

A Hyperspectral Image Classification Based on Spectral Band Graph Convolutional and Attention-Enhanced CNN Joint Network

XU Chenjie, LI Dan*, KONG Fanqiang

College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China

(Received 22 June 2025; revised 30 September 2025; accepted 2 October 2025)

Abstract: Hyperspectral image (HSI) classification is crucial for numerous remote sensing applications. Traditional deep learning methods may miss pixel relationships and context, leading to inefficiencies. This paper introduces the spectral band graph convolutional and attention-enhanced CNN joint network (SGCCN), a novel approach that harnesses the power of spectral band graph convolutions for capturing long-range relationships, utilizes local perception of attention-enhanced multi-level convolutions for local spatial feature and employs a dynamic attention mechanism to enhance feature extraction. The SGCCN integrates spectral and spatial features through a self-attention fusion network, significantly improving classification accuracy and efficiency. The proposed method outperforms existing techniques, demonstrating its effectiveness in handling the challenges associated with HSI data.

Key words: hyperspectral classification; spectral band graph convolutional network; attention-enhance convolutional network; dynamic attention; feature extraction; feature fusion

CLC number: TP751

Document code: A

Article ID: 1005-1120(2025)S-0102-19

0 Introduction

Hyperspectral remote sensing integrates spatial and spectral imaging technology, capturing continuous spectrum image data that encompasses both spatial and spectral information of the ground^[1]. Hyperspectral image (HSI) classification is one of crucial research in hyperspectral interpretation, and the category division map obtained through the classification technology has played a significant role in the fields of military target recognition^[2-3], land resources^[4-6] and situation monitoring^[7]. However, HSI often contains redundant information, presenting challenges such as the same objects with different spectrum and the Hughes phenomenon^[8]. These factors can contribute to noise in the classification results. And challenges such as mixed pixel problems and training under conditions with limited samples add complexity to the classification task.

Traditional HSI classification algorithms are based on machine learning, including K-nearest neighbor (KNN)^[9], decision tree^[10], random forest^[11], and support vector machine(SVM)^[12-13]. Feature extraction methods like principal component analysis (PCA)^[14], independent component analysis (ICA)^[15], and maximum noise fraction (MNF)^[16] focus on dimensionality reduction. While spatial feature extraction methods like extended morphological profiles (EMPs)^[17-19] enhance neighborhood pixel structure. Dallamura et al.^[20] used the extended morphological attribute profile (EMAP) to describe the HSI features in a multi-level and multi-attribute manner. The drawback of such methods lies in extraction of single feature.

Deep learning (DL)-based classification methods have demonstrated significant advantages^[21-23]. Chen et al.^[24] combined autoencoders with logistic regression, and then introduced deep belief networks

*Corresponding author, E-mail address: danli@nuaa.edu.cn.

How to cite this article: XU Chenjie, LI Dan, KONG Fanqiang. A hyperspectral image classification based on spectral band graph convolutional and attention-enhanced CNN joint network[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2025, 42(S): 102-120.

<http://dx.doi.org/10.16356/j.1005-1120.2025.S.009>

(DBN)^[25] for classification to enhance performance. Since then convolutional neural networks (CNNs) are widely employed to enhance classification performance^[26-28]. The one-dimensional CNN (1DCNN)^[29] treats the pixels of each band as a sequence. In contrast, the two-dimensional CNN (2DCNN)^[30] regards the pixels of each spectral band in the HSI data as a feature matrix for classified. Chen et al.^[31] established a deep finite element model using three-dimensional CNN (3DCNN), which realized the extraction of spatial-spectral features and higher classification accuracy. More CNN-based methods are employed to extract a variety of features. Guo et al.^[32] tried to use CNN to extract spectral features and multi-scale spatial features for HSI classification. Zhong et al.^[33] proposed the spectral-spatial residual network (SSRN). It constructs both spectral and spatial residual module to enhance the gradient backpropagation and facilitate the extraction of deeper spectral features extraction. 3DCNN can extract more feature information, while it increases the computational cost. To balance spectral and spatial characteristics, Roy et al.^[34] introduced a HybridSN network that integrates 2DCNN and 3DCNN to extract richer spatial-spectral features and reduce computational complexity. The convolution operation in CNN is implemented by shifting the convolution kernel, so the use of CNN is limited to Euclidean data, that is, regularly arranged data such as images and text. The constraints of the convolution kernel pose challenges for the CNN in capturing the relationship features between pixels and the contextual features of the HSI.

Graph neural networks (GNNs) have advanced in processing special graph-structured data such as images and speech^[35-38]. Kipf et al.^[39] proposed the graph convolutional network (GCN) model, which has been instrumental in extracting structural relationships in HSI data^[40-41]. Qin et al.^[42] utilized a spatial-spectral GCN for HSI classification, focusing on vertex relationships influenced by spectral and spatial similarities. He et al.^[43] further enhanced classification accuracy with a dual graph convolution network. To address computational efficiency, Hong et al.^[44] proposed miniGCN, which trained in mini-batch fashion and extracted the features in the sub-

graph of the entire HSI. Wan et al.^[45] presented a multi-scale dynamic graph convolutional neural network (MDGCN), which employed superpixel segmentation to reduce graph size. Superpixel-based graph has been further used to enhance computational efficiency and integrate weighted features^[46-47]. Attention mechanisms have also been incorporated into GCNs^[48-51]. Velickovic et al.^[52] proposed the graph attention network (GAT), which integrated attention mechanisms into GCNs, enabling variable weighting of vertex relationships and more effectively capturing global dependencies. Wang et al.^[53] used GAT to construct a multi-scale pyramid in HSI's spectral dimension for feature extraction. Despite these advancements, GCN-based HSI classification methods face challenges in balancing computational costs and effective utilization of correlation features. Spatial-based graphs in GCNs, which leverage long-range spatial information, require substantial training samples and computational resources, especially for high-resolution HSI data. Methods focusing on reducing graph structures for network training have improved efficiency. However, they may overlook pixel-level features and fail to fully utilize spectral channel correlations in HSI.

This paper introduces the spectral band graph convolutional and attention-enhanced CNN joint network (SGCCN) for HSI classification. The method constructs a spectral graph from the HSI's band dimension to model long-distance relationships more efficiently than spatial-based graphs. It includes a lightweight spectral band graph convolutional network with attention (SGCN-A) designed to exploit spectral similarities and reduce computational complexity. The SGCN-A features three feature extraction blocks, each with a dynamic attention module (DAM) that adjusts feature weights from spectral band graph convolutions to emphasize discriminative features. Concurrently, an attention-enhanced multi-level CNN (A-MLCN) extracts local spatial features from 2D neighboring patches. It captures pixel-level contextual information along the spatial dimension through multi-level convolution, thereby complementing spectral features. The attention module is incorporated in parallel in each layer to

concentrate on crucial feature and improve representation performance of features for different categories across different areas of the HSI. Finally, the spectral similarity features and spatial context features are fused through the self-attention mechanism within the feature fusion network. This process enhances feature diversity, ensures stable statistical properties, and improves computational efficiency. Experiments conducted on three HSI data sets demonstrate that SGCCN improves efficiency and robustness, and reduces computational costs compared with state-of-the-art methods. In summary, the main innovative contributions of our proposed SGCCN are as follows.

(1) A novel spectral band graph convolutional network with attention is proposed to extract global spectral similarity features between bands and enhance the long-distance feature modeling capabilities, which significantly reduces computational cost of network by constructing a band graph.

(2) An innovative dynamic attention-enhanced multi-level convolutional network to effectively captures local spatial context features from different receptive field by multi-scale kernel, which utilizes local perception and dynamic focus on discriminative features to address information redundancy.

(3) A robust self-attention feature fusion mechanism is employed to enhance classification performance by effectively learning, fusing and improving discrimination ability of spatial-spectral features with limited samples.

1 Related Work

In this section, we introduce some relevant knowledge, covering the network structure of CNN and the fundamental definitions, notations, and network structure of GCN.

1.1 2D convolution neural network

The utilization of convolutional neural networks has become prevalent in the realm of computer vision, enhancing feature extraction capabilities^[54]. The CNN architecture incorporates local connections and shared weights. In HSI classification, 2DCNN is always used for extracting spatial features.

A comprehensive 2DCNN comprises a convolution layer, a pooling layer, and a fully connected layer. The presentation of a neuron F_{ij}^{xy} at position (x, y) within the j th feature map of the i th layer is expressed as follows

$$F_{ij}^{xy} = h \left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} W_{ijm}^{pq} F_{(i-1)m}^{(x+p)(y+q)} \right) \quad (1)$$

$$h(x) = \text{ReLU}(x) = \max(0, x) \quad (2)$$

where $h(x)$ denotes the ReLU activation function, b_{ij} the bias of the j th feature map in the i th layer, m the number of feature map in the $(i-1)$ th layer connected to the j th feature map, W_{ijm}^{pq} the weight at position (p, q) connected to the m th feature map; P_i and Q_i represent the height and the width of the 2D convolution kernel, respectively.

In the pooling layer, the preceding feature map undergoes sub-sampling to decrease the spatial size. Then the fully connected (FC) layer integrates all the extracted features and the classification results are obtained using the SoftMax classifier^[55]. However, the feature extracted from CNN captures local spatial information. Since the network requires the construction of a neighborhood window around a central pixel for input, the HSI is divided into several windows, and each window is relatively independent. This implies that only spatial information contained within the neighborhood is obtained, which may not be sufficient to enhance the classification effect.

1.2 Graph convolution network

Graphs serve as a modeling tool for capturing pairwise relationships between objects. As shown in Fig.1, a graph is made up of vertices, which are connected by edges^[56]. Consider the undirected graph denoted by an ordered pair $G=(V, E)$, where V represents a set of vertices $v_i \in V$ and E a set of edges $(v_i, v_j) \in E$.

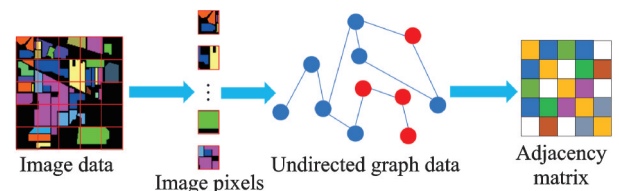


Fig.1 Illustration of graph data and adjacency matrix

Adjacency matrix $A \in \mathbf{R}^{n \times n}$ is a symmetrical matrix that records the relationship between vertices, where n is the number of the vertices. The degree matrix D is a diagonal matrix, where the diagonal elements are the degrees of each vertex, indicating the number of edges associated with each vertex.

The continuous aggregation of adjacent vertices enables the graph convolution layer to facilitate the transfer of neighborhood relationships. The graph convolution formula is represented as the multiplication of the input graph data with a filter g_θ , shown as

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

where U denotes the eigenvectors matrix of the normalized graph Laplacian matrix $L = U\Lambda U^T$, and Λ a diagonal matrix of its eigenvalues. Here, the operator “ $\star$ ” represents the graph convolution operation, which performs filtering in the spectral domain. In addition, g_θ can be thought as a function of the eigenvalues of L , i.e., $g_\theta(\Lambda)$. However, it will cost a lot to decompose the eigenvalue. To address the problem, the Chebyshev polynomial is employed to approximate and restrict the layer-wise convolution operation to 1.

$$g_\theta \star X \approx \theta'_0 X + \theta'_1 (L - I) X = \theta'_0 X - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X \quad (4)$$

To prevent overfitting and reduce operational cost in each layer, the two parameters θ'_0 and θ'_1 are constrained to one single parameter $\Theta = [\theta'_0, -\theta'_1]$. The expression is rewritten as follows

$$g_\theta \star X \approx (I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) X \Theta \quad (5)$$

where the largest eigenvalue of $I + D^{-1/2} A D^{-1/2}$ is 2 and Θ the matrix of filter parameters. To address the problem of numerical instabilities and exploding or vanishing gradients, renormalization is set for adjacency matrix $\tilde{A} = A + I$ and the degree matrix $\tilde{D}_{i,i} = \sum_j \tilde{A}_{i,j}$. Consequently, the commonly used graph convolution formula can be improved to

$$Z = g_\theta \star X \approx \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta \quad (6)$$

where Z is the convolved signal matrix.

The GCN introduces filters based on principles of graph signal processing, providing it with a robust mathematical foundation. In HSI classification, the relationship between pixels can be obtained by GCN. However, in many models, constructing a graph for all pixels in HSI. Given HSI data $I \in \mathbf{R}^{M \times N \times B}$, a adjacency matrix with the size of $(M \times N) \times (M \times N)$ is constructed. This will encounter a high computational cost particularly when applied to extensive graph structures, posing a significant limitation.

2 Proposed Method

This section provides detailed information about the three modules of the proposed SGCCN, as depicted in Fig.2. It consists of three main components. First, a spectral band graph convolution with attention is used to capture spectral similarity features. Second, a spatial context extraction branch

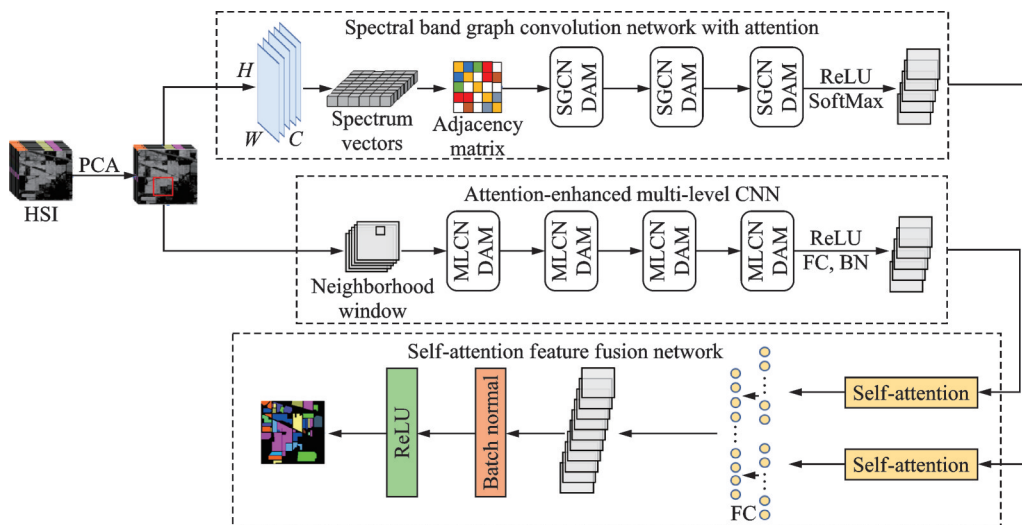


Fig.2 Overview of the spectral band graph convolutional and attention-enhanced CNN joint network

employs an attention-enhanced multi-level convolution network. Finally, a self-attention-based feature fusion and classification network integrates the extracted features for final prediction.

2.1 Spectral band graph convolutional network with attention

In this section, we introduce a novel SGCN-

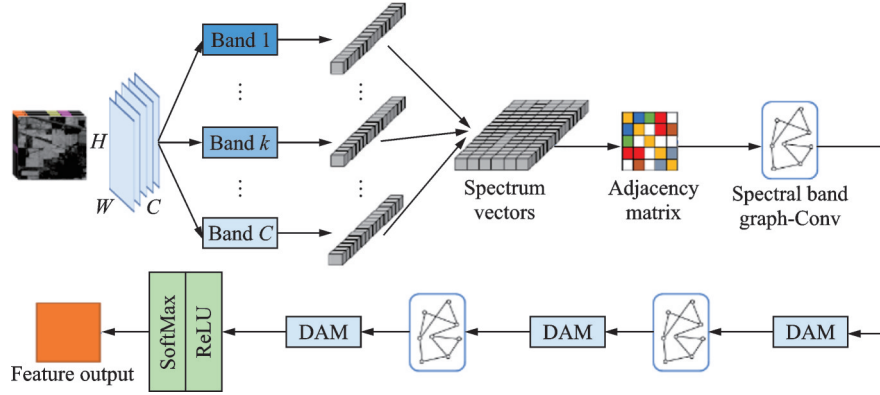


Fig.3 Flowchart of the spectral band graph convolutional network with attention

Let $I \in \mathbb{R}^{M \times N \times B}$ represents the original HSI dataset, where M , N , and B denote the height, width, and channel bands of the HSI, respectively. Firstly, PCA^[14] is used to reduce the dimensionality of the HSI data to $X \in \mathbb{R}^{M \times N \times C}$, effectively removing redundant information and lowering computational complexity. Here, C is the number of spectral dimensions after PCA.

Different from commonly used graph construction methods, SGCN-A adopts a spectral channel perspective in the band graph construction stage. This approach captures global spectral relationships within the HSI. The SGCN-A framework begins by transforming the HSI cube from the channel dimension into a 2D matrix $X \in \mathbb{R}^{S \times C}$, where $S = M \times N$. Each frame of the HSI is denoted as a band vector, represented as a vertex $X_i \in X$, with a size of $M \times N \times 1$, and forms the vertex set of the HSI band graph. This representation facilitates the extraction of spectral information from limited training pixels.

Edges between band vector vertices represent spectral relations in two dimensions. Given HSI's high-dimensional nature, we compute the cosine similarity between any two band vectors to accurately capture spectral similarity features, shown as

A , as depicted in Fig.3. This novel network introduces a new band graph data construction method. It includes a spectral band graph convolution network and incorporates a dynamic attention module. Together, these components can exploit spectral similarity features between any bands, even with small training samples. Additionally, they significantly reduce computational complexity.

$$\cos(X_i, X_j) = \frac{\langle X_i, X_j \rangle}{\|X_i\|_2 \|X_j\|_2} \quad i, j = 1, 2, \dots, C \quad (7)$$

where $\cos(X_i, X_j) \in [-1, 1]$. For the HSI classification task, the correlation of spectral channel forms an undirected graph. Thus, cosine similarities are normalized and constrained to $A_{i,j} \in [0, 1]$ to construct the adjacency matrix A , shown as

$$A_{ij} = \frac{1 - \cos(X_i, X_j)}{2} \quad i, j = 1, 2, \dots, C \quad (8)$$

Since the cosine similarity of two vectors to each other is the same, i.e., $A_{i,j} = A_{j,i}$. Additionally, diagonal elements of the matrix that represent the similarity between a vector and itself are 0. From these, the complete adjacency matrix A of band vectors can be obtained as

$$A = \begin{pmatrix} A_{11} & \dots & A_{1C} \\ \vdots & & \vdots \\ A_{C1} & \dots & A_{CC} \end{pmatrix} = \begin{pmatrix} 0 & \dots & A_{C1} \\ \vdots & & \vdots \\ A_{C1} & \dots & 0 \end{pmatrix} \quad (9)$$

By treating each band vector (i.e., each frame of HSI) as a vertex, the adjacency matrix size is reduced from $(M \times N) \times (M \times N)$ to $C \times C$, significantly lowering computational costs.

To mitigate the influence of the largest similarity between a vector and itself in convolution, the diagonal elements of A are set to 1. Then the adjacency matrix is transformed as follows

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I} = \begin{pmatrix} 1 & \cdots & \mathbf{A}_{c1} \\ \vdots & & \vdots \\ \mathbf{A}_{c1} & \cdots & 1 \end{pmatrix} \quad (10)$$

where $\mathbf{I}$ is the identity matrix. Apparently, the degree matrix that contains the degrees of each vertex $\mathbf{A}_{i,j}$ can be calculated as $\tilde{\mathbf{D}}_{i,i} = \sum_j \tilde{\mathbf{A}}_{i,j}$. Then the normalized similarity matrix can be obtained as

$$\tilde{\mathbf{S}} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \quad (11)$$

The cornerstone of the SGCN-A lies in its unique spectral band graph construction method. By treating each spectral band of the HSI as a graph vertex, this innovative network captures global spectral relationships for accurate classification.

The SGCN-A for the graph-structured HSI data can be formulated as

$$\mathbf{X}^{i+1} = h(\tilde{\mathbf{S}}\mathbf{X}^i\mathbf{W}^i + b^i) \quad (12)$$

where $\mathbf{X}^i$ represents the output in the i th layer, $h(\cdot)$ the graph activation function, $\mathbf{W}^i$ the weights, and b^i the bias of the i th layer.

A lightweight three-layer SGCN-A architecture is designed to enhance spectral correlation feature extraction. Multi-SGCN-A layers are capable of aggregating information from neighboring vertices, thereby improving the representation of each vertex in the spectral space. In the first layer, the nonlinear activation function ReLU^[57] is used in conjunction with convolution to preserve and enhance the spectral correlations, shown as

$$\mathbf{X}^{i+1} = \text{ReLU}(\tilde{\mathbf{S}}\mathbf{X}^i\mathbf{W}^i + b^i) \quad (13)$$

where $\mathbf{X}^{i+1}$ is the output feature map in this layer. In addition, a dropout process is employed to prevent overfitting.

A key innovation of SGCN-A is the integration of a DAM within the channel graph convolution layers. The DAM based on SimAM^[58] enhances the discriminative power of the network by dynamically focusing on the most informative spectral features. The feature maps can be seen as neurons and the optimal closed-form energy function for every neuron is formulated as follows

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (14)$$

where $\hat{\mu}$ is the average value of the HSI feature input with the convolution and $\hat{\sigma}^2$ the variance. The final output is gotten by an element-product with a

sigmoid function, which is to control the output range of the attention vector.

$$\tilde{\mathbf{X}} = \text{sigmoid}\left(\frac{1}{\mathbf{E}}\right) \odot \mathbf{X} \quad (15)$$

where $\mathbf{E}$ denotes the set of all e_i^* grouped across spatial and spectral dimensions. And $\mathbf{E}$ is restricted by sigmoid function to prevent too large.

The output of the first layer in the SGCN-A network is

$$\tilde{\mathbf{X}}^{i+1} = F^{\text{DA}}(\text{ReLU}(\tilde{\mathbf{S}}\mathbf{X}^i\mathbf{W}^i + b^i), \mathbf{A}_{\text{DA}}) \quad (16)$$

where $F^{\text{DA}}(\cdot)$ represents the DAM and $\mathbf{A}_{\text{DA}}$ the attention map. Each SGCN-A layer enhances spectral information extraction and utilization.

The subsequent layers are cascaded and identical, except without the ReLU activation function. These layers refine features to highlight the most discriminative patterns for classification.

In the final stage of SGCN-A, the Softmax function is adopted to transform spectral similarity features into a probability distribution for classification, shown as

$$\tilde{\mathbf{X}}_{\text{SGCN-A}}^{i+1} = \text{Softmax}(F^{\text{DA}}(\tilde{\mathbf{X}}^i, \mathbf{A}_{\text{DA}})) \quad (17)$$

This network ensures that the extracted features are compatible with spatial features from the dynamic attention-enhanced multi-level CNN. This alignment sets the stage for a robust and comprehensive classification process.

2.2 Attention-enhanced multi-level CNN

In this innovative branch network, we introduce A-MLCN designed for advanced spatial feature extraction in HSI. This network designs multi-level convolutions by integrating DAM, which adaptively emphasizes critical features for superior representation and discrimination. The flowchart of the A-MLCN network is depicted in Fig.4.

To address the high dimensionality and redundancy in HSI data, PCA is adopted for reducing the original HSI data $\mathbf{I} \in \mathbf{R}^{M \times N \times B}$ to $\mathbf{X} \in \mathbf{R}^{M \times N \times C}$. This step not only distills representative low-dimensional spectral features but also significantly lowers computational costs.

Prior to the 2D convolution operation, neighborhood patches $\mathbf{P} \in \mathbf{R}^{T \times T}$ are extracted from the HSI data. Each pixel is centered within a $T \times T$ square window, defining the patch category by the

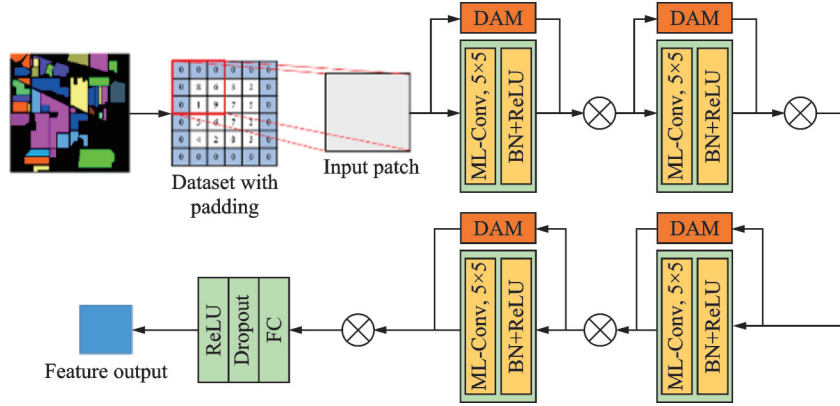


Fig.4 Flowchart of attention-enhanced multi-level CNN

central pixel. For boundary pixels lacking adjacent data, padding ensures consistent patch generation, as illustrated in the first step of Fig.4.

Following padding, the HSI data spatial dimensions expand to $(M + T - 1) \times (N + T - 1)$, facilitating the generation of $M \times N$ patches for convolution. As a result, comprehensive local spatial information can be effectively extracted.

The preprocessed patches are processed through a series of multi-scale kernel convolutional layers. In the first layer, a 2D convolutional kernel of size (7×7) is set for larger receptive field to capture relatively global spatial information. To enhance the nonlinear expressiveness and generalization of the model, batch normalization^[59] and activation functions (such as ReLU) are interspersed between the convolutional layers. These components synergistically enable the A-MLCN network to extract vital spatial feature necessary for distinguishing various ground object categories in HSI. The output of each convolutional block is given by

$$\mathbf{X}^{i+1} = \text{ReLU}(\text{BN}(b_i + \mathbf{X}^i \times \mathbf{W}^{i+1})) \quad (18)$$

The network's layers are further augmented with the DAM, which dynamically emphasizes salient features and suppresses irrelevant information. Operating in parallel with each convolutional layer, the DAM refines feature maps at every stage. It adjusts attention weights across channels and spatial domains, calculated as

$$\mathbf{Y}^{i+1} = F^{\text{DA}}(\mathbf{X}^i, \mathbf{A}_{\text{DA}}) \quad (19)$$

The outputs from the convolutional layer and the DAM are multiplied to produce the enhanced feature map, shown as

$$\tilde{\mathbf{X}}^{i+1} = \mathbf{X}^{i+1} \times \mathbf{Y}^{i+1} \quad (20)$$

where $\tilde{\mathbf{X}}^{i+1}$ represents the dynamically weighted features.

The feature extraction process spans multiple layers, with four cascaded convolutional layers enhanced by the DAM. For the second convolutional layer, we begin to focus on more local information and employ 5×5 convolutional kernel. In the last two convolutional layers, the convolutional kernel is set as 3×3 to reduce computation cost and parameters, speed up model training, and enhance nonlinear capabilities. This multi-scale kernel hierarchical approach captures both fine-grained details and broad spatial contexts, effectively distinguishing various categories within HSI data. The final convolutional output is aggregated through fully connected layers, integrating spatial features for further processing. Dropout layers are included to prevent overfitting and enhance generalization. The final feature output is given by

$$\tilde{\mathbf{X}}_{\text{A-MLCN}}^{i+1} = \text{ReLU}(\text{Dropout}(\text{FC}(\tilde{\mathbf{X}}^i))) \quad (21)$$

where $\tilde{\mathbf{X}}_{\text{A-MLCN}}^{i+1}$ is the output of the A-MLCN network.

This architecture ensures that the network focuses on significant features from the initial layers, with the DAM refocusing these features throughout the subsequent layers. Upon completing the four-layer multi-scale kernel convolution extraction process, the spatial context features are obtained. These features are then fused with spectral features for classification, resulting in a highly effective representation for distinguishing different ground objects in HSI.

2.3 Self-attention feature fusion network

In the SGCCN framework, the feature fusion and classification network constitute a pivotal component for effective HSI classification. This section delves into the fusion strategies employed in our model to integrate spectral and spatial features extracted from the SGCN-A and A-MLCN branches, as depicted in Fig. 5. The goal of these strategies is to enhance the model's classification capabilities.

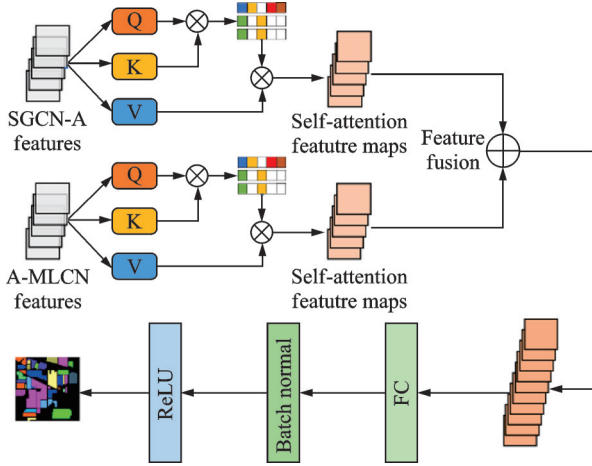


Fig.5 Flowchart of self-attention feature fusion network

An advanced feature fusion strategy that leverages multi-level integration to augment the model's discriminative power is introduced. This fusion network utilizes a self-attention mechanism to adjust the importance of features from the two branches. This ensures that the resulting feature representation captures the most class-discriminatory information.

$$Y = A_{SA} V = \text{Softmax}(QK^T)V \quad (22)$$

where A_{SA} is the attention matrix, and the queries, keys and values Q, K, V are computed from the branch features $\tilde{X}_{SGCN-A}^{i+1}$, $\tilde{X}_{A-MLCN}^{i+1}$ by the self-attention module.

To further enhance the integration of spectral and spatial features, a novel modular approach is incorporated within the fusion network. This allows the model to adaptively adjust the fusion ratio based on the distinctiveness and relevance of the features. This adaptive fusion method not only enhances the model's sensitivity to diverse data features but also bolsters its robustness when tackling complex classification tasks.

$$Y_C^{i+1} = \tilde{Y}_{SGCN-A} \oplus \tilde{Y}_{A-MLCN} \quad (23)$$

where $\tilde{Y}_{SGCN-A}$ represents the spectral similarity fea-

tures from SGCN-A and self-attention and $\tilde{Y}_{A-MLCN}$ represents the spatial context features from A-MLCN and self-attention. The operator $\oplus$ represents features fusion module.

Following feature fusion, we implement a hierarchical classifier that refines the integrated features for precise classification. This classifier employs a series of fully connected layers and non-linear activation functions to learn complex relationships between features. It also uses techniques such as drop-out and regularization to prevent overfitting, ensuring the model's generalizability.

Lastly, to further improve classification performance, we introduce an optimization strategy based on gradient clipping. This technique constrains the growth of gradients during training, helping to maintain training stability. It also prevents gradient explosion, which is particularly important for deep learning models. The combination of these methods results in a robust and highly effective classification framework for HSI data.

3 Experiments

In this chapter, the classification performance of the proposed HSI classification method SGCCN is analyzed on the three HSI data sets. First, the three HSI data sets for following experiments are described briefly. Then, analysis of several main parameters for the proposed SGCCN are conducted. Additionally, ablation study is carried to analyze the effectiveness of three modules in the SGCCN. Next, comparison experiments including experiments under different training samples is conducted for our proposed method and various state-of-the-art classification methods to verify the superiority. Finally, computational costs of each method are compared and analyzed.

3.1 Data description

To assess the effectiveness of the classification model, experiments were conducted on three sets of HSI data sets, i.e., Indian Pines, University of Pavia and Salinas.

(1) Indian Pines: The HSI data set was obtained in 1992, and the shooting system was

AVIRIS, which imaged an Indian pine forest in northwest Indiana, USA. The data set dimensions are $145 \times 145 \times 220$, including 220 spectral segments, the spatial resolution is 20 m, and the spectral resolution ranges from 400 nm to 2 500 nm. Twenty water absorption bands (104—108, 150—163 and 220) were eliminated. Due to the sensor's low spatial resolution and the high altitude of the imaging area, the HSI data exhibit low spatial resolution. This leads to an increased likelihood of mixed pixels and higher classification difficulty. The real images of this data set include 16 different categories of ground objects, about 66% are crops and farmland, and the rest are vegetation. The image and detailed category of the HSI are depicted in Fig.6 and the Table 1.

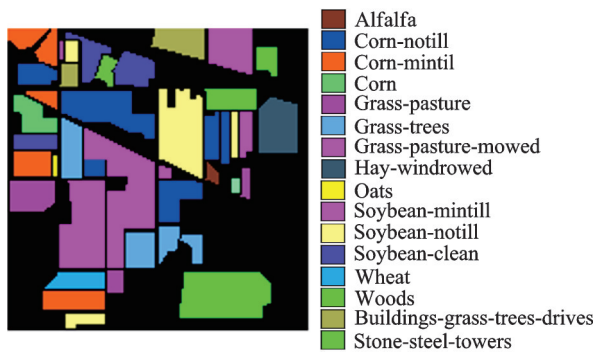


Fig.6 Indian pine data set category and ground-truth map

Table 1 Number of training and test samples on Indian Pines data set

Class	Name	Training	Test	Total
1	Alfalfa	5	41	46
2	Corn-notill	143	1 285	1 428
3	Corn-mintill	83	747	830
4	Corn	24	213	237
5	Grass-pasture	48	434	483
6	Grass-trees	73	657	730
7	Grass-pasture-mowed	3	25	28
8	Hay-windrowed	48	430	478
9	Oats	2	18	20
10	Soybean-notill	97	874	972
11	Soybean-mintill	246	2 209	2 455
12	Soybean-clean	59	534	593
13	Wheat	21	184	205
14	Woods	127	1 138	1 265
15	Buildings-grass-trees-drives	39	347	386
16	Stone-steel-towers	9	84	93
Total		1 027	9 222	10 249

(2) University of Pavia: The HSI data set was obtained by ROSIS in 2003, and it contained images of the university and surrounding Italian urban areas. Its spatial resolution is high, but its spectral resolution is low, and it contains a variety of ground objects. The data set size is $610 \times 310 \times 115$, with 115 bands, a spatial resolution of 1.3 m, and a spectral resolution of 430—860 nm. Twelve bands containing water absorption and noise were eliminated from the raw data, leaving 103 bands for classification experiments. The HSI includes nine categories of ground objects, including asphalt roads, trees, Meadows, gravel, etc. The category information for this data set is displayed in Fig.7 and Table 2.

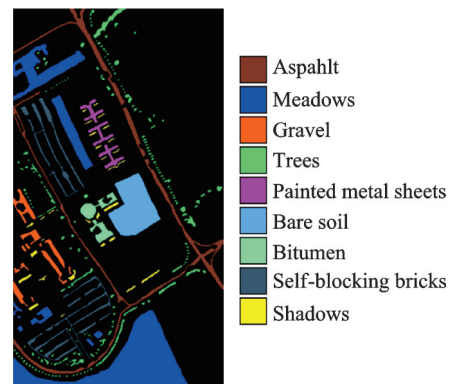


Fig.7 University of Pavia data set category and ground-truth map

Table 2 Number of training and test samples on University of Pavia data set

Class	Name	Training	Test	Total
1	Asphalt	66	6 565	6 631
2	Meadows	186	18 463	18 649
3	Gravel	21	2 078	2 099
4	Trees	31	3 033	3 064
5	Painted metal sheets	13	1 332	1 345
6	Bare soil	50	4 979	5 029
7	Bitumen	13	1 317	1 330
8	Self-blocking bricks	37	3 645	3 682
9	Shadows	9	938	947
Total		426	42 350	42 776

(3) Salinas: The HSI data set was imaged in Salinas Valley, California, USA. The HSI and Indian Pines were also collected in 1998 using the AVIRIS imaging spectrometer, which had a high spatial resolution and a relatively uniform distribution of ground objects. The data set size is $512 \times$

217×224, comprising 224 original bands, with a spatial resolution of 3.7 m and a spectral range of 400—2 500 nm. Twenty water absorption bands (108—112, 154—167, and 224) were eliminated, leaving 204 bands for classification experiments. The HSI contains 16 categories, including crops, farmland, vegetation, etc. Detailed category and ground truth map of this HSI is presented in Fig.8 and Table 3.

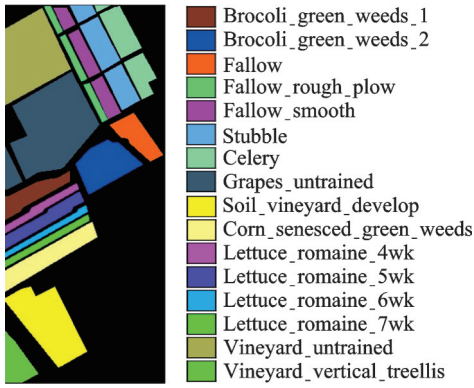


Fig.8 Salinas data set category and ground-truth map

Table 3 Number of training and test samples on Salinas data set

Class	Name	Training	Test	Total
1	Brocoli_green_weeds_1	20	1 989	2 009
2	Brocoli_green_weeds_2	37	3 689	3 726
3	Fallow	20	1 956	1 976
4	Fallow_rough_plow	14	1 380	1 394
5	Fallow_smooth	27	2 651	2 678
6	Stubble	40	3 919	3 959
7	Celery	36	3 543	3 579
8	Grapes_untrained	113	11 158	11 271
9	Soil_vinyard_develop	62	6 141	6 203
10	Corn_senesced_green_weeds	33	3 245	3 278
11	Lettuce_romaine_4wk	11	1 057	1 068
12	Lettuce_romaine_5wk	19	1 908	1 927
13	Lettuce_romaine_6wk	9	907	916
14	Lettuce_romaine_7wk	11	1 059	1 070
15	Vineyard_untrained	73	7 195	7 268
16	Vineyard_vertical_trellis	18	1 789	1 807
Total		543	53 586	54 129

3.2 Experiments setup

A series of experiments were conducted to test and evaluate the classification performance of the proposed method and comparison methods. All experiments were implemented on a computer with a 2.50 GHz Intel Core i5-12400F CPU with 16 GB of

RAM and an NVIDIA GeForce RTX 3060 GPU. The operating system used was Windows 10, and the experiments were conducted using the PyTorch 1.3.0 deep-learning framework and a Python 3.9 compiler. Moreover, the maximum number of epochs in training phase is set to 100 for Indian Pines and Salinas, and 150 for University of Pavia. The batch size is set to 128 in the training phase. For the three data sets, samples are spilled randomly into training set and test set. For Indian Pines 10% of the samples are used for training. While 1% of the samples are selected randomly for training in the case of Salinas and University of Pavia. The detailed number of training and test samples for each class are listed in Tables 1—3. In following experiments, we employ accuracy of each class, overall accuracy (OA), average accuracy (AA), and kappa coefficients as the evaluation metric to quantitatively evaluate the performance of all methods. We make the average of the results obtained from 20 different random training sample experiments as the final experiment results.

3.3 Ablation study

Ablation experiments were conducted on three HSI datasets in this section to evaluate the performance of each feature extraction branch in the proposed SGCCN. The experiments also assessed their performance after incorporating the dynamic attention module. The mean classification results with standard variance (in percentage), including the accuracy of each class, OA, AA, and kappa coefficient, are presented in Tables 4—6. The best one is shown in bold.

Analysis of individual data sets reveals that the classification OA for the spectral band graph convolutional network (SGCN) exceeds 80%. For several classes in the Pavia and Salinas data sets, the accuracy approaches 100%. This highlights the efficacy of the newly designed and lightweight SGCN in capturing spectral similarity features beneficial for classification. Although the SGCN performs not well on the Indian Pines data set, which is due to the low spatial resolution and uneven sample distribution on the data set. Furthermore, the SGCN model exhibits minimal variance, indicating a high

degree of stability of our model.

Results indicate that the multi-level convolutional network (MLCN) branch attains superior classification performance, exemplified by the 98.11% OA achieved on the Salinas data set. The main reason is that the MLCN branch consists four feature extraction blocks, which can extract and strength spatial context features between samples step by step. Additionally, the incorporation of DAM enables the MLCN branch to focus on valuable features, leading to improved accuracy in specific categories.

Furthermore, the SGCCN combined self-atten-

tion feature fusion network to fuse features from SGCN-A and A-MLCN achieves the highest classification accuracy for all three data sets. Notably, the accuracy for Alfalfa and Corn in the Indian data set improves about 5% and reaches 100% and increases nearly 50% on Gravel in Pavia data set. The major reason is that the two types of networks capture spectral and spatial information from different dimension to achieve full utilization of HSI feature information. In conclusion, the combination of these two networks, leveraging the strengths of different feature categories, leads to improved classification accuracy.

Table 4 Overall accuracy, average accuracy and kappa coefficients of ablation experiments on Indian Pines %

Class	SGCN	SGCN-A	MLCN	A-MLCN	SGCCN
1	11.11±0.07	10.67±0.05	95.12±5.78	100±0	100±0
2	64.93±0.03	68.67±0.04	94.79±0.89	92.76±1.17	99.68±0.52
3	52.21±0.01	55.70±0.02	99.73±1.78	100±0	99.20±0.28
4	21.46±0.04	18.72±0.03	94.84±1.93	98.12±2.33	100±0
5	86.59±0.02	84.32±0.03	97.01±0.87	98.62±0.66	98.85±0.57
6	94.93±0.02	92.64±0.01	100±0	100±0	99.39±0.18
7	69.23±0.03	53.85±0.06	100±0	100±0	100±0
8	99.05±0.01	99.76±0.01	100±0	100±0	100±0
9	44.44±0.09	22.22±0.10	83.33±13.56	100±0	100±0
10	63.76±0.03	63.99±0.04	97.83±0.76	99.31±0.98	98.86±0.35
11	80.40±0.02	81.76±0.02	99.05±0.29	99.95±1.27	97.47±0.20
12	63.97±0.03	62.10±0.03	93.26±1.67	94.19±1.85	99.43±0.73
13	90.27±0.03	75.68±0.08	100±0	98.38±2.22	99.46±1.23
14	91.85±0.02	91.42±0.02	100±0	99.82±0.46	99.82±0.07
15	58.08±0.03	55.39±0.02	98.85±3.02	97.12±2.36	99.14±0.09
16	82.14±0.02	90.48±0.01	91.67±3.88	90.48±2.83	96.43±1.31
OA	74.63±0.01	74.91±0.01	97.99±0.46	98.22±0.51	98.94±0.10
AA	67.15±0.01	64.10±0.02	96.59±1.52	98.04±0.91	99.13±1.37
Kappa	70.88±0.01	71.20±0.01	97.71±0.53	97.97±0.58	98.78±0.12

Table 5 Overall accuracy, average accuracy and kappa coefficients of ablation experiments on University of Pavia %

Class	SGCN	SGCN-A	MLCN	A-MLCN	SGCCN
1	86.34±0.03	88.69±0.01	84.66±2.02	86.38±3.24	99.75±0.89
2	93.81±0.01	93.67±0.01	99.66±0.24	99.87±0.34	99.94±0.09
3	45.18±0.05	45.66±0.06	67.42±12.03	56.11±12.26	99.51±2.87
4	76.91±0.01	77.73±0.03	87.08±5.42	82.56±2.99	96.83±0.83
5	98.66±0.01	98.28±0.01	90.54±2.40	94.52±2.40	100±0
6	74.48±0.03	72.55±0.03	100±0	98.23±1.74	99.86±0.08
7	42.22±0.05	41.61±0.06	61.12±10.54	55.35±16.40	95.49±2.44
8	68.71±0.06	67.09±0.05	68.12±8.25	94.21±5.76	97.54±0.57
9	94.11±0.02	92.18±0.02	39.70±11.28	60.51±6.43	85.85±5.70
OA	83.18±0.01	83.13±0.01	89.37±1.14	91.29±1.69	98.95±0.22
AA	75.60±0.01	75.27±0.01	77.59±2.67	80.86±3.99	97.89±0.67
Kappa	77.47±0.01	77.36±0.01	85.87±1.52	88.38±2.27	98.61±0.29

Table 6 Overall accuracy, average accuracy and kappa coefficients of ablation experiments on Salinas %

Class	SGCN	SGCN-A	MLCN	A-MLCN	SGCCN
1	99.90±0.01	99.25±0.01	99.95±1.48	99.75±2.48	99.90±0.01
2	98.45±0.01	99.57±0.01	100±0	100±0	100±0
3	97.79±0.02	96.82±0.01	98.67±0.75	99.49±1.56	100±0
4	80.15±0.03	79.08±0.04	94.13±2.42	97.61±1.64	98.78±0.07
5	96.03±0.02	96.64±0.02	99.13±0.78	87.74±2.72	99.77±1.11
6	99.64±0.01	99.67±0.01	100±0	99.44±0.18	100±0
7	99.60±0.01	99.75±0.01	100±0	99.10±0.85	100±0
8	87.24±0.02	86.63±0.02	97.17±0.94	96.07±1.25	99.08±0.17
9	99.70±0.01	99.59±0.01	100±0	100±0	99.98±0.01
10	96.01±0.01	95.74±0.01	100±0	96.52±1.01	99.57±0.01
11	93.19±0.02	92.24±0.01	86.66±3.78	99.24±3.28	99.91±0.34
12	84.31±0.03	91.32±0.02	98.32±2.57	97.12±3.22	99.90±3.44
13	88.09±0.06	93.38±0.10	90.96±5.83	93.16±2.00	99.89±0.35
14	92.73±0.04	79.13±0.04	99.43±1.51	92.82±1.75	99.90±0.44
15	51.33±0.04	54.46±0.05	95.73±10.67	99.19±6.07	100±0
16	98.43±0.01	98.88±0.01	100±0	100±0	100±0
OA	88.48±0.01	88.84±0.01	98.11±1.60	97.69±0.85	99.72±0.18
AA	91.42±0.01	91.38±0.01	97.51±1.13	97.33±0.60	99.53±0.23
Kappa	87.13±0.01	87.54±0.01	97.89±1.79	97.43±0.95	99.69±0.20

3.4 Comparison with state-of-the-art methods

A comparative analysis was conducted to evaluate our proposed SGCCN model against several state-of-the-art methods, including the 2DCNN^[30], the 3DCNN^[60], the HybridSN^[34], the MiniGCN^[44], the MDGCN^[45], the CNN-enhanced GCN (CEGCN)^[47] and the HybridFormer^[61]. Here, MiniGCN, MDGCN and CEGCN are the representative classification methods based on GCN and HybridFormer is the latest model based on attention mechanism. All the compared methods are designed for HSI classification under limited training sample situation. For fair comparisons and the best performance, the parameter settings of compared methods remained consistent with those detailed in the original papers. The division of the training and testing sets for all classification methods in the subsequent experiments are configured identically as listed in Tables 1—3. The accuracy of each class, OA, AA, and Kappa are employed to evaluate the classification performance of each method. Comprehensive results, including mean values and standard variance, for all methods on Indian Pines, University of Pavia, and Salinas data sets are presented in Tables 7—9 (The best one is shown in bold). Classification maps for these data sets are also shown in Figs. 9—11.

Compared with other competitor methods, our proposed SGCCN achieves the best classification

performance. The SGCCN achieves OA and AA scores of approximately 99% on the Indian Pines data set. In particular, SGCCN achieves the highest classification accuracy of 98.95% and 99.72% with only 1% training samples for the University of Pavia and Salinas data sets. These results demonstrate that SGCCN outperforms other methods. What stands out in this result is that the SGCCN achieves the highest accuracy in most categories across the three data sets. Especially, SGCCN achieves 100% accuracy in the Corn category for Indian Pines and in Fallow, Celery, Vineyard_untrained and Vineyard_vertical_trellis categories for Salinas. This can be attributed to the design of our proposed SGCCN, which incorporates an SGCN-A based on spectral band graph structure construction to extract spectral similarity features. This allows for the excavation of intrinsic connections between the spectrum of HSI, leading to enhanced computational efficiency. Then, the joint application of SGCN-A and A-MLCN allows the model to extract more unseen complementary spectral-spatial relations, facilitating the identification of challenging categories. Additionally, the SGCCN outperforms other spatial-based GCN methods and the joint CNN HybridSN method. The major reason is that the SGCN-A treats each band frame of HSI as a vertex to con-

Table 7 Classification results of different methods on Indian Pines

%

Class	2DCNN	3DCNN	HybridSN	MiniGCN	MDGCN	CEGCN	HybridFormer	SGCCN
1	49.12±1.56	75.99±12.19	96.10±3.48	71.66±0.06	93.75±0	50.69±0.42	97.21±11.35	100±0
2	68.97±0.53	88.10±2.62	94.79±1.21	84.92±0.03	92.63±0.01	90.06±0.08	97.31±6.18	99.68±0.52
3	67.78±0.38	78.34±7.18	98.86±1.28	75.98±0.03	93.12±0.01	98.21±0.02	98.24±0.97	99.20±0.28
4	53.37±0.90	78.51±7.38	93.99±4.04	76.51±0.05	96.14±0.01	82.82±0.14	96.98±1.67	100±0
5	93.57±0.31	89.16±2.95	97.68±1.5	88.88±0.03	96.03±0.01	98.55±0.01	97.84±1.81	98.85±0.57
6	97.20±0.30	97.10±1.36	98.69±0.59	90.26±0.02	97.43±0.01	97.28±0.02	99.19±0.46	99.39±0.18
7	76.19±1.2	68.38±20.96	100±0	85.44±0.05	69.23±0.01	27.87±0.35	98.42±9.86	100±0
8	95.76±0.01	96.93±1.45	99.95±0.09	95.10±0.01	97.99±0.01	89.76±0.15	99.65±0.34	100±0
9	90.91±6.49	64.36±19.49	83.89±13.02	79.00±0.11	100±0	0±0	95.71±7.91	100±0
10	70.02±0.25	81.13±5.84	98.80±0.90	83.30±0.03	84.39±0.01	97.52±0.02	96.04±13.8	98.86±0.35
11	76.97±0.24	87.56±3.22	99.32±0.65	86.17±0.02	94.93±0.01	92.79±0.03	98.32±2.97	97.47±0.20
12	68.10±1.47	80.52±6.13	93.48±3.60	87.64±0.02	90.05±0.02	98.33±0.02	96.91±0.72	99.43±0.73
13	99.46±0.16	98.52±1.44	99.24±0.69	84.75±0.05	100±0	77.65±0.34	99.39±0.62	99.46±1.23
14	94.62±0.12	96.42±0.69	99.70±0.21	96.28±0.01	99.35±0.01	92.23±0.14	99.70±0.04	99.82±0.07
15	82.72±0.59	73.42±2.28	98.56±1.58	80.56±0.02	98.88±0.01	77.75±37	98.67±0.55	99.14±0.09
16	97.11±0.02	89.01±14.73	93.45±4.77	93.60±0.04	98.41±0.01	97.83±0.03	95.52±0.56	96.43±1.31
OA	79.63±0.17	86.60±2.70	97.97±0.41	86.48±0.01	94.34±0.01	92.40±0.03	98.12±1.41	98.94±0.10
AA	77.49±0.35	81.08±4.29	96.65±1.00	85.06±0.01	93.90±0.01	79.33±0.08	98.04±0.56	99.13±1.37
Kappa	76.75±0.19	84.70±3.13	97.68±0.47	84.56±0.01	93.51±0.01	91.36±0.04	97.86±1.82	98.78±0.12

Table 8 Classification results of different methods on University of Pavia

%

Class	2DCNN	3DCNN	HybridSN	MiniGCN	MDGCN	CEGCN	HybridFormer	SGCCN
1	91.30±0.12	89.59±1.19	95.27±3.94	96.99±0.01	59.37±0.06	96.41±0.02	98.67±1.49	99.75±0.89
2	93.71±0.04	91.97±0.89	99.29±1.38	94.50±0.01	78.61±0.03	95.95±0.08	98.85±0.11	99.94±0.09
3	61.92±0.36	75.88±3.84	91.77±5.32	84.27±0.03	77.48±0.02	89.43±0.07	94.21±25.86	99.51±2.87
4	89.28±0.06	94.55±2.00	85.38±4.74	85.89±0.01	73.47±0.03	94.56±0.01	96.93±1.01	96.83±0.83
5	100±0	96.91±1.21	97.35±2.68	99.73±0.01	95.67±0.02	98.01±0.02	99.72±0.04	100±0
6	81.29±0.12	84.12±3.54	95.17±6.24	92.04±0.01	80.76±0.02	92.06±0.06	99.49±0.19	99.86±0.08
7	81.89±0.31	82.69±3.04	97.96±4.85	74.61±0.04	78.92±0.04	90.51±0.09	98.51±7.63	95.49±2.44
8	72.65±0.24	84.42±1.86	87.54±3.29	69.78±0.03	43.78±0.06	94.87±0.03	95.79±7.28	97.54±0.57
9	96.80±0.18	91.55±1.76	77.74±9.89	94.70±0.01	67.94±0.03	84.17±0.11	98.11±0.48	85.85±5.70
OA	88.26±0.04	87.51±0.82	95.13±0.91	90.21±0.01	72.77±0.03	94.68±0.03	97.91±0.87	98.95±0.22
AA	83.92±0.05	85.51±1.28	91.62±1.41	88.06±0.01	72.89±0.02	92.89±0.12	97.82±0.73	97.89±0.67
Kappa	84.28±0.05	83.63±1.02	93.53±1.22	86.77±0.01	65.36±0.03	93.01±0.04	97.24±1.49	98.61±0.29

Table 9 Classification results of different methods on Salinas

%

Class	2DCNN	3DCNN	HybridSN	MiniGCN	MDGCN	CEGCN	SGCCN
1	99.92±0.02	96.73±0.47	99.97±0.05	99.80±0.01	88.83±0.09	92.89±0.09	99.90±0.01
2	99.96±0.02	99.05±0.31	100±0	99.79±0.01	95.83±0.02	94.77±0.06	100±0
3	99.21±0.01	97.47±0.89	99.97±0.08	99.33±0.01	92.70±0.08	93.93±0.06	100±0
4	98.25±0.17	98.58±1.12	96.38±2.72	97.60±0.01	94.35±0.03	94.18±0.02	98.78±0.07
5	97.50±0.04	98.66±0.97	98.12±1.35	88.66±0.04	63.67±0.10	86.89±0.07	99.77±1.11
6	100±0	98.26±1.07	99.68±0.94	99.95±0.01	93.08±0.08	96.42±0.03	100±0
7	99.77±0.01	98.47±0.75	99.96±0.08	99.21±0.01	80.73±0.10	94.32±0.03	100±0
8	84.52±0.17	86.77±5.37	98.80±1.32	78.08±0.01	69.21±0.24	78.26±0.15	99.08±0.17
9	99.30±0.01	98.99±0.30	99.99±0.03	99.49±0.01	88.74±0.05	88.69±0.08	99.98±0.01
10	96.54±0.06	93.15±2.36	99.16±0.51	88.29±0.02	88.98±0.12	97.49±0.01	99.57±0.01
11	98.33±0.15	94.30±0.92	96.77±2.43	82.50±0.04	87.28±0.08	81.55±0.21	99.91±0.34
12	99.04±0.06	97.73±1.46	97.86±3.34	96.67±0.01	92.30±0.05	98.54±0.02	99.90±3.44
13	92.20±1.42	97.30±1.89	98.53±1.30	93.61±0.01	77.77±0.01	83.71±0.18	99.89±0.35
14	93.60±0.13	96.89±2.67	99.11±1.07	99.51±0.01	80.48±0.02	86.93±0.06	99.90±0.44
15	77.81±0.22	78.33±7.78	96.32±5.28	71.57±0.04	81.86±0.06	74.66±0.15	100±0
16	99.41±0.03	76.45±1.65	99.94±0.01	99.29±0.01	96.45±0.06	91.54±0.04	100±0
OA	92.92±0.03	91.48±1.78	98.81±0.76	89.54±0.01	83.08±0.07	87.11±0.03	99.72±0.18
AA	95.94±0.08	93.67±0.90	98.79±0.51	93.33±0.01	85.77±0.05	89.67±0.02	99.53±0.23
Kappa	92.13±0.04	90.53±1.96	98.67±0.85	88.36±0.01	81.30±0.07	85.70±0.03	99.69±0.20

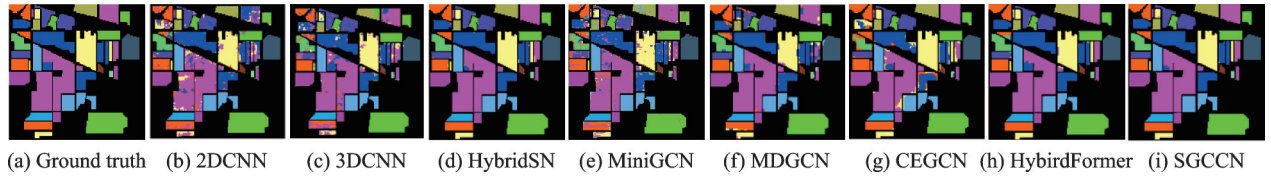


Fig.9 Ground truth and classification maps acquired by different methods on Indian Pines

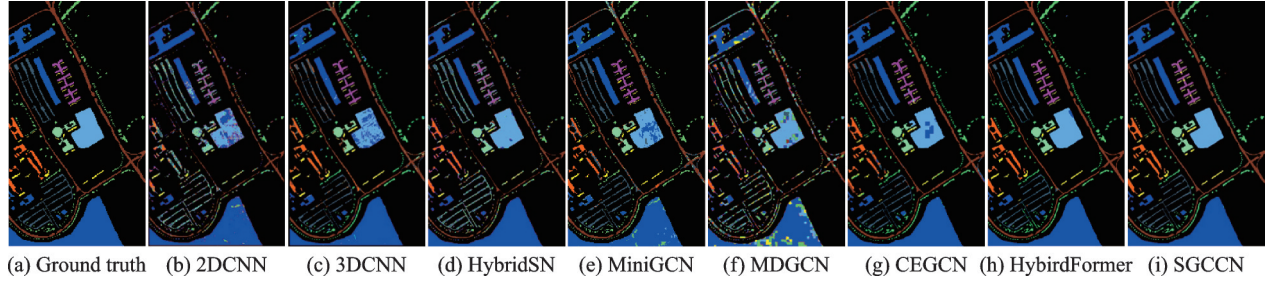


Fig.10 Ground truth and classification maps acquired by different methods on University of Pavia

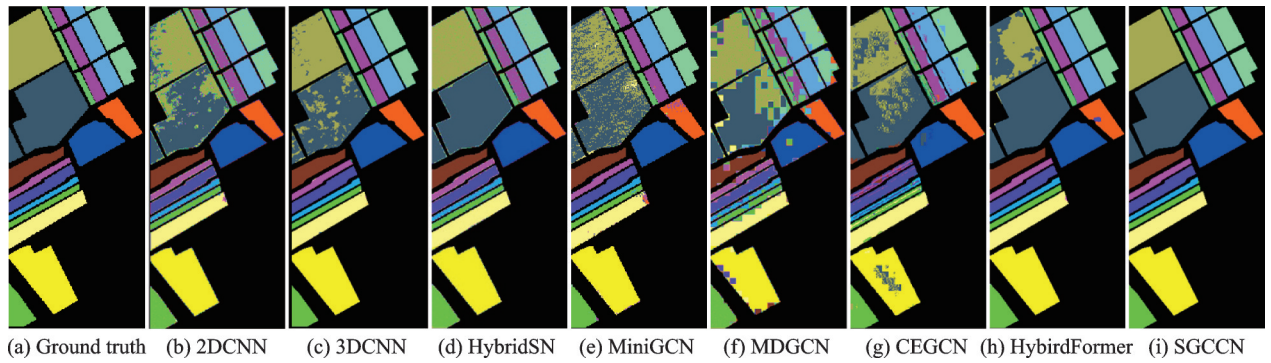


Fig.11 Ground truth and classification maps acquired by different methods on Salinas

struct HSI graph data. This approach extracts spectral features while preserving pixel-level, fine-grained spatial-spectral information. While subgraph represents local region information or superpixel graph ignores detailed spatial information, resulting less effective in capturing detailed features on complex distribution data sets. Despite HybridSN also adopts the hybrid CNN, the ability in modeling channel spectral relations between any vertex of SGCCN-A outperforms CNNs that can only extract information in a local region. Furthermore, the accuracy for each category, OA, AA and Kappa of the SGCCN exhibit relatively small variance, indicating a high degree of stability in the classification outcomes. This verifies the efficiency and superiority of our classification network.

Upon inspecting the classification maps of each classification method on the three data sets, as presented in Figs.9—11. It is evident that the classification maps of 2DCNN, 3DCNN, and MiniGCN contain pepper noise. In contrast, the maps obtained

by MDGCN and CEGCN show jagged boundaries and large areas of window-level misclassifications. Although HybridSN and HybridFormer acquire relatively accurate classification maps, they still have classification errors in the boundary area. The classification maps demonstrate the superior performance of SGCCN, characterized by smoother boundaries and accurate pixel-level maps. This could be attributed to the effective feature fusion network, which introduces an attention mechanism to adjust weights for more valuable information. It also employs multiple fusion processes to combine complex relational features from diverse layers.

In summary, the proposed SGCCN integrates global spectral features and local spatial features derived from SGCCN-A and A-MLCN. By emphasizing fused features through attention modules, SGCCN demonstrates superior classification performance. It also shows greater robustness across various HSI data sets compared with other advanced classification methods.

3.5 Effect of the training sample rate

To analyze the impact of the training sample rate, experiments are conducted for various classification methods on the three data sets, and the results are shown in Fig. 12. Notably, the abscissas of (b) and (c) in Fig. 12 adopt a non-arithmetic progression distribution. This is because in the key low-proportion interval with scarce training samples, model performance fluctuates significantly. Dense sampling in this interval can accurately capture performance variations. As the number of training samples increases, model performance gradually stabilizes. A larger step size is therefore adopted to reduce redundant experimental points, optimizing data presentation efficiency while ensuring the integrity of the performance trend. In the following experiments, the selected training samples per class range from 2% to 12% in steps of 1% for the Indian Pines data set. For the Pavia and Salinas data sets, the training samples per class are set as 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.8%, 1%, 1.25% and 1.5%.

As is evident, the proposed SGCCN consistently achieves the highest classification accuracy on all three data sets, particularly in scenarios with fewer training samples. The SGCCN achieves an accuracy of 90% with only 2% training samples on the Indian Pines data set. It surpasses other methods by more than 10% with only 0.1% training samples on the University of Pavia data set. On the Salinas data set, SGCCN consistently outperforms most methods by at least 5% in OA under varying training sample rates. The main reason is that SGCCN employs a dual-branch feature extraction network. It extracts spectral similarity of any band and local context information for every neighboring patch, which comprehensively utilizes global and local feature information of the HSI. Although HybridSN and HybridFormer achieve relatively high accuracy with more training samples, the limitation of CNNs in extracting information only from local regions restricts their performance. This issue is particularly evident in complex data sets, such as Pavia, and in small-sample scenarios. In summary, experimental results validate the superior learning ability of our proposed SGCCN method. Its classification perfor-

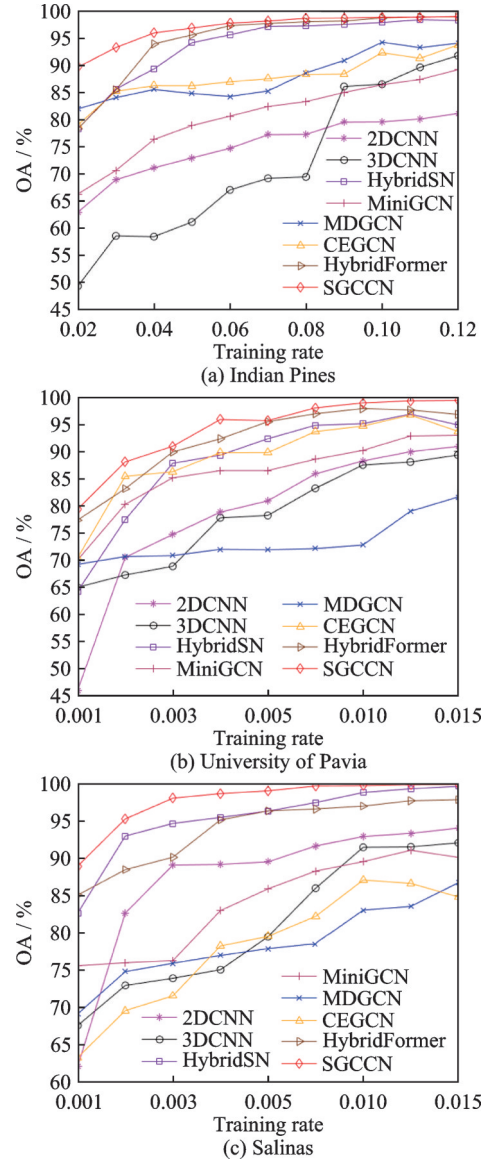


Fig.12 OA under different training sample rates on three data sets

mance is particularly strong under conditions of limited training samples.

3.6 Comparison of training time and test time

To further assess the performance of this proposed SGCCN model, the training time and test time of different methods on the three HSI data sets are compared in Table 10. All experiments are conducted on the same computer.

As evident from the table, our novel SGCCN requires less computational time than other GCN-based classification methods. This is mainly due to that we construct a graph from the channel dimension and design an efficient channel graph convolutional network. Additionally, the feature extraction network with dynamic attention module yields fewer

Table 10 Training time and test time of different methods on three data sets

s

Data set	Method	2DCNN	3DCNN	MiniGCN	MDGCN	CEGCN	HybridFormer	SGCCN
Indian Pines	Training time	38.71	109.48	131.53	269.88	16.31	312.56	97.98
	Test time	5.33	0.99	1.32	1.11	0.37	4.72	2.72
	Total time	44.04	110.47	132.85	270.99	16.61	317.28	100.70
University of Pavia	Training time	17.81	69.84	579	1205.24	163.7	84.37	119.304
	Test time	21.47	5.89	45.29	2.94	2.08	25.38	19.696
	Total time	39.28	75.73	624.29	1208.18	165.79	109.75	139.00
Salinas	Training time	18.66	159.85	300.32	242.96	79.6	171.51	94.25
	Test time	27.56	5.48	47.50	1.39	1.53	22.97	24.65
	Total time	46.22	165.33	347.82	244.35	43.98	194.48	118.9

parameters to compute, which improves computational efficiency and classification accuracy. The longer training time compared with CNN-based methods is attributed to SGCCN creating neighboring patches for each pixel to extract local context features. It also incorporates an additional spectral similarity feature extraction branch, which adds to the computational cost. In addition, CNN-based methods are not always the fastest on different data sets. On the Indian Pines and Salinas data sets, the SGCCN outperforms the 3DCNN, which indicates higher classification ability and robustness on various data sets. Considering the complexity and classification accuracy of networks, all these results demonstrate that the proposed SGCCN performs relatively better and costs less computational time.

4 Conclusions

We propose an innovative HSI classification method and the SGCCN model, which combines the advantages of spectral band graph convolutional networks with attention and attention-enhanced multi-level CNN. The model also incorporates a self-attention feature fusion network to effectively extract both spectral and spatial features of the HSI. Our model constructs a novel spectral band graph structure by treating each band of the HSI as a vertex of the graph, which significantly improves computational efficiency. The SGCCN-A enhances the expressive ability of features by lightweight spectral band graph convolutions. At the same time, the A-MLCN branch in the model focuses on local spatial information to supplement the spectral features. It enhances classification accuracy through a multi-level convolutional network. The dynamic attention al-

so plays a huge role in extracting key features.

After a series of experimental verifications, SGCCN has achieved excellent classification results on multiple standard HSI data sets, surpassing various existing advanced methods. Especially when the number of samples is limited, SGCCN exhibits excellent learning ability and robustness. Furthermore, our model reduces the computational time required for training and testing. It maintains high classification accuracy, demonstrating its potential and efficiency in practical applications.

In future research, we plan to further explore and optimize the SGCCN model from the improvement of feature fusion and model generalization ability. We will research more advanced feature fusion technology to more effectively integrate features from different network branches and further improve classification accuracy. Introducing more types of attention mechanisms and regularization techniques is another direction to improve the model's generalization ability on unseen data. This allows the model to adapt to more diverse HSI data sets. Through these future research directions, we expect that the SGCCN model will not only achieve greater breakthroughs in the field of HSI classification.

References

- [1] GHAMISI P, PLAZA J, CHEN Y, et al. Advanced spectral classifiers for hyperspectral images: A review [J]. IEEE Geoscience and Remote Sensing Magazine, 2017, 5(1): 8-32.
- [2] BESSON O, VINCENT F. Sub-pixel detection in hyperspectral imaging with elliptically contoured t-distributed background[J]. Signal Processing, 2020, 175: 107662.
- [3] LI K, WAN G, CHENG G, MENG L, et al. Object detection in optical remote sensing images: A survey

- and a new benchmark[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 296-307.
- [4] GAO L, YAO D, LI Q, et al. A new low-rank representation based hyperspectral image denoising method for mineral mapping[J]. *Remote Sensing*, 2017, 9(11): 1145.
- [5] GOVENDER M, CHETTY K T, BULCOCK H H. A review of hyperspectral remote sensing and its application in vegetation and water resource studies[J]. *Water SA*, 2009, 33: 145-151.
- [6] TRANSON J, D'ANDRIMONT R, MAUGNARD A, et al. Survey of hyperspectral earth observation applications from space in the sentinel-2 context[J]. *Remote Sensing*, 2018, 10(2): 157.
- [7] PAOLETTI M, HAUT J, PLAZA J, et al. Deep learning classifiers for hyperspectral imaging: A review[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 158: 279-317.
- [8] HUGHES G. On the mean accuracy of statistical pattern recognizers[J]. *IEEE Transactions on Information Theory*, 1968, 14(1): 55-63.
- [9] MA L, CRAWFORD M M, TIAN J. Local manifold learning-based K-nearest-neighbor for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(11): 4099-4109.
- [10] SCOTT C, NOWAK R. Minimax-optimal classification with dyadic decision trees[J]. *IEEE Transactions on Information Theory*, 2006, 52(4): 1335-1353.
- [11] HAM J, CHEN Y C, CRAWFORD M M, et al. Investigation of the random forest framework for classification of hyperspectral data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(3): 492-501.
- [12] GUO B, GUNN S R, DAMPER R I, et al. Customizing kernel functions for SVM-based hyperspectral image classification[J]. *IEEE Transactions on Image Processing*, 2008, 17(4): 622-629.
- [13] TARABALKA Y, BENEDIKTSSON J A, CHANUSSOT J. Spectral spatial classification of hyperspectral imagery based on partitional clustering techniques[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(8): 2973-2987.
- [14] BRO R, SMILDE A K. Principal component analysis [J]. *Analytical Methods*, 2014, 6(9): 2812-2831.
- [15] GEPSHTEIN S, KELLER Y. Iterative spectral independent component analysis[J]. *Signal Processing*, 2019, 155: 368-376.
- [16] LIU X, GAO L, ZHANG B, et al. An improved MNF transform algorithm on hyperspectral images with complex mixing ground objects[C]//*Proceedings of the 2008 Congress on Image and Signal Processing (CISP)*. Sanya, China: IEEE, 2008, 3: 479-483.
- [17] BENEDIKTSSON J A, PALMASON J A, SVEINSSON J R. Classification of hyperspectral data from urban areas based on extended morphological profiles [J]. *IEEE Transactions on Geoscience Remote Sensing*, 2005, 43(3): 480-491.
- [18] ASGHARI BEIRAMI B, MOKHTARZADE M. Band grouping superpca for feature extraction and extended morphological profile production from hyperspectral images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(11): 1953-1957.
- [19] LIU X, YIN X, CAI Y, et al. Visual saliency-based extended morphological profiles for unsupervised feature learning of hyperspectral images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(11): 1963-1967.
- [20] DALLAMURA M, BENEDIKTSSON J A, WASKE B, et al. Morphological attribute profiles for the analysis of very high resolution images[J]. *IEEE Transactions on Geoscience Remote Sensing*, 2010, 48(10): 3747-3762.
- [21] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [23] LUO X, LI S, SHI X, et al. Learning from small data for hyperspectral image classification[J]. *Signal Processing*, 2023, 213: 109183.
- [24] CHEN Y, LIN Z, ZHAO X, et al. Deep learning-based classification of hyperspectral data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6): 2094-2107.
- [25] CHEN Y, ZHAO X, JIA X. Spectral spatial classification of hyperspectral data based on deep belief network[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 2381-2392.
- [26] LEE H, KWON H. Going deeper with contextual CNN for hyperspectral image classification[J]. *IEEE Transactions on Image Processing*, 2017, 26(10): 4843-4855.
- [27] GONG H, LI Q, LI C, et al. Multiscale information fusion for hyperspectral image classification based on hybrid 2D-3D CNN[J]. *Remote Sensing*, 2021, 13(12): 2268.
- [28] SUN H, ZHENG X, LU X. A supervised segmentation network for hyperspectral image classification[J]. *IEEE Transactions on Image Processing*, 2021, 30: 2810-2825.
- [29] HU W, HUANG Y, LI W, et al. Deep convolutional neural networks for hyperspectral image classification

- [J]. *Journal of Sensors*, 2015, 2015: 1-12.
- [30] MAKANTASIS K, KARANTZALOS K, DOULAMIS A, et al. Deep supervised learning for hyperspectral data classification through convolutional neural networks[C]//*Proceedings of 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. [S.l.]: IEEE, 2015: 4959-4962.
- [31] CHEN Y, JIANG H, LI C, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(10): 6232-6251.
- [32] GUO Y, CAO H, BAI J, et al. High efficient deep feature extraction and classification of spectral-spatial hyperspectral image using cross domain convolutional neural networks[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(1): 345-356.
- [33] ZHONG Z, LI J, LUO Z, et al. Spectral spatial residual network for hyperspectral image classification: A 3-D deep learning framework[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2): 847-858.
- [34] ROY S K, KRISHNA G, DUBEY S R, et al. HybridSN: Exploring 3D-2D CNN feature hierarchy for hyperspectral image classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(2): 277-281.
- [35] MAGNER A, BARANWAL M, HERO A O. Fundamental limits of deep graph convolutional networks for graph classification[J]. *IEEE Transactions on Information Theory*, 2022, 68(5): 3218-3233.
- [36] WU W B, SUN J J, CHEN S B, et al. Combinatorial online highorder interactive feature selection based on dynamic graph convolution network[J]. *Signal Processing*, 2023, 212: 109133.
- [37] LIANG G, KINTAK U, YIN H, et al. Multi-scale hybrid attention graph convolution neural network for remote sensing images super-resolution[J]. *Signal Processing*, 2023, 207: 108954.
- [38] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications[J]. *AI Open*, 2020, 1: 57-81.
- [39] KIPF T N, WELLMING M. Semi-supervised classification with graph convolutional networks[C]//*Proceedings of International Conference on Learning Representations (ICLR)*. [S.l.]: ICLR, 2017.
- [40] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4-24.
- [41] MA Z, JIANG Z, ZHANG H. Hyperspectral image classification using feature fusion hypergraph convolution neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-14.
- [42] QIN A, SHANG Z, TIAN J, et al. Spectral spatial graph convolutional networks for semisupervised hyperspectral image classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16(2): 241-245.
- [43] HE X, CHEN Y, GHAMISI P. Dual graph convolutional network for hyperspectral image classification with limited training samples[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-18.
- [44] HONG D, GAO L, YAO J, et al. Graph convolutional networks for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(7): 5966-5978.
- [45] WAN S, GONG C, ZHONG P, et al. Multiscale dynamic graph convolutional network for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(5): 3162-3177.
- [46] ZENG H, LIU Q, ZHANG M, et al. Semi-supervised hyperspectral image classification with graph clustering convolutional networks[EB/OL]. (2020-12-20). <https://arxiv.org/abs/2012.10932>.
- [47] LIU Q, XIAO L, YANG J, et al. CNN-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(10): 8657-8671.
- [48] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(11): 1254-1259.
- [49] LIU M, PAN H, GE H, et al. MS3Net: Multiscale stratified-split symmetric network with quadra-view attention for hyperspectral image classification[J]. *Signal Processing*, 2023, 212: 109153.
- [50] DONG Y, LIU Q, DU B, et al. Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification[J]. *IEEE Transactions on Image Processing*, 2022, 31: 1559-1572.
- [51] SHA A, WANG B, WU X, et al. Semisupervised classification for hyperspectral images using graph attention networks[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(1): 157-161.
- [52] VELIKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//*Proceedings of International Conference on Learning Representations (ICLR)*. [S.l.]: ICLR, 2018.
- [53] WANG T, WANG G, TAN K E, et al. Spectral

- Pyramid graph attention network for hyperspectral image classification[EB/OL]. (2020-01-20). <https://arxiv.org/abs/2001.07108>.
- [54] GU J, WANG Z, KUEN J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77: 354-377.
- [55] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal Computer Vision, 2015, 115(3): 211252.
- [56] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains[J]. IEEE Signal Processing Magazine, 2013, 30(3): 83-98.
- [57] GAMA F, MARQUES A G, MATEOS G, et al. Rethinking sketching as sampling: A graph signal processing approach[J]. Signal Processing, 2020, 169: 107404.
- [58] YANG L X, ZHANG R Y, LI L D, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the 38th International Conference on Machine Learning (PMLR). [S.l.]:PMLR,2021: 11863-11874.
- [59] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on Machine Learning (ICML). [S.l.]: ICML,2015: 448-456.
- [60] BEN HAMIDA A, BENOIT A, LAMBERT P, et al. 3-D deep learning approach for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8): 4420-4434.
- [61] OUYANG E, LI B, HU W, et al. When multigranularity meets spatialspectral attention: A hybrid transformer for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-18.

Acknowledgements This work was supported in part by the National Natural Science Foundations of China (No. 61801214), and the Postgraduate Research Practice Innovation Program of NUAA (No.xcxjh20231504).

Authors

The first author Mr. XU Chenjie received the M.S. degree in optical engineering from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2025. His research interests include hyperspectral image classification and deep learning.

The corresponding author Dr. LI Dan received the B.S., M.S., and Ph.D. degrees in control science and engineering from Harbin Institute of Technology (HIT), Harbin, China, in 2012, 2014, and 2018, respectively. She is currently an associate professor with College of Astronautics, NUAA, Nanjing, China. Her research interests include hyperspectral image classification, signal processing, sparse sampling, and reconstruction technology.

Author contributions Mr. XU Chenjie was responsible for technical analysis, experimental validation, and drafting the initial manuscript. Dr. LI Dan designed the study, participated in method development, and revised the manuscript. Dr. KONG Fanqiang provided overall guidance for the project and offered expert insights on the interpretation of results. All authors commented on the manuscript and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: LIU Yandong)

基于谱段图卷积与注意力增强卷积联合网络的高光谱图像分类方法

徐陈捷, 李 丹, 孔繁铨

(南京航空航天大学航天学院, 南京 211106, 中国)

摘要: 高光谱图像(HSI)分类对于众多遥感应用至关重要。传统的深度学习方法可能会丢失像元间上下文信息, 从而导致分类效果不佳。本文介绍了谱段图卷积与注意力增强卷积联合网络(SGCCN), 这是一种利用谱段图卷积来捕获高光谱图像的长程全局特征利用注意力增强多级卷积的局部感知来获取局部空间特征并采用动态注意力机制来增强特征提取的新方法。SGCCN通过自注意力融合网络集成光谱和空间特征, 显著提高了分类精度和效率。通过实验证明了本文方法在处理与高光谱图像分类任务中的有效性。

关键词: 高光谱图像分类; 谱段图卷积; 注意力增强卷积; 动态注意力; 特征提取; 特征融合