extraction model, respectively.

One fully connected layer is applied after the fusion model, and it is used for dimensionality reduction of features to increase computational efficiency. Meanwhile, it can also introduce different weights for fusion features to acquire effective classification performance.

The loss function is implemented on the DGU-HIS to optimize the learning process. The cross entropy is adopted for loss function calculation

$$Loss = - \sum_{i=1}^n y_i \cdot \log(p_i) \quad (23)$$

where  $n$  is the number of pixels,  $y_i$  is the one-hot representation of the ground truth of pixel,  $p_i$  is the probability distribution of prediction.

## IV. EXPERIMENTS

### A. Data Description

In this section, experiments are conducted on four publicly available datasets: Indian Pines, Pavia University, Salinas, and Trento. In light of each dataset, 200 labeled pixels per class are randomly selected for training, and other pixels in the ground-truth map for testing. The detailed descriptions of these datasets are listed as follows.

- 1) The Indian Pines dataset was collected by AVIRIS sensor over Northwestern Indiana, which contains  $145 \times 145$  pixels. This dataset contains 200 spectral bands with a wavelength ranging from 0.4 to 2.5  $\mu m$ . There are 16 land-cover classes in the ground-truth of this dataset. The classes with fewer samples are removed, and 8 classes with more samples are selected [39]–[41]. The numbers of training and testing samples are listed in Table I.
- 2) The Pavia University dataset was captured by the ROSIS sensor over Pavia, northern Italy. This image contains

TABLE I  
NUMBERS OF TRAINING AND TESTING SAMPLES FOR THE INDIAN PINES DATASET



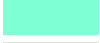














No.	color	Class	Training	Test
1		Corn-notill	200	1228
2		Corn-mintill	200	630
3		Grass-pasture	200	283
4		Hay-windrowed	200	278
5		Soybean-notill	200	772
6		Soybean-mintill	200	2255
7		Soybean-clean	200	393
8		Woods	200	1065
Total			1600	6904

TABLE II  
NUMBERS OF TRAINING AND TESTING SAMPLES FOR THE PAVIA UNIVERSITY DATASET

No.	color	Class	Training	Test
1		Asphalt	200	6341
2		Meadows	200	18449
3		Gravel	200	1899
4		Trees	200	2864
5		Sheets	200	1145
6		Baresoil	200	4829
7		Bitumen	200	1130
8		Bricks	200	3482
9		Shadows	200	747
Total			1800	40976

610 × 340 pixels and 103 spectral bands with the spatial resolution of 1.3 m. In total, there are 9 classes and 42 776 samples. Table II indicates the numbers of training and testing samples.

- The third dataset is the Salinas dataset, and it is the most sampled dataset. The Salinas dataset was gathered with the AVIRIS sensor over Salinas Valley, which contains 512 × 217 pixels, 204 spectral bands. The numbers of training and testing samples are listed in Table III.
- The last dataset is the Trento dataset. The Trento dataset was gathered with the AISA Eagle sensor over a rural area in the south of Trento, Italy, which contains 660 × 166 pixels, 63 spectral bands. The numbers of training and testing samples are listed in Table IV.

### B. Effect of Different Window Sizes

The proposed DGU-HSI utilizes spectral similarity across different pixels to construct a spectral graph for HSIs classification. The window size decides the number of nodes that can be utilized in spectral graph construction.

TABLE III  
NUMBERS OF TRAINING AND TESTING SAMPLES FOR THE SALINAS DATASET

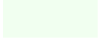














No.	color	Class	Training	Test
1		Broccoli-green-weeds	200	3526
2		Fallow	200	1776
3		Fallow-rough-plow	200	1194
4		Fallow-smooth	200	2478
5		Stubble	200	3759
6		Celery	200	3379
7		Grapes-untrained	200	11071
8		Soil-vinyard-develop	200	6003
9		Corn-senesced-green-weeds	200	3078
10		Lettuce-romaine-4wk	200	868
11		Lettuce-romaine-5wk	200	1727
12		Lettuce-romaine-6wk	200	716
13		Lettuce-romaine-7wk	200	870
14		Vinyard-untrained	200	7068
15		Vinyard-vertical-trellis	200	1607
Total			3000	50929

TABLE IV  
NUMBERS OF TRAINING AND TESTING SAMPLES FOR THE TRENTO DATASET

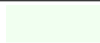





No.	color	Class	Training	Test
1		Apple trees	200	3834
2		Buildings	200	2703
3		Ground	200	279
4		Woods	200	8923
5		Vineyard	200	10301
6		Roads	200	2974
Total			1200	29014

Fig. 4 illustrates the classification performance with different window sizes in four benchmark datasets. The window sizes range from 7 × 7 to 13 × 13. When the window size is 9 × 9, the classification performance of DGU-HSI tends to be satisfactory. The small window size is unable to extract sufficient spectral information and its performance is poor, such as the Pavia University dataset. The large window size introduces the redundancy of spectral information, which plays a negative role in classification performance. The phenomenon is evident in the Salinas dataset.

### C. Influence of Different Graph Nodes

The effectiveness of spectral graph data depends on the window size and is related to the number of graph nodes. The number of graph nodes affects the similarity relationship of pixels and

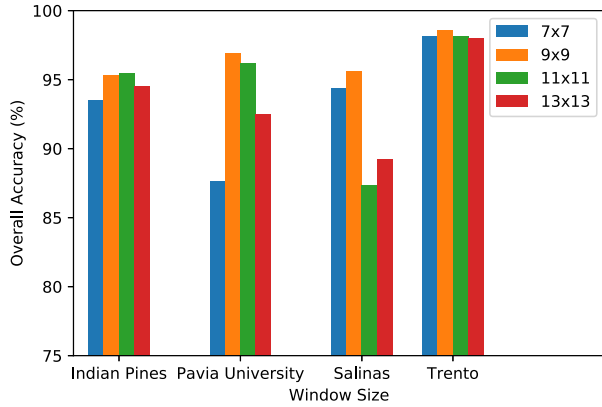


Fig. 4. Classification performance with different window sizes.

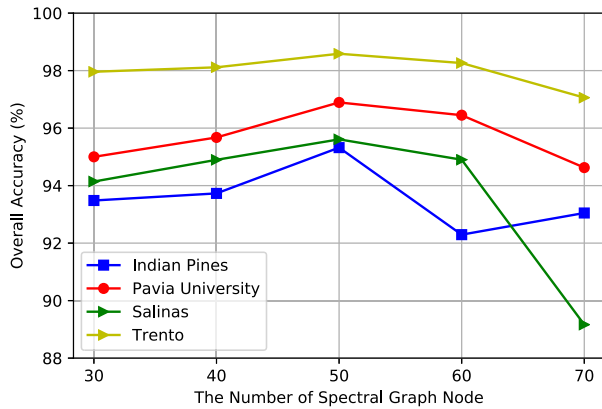


Fig. 5. Classification performance with different window sizes.

indirectly affects the classification performance. Fig. 5 demonstrates the classification performance with a different number of graph nodes. The number of graph nodes varies from 30 to 70, and the classification performance presents an increase followed by a decrease. When the number of graph nodes is 50, the classification gets the optimum performance. The spectral graph with fewer nodes is unable to represent spectral information and spectral similarity adequately. The excess nodes introduce negative effects for the spectral graph, which is probably related to the phenomenon of different materials with the same spectra characteristics. Thus, the number of graph nodes is set as 50.

#### D. Effect of PCA

Since HSIs contain hundreds of spectral bands with the nanoscale spectral resolution, the spectral information is so rich that redundancy information is inevitable. Therefore, it is essential to analyze effective spectral information to solve redundancy information. To address this problem, PCA is adopted to resolve the spectral redundancy information. The purpose of PCA is to design effective spectral and spatial graphs, and it is expected to affect HSIs classification performance.

Table V shows the effects of PCA on classification performance. The performance of DGU-HSI with PCA has an improvement of 10.96%, 3.12%, 12.76%, 1.55% on Indian Pines,

TABLE V  
EFFECTS OF PCA ON CLASSIFICATION

	Indian Pines (OA%)	Pavia University (OA%)	Salinas (OA%)	Trento (OA%)
Without PCA	84.36	93.77	82.85	97.04
With PCA	95.32	96.89	95.61	98.59

TABLE VI  
EFFECTS OF DUAL ARCHITECTURE ON CLASSIFICATION

	Indian Pines (OA%)	Pavia University (OA%)	Salinas (OA%)	Trento (OA%)
Only Spatial	93.05	95.69	94.42	97.58
Only Spectral	92.29	94.65	91.99	97.41
Spectral and Spatial	95.32	96.89	95.61	98.59

TABLE VII  
EFFECTS OF ATTENTION MECHANISM ON CLASSIFICATION

	Indian Pines (OA%)	Pavia University (OA%)	Salinas (OA%)	Trento (OA%)
Without Attention	91.89	94.51	93.51	98.29
With Attention	95.32	96.89	95.61	98.59

Pavia University, Salinas, and Trento, respectively. Compared to no PCA, PCA retains the main information of HSIs and reduces the redundancy information. On the whole, the PCA plays a positive role in classification performance.

#### E. Effect of Dual Architecture

In this section, three experiments are conducted to verify the effectiveness of dual architecture. The effects of dual architecture on classification are listed in Table VI. The performance on Indian Pines, Pavia University, Salinas, and Trento datasets using only the spatial feature is 93.05%, 95.69%, 94.42%, and 97.58%, respectively, and the spectral feature is 92.29%, 94.65%, 91.99%, and 97.41%, respectively. There is a 1.1% to 3.7% improvement in classification performance using spectral and spatial union features. Spectral feature retains abundant spectral similarity information among pixels but does not contain the neighborhood information of contiguous pixels. The spatial feature has the neighborhood relationship of the center pixel while spectral information is less. For HSIs classification, it is inadequate to utilize only utilizing the spatial feature or the spectral feature. The union feature contains both spectral and spatial information. In a word, the progress of dual architecture bring for HSIs classification is considerable.

#### F. Effect of Attention Mechanism

In the proposed DGU-HIS, an attention mechanism is introduced to screen the weight of the spectral and spatial union feature. The classification performance without an attention mechanism and with an attention mechanism is evaluated. The results are shown in Table VII.

The classification performance with attention mechanism has a 3.43% improvement on the Indian Pines dataset, which is the most obvious among the four datasets. The Pavia University, Salinas, and Trento datasets also show 96.89%, 95.61%, and 98.59% performance with attention mechanism, against

TABLE VIII  
EFFECTS OF DIFFERENT FUSION STRATEGY

	Indian Pines (OA%)	Pavia University (OA%)	Salinas (OA%)	Trento (OA%)
Additive(A)	94.25	96.89	94.84	98.31
Concatenation(C)	95.32	96.89	95.61	98.59

TABLE IX  
EFFECTS OF DIFFERENT NUMBERS OF TRAINING SAMPLES

	Indian Pines (OA%)	Pavia University (OA%)	Salinas (OA%)	Trento (OA%)
50	80.98	92.38	84.18	95.10
100	84.55	94.74	89.48	97.12
200	95.32	96.89	95.61	98.59
300	97.67	98.03	95.71	98.81

TABLE X  
COMPARISON OF THE OVERALL CLASSIFICATION ACCURACY (%) AMONG THE PROPOSED METHOD AND THE BASELINES USING THE INDIAN PINES DATASET

Class	Performance					
	SVM	1D-CNN	2D-CNN	3D-CNN	GCN	DGU-HSI
1	72.31	89.98	87.05	87.95	78.85	<b>92.43</b>
2	75.87	77.78	95.56	94.60	80.60	<b>97.62</b>
3	96.11	93.29	<b>99.65</b>	98.94	96.69	96.11
4	99.28	<b>100</b>	<b>100</b>	<b>100</b>	99.58	<b>100</b>
5	81.48	77.72	93.13	94.43	92.08	<b>96.24</b>
6	67.10	73.70	86.61	88.82	68.84	<b>92.59</b>
7	85.24	67.68	97.96	<b>98.22</b>	90.56	97.96
8	95.87	98.50	99.72	99.81	97.55	<b>100</b>
AA(%)	84.16	84.78	94.96	95.35	88.09	<b>96.62</b>
OA(%)	78.39	82.63	91.98	92.92	83.41	<b>95.32</b>
Kappa	0.7395	0.7870	0.9019	0.9134	0.8041	<b>0.9427</b>

The bold entities indicates the maximum accuracy of the classification of each category in the comparison experiment.

TABLE XI  
COMPARISON OF THE OVERALL CLASSIFICATION ACCURACY (%) AMONG THE PROPOSED METHOD AND THE BASELINES USING THE PAVIA UNIVERSITY DATASET

Class	Performance					
	SVM	1D-CNN	2D-CNN	3D-CNN	GCN	DGU-HSI
1	66.52	92.33	<b>95.43</b>	94.54	77.71	93.45
2	73.36	87.54	<b>99.39</b>	96.93	81.83	98.76
3	81.83	90.31	86.73	91.42	80.47	<b>95.58</b>
4	96.40	96.30	<b>97.70</b>	97.24	93.18	96.79
5	99.39	100	<b>99.91</b>	99.83	99.48	99.91
6	89.42	94.93	95.15	96.36	87.23	<b>96.71</b>
7	95.84	79.47	93.19	<b>97.17</b>	92.41	96.02
8	79.55	60.20	92.73	<b>95.52</b>	85.82	93.08
9	100	99.87	<b>100</b>	<b>100</b>	99.89	99.87
AA(%)	86.92	88.99	95.58	96.56	88.68	<b>96.68</b>
OA(%)	78.54	87.93	96.85	96.28	84.20	<b>96.89</b>
Kappa	0.7268	0.8413	0.9576	0.9501	0.7970	<b>0.9583</b>

The bold entities indicates the maximum accuracy of the classification of each category in the comparison experiment.

94.51%, 93.51%, 98.29% without attention mechanism. The attention mechanism finds the more important ones in the current feature map and distributes them with higher weights. This operation makes the feature more effective in classification.

### G. Effect of Fusion Strategy

In this section, two fusion strategies are evaluated to verify their effectiveness in terms of classification performance. Table VIII illustrates the effectiveness of different fusion strategies. Specifically, the comparison between the two commonly

TABLE XII  
COMPARISON OF THE OVERALL CLASSIFICATION ACCURACY (%) AMONG THE PROPOSED METHOD AND THE BASELINES USING THE SALINAS DATASET

Class	Performance					
	SVM	1D-CNN	2D-CNN	3D-CNN	GCN	DGU-HSI
1	99.78	99.66	<b>99.93</b>	99.69	99.38	99.60
2	99.66	<b>99.94</b>	99.38	99.61	91.95	98.99
3	99.58	99.50	<b>99.93</b>	99.58	97.84	<b>100</b>
4	98.31	97.58	<b>99.56</b>	99.11	98.06	<b>99.56</b>
5	99.79	99.95	<b>99.97</b>	99.87	99.00	<b>99.97</b>
6	99.70	99.91	99.94	99.91	99.29	99.85
7	82.50	78.66	78.16	68.70	82.25	<b>88.02</b>
8	99.07	99.72	99.80	99.50	97.11	<b>99.87</b>
9	94.25	94.12	96.30	94.50	91.60	<b>97.03</b>
10	98.27	99.08	<b>100</b>	99.88	90.77	99.77
11	99.65	99.71	<b>100</b>	<b>100</b>	<b>100</b>	99.94
12	98.60	99.44	<b>100</b>	<b>100</b>	98.96	<b>100</b>
13	96.09	98.74	<b>100</b>	97.13	97.35	<b>100</b>
14	66.89	76.02	78.41	86.12	70.44	<b>89.13</b>
15	98.94	99.63	99.19	98.26	97.10	<b>100</b>
AA(%)	95.41	96.11	96.70	96.12	94.07	<b>98.13</b>
OA(%)	90.86	91.43	91.96	90.69	94.34	<b>95.61</b>
Kappa	0.8971	0.9037	0.9096	0.8958	0.8928	<b>0.9506</b>

The bold entities indicates the maximum accuracy of the classification of each category in the comparison experiment.

TABLE XIII  
COMPARISON OF THE OVERALL CLASSIFICATION ACCURACY (%) AMONG THE PROPOSED METHOD AND THE BASELINES USING THE TRENTO DATASET

Class	Performance					
	SVM	1D-CNN	2D-CNN	3D-CNN	GCN	DGU-HSI
1	94.31	90.11	99.43	<b>99.87</b>	90.01	99.58
2	88.12	89.68	88.79	92.93	75.37	<b>93.96</b>
3	99.28	97.85	<b>100</b>	<b>100</b>	97.29	<b>100</b>
4	96.94	97.03	99.42	<b>99.78</b>	95.52	99.63
5	84.49	88.61	97.46	<b>99.75</b>	76.69	99.43
6	75.12	85.94	89.04	89.21	82.73	<b>95.33</b>
AA(%)	89.71	91.54	95.69	96.92	86.27	<b>97.99</b>
OA(%)	89.14	91.31	96.67	98.06	84.99	<b>98.59</b>
Kappa	0.8556	0.8838	0.9553	0.9739	0.8029	<b>0.9810</b>

The bold entities indicates the maximum accuracy of the classification of each category in the comparison experiment.

used fusion strategies indicates that Concatenation(C) has better classification performance than Additive(A), especially on the Indian Pines and Salinas datasets, with 1.07% and 0.77% improvement in classification performance. The result suggests that the Concatenation(C) strategy is able to contain more effective features. The Additive(A) strategy may lose some information when performing the addition operation.

### H. Influence of the Number of Training Samples

To further demonstrate the performance of the proposed DGU-HSI, experiments are conducted on the three datasets with different numbers of training samples per class. As is shown in Table IX, the number of training samples varies from 50 to 300. Too few samples have poor performance in HSI classification, such as training samples number is 50 or 100. However, the DGU-HSI gets a satisfying performance when the training samples is 200. The improvement is limited when training samples are between 200 and 300 per class. The result indicates that the performance is relatively stable under the influence of a small number of samples.



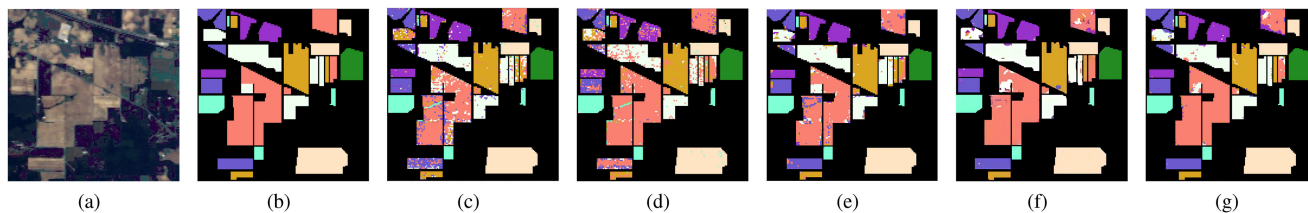


Fig. 6. False color composite, the ground truth map and classification maps of Indiana Pines. (a) False color composite. (b) Ground truth map. (c) Classification maps of SVM. (d) Classification maps of 1-D-CNN. (e) Classification maps of 2-D-CNN. (f) Classification maps of 3-D-CNN. (g) Classification maps of DGU-HSI.

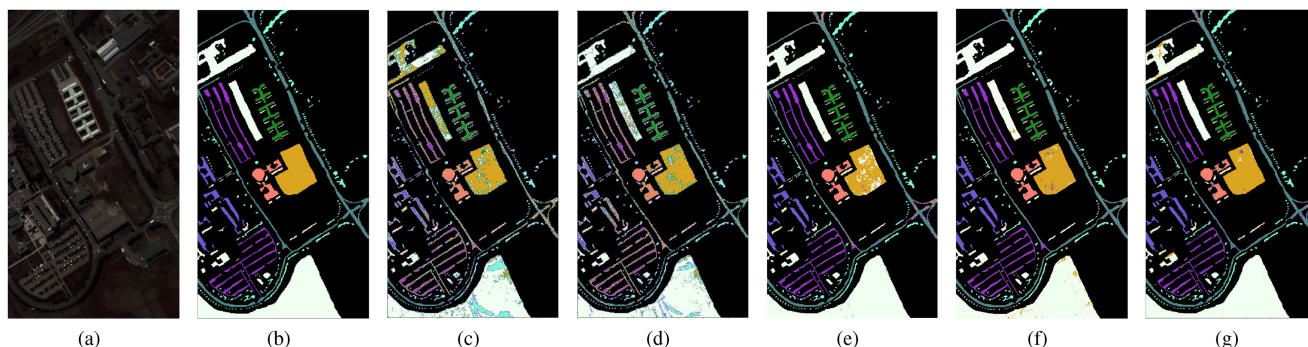


Fig. 7. False color composite, the ground truth map and classification maps of Pavia University. (a) False color composite. (b) Ground truth map. (c) Classification maps of SVM. (d) Classification maps of 1-D-CNN. (e) Classification maps of 2-D-CNN. (f) Classification maps of 3-D-CNN. (g) Classification maps of DGU-HSI.

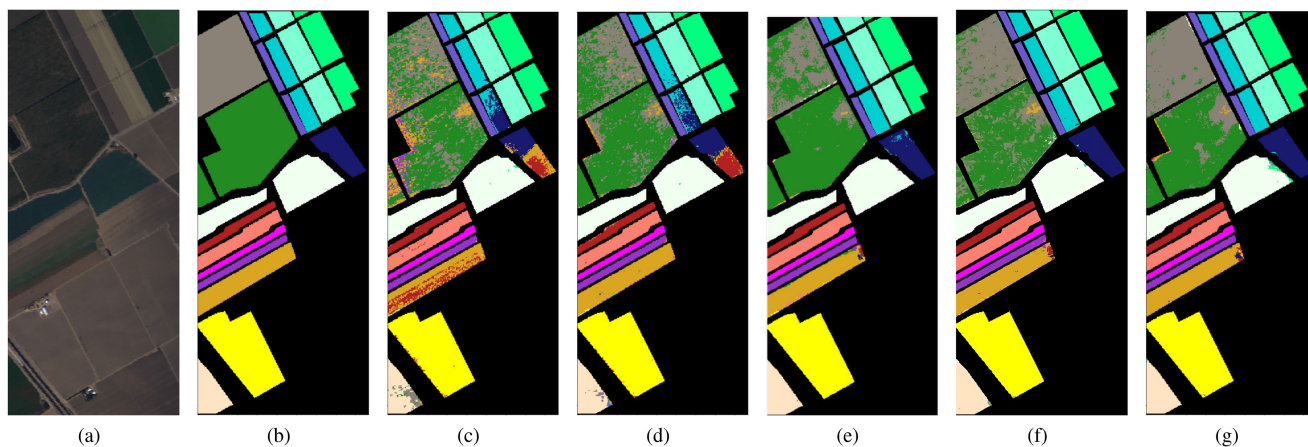


Fig. 8. False color composite, the ground truth map and classification maps of Salinas. (a) False color composite. (b) Ground truth map. (c) Classification maps of SVM. (d) Classification maps of 1-D-CNN. (e) Classification maps of 2-D-CNN. (f) Classification maps of 3-D-CNN. (g) Classification maps of DGU-HSI.

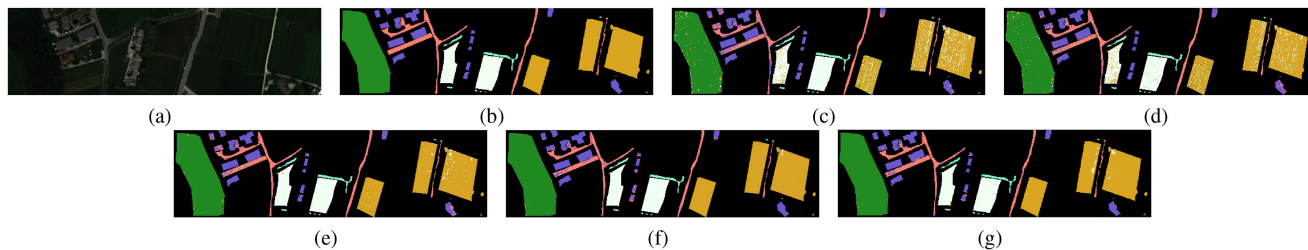


Fig. 9. False color composite, the ground truth map and classification maps of Trento. (a) False color composite. (b) Ground truth map. (c) Classification maps of SVM. (d) Classification maps of 1-D-CNN. (e) Classification maps of 2-D-CNN. (f) Classification maps of 3-D-CNN. (g) Classification maps of DGU-HSI.

### I. Classification Performance

To validate the effectiveness of the proposed DGU-HSI, the method was compared with some related HSI classification methods, such as SVM, 1D-CNN, 2D-CNN, 3D-CNN, GCN. The results about per-class accuracies, overall accuracy (OA), average accuracy (AA), and kappa statistic (K) are presented in Tables X–XIII for evaluating the classification efficacy of each method.

Overall, 1-D-CNN, 2-D-CNN, 3-D-CNN, GCN, DGU-HSI outperform SVM on four datasets, indicating the superiority of the neural network in image classification. In addition, the classification performance of 2-D-CNN, 3-D-CNN, DGU-HSI is also better than 1-D-CNN on Indian Pines, Pavia University, and the Trento dataset, which suggests that the spectral and spatial union information is superior to the only spectral information in HSIs classification.

In general, the classification performance of the proposed DGU-HSI is superior to the other methods in the four datasets. As shown in Table XII, the proposed DGU-HSI yields OA 95.61%, about 1.3% higher than that of GCN (i.e., 94.34%), 2.4% higher than that of 3-D-CNN (i.e., 92.92%), 4% higher than that of the 2-D-CNN (i.e., 91.96%), 4.2% higher than that of the 1-D-CNN (i.e., 91.43%) and approximately 5% improvement compared to the SVM (i.e., 90.86%). The similarity phenomenon can also be found on the Indian Pines, Pavia University, and the Trento dataset. The main reason is that the spectral graph data of the proposed method DGU-HSI relieve the defect of fixed CNN kernel receptive field and provide a bigger receptive field to HSIs classification. Besides, compared with GCN, the proposed DGU-HSI has a deeper network, which improves the feature extraction ability of graph convolution block, and the classification results verify this thesis. For example, DGU-HIS Kappa is 0.9427, 0.9583, 0.9506, 0.9810 on four dataset, higher than that of GCN (i.e., 0.8041, 0.7970, 0.8928, 0.8029).

Furthermore, Figs. 6–9 illustrate the full classification maps, where the visual classification results are consistent with the performance listed in Tables X–XIII.

All the experiments were conducted on a computer with an NVIDIA GeForce RTX-2070 SUPER GPU (8 GB GDDR5) and 32 GB memory. Based on the stand back-propagation algorithms, the Adam algorithm is adopted to learn the network parameters, where the batch size is set to 120, and the learning rate is 0.0001.

### V. CONCLUSION

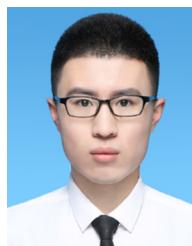
In recent years, deep learning methods have attracted more attention on HSIs classification especially the GCN-based method. However, the traditional GCN-based methods only use spectral similarity to design graph and neglect spatial information. In addition, due to the over-smooth phenomenon, most GCN models are unable to extract effective feature. To tackle this issue, a dual graph u-nets model is proposed for HSIs classification, which resolves the over-smooth problem of deeper GCN. This model contains three main parts: first, spectral similarity and spatial adjacency information are used to design spectral and spatial graph. Second, local information and high-level features were fused by graph u-nets, and an attention mechanism was used to

get the discriminable spectral and the spatial feature. Finally, the spectral and the spatial feature were fused for HSIs classification. From the results in four public datasets, it can be concluded that the proposed DGU-HSI achieved convincing results compared with the state-of-the-art. It is worth noting that the performance of DGU-HSI is not satisfying when using few training samples. One of the future works is expected to improve the DGU-HSI to adapt HSI classification with fewer training samples.

### REFERENCES

- [1] A. W. Nolin and J. Dozier, "A hyperspectral method for remotely sensing the grain size of snow," *Remote Sens. Environ.*, vol. 74, no. 2, pp. 207–216, 2000.
- [2] E. Torrecilla, D. Stramski, R. A. Reynolds, E. Millán-Núñez, and J. Píera, "Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2578–2593, 2011.
- [3] R. J. Ellis and P. W. Scott, "Evaluation of hyperspectral remote sensing as a means of environmental monitoring in the St. Austell China clay (kaolin) region, Cornwall, U.K.," *Remote Sens. Environ.*, vol. 93, no. 1/2, pp. 118–130, 2004.
- [4] X. Yang and Y. Yu, "Estimating soil salinity under various moisture conditions: An experimental study," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2525–2533, May 2017.
- [5] J. Franke, D. A. Roberts, K. Halligan, and G. Menz, "Hierarchical multiple endmember spectral mixture analysis (MESMA) of hyperspectral imagery for urban environments," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1712–1723, 2009.
- [6] S. Roessner, K. Segl, U. Heiden, and H. Kaufmann, "Automated differentiation of urban surfaces based on airborne hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1525–1532, Jul. 2001.
- [7] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.
- [8] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [9] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.
- [10] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
- [11] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [12] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial-spectral fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 4, pp. 2485–2501, Apr. 2020.
- [13] Z. Li *et al.*, "Deep multilayer fusion dense network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 3, pp. 1258–1270, Mar. 2020.
- [14] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Pyramid fully convolutional network for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 5, pp. 1549–1558, May 2019.
- [15] L. Zou, X. Zhu, C. Wu, Y. Liu, and L. Qu, "Spectral-spatial exploration for hyperspectral image classification via the fusion of fully convolutional networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 2, pp. 659–674, Feb. 2020.
- [16] J. Feng *et al.*, "CNN-based multilayer spatial-spectral feature fusion and sample augmentation with local and nonlocal constraints for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1299–1313, Apr. 2019.
- [17] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.

- [18] Z. Li, X. Cui, L. Wang, H. Zhang, X. Zhu, and Y. Zhang, "Spectral and spatial global context attention for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 771.
- [19] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.
- [20] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.
- [21] Y. Cai, X. Liu, and Z. Cai, "BS-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1969–1984, Mar. 2020.
- [22] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, Art. no. 258619.
- [23] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.
- [24] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.
- [25] B. Pan, Z. Shi, and X. Xu, "MugNet: Deep learning for hyperspectral image classification using limited samples," *ISPRS J. Photogramm.*, vol. 145, pp. 108–119, 2018.
- [26] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4590–4603, Jul. 2020.
- [27] R. Li, Z. Pan, Y. Wang, and P. Wang, "A convolutional neural network with mapping layers for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3136–3147, May 2020.
- [28] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 593–597, Apr. 2019.
- [29] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [30] L. Mou, X. Lu, X. Li, and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8246–8257, Dec. 2020.
- [31] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, May 2020.
- [32] L. Tong, J. Zhou, B. Qian, J. Yu, and C. Xiao, "Adaptive graph regularized multilayer nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 1, pp. 434–447, Jan. 2020.
- [33] Y. Cai, Z. Zhang, X. Liu, and Z. Cai, "Efficient graph convolutional self-representation for band selection of hyperspectral image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 8, pp. 4869–4880, Aug. 2020.
- [34] J. Bai, B. Ding, Z. Xiao, L. Jiao, H. Chen, and A. C. Regan, "Hyperspectral image classification based on deep attention graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2021.3066485](https://doi.org/10.1109/TGRS.2021.3066485).
- [35] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, and Y. Y. Tang, "Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 241–245, Feb. 2019.
- [36] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, 2011.
- [37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–14.
- [38] H. Gao and S. Ji, "Graph U-Nets," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2083–2092.
- [39] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.
- [40] Z. Li, F. Guo, Q. Li, G. Ren, and L. Wang, "An encoder-decoder convolution network with fine-grained spatial information for hyperspectral images classification," *IEEE Access*, vol. 8, pp. 33 600–33 608, 2020.
- [41] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2013.



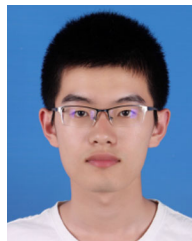
**Fangming Guo** received the B.E. degree in computer science and technology from the China University of Petroleum (East China), Qingdao, China, in 2018. He is currently working toward the doctoral degree in computer technology and resources information engineering with the College of Oceanography and Space Informatics, China University of Petroleum, Beijing, China.

His current research interest includes hyperspectral images classification.



**Zhongwei Li** received the Ph.D. degree in oil and gas well engineering from the China University of Petroleum, Beijing, China.

He is currently a Professor with the College of Oceanography and Space Informatics, China University of Petroleum. His current research interests include remote sensing image processing and ocean numerical forecasting and cloud computing.



**Ziqi Xin** received the B.E. degree in communication engineering from the China University of Petroleum (East China), Qingdao, China, in 2020. He is currently working toward the master's degree in computer science and technology with the College of Computer Science and Technology, China University of Petroleum.

His current research interests include hyperspectral images classification and computer vision.



**Xue Zhu** received the B.E. degree in computer science and technology from the China University of Petroleum (East China), Qingdao, China, in 2019. At present, she is in the second year of the graduate school.

Her research interests include remote sensing image processing and computer vision application.



**Leiquan Wang** (Member, IEEE) received the Ph.D. degree in communication and electrical systems from BUPT, Beijing, China.

He is currently a Lecturer with the College of Computer and Communication Engineering, China University of Petroleum, Beijing. His current research interests include multimodal fusion, cross modal retrieval, and image/video caption.



**Jie Zhang** received the Ph.D. degree in applied mathematics from Tsinghua University, Beijing, China.

He is currently a Professor with the College of Oceanography and Space Informatics, China University of Petroleum, Beijing.