Adaptive Cross-Attention-Driven Spatial–Spectral Graph Convolutional Network for Hyperspectral Image Classification

Jin-Yu Yang, Student Member, IEEE, Heng-Chao Li, Senior Member, IEEE, Wen-Shuai Hu, Lei Pan, and Qian Du, Fellow, IEEE

Abstract—Recently, graph convolutional networks (GCNs) have been developed to explore the spatial relationship between pixels, achieving better classification performance of hyperspectral images (HSIs). However, these methods fail to sufficiently leverage information about the spectral bands in spatial domain. As such, we propose an adaptive cross-attention-driven spatial–spectral graph convolutional network (ACSS-GCN), which is composed of a spatial GCN (Sa-GCN) subnetwork, a spectral GCN (Se-GCN) subnetwork, and a graph cross-attention fusion module (GCAFM). Specifically, Sa-GCN and Se-GCN are proposed to extract the spatial and spectral features by modeling the correlations between spatial pixels and between spectral bands, respectively. Then, by integrating attention mechanism into information aggregation of the graph, the GCAFM, including three parts, i.e., the spatial graph attention block, the spectral graph attention block, and the fusion block, is designed to fuse the spatial and spectral features, and suppress noise interference in Sa-GCN and Se-GCN. Moreover, the idea of the adaptive graph is introduced to explore an optimal graph through backpropagation during the training process. Experiments on two HSI datasets show that the proposed method achieves better performance than other classification methods.

Index Terms—Attention mechanism, feature extraction, graph convolutional networks (GCNs), hyperspectral image (HSI) classification.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are the 3-D data recording the spatial–spectral information of land covers. HSI classification, as an active research area, has received extensive attention in many fields. Recently, deep learning methods exhibit good classification performance in HSIs [1], [2].

The graph theory is an effective manner to represent the similar relationship of HSI data for revealing their intrinsic geometric properties. Specifically, graph embedding methods have been used to learn graph representation in the low-dimensional space for reducing the redundant information of HSI data [3], [4]. Inspired by the success of deep learning [5], Kipf and Welling [6] proposed a graph convolutional network (GCN) to aggregate and transform the node feature with the neighborhood structure of graph data. Moreover, aiming at the case of limited labeled samples, a spectral–spatial GCN [7] was utilized to construct a semisupervised framework for HSI classification. Subsequently, Wan et al. proposed a multiscale dynamic GCN (MDGCN) [8] and a dual-interactive GCN (DIGCN) [9] to capture the spatial information at different scales. A minibatch GCN model was developed to reduce the complexity of training [10]. Recently, the graph topological consistent was utilized to learn the underlying spatial context information of HSIs [11].

In [12], GCN was applied to model the temporal correlation between different frames in a video. Similar to the temporal relationship, the spectral correlation exists in HSI data, which has been applied for HSI classification [13]. Moreover, to produce high-resolution HSI, the spectral structures in low-spatial-resolution HSI and high-spatial-resolution multispectral images were inherited according to the construction of the spectral graph along spectral dimension [14]. Therefore, graph learning or GCN might be one of the methods to capture the spectral correlation in the spectral domain.

In recent years, with the development of attention mechanism in deep learning, it has been gradually applied for GCNs, which enables them more learnable with stronger generalization capability. Pu et al. [15] designed a multiscale attention mechanism by distinguishing the importance of different spectral bands. Moreover, by allocating weights jointly in the horizontal and vertical directions, the effectiveness of cross attention was demonstrated in [16]. Nevertheless, there are still some issues in the GCN-based models for HSI classification. First, the initial adjacent matrix may fail to reveal the effective spatial relationship. Second, the relationship of spectral bands is undersized. Third, existing GCNs usually utilize the addition, multiplication, or concatenation operations to fuse features, which fail to explore the complementary between features. As such, in this letter, an adaptive cross-attention-driven spatial–spectral graph convolutional network (ACSS-GCN) is proposed to jointly extract the spatial–spectral features for HSI classification. The contributions are summarized as follows.

1) Considering the spatial–spectral characteristic of HSI data, a dual-branch GCN-based spatial–spectral structure is proposed, containing the spatial GCN (Sa-GCN) and spectral GCN (Se-GCN) subnetworks to extract the spatial and spectral features by exploring correlations from the spatial and spectral dimensions, respectively.
2) A novel graph cross-attention fusion module (GCAFM) with a spatial graph attention block (SAGB) and a spectral graph attention block (SEGB) is developed by integrating attention mechanism into information aggregation over graph-structured data to explore the complementary information of Sa-GCN and Se-GCN.

3) With the idea of the adaptive graph, a novel ACSS-GCN framework is constructed for the spatial–spectral feature extraction and classification of HSIs.

Experimental results demonstrate the superiority of the proposed method over existing GCNs in HSI classification.

II. METHODOLOGY

The overall architecture of our ACSS-GCN framework is depicted in Fig. 1. Fig. 1(a) shows data preprocessing on the original HSI data. The backbone of the proposed framework is composed of Sa-GCN and Se-GCN, as shown in Fig. 1(b) and (c), which can capture spatial and spectral information from HSI data. Moreover, Fig. 1(d) displays the GCAFM, followed by the classification result in Fig. 1(e).

A. Data Preprocessing

Simple linear iterative clustering [8] and principal component analysis (PCA) are imposed on the original HSI data to reduce the computational complexity, as shown in Fig. 1(a). The feature x of each superpixel is defined by the average of its all pixels, and the features of all superpixels and spectral bands, respectively. A set of superpixels for the i-th spectral band is represented as $Z_{Sa} = \{x_{1i}, x_{2i}, \ldots, x_{ni}\}$, where n and s are the numbers of superpixels and spectral bands, respectively. A set of superpixels for the j-th superpixel is denoted as $Z_{Se} = \{x_{j1}, x_{j2}, \ldots, x_{jn}\}$ for $j = 1, 2, \ldots, n$.

B. Spatial–Spectral Feature Extraction

From Fig. 1, for our ACSS-GCN framework, a dual-branch GCN-based spatial–spectral structure forms its backbone, which is composed of Sa-GCN and Se-GCN to model the correlations between spatial pixels and between spectral bands. Specifically, Sa-GCN extracts spatial features by exploring the spatial relationship, while Se-GCN learns spectral information by modeling the spectral correlation.

1) Spatial Graph Convolutional Network: The relationship of pixels is distinctive in multiple spectral bands. As shown in Fig. 1(b), a spatial graph corresponding to each spectral band is constructed individually to explore the suitable spatial relationship. For convenience, a set of spatial graphs is written as $\{G_{Sa1}, G_{Sa2}, \ldots, G_{San}\}$. The spatial graph $G_{Sa1}$ of the i-th spectral band is defined as $(V_{Sa1}, E_{Sa1}, A_{Sa1})$, where each superpixel is treated as a graph node, $V_{Sa1}$ is a set of graph nodes, and $E_{Sa1}$ is a set of edges. $A_{Sa1}$ is the adjacent matrix, indicating whether each pair of superpixels is connected. The Gaussian distance is employed to measure the pairwise similarity between these superpixels. The weight of graph nodes $x_{mi}$ and $x_{hi}$ in $A_{Sa1}$ is written as

$$A_{Sa1}_{x_{mi}x_{hi}} = \begin{cases} e^{-\gamma \|x_{mi} - x_{hi}\|^2}, & \text{if } x_{mi} \in N(x_{hi}) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $N(x_{hi})$ is the set of neighbors of $x_{hi}$, which includes all superpixels directly connected to $x_{hi}$. The value of $\gamma$ is empirically set as 0.5 in the experiments.

After that, the Sa-GCN individually processes each spectral band by the corresponding spatial graph, which has two graph convolutional layers with the ReLU function. Thus, for the i-th band, the spatial information is learned by

$$H_{Sa1}^{l+1} = \text{ReLU}(L_{Sa1} H_{Sa1}^l W_{Sa1}^l) \tag{2}$$

where $L_{Sa1} = \tilde{D}_{Sa1}^{-1/2} \tilde{A}_{Sa1} \tilde{D}_{Sa1}^{-1/2}$ with $\tilde{D}_{Sa1}$ being the diagonal degree matrix of $\tilde{A}_{Sa1} = A_{Sa1} + 1$. $H_{Sa1}^l \in \mathbb{R}^{N \times (F^l/3)}$ denotes the spatial information of the i-th band in the l-th layer with $H_{Sa1}^0 = Z_{Sa1}$, and $F^l$ is the dimension of feature in the l-th layer. $W_{Sa1}^l$ represents the trainable weight for the i-th band in the l-th layer, and ReLU(·) is the activation function.

Finally, the spatial features in all bands are concatenated as

$$H_{Sa} = [H_{Sa1}, H_{Sa2}, \ldots, H_{San}] \tag{3}$$

2) Spectral Graph Convolutional Network: Considering the spectral characteristic of HSIs, an Se-GCN subnetwork is further designed. Similar to [14], a set of initial spectral graphs $\{G_{Se1}, G_{Se2}, \ldots, G_{Sen}\}$ is built to model the spectral relation by using s-nearest neighbor strategy in Fig. 1(c). Different from the spatial graph, $G_{Se1} = (V_{Se1}, E_{Se1}, A_{Se1})$ is the spectral graph of the j-th superpixel. The construction of $A_{Se1}$ is similar to (1); however, the similarity calculation is implemented between each pair of bands in the j-th superpixel.

Authorized licensed use limited to: Qian Du. Downloaded on January 13,2022 at 05:55:27 UTC from IEEE Xplore. Restrictions apply.
Then, the Se-GCN is built by stacking two convolutional layers with the ReLU function, in which each superpixel is separately processed to learn spectral information. The output for the $j$th superpixel is expressed as

$$H^1_{Se_j} = \text{ReLU} \left( L_{Se_j} H^1_{Sa_j} W^1_{Se} \right)$$  \hspace{1cm} (4)$$

where $L_{Se_j} = D^{-1/2}_{Se_j} \tilde{A}_{Se_j} D^{-1/2}_{Se_j}$, $H^1_{Se_j} \in \mathbb{R}^{1 \times F_1}$ is the spectral information of the $j$th superpixel in the $l$th layer with $H^0_{Se_j} = Z_{Se_j}$, and the dimension in the $l$th layer of Se-GCN and Sa-GCN is the same. $W^1_{Se}$ is the trainable weight in the $l$th layer shared by all superpixels, which alleviates the overfitting of the proposed network to a certain extent. Finally, similar to (3), by cascading the outputs of all superpixels, the spectral features $H_{Se} \in \mathbb{R}^{N \times F_1}$ are extracted in Se-GCN.

C. Graph Cross-Attention Fusion Module

Based on the above analysis, Sa-GCN and Se-GCN can extract different and complementary features by modeling the spatial and spectral correlations of HSIs, respectively. As such, a novel GCAFM is designed by introducing an attention mechanism into GCN to fully exploit this complementary information, whose backbone contains an SAGB, an SEGB, and a fusion block, as illustrated in Fig. 1(d).

1) Spatial Graph Attention Block: From Fig. 1(c), Se-GCN only uses the spectral information such that the spatial correlation of HSI data is not effectively explored. Therefore, an SAGB module with a spatial graph convolutional layer and two fully connected (FC) layers is designed in Fig. 1(d). Therefore, spatial attention weights are generated from the first spectral graph convolutional layer of the Se-GCN to promote the induction of effective spatial feature extraction from Sa-GCN. Specifically, the spatial graph convolutional layer is applied to learn spatial information, and two FC layers are utilized to acquire more representation information. The feature from the second FC layer is denoted by $I_{Sa} \in \mathbb{R}^{N \times F_2}$, which is further fed into a softmax function to generate the normalized spatial attention map. The spatial-enhanced features $H_1 \in \mathbb{R}^{N \times F_2}$ can be computed as

$$I_{Sa} = f_{FC}(f_{FC}(\text{ReLU}(L_{Sa} H^1_{Sa} W_{Sa})))$$

$$H_1 = H_{Sa} \odot \text{softmax}(I_{Sa})$$  \hspace{1cm} (5)$$

where $H^1_{Sa} \in \mathbb{R}^{N \times F_1}$ is output of the first layer in Se-GCN. $f_{FC}(\cdot)$ is the FC layer, $\odot$ is the elementwise multiplication operation, and softmax(·) function is used for normalization.

2) Spectral Graph Attention Block: The special structure of Sa-GCN makes it contain the primary spectral relationship. Therefore, an SEGB module is built to generate spectral weight coefficients for adaptively selecting the important spectral features from the Se-GCN. From Fig. 1(d), the SEGB contains a spectral graph convolutional layer, two FC layers, and a softmax function, which, however, calculates the spectral attention map along with the spectral dimension from the output of the first spatial graph convolutional layer in Sa-GCN. In addition, the output of the second FC layer is expressed as $I_{Se} \in \mathbb{R}^{N \times F_2}$. Therefore, the spectral-enhanced features $H_2 \in \mathbb{R}^{N \times F_2}$ are written as

$$I_{Se} = f_{FC}(f_{FC}(\text{ReLU}(L_{Se} H^1_{Se} W_{Se})))$$

$$H_2 = H_{Se} \odot \text{softmax}(I_{Se})$$  \hspace{1cm} (6)$$

where $H^1_{Sa} \in \mathbb{R}^{N \times F_1}$ are the features from the first layer in Sa-GCN.

Finally, in the fusion block, the addition and concatenation operations are applied to fuse the spatial and spectral attention features of Sa-GCN and Se-GCN. Inspired by the residual learning, two fusion methods can be written as

$$H_{add} = (H_1 \oplus H_{Sa}) \oplus (H_2 \oplus H_{Se})$$

$$H_{con} = [H_1 \oplus H_{Sa}, H_2 \oplus H_{Se}]$$  \hspace{1cm} (7)$$

where $\oplus$ is the elementwise addition. $H_{add} \in \mathbb{R}^{N \times F_2}$ and $H_{con} \in \mathbb{R}^{N \times 2F_2}$ mean the extracted spatial–spectral features by these two methods.

D. Adaptive Graph Refinement

The performance of graph convolution operation mainly depends on the quality of graph structure, which is important to explore the optimal graph for HSI classification. Due to the effects of the noise and complex scenes on HSI data, the spatial and spectral graphs initially constructed in (1) are not accurate, which are not completely suitable for the HSI classification. As such, the adaptive matrix $W_p$ is designed to dynamically refine these two graphs, and these corresponding adjacency matrices can be updated as

$$A_o = A_{in} + \beta W_p A_{in}$$  \hspace{1cm} (8)$$

where $A_{in} = A_{Sa}$ or $A_{in} = A_{Se}$, and $\beta$ is a tuning parameter. Specifically, each spatial graph with its own $W_p$ makes the model more flexible, while all spectral graphs share the same $W_p$ to avoid the overfitting, all of which are randomly initialized. Similar to [8], by optimizing the cross-entropy loss function with full-batch gradient descent, all $W_p$ together with other trainable parameters of ACSS-GCN can be updated during training, thus producing the refined spatial and spectral graphs, and extracting the more effective spatial–spectral features for HSI classification.

III. EXPERIMENTAL RESULTS

In this section, the performance of the ACSS-GCN is validated on two public HSI datasets, i.e., Indian Pines and University of Pavia. The Indian Pines dataset recorded northwestern Indiana with the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in 1992, which is made up of $145 \times 145$ pixels with 200 bands. The University of Pavia data were captured in the Pavia University in Italy by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor in 2001, consisting of $610 \times 340$ pixels and 103 spectral bands. In the experiments, 30 pixels of each class are selected as the training set, while 15 pixels will be extracted for the class where the number of samples does not reach 30. In addition, the overall accuracy (OA), average accuracy (AA), and kappa coefficient ($\kappa$) are used to study the performance of all methods.

To evaluate the performance of our proposed method, seven methods, i.e., SVM [17], 2-D CNN [1], GCN [6], FuNet [10], AMDPCN [15], MDGCN [8], and DIGCN [9], are utilized for comparison. Moreover, our ACSS-GCN model with two fusion strategies (i.e., addition and concatenation) is represented as ACSS-GCN-A and ACSS-GCN-C, respectively.
TABLE I
SENSITIVITY ANALYSIS UNDER DIFFERENT VALUES OF THE PARAMETER $\beta$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>0</th>
<th>0.0001</th>
<th>0.0005</th>
<th>0.001</th>
<th>0.005</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Pines</td>
<td>64.40</td>
<td>85.78</td>
<td>92.30</td>
<td>94.06</td>
<td>95.02</td>
<td>94.19</td>
<td>94.64</td>
<td>94.59</td>
</tr>
<tr>
<td>University of Pavia</td>
<td>69.97</td>
<td>91.93</td>
<td>95.77</td>
<td>96.50</td>
<td>96.38</td>
<td>94.40</td>
<td>93.41</td>
<td>93.63</td>
</tr>
</tbody>
</table>

A. Parameter Settings
In the experiments, the spectral dimension after PCA is set to 20. In addition, dropout $= 0.5$ for each graph convolutional layer to alleviate the overfitting. For Sa-GCN and Se-GCN, the dimensions of the two graph convolutional layers are set as 40 and 20, respectively. The output dimensions of the graph convolutional layer and two FC layers in the SEGB and SAGB modules are set to 40, 25, and 20, respectively. Moreover, an FC layer and a softmax function are applied to classify the HSI samples. To eliminate the deviation caused by random initialization, the average value after five repetitions is given for each quantitative metric. The epoch is set to 3000. The adaptive momentum optimizer with a learning rate of 0.005 is used to optimize the loss function of cross-entropy for the proposed model.

Besides, we further investigate the effect of the tuning parameter $\beta$ on the classification performance of HSIs. Table I reports the OA results for different $\beta$, from which we can observe that the proposed model can obtain the quasi-optimal performance when $\beta$ is set to 0.005 and 0.01 for these two HSI datasets, respectively.

B. Classification Performance
The classification results of all methods on these two HSI datasets are reported in Tables II and III. We can find that the proposed method outperforms other methods. On the one hand, Sa-GCN and Se-GCN effectively learn the spatial and spectral correlations of HSI data. On the other hand, GCAFM adaptively explores complementary information of the extracted spatial and spectral features, and suppresses noise interference in Sa-GCN and Se-GCN. From Tables II and III, compared with the second best classification performance achieved by DIGCN or MDGCN, the OA results of ACSS-GCN-A on two datasets are increased by 1.00% and 1.28%, respectively. In particular, the comparison between two fusion strategies reveals that ACSS-GCN-A can achieve the best classification results and bring at least 0.45% increments in OA over the ACSS-GCN-C. In addition, the total sum of the training and testing computational time [seconds (s)] of all methods on two datasets is summarized in the last row of Tables II and III, from which, compared with other GCN-based methods, the GCN method [6] consumes more time owing to the participation of all pixels in training step; however, our method only costs a little more time to construct the spectral graph and run the GCAFM, but achieves the best performance for HSI classification.

To analyze the classification performance more intuitively, the visualized results of all methods on the Indian Pines dataset are shown in Fig. 2, from which it can be observed that the classification map of our proposed model is closest to the ground-truth map, and there are least noise and misclassifications. Particularly, for the GCN, FuNet, and AMDPCN methods, the pepper-noise-like mistakes in certain regions can be observed due to the loss of spatial context information, leading to poor classification results. The above experimental results further illustrate the superiority of ACSS-GCN.

C. Sensitivity Comparison and Analysis Under Small Samples
To further illustrate the classification performance under the small number of training samples, 5, 10, 15, 20, and 30 samples of each class are randomly selected from these two HSI datasets for training. Fig. 3 shows the OA curves with different numbers of training samples. As expected, the performance of all methods is improved with the increase of training samples. The ACSS-GCN method outperforms other methods.
methods, which verifies the advantages of our ACSS-GCN method. It is noted that the ACSS-GCN-A is superior to the ACSS-GCN-C model in most cases, which may be because the concatenation operation generates more redundant features than the addition operation affecting the final classification performance. These experimental results also show that the ACSS-GCN model is superior to other considered models.

D. Ablation Study

To highlight the effectiveness of Se-GCN, Sa-GCN, and GCAFM in our ACSS-GCN-A, detailed ablation studies are conducted to see how they contribute to the classification accuracy, where the ASS-GCN-A is a variant of ACSS-GCN-A by replacing the GCAFM with the elementwise addition operation. From Table IV, by directly adding the features of Se-GCN and Sa-GCN, the ASS-GCN-A can extract the joint spatial–spectral features for good classification performance of HSIs. Furthermore, by introducing the idea of cross attention, a GCAFM is designed to fully explore the complementary information between Se-GCN and Sa-GCN. Compared with the ASS-GCN-A, GCAFM can bring 0.76% and 0.49% gains to the ACSS-GCN-A for these two datasets, respectively. Experimental results show that the Se-GCN and Sa-GCN are effective and complementary, and the joint learning of spatial–spectral information can further improve the performance of HSI classification. Besides, the GCAFM can be beneficial for improving the classification accuracy of the ACSS-GCN.

IV. CONCLUSION

In this letter, a novel ACSS-GCN method has been proposed for HSI classification. First, a dual-branch GCN-based spatial–spectral structure is proposed as a backbone of ACSS-GCN to jointly learn the spatial and spectral information of HSI data. Then, the GCAFM is designed to explore complementary information of Sa-GCN and Se-GCN for better HSI classification. Finally, an adaptive graph is proposed to dynamically update the spatial and spectral graphs during the backpropagation of the whole method. Experimental results on two HSI datasets show that our method offers better classification performance than other GCN-based methods.

REFERENCES