Pattern Recognition 129 (2022) 108705

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Multi-level graph learning network for hyperspectral image classification

Sheng Wan[a], Shirui Pan[b], Shengwei Zhong[a], Jie Yang[c], Jian Yang[a], Yibing Zhan[d], Chen Gong[a,*]

[a] PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, and Jiangsu Key Lab of Image and Video Understanding for Social Security, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China
[b] Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia
[c] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China
[d] JD Explore Academy, Beijing 100176, China

ARTICLE INFO

Article history:
Received 30 August 2021
Revised 24 January 2022
Accepted 7 April 2022
Available online 14 April 2022

Keywords:
Graph convolutional network
Graph-based machine learning
Hyperspectral image classification
Remote sensing
Graph structural learning

ABSTRACT

Graph Convolutional Network (GCN) has emerged as a new technique for hyperspectral image (HSI) classification. However, in current GCN-based methods, the graphs are usually constructed with manual effort and thus is separate from the classification task, which could limit the representation power of GCN. Moreover, the employed graphs often fail to encode the global contextual information in HSI. Hence, we propose a Multi-level Graph Learning Network (MGLN) for HSI classification, where the graph structural information at both local and global levels can be learned in an end-to-end fashion. First, MGLN employs attention mechanism to adaptively characterize the spatial relevance among image regions. Then localized feature representations can be produced and further used to encode the global contextual information. Finally, prediction can be acquired with the help of both local and global contextual information. Experiments on three real-world hyperspectral datasets reveal the superiority of our MGLN when compared with the state-of-the-art methods.

1. Introduction

The past decades have witnessed a great surge of interest in employing hyperspectral imaging for earth observations [1]. Owing to the capacity to detect subtle spectral information, hyperspectral imaging is quite effective in discriminating different geographic objects [2]. Therefore, hyperspectral image (HSI) classification, which aims to categorize each image pixel into a certain meaningful class according to the image contents, has been attracting a growing interest in real-world applications, such as urban planning, vegetation monitoring, and disaster prevention and control [3,4]. However, the similarity occurring in the spectral bands among different land covers makes the classification task challenging. Confronted with this circumstance, spatial context is incorporated to generate discriminative spectral-spatial features [5], such as spatial-based filtering techniques [6], which generally employ handcrafted features conjuncted with a classifier with predefined hyperparameters, and thereby requiring massive experts' experience [7].

Recently, deep learning methods, which act automatically to generate more robust and expressive feature representations than the handcrafted ones, have demonstrated their potentials in modeling the spectral-spatial features of HSI [8]. Especially, Convolutional Neural Network (CNN) has shown great promise and is widely applied to HSI classification tasks [9]. For instance, in Zhang et al. [7], a diversity of discriminative appearance factors are incorporated into CNN, in order to encode the semantic context-aware representations for generating representative features. Although much progress has been made on developing CNN-based HSI classification methods, the power of CNN is still limited in some irregular regions, such as class boundaries [10]. Concretely, CNN cannot perceive the geometric variations between different object regions, since its convolution kernel is designed to only perform in regular squared regions. In addition, the weights of a certain convolution kernel are kept identical when convolving all HSI patches, which inevitably causes a great loss of information around class boundaries, and thus decreasing the expressive power of the generated features. Different from CNN, Graph Convolutional Network (GCN) [11,12] can be naturally applied to irregular image regions, since its convolution operation is adaptively governed by

* Corresponding author.
E-mail address: chen.gong@njust.edu.cn (C. Gong).

https://doi.org/10.1016/j.patcog.2022.108705
0031-3203/

the neighborhood structure [10,13]. Recently, GCN is gradually applied to HSI classification and achieves some success. For example, in Qin et al. [14], GCN with the classical structure has been applied to HSI classification and achieved satisfactory results. Besides, Hong et al. [15] designed a new mini-batch GCN to allow network training on large-scale data in a mini-batch fashion, which outperformed several CNN-based methods.

Nevertheless, there still exist some common defects in most existing GCN-based methods. First, the structural information of the graph is not originally available in hyperspectral data, and it is nature to manually construct an input graph based on pairwise Euclidean distance [10,14]. However, the graph construction procedure is separate from network training, and thus is difficult to manipulate, which might lead to suboptimal performance. To be specific, the constructed graph can be either noisy or have edges that do not correspond to label agreement, as Euclidean distance may not reveal the intrinsic relationships among graph data [16]. This will ultimately weaken the expressive power of the generated representations. Second, in HSI, the regions that are far away in the 2D space may belong to the same land-cover class. However, current GCN-based methods often fail to incorporate the global contextual information, since they mainly focus on encoding the pairwise importance among local regions and thereby disregarding the long range dependencies.

To address the aforementioned problems, we propose a new GCN model, termed Multi-level Graph Learning Network (MGLN), where local spatial importance and global contextual information among image regions can be flexibly learned in a unified framework. Specifically, in our MGLN, the pairwise importance among local regions is automatically learned via attention mechanism, which can help reduce the negative effect of an inaccurate precomputed graph. As such, our model is able to focus on the most relevant spatial information of each region to make decisions. In addition, graph convolution governed by different spatial levels is performed to comprehensively capture the contextual information at different spatial levels. Consequently, the expressive power of the generated feature representations can be enhanced to better represent the regions with diverse object appearances. Meanwhile, in our MGLN, the global contextual information is incorporated by reconstructing the topological information of the graph. Here, the powerful representations produced from local graph convolution are leveraged to encode the long range dependencies among image regions. Then the feature representations can be progressively updated by aggregating the global contextual information appropriately. Finally, the representations generated from local and global levels are integrated, balanced by learning, to obtain comprehensive results. It is noteworthy that our MGLN model jointly optimizes the representation learning and graph reconstruction process to achieve the mutual benefit of both components. Specifically, the topological information of the graph can be refined with the expressive feature representations at local level, which will in turn generate enhanced representations at global level.

The key contributions of this paper can be summarized as follows:

- Attention mechanism is adopted to adaptively capture the most meaningful contextual information at different spatial scales. As a result, the image regions with diverse object appearances can be well represented by our proposed model.
- Graph structural information at both local and global levels is jointly encoded to perform graph convolution, by which the long range dependencies among image regions can be exploited.
- The representation learning and graph reconstruction modules work collaboratively in a unified framework to produce accurate classification results. Experimental results on various typical real-world datasets confirm the superiority of our proposed MGLN, when compared with other state-of-the-art methods.

2. Related work

In this section, we will review some representative works on CNN-based HSI classification and Graph Convolutional Network, since they are closely related to this paper.

2.1. CNN-based hyperspectral image classification

Deep learning [17] has attracted increasing attention for its application to conventional computer vision tasks. One main advantage is that deep learning techniques can automatically learn effective feature representations for a problem domain, thereby avoiding the complicated hand-crafted feature engineering. In recent years, deep learning methods have also revolutionized the field of HSI classification [18]. Particularly, CNN, which is a class of neural networks with fewer parameters than fully-connected networks under the same number of hidden units, has demonstrated its superior performance. For example, in Slavkovikj et al. [19], the spatial dimensions of original hyperspectral data are flattened to generate a 2D image. Besides, the architecture combing 1D CNN and 2D CNN (namely, 1D+2D CNN) has also shown their potentials in HSI classification, such as [20] which uses a CNN to perform spatial-spectral convolution in the first layer for dimensionality reduction and a 2D CNN to form the deeper layers. In addition, 3D CNN is developed to further boost the performance of HSI classification, which is capable of learning to recognize more complex 3D patterns of reflectances with fewer parameters and layers than 1D+2D CNN [21]. Although CNN-based methods have achieved promising performance for HSI classification, they simply apply fixed convolution kernels to different image regions, which will inevitably cause information loss in complex situations. Moreover, traditional CNN-based methods often fail to consider the long-range dependencies due to the localized convolution operation [15], and thus leading to imperfect classification results.

2.2. Graph convolutional network

As one of the hottest topics in graph-based deep learning, GCN defines convolutions and readout operations on irregular graph-structured data [22]. The convolutions on graphs can be roughly divided into two groups, namely spectral convolutions which perform convolution by transforming node representations into spectral domain with graph Fourier transform or its extensions, and spatial convolutions which are based on neighborhood aggregation [11]. Spectral CNN [23], which is the pioneering work of spectral methods, converts signals defined in the vertex domain into spectral domain by leveraging graph Fourier transform. Afterwards, Kipf and Welling [12] proposed a localized first-order approximation to ChebyNet, which contributes to more efficient filtering operation. Different from spectral methods, the spatial methods directly define convolution for each node as a weighted average function over its neighboring nodes, where the weighting function characterizes the impact exerting to the target node by its neighboring nodes [24,25].

Over the past few years, GCN has demonstrated its superior performance in several fields, such as social network mining and natural language processing [26]. More recently, GCN has also been applied to HSI classification [10,14,15,27]. For instance, Wan et al. [10] developed a dynamic GCN to exploit the diverse spectral-spatial information, which constructs the input graphs based on local spatial context in advance. Apart from this, Mou et al. [27] proposed a non-local GCN which learns a dense fully-connected graph based on all pixel features for graph convolution. However, the

fully-connected graph utilized in this non-local GCN may contain some improper inter-class edges, which will bring uncorrelated information to the target pixel during graph convolution. Unlike these early-staged methods, our proposed MGLN can exploit local spatial information and long range dependencies among image regions simultaneously without aggregating uncorrelated information during graph convolution. Besides, the pre-computed fixed graph can also be refined with the adaptive graph learning operations.

3. Proposed method

This section details our proposed MGLN algorithm, of which the schematic is exhibited in Fig. 1. We first segment the input HSI (Fig. 1(a)) into a set of compact regions (Fig. 1(b)). Then graph convolution at multiple spatial (*i.e.,* local) levels (Fig. 1(c)) are performed to obtain expressive feature representations. Subsequently, by reconstructing the topological information of the graph, we perform global-level convolution (Fig. 1(d)) to capture long range dependencies among image regions. Finally, the classification result is produced by adaptively combining the outputs generated at local and global levels (Fig. 1(e)). In the following, the critical steps will be detailed by explaining the region-based segmentation technique (Section 3.1), elaborating the graph learning modules (Section 3.2), and describing the integration of multi-level contextual information (Section 3.3).

3.1. Region-based segmentation

To construct a graph for HSI, treating each image pixel as a graph node is a common and simple approach. However, the efficiency of the subsequent graph convolution operations will be severely restricted due to the huge number of pixels. Consequently, in advance of classification, we employ a segmentation technique named 'Simple Linear Iterative Clustering' (SLIC) [28] to divide the original HSI into a set of compact homogeneous image regions, each of which consists of a small amount of pixels with strong spectral-spatial correlations. To be specific, the SLIC algorithm performs segmentation via iteratively growing the local clusters using a k-means algorithm. Afterwards, each image region is regarded as a graph node, and thus the number of graph nodes can be greatly reduced, which will accelerate the subsequent graph convolution operations. Here, the region features can be obtained by calculating the average spectral signatures of the involved pixels.

3.2. Graph learning modules

As mentioned in the introduction, the structural information of graphs is not available in most hyperspectral data. A common approach to obtain the structural information is calculating the pairwise Euclidean distance among the graph nodes (namely, image regions) in advance [10,14]. However, the existence of different types of noise in HSI may degrade the quality of the generated graph. Meanwhile, since the model training and graph construction are isolated steps, the obtained graph may not always fit the subsequent classification task. To ameliorate this issue, we propose to automatically learn the structral information of the graphs at both the local and global levels. Such a manner can be naturally integrated to the classification model.

To model the local spatial context of HSI, we resort to the attention mechanism to adaptively capture the contextual relations among image regions, instead of using pre-computed fixed weight (*e.g.,* Euclidean distance) as the measurement of pairwise importance. Specifically, a shared linear transformation parametrized by a weight matrix $\mathbf{W}$ is first applied to each node (*i.e.,* region) $\mathbf{x}_i$ as an encoder, aiming at producing feature representations with sufficient expressive power. Then we perform self-attention on the encoded node features as follows:

$$c_{ij} = \mathbf{a}^{\top}[\mathbf{W}\mathbf{x}_i || \mathbf{W}\mathbf{x}_j], \tag{1}$$

where the attention coefficient c_{ij} reveals the pairwise importance between $\mathbf{x}_i$ and $\mathbf{x}_j$, $\mathbf{a}$ is a learnable weight vector, and $||$ denotes the operation to concatenate two vectors. In the most general formulation of Eq. (1), each node is allowed to attend over all the other nodes [25].

In this paper, we inject the local spatial structure into the attention mechanism to exploit the spatial context of HSI. In other words, we only compute the c_{ij} for the regions $\mathbf{x}_j \in N(\mathbf{x}_i)$, where $N(\mathbf{x}_i)$ denotes the spatial neighborhood of $\mathbf{x}_i$. Then the attention coefficient c_{ij} is normalized across all the spatial neighbors of $\mathbf{x}_i$ with a softmax function, namely

$$\alpha_{ij} = \frac{\exp(c_{ij})}{\sum_{\mathbf{x}_k \in N(\mathbf{x}_i)} \exp(c_{ik})}, \tag{2}$$

to be easily comparable across different nodes. From the perspective of attribute values, the normalized attention coefficient α_{ij} can be represented as a single-layer feedforward neural network, which is parametrized by a weight vector $\mathbf{a}$ with the LeakyReLU

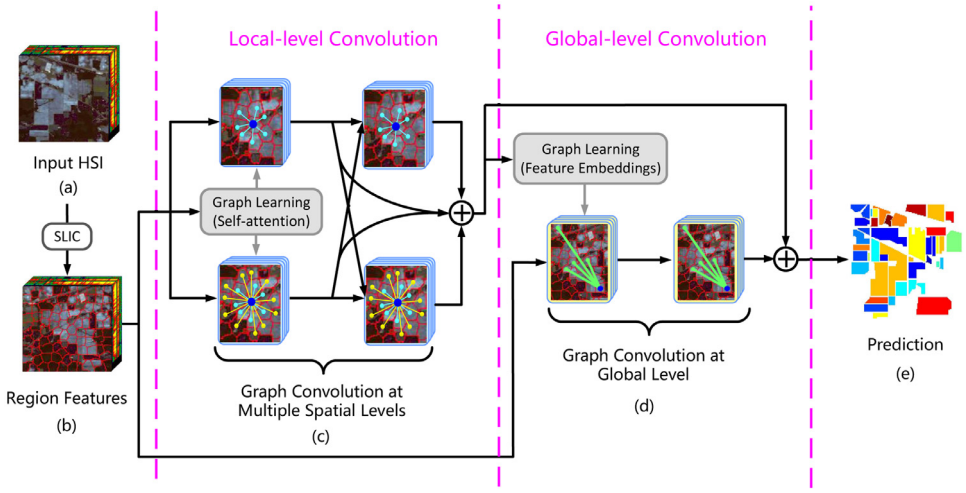


Fig. 1. The framework of our algorithm. (a) is the input hyperspectral data. (b) represents the regions obtained by over-segmenting the original HSI. (c) denotes the graph convolution at multiple spatial (*i.e.,* local) levels, where the pairwise importance among the regions can be learned with attention mechanism automatically. (d) shows the global-level graph convolution, where the topological information of the graph is reconstructed based on the representations generated at local level. In (e), the classification result is acquired by adaptively integrating the outputs at local and global levels.

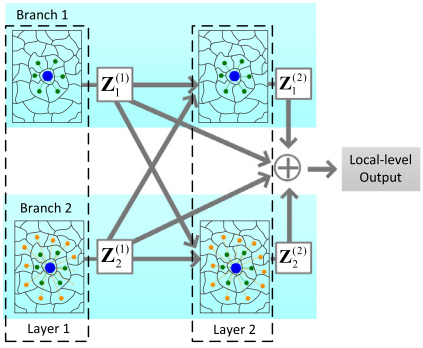


Fig. 2. Exploitation of multi-level spatial information in our method. Graphs used in different branches comprise different neighborhood scales. $\mathbf{Z}_1^{(l)}$ and $\mathbf{Z}_2^{(l)}$ denote the representations generated from the l^{th} layer ($l = 1$ or 2) in branch 1 and 2, respectively. The green and orange nodes represent the spatial neighbors of the central blue one.

nonlinearity. The attention mechanism can be finally expressed as

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^\top[\mathbf{W}\mathbf{x}_i||\mathbf{W}\mathbf{x}_j]))}{\sum_{\mathbf{x}_k \in N(\mathbf{x}_i)} \exp(\text{LeakyReLU}(\mathbf{a}^\top[\mathbf{W}\mathbf{x}_i||\mathbf{W}\mathbf{x}_k]))}. \tag{3}$$

By utilizing the attention mechanism, our proposed model can automatically aggregate the important feature information from local spatial neighborhood. As a consequence, our graph learning model becomes less sensitive to the noise contained in hyperspectral data, compared with [14] and [10], which compute the pairwise similarities with fixed Euclidean distance.

Nonetheless, there often exist various types of object appearances in HSI, where the object regions of the same land-cover class may even have diverse sizes and shapes [7]. Consequently, it is insufficient to incorporate contextual information from only a single spatial level. To deal with this issue, we leverage the merits from multiple spatial levels, so that the image regions can be better represented, as illustrated in Fig. 2. In branch 1, the green nodes constitute the 1-hop spatial neighbors of the central one. Meanwhile, the orange and green nodes form the 2-hop spatial neighbors, as shown in branch 2. Here, the representations generated from the 1st layer in branch 1 (namely, $\mathbf{Z}_1^{(1)}$) can be propagated along three pathways. Under this circumstance, when $\mathbf{Z}_1^{(1)}$ is propagated to branch 1, branch 2, or directly to the output, contextual information covering the 2-hop, 3-hop, or 1-hop spatial neighborhood will be incorporated with the successive graph convolution, respectively. Analogously, $\mathbf{Z}_2^{(1)}$ is allowed to propagate along different pathways to involve the contextual information from different spatial levels. Note that the neighborhood sizes employed at branch 1 and branch 2 (namely, s_1 and s_2) can be adjusted for different HSI, which will be discussed in Section 4.6. Then our local-level graph convolution can be presented as

$$\left[\mathbf{Z}_b^{(1)}\right]_{i,:} = \sigma\left(\sum_{\mathbf{x}_j \in N_b(\mathbf{x}_i)} \alpha_{ij}\mathbf{W}_b^{(1)}\mathbf{x}_j\right), \tag{4}$$

$$\left[\mathbf{Z}_b^{(2)}\right]_{i,:} = \sigma\left(\sum_{\mathbf{x}_j \in N_b(\mathbf{x}_i)} \alpha_{ij}\mathbf{W}_b^{(2)}\left(\left[\mathbf{Z}_1^{(1)}\right]_{j,:} + \left[\mathbf{Z}_2^{(1)}\right]_{j,:}\right)\right), \tag{5}$$

where the branch index $b = 1$ or 2, $\sigma(\cdot)$ is an activation function (*e.g.,* the ReLU function), and $\left[\mathbf{Z}_b^{(1)}\right]_{i,:}$ denotes the i-th row of $\left[\mathbf{Z}_b^{(1)}\right]$. Besides, in branch b, $N_b(\mathbf{x}_i)$ denotes the corresponding spatial neighborhood of $\mathbf{x}_i$ and $\mathbf{W}_b^{(l)}$ is the learnable weight matrix used in the l^{th} layer. Intuitively, the representations $\mathbf{Z}_b^{(l)}$ produced at different spatial levels focus on different types of object appearances, which should be distinguished reasonably. To achieve this, the local-level output can be calculated as

$$\mathbf{Z}_{loc} = \sum_{l=1}^{2}\sum_{b=1}^{2} \lambda_b^{(l)}\mathbf{Z}_b^{(l)} \tag{6}$$

by assigning a parameter $\lambda_b^{(l)}$ to each $\mathbf{Z}_b^{(l)}$, where $\lambda_b^{(l)}$ can be learned via gradient descent. In this manner, contextual information from different spatial levels can adaptively contribute to the local-level output, by which the object regions with diverse shapes and sizes can be represented appropriately.

Nevertheless, the local convolution module fails to exploit the long range dependencies among image regions, thereby lacking the capacity to model the global context of HSI. Hence, we intend to explore the contextual relations beyond local level by reconstructing the topological information of the graph. To be specific, with the learned feature representations $\mathbf{Z}_{loc}$, the reconstructed graph adjacency matrix can be obtained as

$$\widetilde{\mathbf{A}}_{ij} = \exp\left(-\left\|[\mathbf{Z}_{loc}]_{i,:} - [\mathbf{Z}_{loc}]_{j,:}\right\|^2\right). \tag{7}$$

Here, $\widetilde{\mathbf{A}}$ is able to encode the contextual relations between all region pairs and the long range dependencies among faraway regions can also be captured. Meanwhile, the strong expressive power of $\mathbf{Z}_{loc}$ helps characterize the pairwise importance $\widetilde{\mathbf{A}}_{ij}$ precisely. Furthermore, we also propose a reconstruction loss to improve the discriminative power of the learned graph, namely

$$\mathcal{L}_r = \sum_{i,j \in \mathbf{y}_G} \ell\left(\widetilde{\mathbf{A}}_{ij} - \mathbb{1}_{[y_i = y_j]}\right), \tag{8}$$

where $\mathbf{y}_G$ denotes the set of indices corresponding to the labeled examples, y_i represents the class label of $\mathbf{x}_i$, $\mathbb{1}_{[y_i=y_j]}$ is an indicator function evaluating to 1 if $y_i = y_j$ and 0 otherwise, and ℓ is the squared error.

In practice, a densely connected graph often leads to degraded classification performance, as the nodes that do not correspond to the class agreement can be connected. To address this issue, we only retain the graph edges with strong importance and remove the others, which can be presented as

$$\mathbf{A}_{ij} = \begin{cases} \widetilde{\mathbf{A}}_{ij} & \text{if } \widetilde{\mathbf{A}}_{ij} \geq \beta \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

where the parameter β is fixed to 0.75 throughout the experiments. Then the graph convolution at global level can be performed as

$$\mathbf{Z}_{glo}^{(l)} = \sigma\left(\mathbf{A}\mathbf{X}\mathbf{W}_{glo}^{(l)}\right), \tag{10}$$

where $\mathbf{X}_{i,:} = \mathbf{x}_i$, $\mathbf{W}_{glo}^{(l)}$ is the learnable weight matrix used in the l^{th} graph convolutional layer, and $\mathbf{Z}_{glo}^{(l)}$ denotes the representations generated from the l^{th} layer. Since two graph convolutional layers have been utilized, the global-level output $\mathbf{Z}_{glo}$ can be acquired as

$$\mathbf{Z}_{glo} = \mathbf{Z}_{glo}^{(2)}. \tag{11}$$

3.3. Integration of multi-level contextual information

After performing graph convolution at local and global levels, the network output can be generated by fusing the multi-level outputs. Specifically, we first apply a linear transformation parametrized by a weight matrix $\widehat{\mathbf{W}}$ to the local-level output $\mathbf{Z}_{loc}$ as

$$\widehat{\mathbf{Z}}_{loc} = \mathbf{Z}_{loc}\widehat{\mathbf{W}}, \tag{12}$$

in order to make the feature dimension of the output representations consistent with the number of classes. Then the prediction of our model can be computed as

$$\mathbf{O} = \widehat{\mathbf{Z}}_{loc} + \lambda_{glo}\mathbf{Z}_{glo}, \tag{13}$$

where λ_{glo} is used to learn the contribution of global-level output.

Apart from optimizing the reconstruction error of Eq. (8), the cross-entropy error is adopted to penalize the differences between the network output and the labels of the originally labeled regions, namely

$$\mathcal{L}_c = -\sum_{i\in\mathbf{y}_G}\sum_{j=1}^{C}\mathbf{Y}_{ij}\log\mathbf{O}_{ij}, \tag{14}$$

where C is the number of classes, and $\mathbf{Y}$ represents the label matrix. Here, we let $\mathbf{Y}_{ij}=1$ if $\mathbf{x}_i$ belongs to the j^{th} class, and 0 otherwise. Then the overall loss function can be expressed as

$$\mathcal{L} = \mathcal{L}_r + \zeta\mathcal{L}_c, \tag{15}$$

where ζ is the coefficient assigned to the cross-entropy error and can be learned via gradient descent. In our proposed method, all the network parameters are updated through full-batch gradient descent, and the implementation details are summarized in Algorithm 1.

Algorithm 1 The Proposed MGLN for HSI Classification.

Require: Input image; number of iterations $\mathcal{T}$; learning rate η; the neighborhood sizes s_1 and s_2;
1: Segment the whole image into superpixels via SLIC algorithm;
2: // Train the MGLN model
3: **for** $t=1$ to $\mathcal{T}$ **do**
4: **// Graph convolution at local level**
5: Perform graph learning at local spatial level by Eq.~(3);
6: Calculate the local-level output $\mathbf{Z}_{loc}$ through Eq.~(6);
7: **// Graph convolution at global level**
8: Reconstruct the contextual relations at global level based on $\mathbf{Z}_{loc}$ via Eqs.~(7) and (9);
9: Generate the output $\mathbf{Z}_{glo}$ by global-level convolution via Eqs.~(10) and (11);
10: Integrate the multi-level outputs via Eq.~(13);
11: Calculate the error terms according to Eq.~(15), and update the network parameters using full-batch gradient descent;
12: **end for**
13: Conduct label prediction based on the trained network;
Ensure: Predicted label for each image region.

4. Experimental results

In this section, extensive experiments will be conducted to prove the effectiveness of our proposed method, and the corresponding analyzes will also be provided. First, we compare MGLN with other state-of-the-art approaches on three real-world HSI datasets, where four metrics including per-class accuracy, Overall Accuracy (OA), Average Accuracy (AA), and kappa coefficient are used for performance evaluation [3]. Afterwards, we investigate the impact of the number of labeled examples on OA. Then, the impact of the spatial neighborhood sizes on OA is analyzed. Subsequently, we demonstrate that the long range dependencies are advantageous for our model to improve the classification result. Finally, we compare the computational cost of various methods to show the efficiency of our method.

4.1. Datasets

The performance of our proposed MGLN is evaluated on three real-world benchmark datasets, *i.e.,* the Indian Pines, the Salinas, and the Houston University [29] which will be introduced below.

The well-known Indian Pines scene was gathered by AVIRIS sensor in 1992 and records north-western India. This dataset consists of 145×145 pixels with a spatial resolution of 20 m $\times$ 20 m, and there are 220 spectral channels with wavelength varying from 0.4 to 2.5 μm. As a usual step, 20 water absorption and noisy bands are removed, and the remaining 200 bands are retained. This dataset is challenging for traditional HSI classification methods due to the existence of highly mixed examples [30]. The original ground truth of the Indian Pines dataset includes 16 land-cover classes, such as 'Alfalfa', 'Corn-notill', 'Corn-mintill', *etc.*

The Salinas dataset is another classic HSI which was also collected by the AVIRIS sensor, but over a different location in Salinas Valley, California. This dataset comprises 204 spectral bands (20 water absorption bands are removed) and 512×217 pixels with a spatial resolution of 3.7 m. The Salinas dataset contains 16 land-cover classes, including 'Fallow' and 'Celery'.

The Houston University dataset was acquired by the NSF-funded Center for Airborne Laser Mapping over the Houston University campus and its neighboring areas in 2012. This dataset was first distributed in the 2013 IEEE Geoscience and Remote Sensing Society Data Fusion Contest [30]. It contains 349×1905 pixels at a spatial resolution of 2.5 m with 15 classes of interest and 144 spectral bands in the range of 380–1050 nm.

4.2. Experimental settings

In the experiments, our proposed MGLN algorithm is implemented via TensorFlow with Adam optimizer. For all datasets mentioned in Section 4.1, we randomly select 30 labeled pixels per class for network training. If the corresponding class contains less than 30 pixels, 15 will be randomly chosen. During training, 90% of the labeled examples are utilized to learn the network parameters, and the remaining 10% are used as the validation set for hyperparameter tuning. Considering that GCN usually does not require deep structure to achieve promising performance [14], the number of graph convolutional layers is fixed to 2 at all levels. The selection of other hyperparameters in our MGLN, including the learning rate η, the number of iterations $\mathcal{T}$, the number of hidden units u, and the neighborhood scales s_1 and s_2 can be found supplementary materials. In Section 4.6, the parametric sensitivity of the neighborhood scales s_1 and s_2 will be investigated in detail.

In order to justify the effectiveness of our proposed MGLN, various current state-of-the-art HSI classification methods are employed to conduct the comparison. To be specific, we utilize five CNN-based methods, *i.e.,* Diverse Region-based deep CNN (DR-CNN) [7], CNN-Pixel-Pair Features (CNN-PPF) [31], Hybrid-Spectral-Net (HybridSN) [32], Attention-Based Adaptive Spectral-Spatial Kernel ResNet (A^2S^2K) [33], and Spectral-Spatial Attention Network (SSAN) [34], together with three GCN-based methods, *i.e.,* Spectral-Spatial Graph Convolutional Network (S^2GCN) [14], Multi-scale Dynamic Graph Convolutional Network (MDGCN) [10], and mini-batch Graph Convolutional Network (miniGCN) [15]. Additionally, the classification result of our MGLN is also compared with a traditional machine learning method, namely, Multiple Feature Learning (MFL) [35]. All the methods are conducted ten times. The mean accuracies and standard deviations over these ten independent implementations are exhibited.

4.3. Classification results

To reveal the effectiveness of our proposed MGLN, we quantitatively and qualitatively evaluate the classification performance via comparing MGLN with the aforementioned baseline methods.

4.3.1. Results on the Indian Pines dataset

Table 1 reports the quantitative results of different methods on the Indian Pines dataset, where the best performance of all the methods has been highlighted in boldface. It is obvious that our proposed MGLN achieves better classification results than the

Table 1
Per-Class Accuracy, OA, AA (%), and Kappa Coefficient of Different Methods Achieved on Indian Pines Dataset.

Methods	MFL [35]	DR-CNN [7]	CNN-PPF [31]	HybridSN [32]	A^2S^2K [33]	SSAN [34]	S^2GCN [14]	MDGCN [10]	miniGCN [15]	MGLN
OA	80.22±0.20	86.65±0.59	80.09±1.56	84.49±2.25	86.15±1.70	76.14±2.98	89.49±1.08	93.47±0.38	73.45±1.34	**94.27±0.92**
AA	87.85±0.19	93.99±0.25	87.80±1.53	92.58±1.18	76.97±1.05	86.70±1.27	92.99±1.04	**96.24±0.21**	84.10±0.90	95.58±1.18
Kappa	77.59±0.22	84.88±0.67	77.52±1.74	82.45±2.50	84.34±1.87	73.19±3.22	88.00±1.23	92.55±0.43	69.92±1.51	**93.46±1.04**

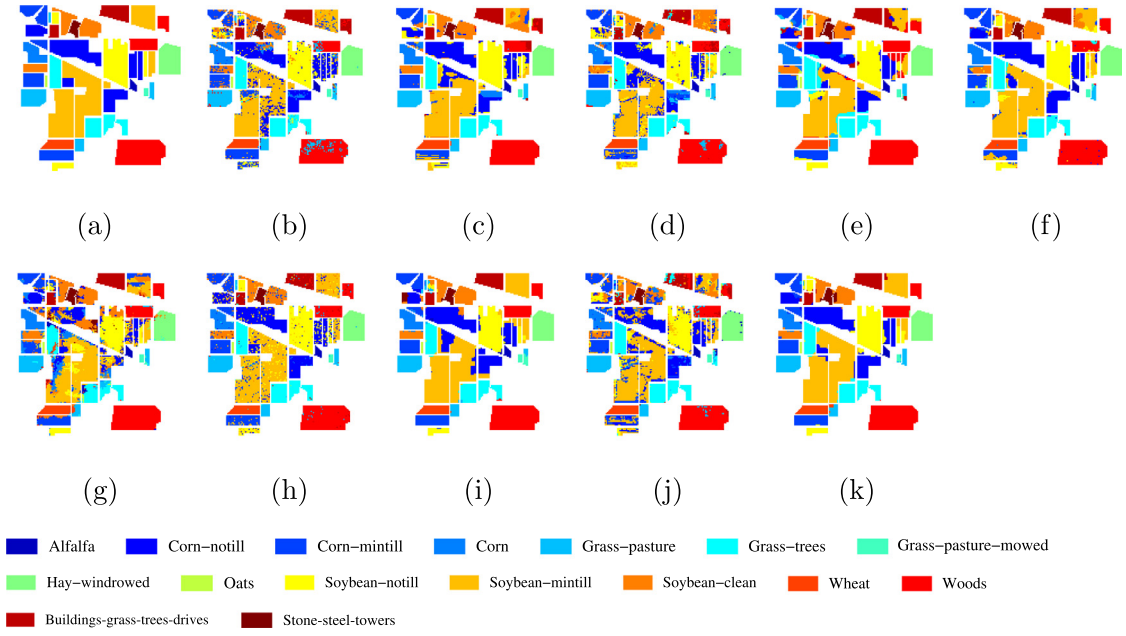


Fig. 3. Classification maps obtained by different methods on Indian Pines dataset. (a) Ground truth map; (b) MFL; (c) DR-CNN; (d) CNN-PPF; (e) HybridSN; (f) A^2S^2K; (g) SSAN; (h) S^2GCN; (i) MDGCN; (j) miniGCN; (k) MGLN.

baseline methods in terms of OA, AA, and Kappa coefficient on Indian Pines dataset. In particular, the improvements of MGLN over all of the CNN-based methods (*i.e.*, DR-CNN, CNN-PPF, HybridSN, A^2S^2K, and SSAN) can be even higher than 7% in terms of OA, which indicates the great potential of GCN in HSI classification. We also find MGLN and MDGCN perform better than other baseline methods, by which the effectiveness of contextual information at multiple spatial levels can be validated. Moreover, it is notable that the standard deviations of our MGLN are relatively small, compared with the deep learning-based methods. This demonstrates the stability of our proposed method.

Fig. 3 visualizes the results generated via using different methods, where the classification map obtained by our proposed MGLN is noticeably closer to the ground-truth map (see Fig. 3(a)) than those of other methods. In addition, we see that most of the GCN-based methods (namely, S^2GCN, MDGCN, and MGLN) can produce fewer errors around class boundaries than the other ones, which confirms the good discriminability of the GCN in boundary regions. Observing the second class (namely, 'Corn-notill'), fewer misclassifications can be found in the results of DR-CNN, MDGCN, and out MGLN. This could be due to the utilization of multiple types of convolution kernels, which might be beneficial to identifying the land covers like 'Corn-notill'.

4.3.2. Results on the Salinas dataset

In Table 2, we quantitatively evaluate the classification performance of different methods on the Salinas dataset and the highest values are marked in bold. Clearly, the five CNN-based baseline methods (namely, DR-CNN, CNN-PPF, HybridSN, A^2S^2K, and SSAN) achieve higher OA, AA, and Kappa coefficient than MFL, indicating the power of CNN in extracting representative features. Besides, the mean classification accuracies (i.e., AAs) of the CNN-based baseline methods can reach 95% or more, which are generally promising. Nevertheless, our proposed MGLN yields OA 98.39%, which is about 2.6% higher than that of the best CNN-based baseline method (*i.e.*, 95.82%). We infer that the performance enhancement is achieved mainly by learning the refined graphs to adaptively exploit the spectral-spatial information of HSI, which can be more effective than the fixed convolution operation of CNN. Consequently, the statistical results in Table 2 again validate the effectiveness of our proposed MGLN.

A visual comparison can be found in Fig. 4. We can observe that some pixels in class 'Grapes untrained' are misclassified into class 'Vineyard untrained'. It can be inferred that these two land-cover classes have very similar spectral signatures and are a bit difficult to distinguish. However, as is shown in Fig. 4(k), our proposed MGLN still yields a smoother visual effect than all the competitors in these two classes.

4.3.3. Results on the Houston University dataset

Table 3 shows the quantitative classification results achieved by different methods on the Houston University dataset, where the highest values are marked in bold. It is noticeable that the GCN-based methods can generally achieve good classification results, especially MDGCN and our MGLN, which can outperform all the CNN-based baseline methods in regard to OA, AA, and Kappa coefficient, respectively. This indicates the effectiveness of GCN models in HSI classification. What's more, our proposed MGLN can achieve OA increments of 3.72%, 1.63%, and 6.03% in comparison with S^2GCN, MDGCN, and miniGCN, respectively. It can be inferred that the multi-level graph learning module designed by our MGLN plays an important role in improving the classification performance of traditional GCN models.

Table 2
OA, AA (%), and Kappa Coefficient of Different Methods Achieved on Salinas Dataset.

Methods	MFL [35]	DR-CNN [7]	CNN-PPF [31]	HybridSN [32]	A^2S^2K [33]	SSAN [34]	S^2GCN [14]	MDGCN [10]	miniGCN [15]	MGLN
OA	88.36±0.22	90.35±1.14	90.52±0.77	95.82±1.54	94.22±0.88	91.79±1.70	88.39±1.01	97.25±0.87	87.38±1.34	**98.39±0.63**
AA	94.21±0.08	95.72±0.39	95.22±0.34	98.34±0.41	96.58±0.34	95.20±0.84	94.30±0.47	98.21±0.30	94.36±0.58	**98.60±0.24**
Kappa	87.06±0.24	89.26±1.26	89.46±0.85	95.36±1.70	93.56±0.97	90.89±1.87	87.10±1.12	96.94±0.96	85.99±1.48	**98.21±0.70**

Table 3
OA, AA (%), and Kappa Coefficient of Different Methods Achieved on Houston University Dataset.

Methods	MFL [35]	DR-CNN [7]	CNN-PPF [31]	HybridSN [32]	A^2S^2K [33]	SSAN [34]	S^2GCN [14]	MDGCN [10]	miniGCN [15]	MGLN
OA	82.41±0.15	88.08±1.09	87.54±1.03	90.67±1.26	91.12±0.96	67.69±4.44	89.31±1.00	91.40±0.92	87.00±0.71	**93.03±1.02**
AA	84.90±0.10	89.66±0.95	88.57±0.77	92.11±1.15	91.56±0.78	70.07±4.21	90.33±1.06	92.37±0.89	88.73±0.58	**93.65±0.94**
Kappa	80.97±0.16	87.10±1.18	86.53±1.12	89.90±1.36	90.41±1.04	65.11±4.78	88.44±1.08	90.70±1.00	85.94±0.77	**92.46±1.10**

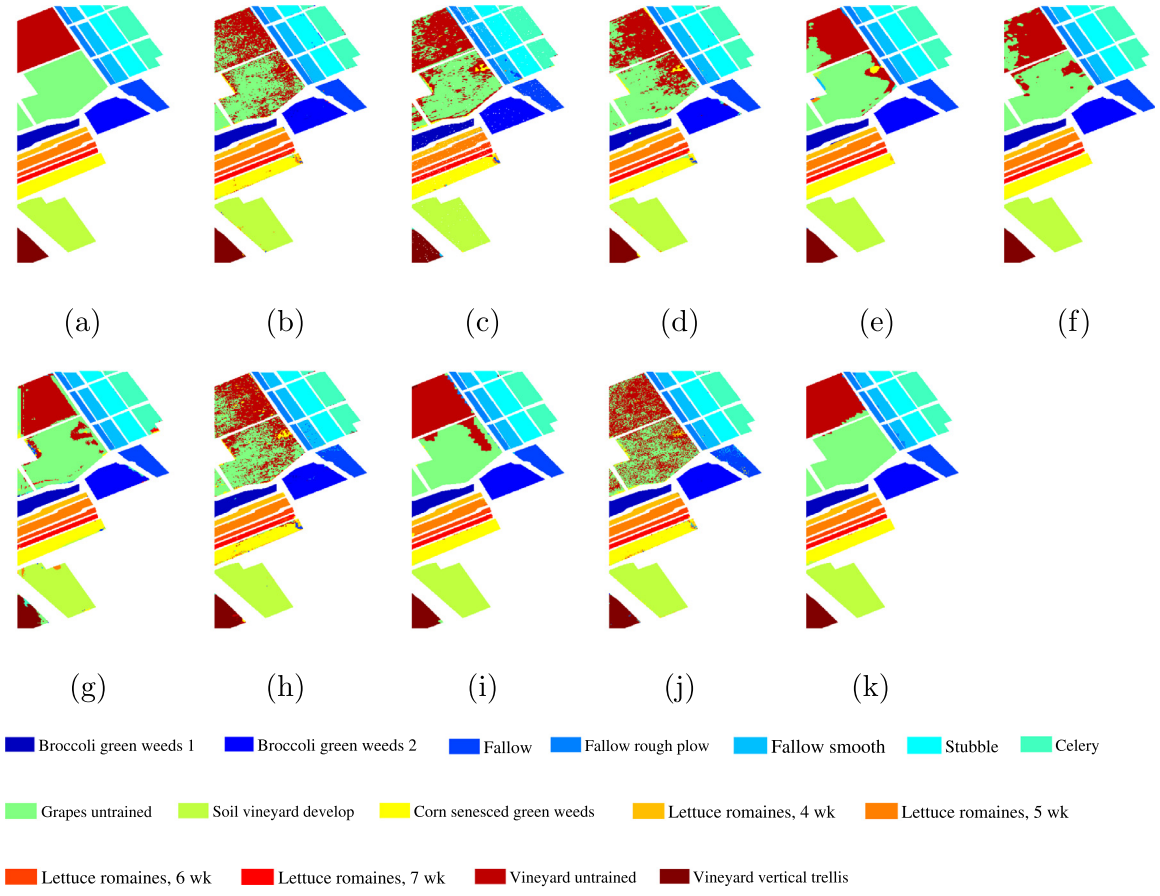


Fig. 4. Classification maps obtained by different methods on Salinas dataset. (a) Ground truth map; (b) MFL; (c) DR-CNN; (d) CNN-PPF; (e) HybridSN; (f) A^2S^2K; (g) SSAN; (h) S^2GCN; (i) MDGCN; (j) miniGCN; (k) MGLN.

Fig. 5 presents a visual comparison of the classification results produced by ten different methods on the Houston University dataset, where Fig. 5(a) is the ground-truth map. As can be seen, the proposed MGLN can generate more precise results than the competitors. Meanwhile, there are fewer errors in the zoomed-in regions, compared with the results of other methods. Consequently, we can reasonably infer that the proposed MGLN is more effective than the compared methods.

4.4. Impact of the number of labeled examples

Fig. 6 presents the classification performance of our proposed MGLN and the baseline methods under different numbers of initially labeled examples. Specifically, we vary the number of labeled examples per class from 5 to 30 with an interval of 5 and report the OA obtained by all the ten methods on three datasets, *i.e.,* the Indian Pines, the Salinas, and the Houston University. We can clearly observe from these results that MDGCN and our MGLN generally perform better than the other methods, illustrating the importance of multi-level spatial information in HSI classification. Meanwhile, our proposed MGLN allows learning improved graph structural information automatically, which is more robust than using a pre-computed fixed graph. Therefore, the proposed MGLN is able to achieve the best performance among all the methods. It is also worth mentioning that the performance of the proposed MGLN is relatively stable with the changed numbers of labeled examples. All these observations demonstrate the effectiveness and stability of our MGLN.

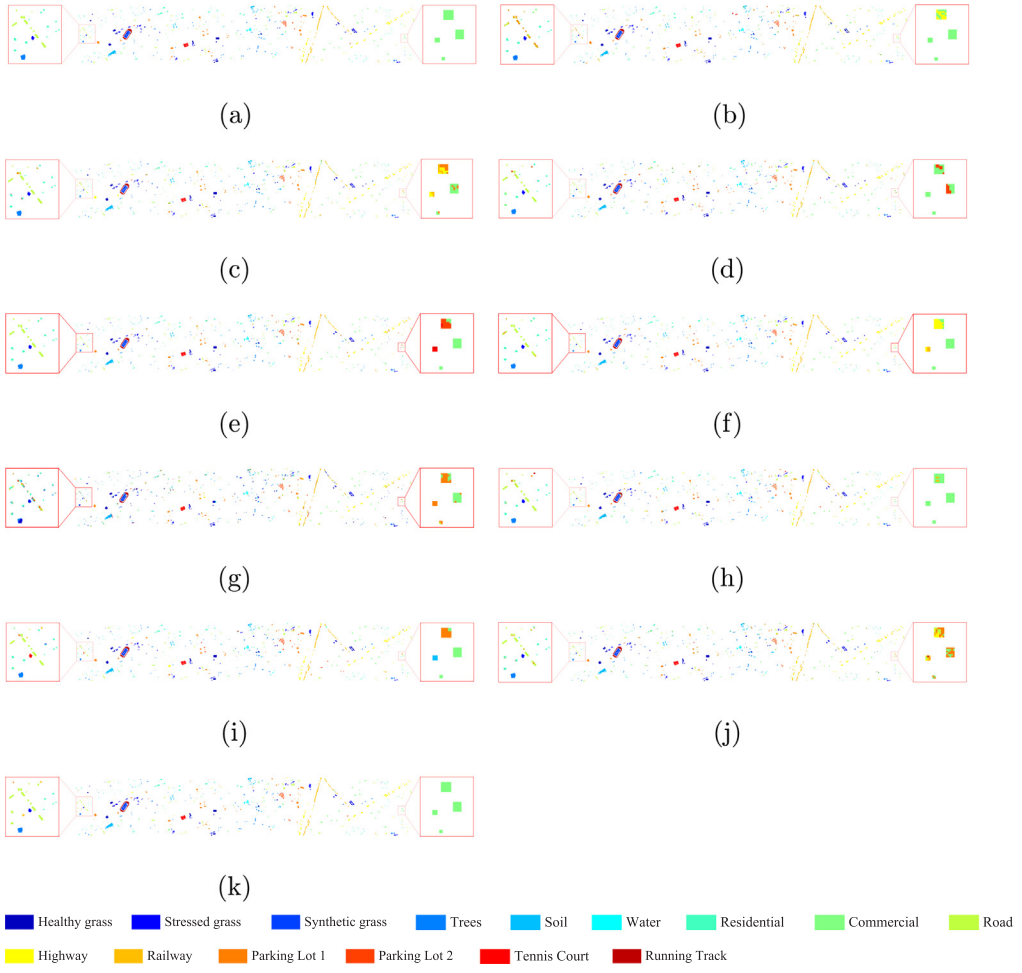


Fig. 5. Classification maps obtained by different methods on Houston University dataset. (a) Ground truth map; (b) MFL; (c) DR-CNN; (d) CNN-PPF; (e) HybridSN; (f) A^2S^2K; (g) SSAN; (h) S^2GCN; (i) MDGCN; (j) miniGCN; (k) MGLN. In (a)–(k), zoomed-in views of the regions are denoted by red boxes.

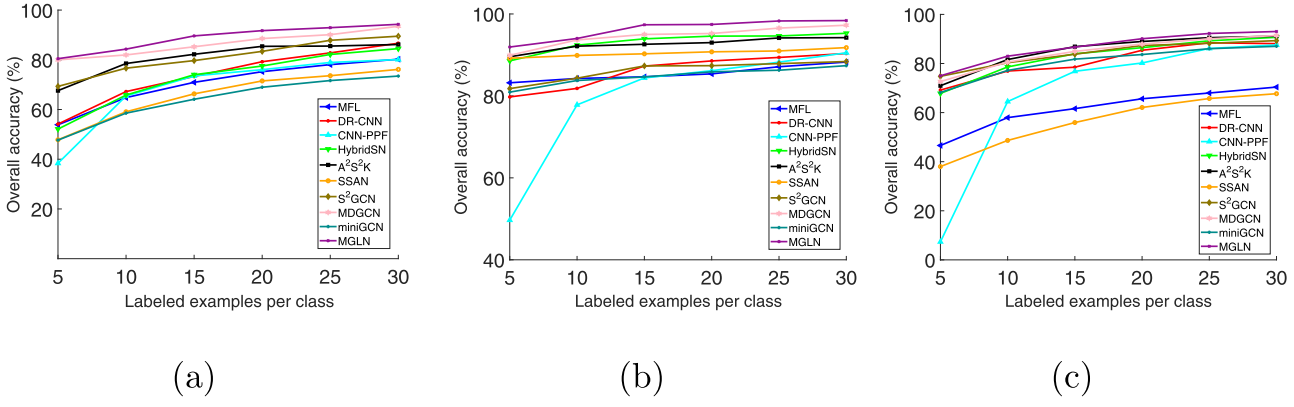


Fig. 6. Overall accuracies of various methods under different numbers of labeled examples per class. (a) Indian Pines dataset; (b) Salinas dataset; (c) Houston University dataset.

4.5. Ablation study

In our proposed MGLN, the global contextual information is incorporated to improve the representative ability of the model. Here, we investigate the ablative effect of the global-level convolution. For the sake of comparison, we report the classification result obtained without using the global contextual information, and the reduced model is denoted as 'MGLN-Loc' for simplicity. Table 4 exhibits the statistical results on the Indian Pines, the Salinas, and the Houston University datasets, respectively, where the number of labeled pixels per class is kept identical to the above experiments in Section 4.3. It is apparent that the overall classification accuracy will decrease when the component of global-level convolution is dropped. The results reveal the essential contribution of the global-level convolution in our MGLN. Figs. 7–9 visualize the embedding results of MGLN-Loc and MGLN on the Indian Pines, the Salinas, and the Houston University datasets via using t-SNE method [36]. As can be observed, the 2D projections of the embeddings gener-

Table 4
Overall classification accuracies (%) Achieved by Different Model Settings on Indian Pines, Salinas, and Houston University Datasets.

Methods	MGLN-Loc	MGLN
Indian Pines	91.92±1.66	94.27±0.92
Salinas	96.07±0.72	98.39±0.63
Houston University	88.99±1.64	93.03±1.02

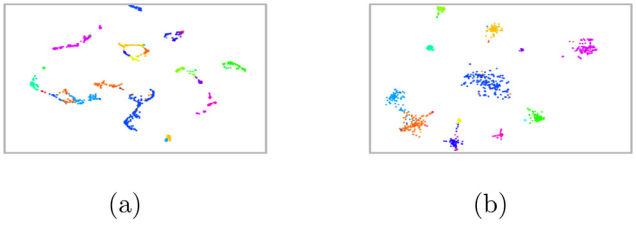


Fig. 7. t-SNE embeddings of the image regions acquired by different model settings on Indian Pines dataset. (a) MGLN-Loc; (b) MGLN.

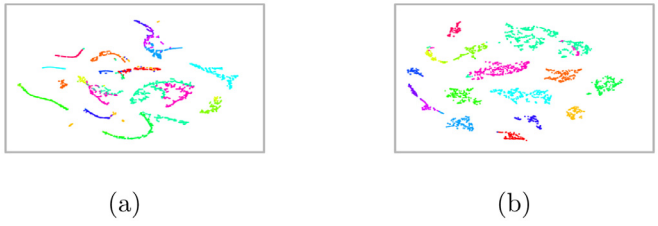


Fig. 8. t-SNE embeddings of the image regions acquired by different model settings on Salinas dataset. (a) MGLN-Loc; (b) MGLN.

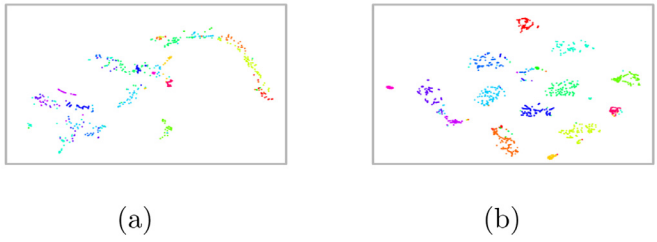


Fig. 9. t-SNE embeddings of the image regions acquired by different model settings on Houston University dataset. (a) MGLN-Loc; (b) MGLN.

ated by our MGLN (see Figs. 7(a), 8(b), and 9(b), respectively) can exhibit more coherent clusters when compared with the reduced model. Therefore, we believe that the global-level contextual information is an important supplement to the local spatial context.

4.6. Parametric sensitivity

In this experiment, we analyze the impact of the neighborhood sizes s_1 and s_2, which are used for incorporating multi-level spatial information, on the classification performance. For parametric simplicity, we fix s_1 to 1 and tune the other parameter s_2 from 2 to 5. Fig. 10 provides detailed classification results obtained by our method with different values of s_2. Observing the curves presented in Fig. 10, it is remarkable that the selection of neighborhood sizes

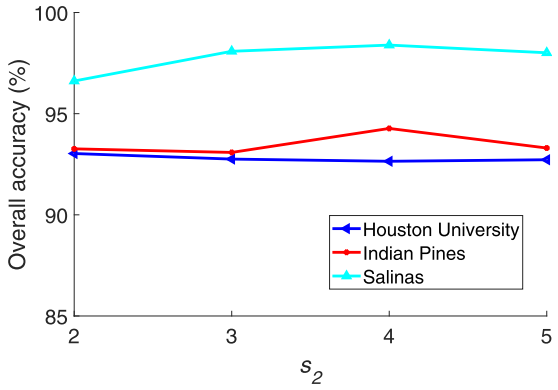


Fig. 10. Parametric sensitivity of the neighborhood size s_2 on different datasets.

is critical for our MGLN to achieve satisfactory performance. For the three hyperspectral datasets, the values of s_2 adopted by our method can be found in supplementary materials.

4.7. Running time

Table 5 shows the running time of our proposed MGLN and the eight deep models, including DR-CNN, CNN-PPF, HybridSN, A^2S^2K, SSAN, S^2GCN, MDGCN, miniGCN, on three datasets. The number of labeled pixels is kept identical to Section 4.3. All the algorithms are implemented on a server with a 3.60 GHz Intel Xeon CPU with 264 GB of RAM and a Tesla P40 GPU. In Table 5, we notice that miniGCN shows the highest efficiency among all the methods on Indian Pines and Houston University datasets, which is mainly due to the utilization of mini-batch strategy during network training. It is also notable that our proposed method is not as efficient as S^2GCN on the Indian Pines and Houston University datasets. We conjecture that the multi-level graph learning module could increase the computational cost moderately. Nevertheless, we observe that MGLN achieves greater efficiency than most of the deep-learning based methods on different dataset, which demonstrates that applying the multi-level graph learning module will not impose a substantial computational burden. Considering the superiority of our proposed MGLN in terms of classification performance, the computational cost can be acceptable. Overall, the proposed MGLN effectively trades off the classification performance and running time.

5. Conclusion

This paper presents a new graph learning network, namely MGLN, to incorporate the global contextual information within the local convolution operation. Two critical techniques contribute to the effectiveness of our MGLN in HSI classification. Firstly, the graph structural information can be adaptively learned by the network, allowing the representation learning and graph reconstruction modules to be jointly optimized. Secondly, our MGLN captures contextual information from multiple levels during graph convolution. Compared with most existing GCN-based HSI classification methods, our proposed MGLN has some advantages. For example, instead of employing a pre-computed fixed graph, our MGLN

Table 5
Running Time Comparison (In Seconds) of Different Methods. 'IP' Denotes Indian Pines Dataset, 'SA' Denotes Salinas Dataset, and 'UH' Denotes Houston University Dataset.

Methods	DR-CNN[7]	CNN-PPF [31]	HybridSN [32]	A^2S^2K [33]	SSAN [34]	S^2GCN[14]	MDGCN [10]	miniGCN [15]	MGLN
IP	2753	1495	343	501	1530	71	95	**65**	82
SA	3477	1769	**398**	1272	1880	3528	1108	594	1670
UH	2969	1223	395	1596	1279	971	360	**179**	401

allows to automatically learn the topological information of the graph, which could result in the improved graph structure. Besides, both the local spatial relevance and global dependencies among image regions have been incorporated to make devisions, which can be more comprehensive than the localized GCN models. Experimental results conducted on three widely-used hyperspectral datasets demonstrate the effectiveness and efficiency of our proposed MGLN when compared with various methods.

However, there still exist one potential limitation in our MGLN. Specifically, different GCN branches are employed to explore the contextual information at both local and global levels, which could be a bit time-consuming. In the future, we plan to incorporate the contextual information at different levels with a simplified architecture, where transformer models could be helpful.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by NSF of China (Nos. 61973162, 61876107, U1803261), NSF of Jiangsu Province (No. BZ2021013), and the Fundamental Research Funds for the Central Universities (Nos. 30920032202, 30921013114).

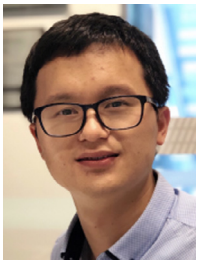
Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2022.108705.

References

[1] A.F. Goetz, Three decades of hyperspectral remote sensing of the earth: a personal view, Remote Sens. Environ. 113 (2009) S5–S16.
[2] Q. Du, C.-I. Chang, A linear constrained distance-based discriminant analysis for hyperspectral image classification, Pattern Recognit. 34 (2) (2001) 361–373.
[3] M. Fauvel, Y. Tarabalka, J.A. Benediktsson, J. Chanussot, J.C. Tilton, Advances in spectral-spatial classification of hyperspectral images, Proc. IEEE 101 (3) (2013) 652–675.
[4] W. Li, Q. Du, B. Zhang, Combined sparse and collaborative representation for hyperspectral target detection, Pattern Recognit. 48 (12) (2015) 3904–3916.
[5] C. Deng, X. Liu, C. Li, D. Tao, Active multi-kernel domain adaptation for hyperspectral image classification, Pattern Recognit. 77 (2018) 306–315.
[6] B. Du, R. Zhao, L. Zhang, L. Zhang, A spectral-spatial based local summation anomaly detection method for hyperspectral images, Signal Process. 124 (2016) 115–131.
[7] M. Zhang, W. Li, Q. Du, Diverse region-based CNN for hyperspectral image classification, IEEE Trans. Image Process. 27 (6) (2018) 2623–2634.
[8] Y. Li, W. Xie, H. Li, Hyperspectral image reconstruction by deep convolutional neural network for classification, Pattern Recognit. 63 (2017) 371–383.
[9] S. Jia, Z. Lin, M. Xu, Q. Huang, J. Zhou, X. Jia, Q. Li, A lightweight convolutional neural network for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 59 (5) (2021) 4150–4163.
[10] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, J. Yang, Multiscale dynamic graph convolutional network for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 58 (5) (2020) 3162–3177.
[11] Z. Zhang, P. Cui, W. Zhu, Deep learning on graphs: a survey, IEEE Trans. Knowl. Data Eng. 34 (1) (2022) 249–270.
[12] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the International Conference on Learning Representations, 2017.
[13] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Proceedings of the Advances in Neural Information Processing Systems, 2017, pp. 1024–1034.
[14] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, Y.Y. Tang, Spectral-spatial graph convolutional networks for semi-supervised hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 16 (2) (2019) 241–245.
[15] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, J. Chanussot, Graph convolutional networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 59 (7) (2021) 5966–5978.
[16] R. Li, S. Wang, F. Zhu, J. Huang, Adaptive graph convolutional neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
[17] X. Liu, L. Jiao, L. Li, L. Cheng, F. Liu, S. Yang, B. Hou, Deep multiview union learning network for multisource image classification, IEEE Trans. Cybern. (2020) 1–13.
[18] W. Xie, J. Lei, S. Fang, Y. Li, X. Jia, M. Li, Dual feature extraction network for hyperspectral image analysis, Pattern Recognit. 118 (2021) 107992.
[19] V. Slavkovikj, S. Verstockt, W. De Neve, S. Van Hoecke, R. Van de Walle, Hyperspectral image classification with convolutional neural networks, in: Proceedings of the ACM International Conference on Multimedia, 2015, pp. 1159–1162.
[20] Y. Luo, J. Zou, C. Yao, X. Zhao, T. Li, G. Bai, HSI-CNN: a novel convolution neural network for hyperspectral image, in: Proceedings of the International Conference on Audio, Language and Image Processing, 2018, pp. 464–469.
[21] Y. Li, H. Zhang, Q. Shen, Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network, Remote Sens. 9 (1) (2017) 67.
[22] Y. Yang, Y. Qi, Image super-resolution via channel attention and spatial graph convolutional network, Pattern Recognit. 112 (2021) 107798.
[23] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, in: Proceedings of the International Conference on Learning Representations, 2013.
[24] X. Bruce, Y. Liu, K.C. Chan, Q. Yang, X. Wang, Skeleton-based human action evaluation using graph convolutional network for monitoring alzheimer's progression, Pattern Recognit. 119 (2021) 108095.
[25] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2018.
[26] H. Zhou, T. Young, M. Huang, H. Zhao, J. Xu, X. Zhu, Commonsense knowledge aware conversation generation with graph attention, in: Proceedings of the International Joint Conference on Artificial Intelligence, 2018, pp. 4623–4629.
[27] L. Mou, X. Lu, X. Li, X.X. Zhu, Nonlocal graph convolutional networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 58 (12) (2020) 8246–8257.
[28] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[29] A. Plaza, J.A. Benediktsson, J.W. Boardman, J. Brazile, L. Bruzzone, G. Camps–Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, Recent advances in techniques for hyperspectral image processing, Remote Sens. Environ. 113 (supp-S1) (2009) S110–S122.
[30] C. Liu, J. Li, L. He, A. Plaza, S. Li, B. Li, Naive gabor networks for hyperspectral image classification, IEEE Trans. Neural Netw. Learn. Syst. 32 (1) (2021) 376–390.
[31] W. Li, G. Wu, F. Zhang, Q. Du, Hyperspectral image classification using deep pixel-pair features, IEEE Trans. Geosci. Remote Sens. 55 (2) (2017) 844–853.
[32] S.K. Roy, G. Krishna, S.R. Dubey, B.B. Chaudhuri, HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 17 (2) (2020) 277–281.
[33] S.K. Roy, S. Manna, T. Song, L. Bruzzone, Attention-based adaptive spectral-spatial kernel ResNet for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 59 (9) (2021) 7831–7843.
[34] X. Mei, E. Pan, Y. Ma, X. Dai, J. Huang, F. Fan, Q. Du, H. Zheng, J. Ma, Spectral-spatial attention networks for hyperspectral image classification, Remote Sens. 11 (8) (2019) 963.
[35] J. Li, X. Huang, P. Gamba, J.M. Bioucas-Dias, L. Zhang, J.A. Benediktsson, A. Plaza, Multiple feature learning for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 53 (3) (2015) 1592–1606.
[36] L.v.d. Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

Sheng Wan is currently pursuing the Ph.D. degree with the PCA Lab, the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, and the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, under the supervision of Dr. C. Gong.

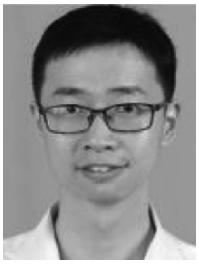


Shirui Pan received the Ph.D. degree in computer science from the University of Technology Sydney, Ultimo, NSW, Australia, in 2015. He is currently a Senior Lecturer with the Faculty of Information Technology, Monash University, Clayton, VIC, Australia. His research interests include data mining and machine learning.

Shengwei Zhong received the B.E. degree in information countermeasure technology, the M.S. and Ph.D. degrees in electronics and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2013, 2015, and 2020, respectively. Her research interests include hyperspectral image processing, remote sensing image fusion, and applications.

Jie Yang received his Ph.D. from the Department of Computer Science, Hamburg University, Germany, in 1994. Currently, he is a professor at the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China. His research interests include object detection and recognition, data mining, and medical image processing.

Jian Yang received the PhD degree from Nanjing University of Science and Technology in 2002. Now, he is a Chang-Jiang professor in the School of Computer Science and Technology of Nanjing University of Science and Technology. His papers have been cited over 25000 times in the Scholar Google.



Yibing Zhan received the B.E. degree in 2012 from the University of Science and Technology of China, Hefei, China, where he is currently working toward the Ph.D. degree in information and communication engineering. His research interests include image processing and perceptual quality assessment.

Chen Gong received dual doctoral degree from Shanghai Jiao Tong University and University of Technology Sydney in 2016 and 2017, respectively. Currently, he is a full professor in the School of Computer Science and Engineering, Nanjing University of Science and Technology.