



Article

Pyramid Cascaded Convolutional Neural Network with Graph Convolution for Hyperspectral Image Classification

Haizhu Pan ^{1,2,*}, Hui Yan ¹, Haimiao Ge ^{1,2} , Ligu Wang ³ and Cuiping Shi ⁴

¹ College of Computer and Control Engineering, Qiqihar University, Qiqihar 161000, China; 2022912329@qqhru.edu.cn (H.Y.); 01557@qqhru.edu.cn (H.G.)

² Heilongjiang Key Laboratory of Big Data Network Security Detection and Analysis, Qiqihar University, Qiqihar 161000, China

³ College of Information and Communication Engineering, Dalian Nationalities University, Dalian 116000, China; wangliguo@hrbeu.edu.cn

⁴ College of Telecommunication and Electronic Engineering, Qiqihar University, Qiqihar 161000, China; shicui ping@qqhru.edu.cn

* Correspondence: panhaizhu@qqhru.edu.cn

Abstract: Convolutional neural networks (CNNs) and graph convolutional networks (GCNs) have made considerable advances in hyperspectral image (HSI) classification. However, most CNN-based methods learn features at a single-scale in HSI data, which may be insufficient for multi-scale feature extraction in complex data scenes. To learn the relations among samples in non-grid data, GCNs are employed and combined with CNNs to process HSIs. Nevertheless, most methods based on CNN-GCN may overlook the integration of pixel-wise spectral signatures. In this paper, we propose a pyramid cascaded convolutional neural network with graph convolution (PCCGC) for hyperspectral image classification. It mainly comprises CNN-based and GCN-based subnetworks. Specifically, in the CNN-based subnetwork, a pyramid residual cascaded module and a pyramid convolution cascaded module are employed to extract multiscale spectral and spatial features separately, which can enhance the robustness of the proposed model. Furthermore, an adaptive feature-weighted fusion strategy is utilized to adaptively fuse multiscale spectral and spatial features. In the GCN-based subnetwork, a band selection network (BSNet) is used to learn the spectral signatures in the HSI using nonlinear inter-band dependencies. Then, the spectral-enhanced GCN module is utilized to extract and enhance the important features in the spectral matrix. Subsequently, a mutual-cooperative attention mechanism is constructed to align the spectral signatures between BSNet-based matrix with the spectral-enhanced GCN-based matrix for spectral signature integration. Abundant experiments performed on four widely used real HSI datasets show that our model achieves higher classification accuracy than the fourteen other comparative methods, which shows the superior classification performance of PCCGC over the state-of-the-art methods.



Citation: Pan, H.; Yan, H.; Ge, H.; Wang, L.; Shi, C. Pyramid Cascaded Convolutional Neural Network with Graph Convolution for Hyperspectral Image Classification. *Remote Sens.* **2024**, *16*, 2942. <https://doi.org/10.3390/rs16162942>

Academic Editor: Saeid Homayouni

Received: 25 June 2024

Revised: 4 August 2024

Accepted: 8 August 2024

Published: 11 August 2024

Keywords: hyperspectral image classification; multiscale features extraction; convolutional neural network; graph convolutional network; mutual-cooperative attention mechanism



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of hyperspectral imaging techniques, hyperspectral images (HSIs) with abundant spectral signatures and spatial features are available. HSIs have hundreds of narrow and continuous electromagnetic spectrum bands, spanning from the visible to the near-infrared ranges. HSI can reflect the area on an earth surface with spectral and spatial information. And, it has been widely utilized in various applications, e.g., mineral exploitation [1], environment surveillance [2], water quality monitoring [3], and urban planning [4,5]. Among various applications, HSI classification, which discriminates each land-cover type at the pixel-by-pixel level [6], plays a critical role and has attracted

more researchers to study it. However, it is a significant challenge for HSI classification that the HSIs contain excessively high spectral dimensions and complex spatial features [7].

The traditional methods based on hand-crafted features for HSI classification mainly fall into two categories, such as spectral-based and spatial spectral-based methods [8,9]. The spectral-based method, e.g., random forest [10], k-nearest neighbor [11], and support vector machine [12], which only learn spectral signatures, produce the classification maps with salt-and-pepper phenomena and unsatisfactory classification performance. At the same time, the principal component analysis [13] method is developed to remove redundant spectral signatures for spectral signatures extraction. To further improve the classification performance, another category of methods based on spectral-spatial feature have emerged, such as extended morphological profile [14] and superpixel segmentation [15]. In these, the classification result is enhanced compared with the spectral-based methods. Although the aforementioned methods can effectively solve the HSI classification tasks, these methods have some limited feature extraction capabilities and fail to learn deep semantic information due to a lack of strong data fitting abilities and being less robust for complex HSI data scenes [16,17].

Over the past few years, deep learning (DL) has been recognized as a powerful data analysis technique [18] for effectively addressing nonlinear problems, and it has been extensively used for HSI processing tasks. Compared with traditional machine learning methods, DL, which includes various deep networks such as CNN [19], RNN [20], GNN [21], Transformer [22–24], and GAN [25], makes significant progress in HSI classification. Among various network architectures, CNN and GCN have received the most attention. In the case of CNN, it is extensively used because it can reuse convolutional kernels with shared weights across input feature maps, enabling it to adeptly learn features of images [26]. For example, Hu et al. [19] employ 1D convolution to learn features along the spectral channel of the HSI data cube, resulting in good classification performance compared with traditional methods. Although the method based on 1D convolution better adapts to spectral signatures, it overlooks spatial information, which limits the capability of the method to describe spatial contextual information. To cope with this problem, in [27], a novel devised dual-branch spectral and spatial model is proposed. This network uses 1D convolution and 2D convolution to learn spectral signatures and spatial features separately. Subsequently, a fully connected layer is utilized to exploit spectral and spatial correlation. The mentioned double-branch network learns the spectral signatures and spatial features separately. To jointly excavate features from a 3D HSI cube, in [28], a newly developed cascaded 3D convolution to learn spectral-spatial features is used. However, this cascaded 3D CNN uses a 3D kernel with a size of $(h \times w \times d)$, which has many parameters. In this article, 3D convolutions with kernel sizes of $(h \times w \times 1)$ and $(1 \times 1 \times d)$ are employed to reduce the network parameters. Meanwhile, too many 3D convolutions will lead to the disappearance of the network gradients and decrease the final classification accuracy. To conquer this phenomenon, the residual network [29] and the dense network [30] are utilized. Although the above-mentioned CNN-based methods can be better used for HSI classification, they overlook the extraction of multiscale features [31], thus not being robust enough for various HSI data scenes. Meanwhile, the CNN method based on grid data has limitations in capturing the relationships between samples in HSI data.

With the emergence of graph neural network (GNN) [32] architecture, GNNs have been used by some researchers as effective tools to deal with HSI data based on non-Euclidean geometry properties. Among various GNN architectures [33–35], GCN has been widely used in HSI processing due to its direct application to arbitrarily shaped graphs, allowing it to learn the graph structure information and node features simultaneously [36]. To proficiently utilize the relationships between different nodes, the methods for graph construction of HSI data typically include superpixel-based and pixel-based [37,38] approaches. Specifically, from the perspective of superpixels, in [33], a novel graph convolutional method that utilizes the GCN to extract features from the constructed graph using superpixel methods was explored. Based on the superpixel graph, Wan et al. [33] employ

multiscale stacked GCN layers to learn context-based spatial information. Although the above-mentioned methods based on GCN can achieve satisfactory feature extraction, the graph constructed by the superpixel method tends to overlook certain local spectral and spatial information. Then, a graph-in-graph method is devised by Jia et al. [39], where each node within a local range structure is called an internal graph, and all the nodes of the HSI form an external graph. This approach can highlight both the local and global information of the entire graph. Although the graph based on superpixels can describe the local features of HSI well, pixels with different labels may be assigned to the same superpixel region, resulting in the misclassification of results in the final classification map [8]. Meanwhile, the constructed superpixel-based graph models may overlook pixel-level feature descriptions, which further leads to poor pixel-level classification results. To cope with these difficulties, from another pixel-based perspective, a mini-batch GCN method is proposed by [40], which employs a batch-by-batch GCN network training approach for pixel-wise feature extraction of HSI data. Based on the patch training strategy, Gao et al. [41] propose a novel model, which proficiently improves neighborhood node aggregation by adaptively learning the weight correlations between different nodes. Zhang et al. [42] devise a proficient GCN method, which regards the HSI data as graph-structured data and systematically aggregates structural information between different nodes for pixel-wise land cover processing.

GCNs use convolution as a weighted function to indicate the influence exerted on a target node by its neighbors and itself, which is beneficial for graph-based data processing. The constructed graph can be updated to adapt to the HSI data representations produced by each GCN layer, which in turn makes the data representations more accurate [43]. Meanwhile, the GCN can handle arbitrary graph-based data and efficiently learn the internal similarity relationships between adjacent nodes in HSI data [44]. Following the benefits of GCN, some researchers try to combine the advantages of CNNs and GCNs for classifying different HSI data scenes [45]. Lu et al. [46] develop a novel model that combines a separable GCN with a CNN for HSI classification. Specifically, the model encodes spectral-spatial features, adaptively learned by the designed attention module, into the structure of a graph. Then, a separable deep GCN is developed to learn long-range contextual structure relationships from the graph. Meanwhile, a local convolutional feature extraction network is employed to extract complementary local features. To make the model compatible with different HSI data scenes, Li et al. [47] design a staged feature fusion model that combines CNNs and GCNs. In the first stage, the model uses CNN to extract non-local features. In the second stage, GCN is employed to optimize the connectivity relationships of the graph constructed based on spectral similarity. Shi et al. [44] propose a novel network that has a graph convolution branch and grouping convolution branch. In the graph branch, a multihop graph rectify attention is proposed to weigh the features extracted by the GCN. In the convolution branch, a spectral intra-group and inter-group signature extraction module is designed to address the problem of high spectral dimensionality. Ghotekar et al. [45] devise an feature segmentation network, consisting of hybrid convolution and graph convolution networks, for HSI classification. First, a CNN is used to extract multi-layer features. Then, the features are fed into GCN module to obtain patch-to-patch correlation feature maps. Finally, the extracted features are concatenated to be fed into the linear layer for the final classification results.

These hybrid collaborative networks, which include CNNs and GCNs, show efficient performance in HSI processing. Nevertheless, some challenges persist in hybrid networks for HSI feature learning. For hybrid networks, the CNN module based on the single-scale kernel exhibits limited effectiveness in extracting features from original spectral and spatial information included in complex HSI data scenes. And the lack of multiscale features will inevitably lead to poor classification results. To extract graph-based high-level features, the GCN-based module employs multiple stacked graph convolutional layers, which inevitably results in oversmoothing issues. Considering that CNNs employ shared kernels across spectral feature maps to learn pixel spectral signatures and GCNs obtain pixel spectral signatures through two different matrix multiplications, the CNN-based spectral signatures

may have a certain degree of incompatibility with the GCN-based spectral signatures. Therefore, directly integrating the two different types of pixel spectral signatures can result in degraded final classification accuracy.

In this article, we propose a novel pyramid cascaded convolutional neural network with graph convolution (PCCGC) for HSI classification. It contains two parallel subnetworks, e.g., a CNN-based subnetwork and a GCN-based subnetwork. Specifically, from the perspective of the CNN-based subnetwork, it includes a spectral pyramid residual cascaded module and a spatial pyramid convolution cascaded module. The spectral module features a designed spectral pyramid hybrid convolution block and multiple 3D spectral convolution layers, which are connected in a cascaded manner for multiscale spectral signature extraction. Moreover, considering that HSIs have rich spectral signatures, a residual connection is employed for spectral signature extraction. The spatial module is similar to the spectral module for multiscale spatial feature extraction. The difference between the two modules is the 3D convolution kernel used. Then, an adaptive feature-weighted fusion strategy is utilized to fuse multiscale spectral and spatial features based on their respective weights. From another perspective of the GCN subnetwork, a band selection network (BSNet) is used to learn the spectral signatures in the HSI using non-linear inter-band dependencies. Then, a spectral-enhanced GCN module is utilized to learn and accentuate the important information in the spectral matrix. And to prevent the oversmoothing problem and learn deep features, multiple graph convolution layers are utilized in a one-shot strategy. Subsequently, a mutual-cooperative attention mechanism is constructed that can align the spectral signatures between a BSNet-based matrix with a spectral-enhanced GCN-based matrix for pixel-wise spectral signature integration. It can transfer the spectral features extracted by BSNet to the GCN-based feature matrix through cross multi-head self-attention blocks, and transfer the spectral features learned by the spectral-enhanced GCN module to the BSNet-based spectral feature matrix through cross multi-head self-attention block. Finally, an additive fusion strategy is utilized to fuse the features extracted by the CNN-based and the GCN-based subnetworks. Our main contributions are as follows:

(1) To extract multiscale spectral and spatial features from complex HSI datasets, the spectral pyramid residual cascaded module and spatial pyramid convolution cascaded module are designed. The spectral module includes a devised spectral pyramid hybrid convolution block and multiple 3D spectral convolution layers, which are connected in a cascaded manner. Moreover, considering that HSIs have rich spectral signatures, a residual connection is employed to enhance spectral signature extraction. The spatial module is similar to the spectral module but is used for spatial feature extraction. Furthermore, the 3D convolution kernels used in spectral and spatial modules are different, which benefits the extraction of spectral and spatial features separately. Then, an adaptive feature-weighted fusion strategy is utilized to fuse multiscale spectral and spatial features based on their respective weights.

(2) To model the important spectral relations of the samples, a spectral-enhanced GCN module is employed. It can strengthen the deep significant spectral relations based on the constructed graph and capture the interconnectivity between pixels as well as the interdependencies among spectral signatures. To prevent the oversmoothing problem, the multiple graph convolution layers in the spectral-enhanced GCN module are stacked in a one-shot strategy.

(3) A mutual-cooperative attention mechanism is constructed to align the spectral signatures between the BSNet-based matrix and the spectral-enhanced GCN-based matrix for spectral signature integration. It transfers the spectral features extracted by BSNet to the GCN-based feature matrix through a cross multi-head self-attention block, and transfers the spectral features learned by the spectral-enhanced GCN module to the BSNet-based spectral feature matrix through another cross multi-head self-attention block. Subsequently, the two aligned matrices are concatenated for spectral signature integration.

(4) A novel method called the PCCGC is proposed to realize hyperspectral image classification. PCCGC can extract CNN-based multiscale spectral and spatial features, which are then fused adaptively. In addition, PCCGC utilizes BSNet and spectral-enhanced GCN for significant pixel-wise spectral signature extraction, and these extracted pixel-wise spectral signatures are integrated using a mutual-cooperation attention mechanism. Furthermore, the integrated spectral signature are added to the CNN-based features, resulting in the proposed model achieving good classification performance.

The remainder of this article is structured as follows: related work is shown in Section 2, the devised PCCGC is characterized in Section 3, the experimental results are listed in Section 4, some parameters of PCCGC are discussed in Section 5, and the conclusion is presented in Section 6.

2. Related Work

2.1. Convolutional Neural Network

An HSI data patch with a size $H \times W \times D$ is specified as input data, where $H \times W$ indicates the spatial size, and D represents the number of spectral bands [48]. In (1), the 3D convolution has p 3D convolution kernels of size $(h \times w \times c)$. Following the 3D convolution process, p feature maps of size $(H - h + 1) \times (W - w + 1) \times (D - c + 1)$ are generated. Moreover, by calculating the dot product between the local area position (x, y, z) and the weight matrix, each feature map (F) is obtained [49]. The output of a neuron $v_{l,i}^{x,y,z}$ at the position (x, y, z) of the i -th F in the l th layer can be calculated by:

$$v_{l,i}^{x,y,z} = \delta \left(\sum_p \sum_{h=0}^{H_l-1} \sum_{w=0}^{W_l-1} \sum_{c=0}^{C_l-1} k_{l,i,p}^{h,w,c} \times v_{(l-1),p}^{(x+h),(y+w),(z+c)} + b_{l,i} \right) \quad (1)$$

where δ indicates the activation function, such as Mish, and $b_{l,i}$ is the bias of the i th F in the l th layer. The p index indicates the connection between the current F and the F in the previous layer. W_l and H_l are indicate the width and height of the 3D convolution kernel in the spatial dimension, respectively. C_l refers to the 3D convolution kernel size of the spectral dimension. The weight $k_{l,i,p}^{h,w,c}$ is used to convolve the input data cube $v_{(l-1),p}^{(x+h),(y+w),(z+c)}$ in 3D convolution kernels, with an offset of (h, w, c) [50].

The (1) in our manuscript is the process of the 3D convolution, which is similar to 1D and 2D convolution. However, it is essential to note that the input format of a 3D convolutional network is $(B_{batch}, C_{channel}, H_{height}, W_{width}, D_{spectral})$. The 3D feature extraction model has demonstrated itself to be very effective in simultaneously capturing the spatial and spectral features of 3D feature maps by applying 3D kernels to 3D hyperspectral image data scenes [8,51]. Compared with the 1D convolution and 2D convolution operation process for HSI data with rich spectral signatures, the 3D convolution can greatly decrease spectral distortion phenomenon and learn more information (e.g., spatial-spectral correlation characteristics and absorption differences between adjacent spectral bands). Moreover, the 3D CNN is theoretically well-suited to excavating 3D feature maps for HSI processing since HSIs are usually denoted as a 3D patch cube.

2.2. Graph Convolutional Network

Graph neural networks (GNNs) can generalize the convolution process from grid-based data to graph-based data. The fundamental concept is to describe a node V with its own feature and neighbors' feature. The model based on graph convolution can learn the high-level node feature representation through multiple stacked convolutional layers. And, GNNs fall into two categories [52], being spectral-based [53] and spatial-based [54] methods. And the spectral-based GNN methods define convolution in the graph signal processing. Spatial-based GNN methods define convolution by the information propagation strategy. Among Spatial-based and Spectral-based GNN methods, the GCN is widely employed due to its generality.

The undirected graph is typically defined as $G = (V, E)$, where V denotes the set of nodes or vertices, and E represents the set of edges. According to the undirected graph G , the adjacency A is constructed. Based on the convolution on the undirected graph, here is a layer-by-layer propagation rule for multiple layers GCN, defined as follows:

$$H^{(i+1)} = \varphi \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^i W^i \right) \quad (2)$$

where $\tilde{A} = A + I$ is termed as renormalization of A , A is the adjacency matrix of G , and I is the identity matrix. \tilde{D} is defined as renormalization of D , and $D_{ii} = \sum_j (A_{ij})$ is a diagonal matrix indicating the degree of A . $\varphi(\cdot)$ indicates an activation function, such as ELU(). $H^{(i+1)} \in \mathbb{R}^{N \times D}$ and $H^{(i)} \in \mathbb{R}^{N \times D}$ is the matrix of $(i+1)$ th and (i) th layer, respectively. Besides, $H^0 = X$, where X indicates a matrix of feature vectors of the input node. W^i represent the trainable weight matrix.

2.3. CNN and GCN for HSI Classifications

Considering that HSI data has rich spectral and spatial information, an effective model is well-suited to HSI data classification. CNNs can effectively extract local features using shared weight kernels. Based on different convolutional kernels, 1D CNNs extract spectral signatures, 2D CNNs extract spatial features, and 3D CNNs extract spectral-spatial features. The use of CNNs can greatly improve HSI processing performance. Meanwhile, GCNs generalize the convolution operation to graph data [8], allowing for the learning of node feature representations. High-level feature representations can be captured by multiple stacked GCN layers [52]. In HSI data, GCNs are employed to capture spatial contextual structure information, which is advantageous for HSI information processing. Based on the above-mentioned, some models combining CNNs with GCNs are designed for HSI classification. Liu et al. [43] design a novel heterogeneous network called CNN-enhanced GCN. Specifically, the 2D CNN is used to extract features from local-range regular regions, while the GCN is employed to learn features from long-range irregular region. The features extracted by both CNN and GCN are then used as complementary features for HSI classification. Lu et al. [46] develop a novel SDGCP method. It employs a separable deep GCN for learning long-range contextual structure features. The learned features are then combined with local complementary features extracted by CNN for HSI classification. Wang et al. [55] design a novel DF2Net for HSI classification, which includes two subnetworks: a spectral-spatial hypergraph convolutional subnetwork for learning long-range and high-order correlations, and a spectral-spatial convolution subnetwork for pixel-wise local feature extraction.

3. Methods

3.1. The Overall Structure of PCCGC

In this article, we propose a novel PCCGC method for HSI classification. As depicted in Figure 1, it contains two parallel subnetworks, e.g., a CNN-based subnetwork and a GCN-based subnetwork. Specifically, from the view of CNN subnetwork, a spectral pyramid residual cascaded module (SpePRCM) is used to extract multiscale spectral signatures. Meanwhile, a spatial pyramid convolution cascaded module (SpaPCCM) is employed to extract multiscale spatial features. And, the features extracted by the CNN subnetwork are more robust for the proposed model in classifying HSIs. Furthermore, an adaptive feature-weighted fusion strategy is utilized to adaptively fuse multiscale spectral and spatial features based on their respective weights. From another perspective of the GCN subnetwork, a BSNet is used to learn the spectral signatures in the HSI using non-linear inter-band dependencies, which also reduces the computational cost of GCN. Then, the spectral-enhanced GCN module is utilized to learn and accentuate the important information in the spectral matrix. Subsequently, a mutual-cooperative attention mechanism is

constructed to align the spectral signatures between a BSNet-based matrix and a spectral-enhanced GCN-based matrix for spectral signature integration. Finally, the additive fusion strategy is utilized to fuse the features extracted from GCN-based and CNN-based sub-networks. In the following, we will elaborate on the functionalities of each module in the proposed model.

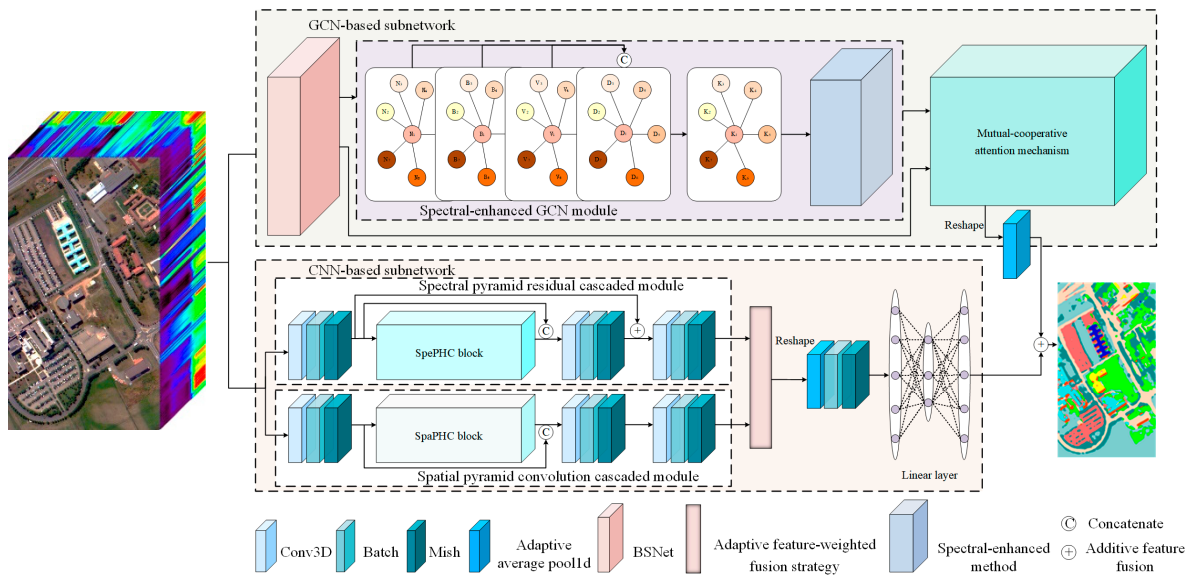


Figure 1. The overall structure of the PCCGC.

3.2. Adaptive Feature-Weighted Feature Fusion Based SpePRCM and SpaPCCM

Considering the HSI data cubes, which contain plentiful spectral signatures and a lot of spatial information, the SpePRCM and SpaPCCM are devised to extract multiscale spectral and spatial features separately. Moreover, the spectral pyramid hybrid convolution (SpePHC) block and spatial pyramid hybrid convolution (SpaPHC) block are included in SpePRCM and SpaPCCM separately. And then an adaptive feature-weighted fusion strategy is employed to fuse the extracted multiscale spectral and spatial information. Furthermore, the 3D convolutional layer used hereinafter refers to 3D convolution, Mish activation function, and Batch Normalization.

3.2.1. Spectral Pyramid Hybrid Convolution Block

The proposed SpePHC, as shown in Figure 2, includes a pyramid architecture with different types of convolutional layers, featuring various sizes of kernels and varying numbers of output feature channels. The processed spectral feature maps $FM_{spe}^i \in \mathbb{R}^{h \times w \times d}$, where i indicates the i th layer, are fed into the SpePHC block, then FM_{spe}^i are processed in parallel by three different steps, e.g., *Step*₁, *Step*₂, and *Step*₃. For *Step*₁, the FM_{spe}^i are convolved by the 3D convolutional layer $Conv_{spe}^{i,3}$ located at the bottom of the pyramid architecture, with the purpose of learning the spectral feature using a kernel size of $(1 \times 1 \times 5)$. This results in the spectral feature maps $FM_{spe}^{(i+1)-3}$, which have an output dimension of 36. For *Step*₂, the FM_{spe}^i are convolved by the 3D Transpose convolutional layer $TransConv_{spe}^{i,2}$ located at the middle of the pyramid architecture, with the purpose of learning the spectral feature using a kernel size of $(1 \times 1 \times 3)$. This yields the spectral feature map $FM_{spe}^{(i+1)-2}$, which has an output dimension of 24. For *Step*₃, the FM_{spe}^i are convolved by the 3D convolutional layer $Conv_{spe}^{i,1}$ located at the top of the pyramid architecture, which uses a kernel size of $(1 \times 1 \times 1)$. The output dimensions of resulting spectral feature maps $FM_{spe}^{(i+1)-1}$ are 12. Then, we concatenate the three different spectral feature

maps using the concatenate operation in the channel dimension, and the output spectral feature maps FM_{spe}^{i+2} are obtained. The detailed operation process is shown in the following:

$$FM_{spe}^{(i+1)-1} = Conv_{spe}^{i-1}(FM_{spe}^i) \tag{3}$$

$$FM_{spe}^{(i+1)-2} = TransConv_{spe}^{i-2}(FM_{spe}^i) \tag{4}$$

$$FM_{spe}^{(i+1)-3} = Conv_{spe}^{i-3}(FM_{spe}^i) \tag{5}$$

$$FM_{spe}^{i+2} = Concat(FM_{spe}^{(i+1)-3}, FM_{spe}^{(i+1)-2}, FM_{spe}^{(i+1)-1})_{dim=channel} \tag{6}$$

where $Concat()_{dim=channel}$ indicates the concatenation operation that operates on the channel dimension.

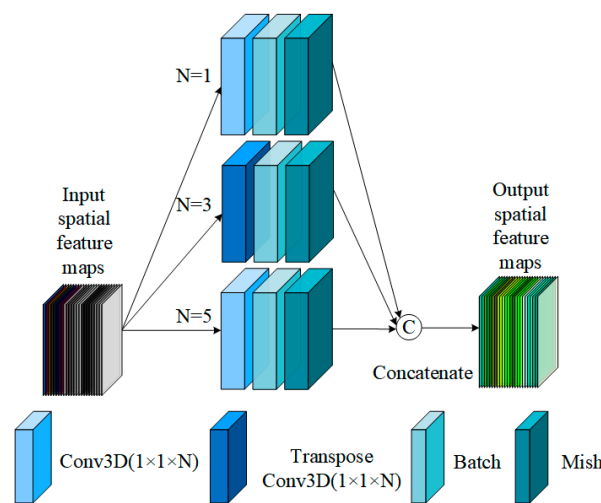


Figure 2. The detailed structure of the SpePHC block.

3.2.2. Spatial Pyramid Hybrid Convolution Block

The proposed spatial pyramid hybrid convolution (SpaPHC), as shown in Figure 3, is employed to learn multiscale spatial features. It has almost the same pyramid architecture as SpePHC and includes three parallel steps for processing the processed spatial feature maps FM_{spa}^i .

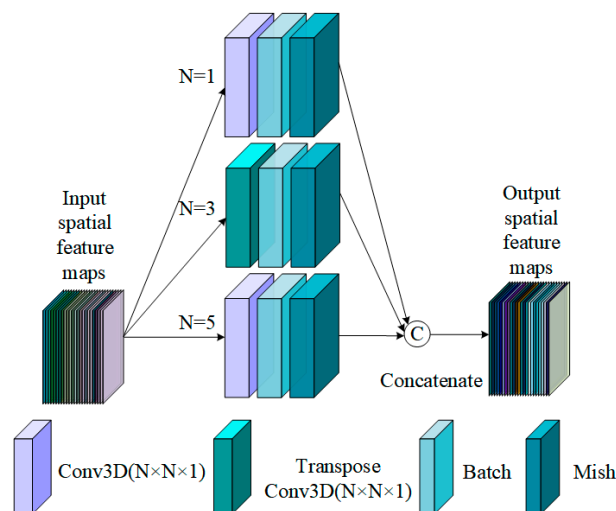


Figure 3. The detailed structure of the SpaPHC block.

For the first and second step, it has almost the same procedures as the *Step_1* and *Step_2* in the SpePHC block, but with different kernel sizes within the 3D convolutional layer $Conv_{spa}^{i-1}$ and 3D Transpose convolutional layer $Conv_{spa}^{i-2}$, which are $(5 \times 5 \times 1)$ and $(3 \times 3 \times 1)$ respectively. And then, two different spatial feature maps are generated, namely, $FM_{spa}^{(i+1)-3}$ with output dimension 12, and $FM_{spa}^{(i+1)-2}$ with output dimension 24. For the third step, the FM_{spa}^i are convolved by the 3D convolutional layer $Conv_{spa}^{i-1}$ located at the top of the pyramid architecture using a kernel size of $(1 \times 1 \times 1)$, resulting in the spatial feature maps $FM_{spa}^{(i+1)-1}$ with an output dimension of 48. Then, we concatenate the three different spatial feature maps in the channel dimension, the output spatial feature maps FM_{spa}^{i+2} are obtained.

$$FM_{spa}^{(i+1)-1} = Conv_{spa}^{i-1} \left(FM_{spa}^i \right) \quad (7)$$

$$FM_{spa}^{(i+1)-2} = TransConv_{spa}^{i-2} \left(FM_{spa}^i \right) \quad (8)$$

$$FM_{spa}^{(i+1)-3} = Conv_{spa}^{i-3} \left(FM_{spa}^i \right) \quad (9)$$

$$FM_{spa}^{i+2} = Concat \left(FM_{spa}^{(i+1)-3}, FM_{spa}^{(i+1)-2}, FM_{spa}^{(i+1)-1} \right)_{dim=channel} \quad (10)$$

The $Concat()_{dim=channel}$ indicates the concatenation operation that operates on the channel dimension of the spatial feature maps.

The proposed SpePHC and SpaPHC can generate spectral and spatial feature maps with various receptive field and corresponding output feature channels, learning information about more granular-level objects with larger output feature channels, as well as capturing more details about context information with smaller output feature channels.

3.2.3. The Multiscale Spectral and Spatial Feature Extraction of SpePRCM and SpaPCCM

In the spectral pyramid residual cascaded module (SpePRCM), the original feature map FM_{HSI}^j , j indicates the j th layer, and is first processed by the 3D convolutional layer $Conv_{spe}^j$ with a kernel size of $(1 \times 1 \times 7)$ to extract the spectral signatures. This generates the spectral feature maps FM_{spe}^{j+1} with the output channel number of 46. To extract the multiscale spectral features, the FM_{spe}^{j+1} are fed into the SpePHC block $P_{spe}(x; \varepsilon)$, the FM_{spe}^{j+2} are generated, where ε is a learnable parameter. Then, the concatenation operation is employed on this FM_{spe}^{j+1} and FM_{spe}^{j+2} along the channel dimension. To further extract the spectral features from the previous feature maps and prevent information loss, the 3D convolutional layer $Conv_{spe}^{j+2}$ with a kernel size of $(1 \times 1 \times 1)$ is employed, to obtain the multiscale spectral feature maps FM_{spe}^{j+3} . At the same time, the residual connection $Res()$ is added to the FM_{spe}^{j+1} and FM_{spe}^{j+3} to assist the spectral pyramid feature extraction module in learning the original spectral signatures, thereby benefiting the improvement of classification accuracy. Finally, to fully extract the multiscale spectral signatures in the HSIs and decrease the depths of the HSI data cube, the 3D convolutional layer $Conv_{spe}^{j+3}$ with a kernel size of $\left(1 \times 1 \times \left(\frac{band-7}{2} + 1\right)\right)$ is utilized to obtain the spectral feature maps FM_{spe}^{j+4} with an output channel number of 72. The detailed operation is presented below:

$$FM_{spe}^{j+1} = Conv_{spe}^j \left(FM_{HSI}^j \right) \quad (11)$$

$$FM_{spe}^{j+2} = P \left(FM_{spe}^{j+1} \right) \quad (12)$$

$$FM_{spe}^{j+3} = Conv_{spe}^{j+2} \left(Cat \left(FM_{spe}^{j+1}, FM_{spe}^{j+2} \right)_{dim=channel} \right) \quad (13)$$

$$FM_{spe}^{j+4} = Conv_{spe}^{j+3} \left(Res \left(FM_{spe}^{j+1}, FM_{spe}^{j+3} \right) \right) \quad (14)$$

The spatial pyramid convolution cascaded module includes the 3D convolutional layer $Conv_{spa}^j$ with a kernel size of $(1 \times 1 \times band)$, which is employed to squeeze the depth of the FM_{HSI}^j , resulting in the spatial feature maps FM_{spa}^{j+1} . Then, the SpaPHC block $P_{spa}(x; \varepsilon)$ is utilized to extract the multiscale spatial feature, yielding the spatial feature maps FM_{spa}^{j+2} . And the subsequent operations in the spatial pyramid feature extraction module are similar to the spectral signature extraction process in the spectral pyramid feature extraction module. The detailed multiscale spatial feature extraction process in the spatial pyramid convolution cascaded module is shown as follows:

$$FM_{spa}^{j+1} = Conv_{spa}^j \left(FM_{HSI}^j \right) \quad (15)$$

$$FM_{spa}^{j+2} = P \left(FM_{spa}^{j+1} \right) \quad (16)$$

$$FM_{spa}^{j+3} = Conv_{spa}^{j+2} \left(Cat \left(FM_{spa}^{j+1}, FM_{spa}^{j+2} \right)_{dim=channel} \right) \quad (17)$$

The FM_{spa}^{j+3} are the output spatial feature maps generated from the spatial pyramid convolution cascaded module (SpaPCCM). The SpePHC and SpaPHC are included in CNN-based subnetwork which make it more generalized and robust while learning from different datasets.

3.2.4. The Multiscale Spectral and Spatial Feature Fusion with the Adaptive Feature-Weighted Fusion Strategy

Considering the importance of the feature for classification results, the extracted multiscale spectral and spatial features play a significant but unequal role. Meanwhile, to fully harness the extracted spectral and spatial features, inspired by [56], the adaptive feature-weighted fusion strategy is employed. It can fuse the spectral signatures and spatial features adaptively. In this strategy, the extracted features located at the same location are added element-wise, aggregating information across spectral signatures and spatial features, thereby enhancing the features that are important to the classification accuracy. Meanwhile, to dynamically allocate weights, two different weight coefficients, namely, α_1 and α_2 , are used. To avoid the poor classification accuracy that could result from these two values being too large or too small, a softmax function is applied to adjust the values of α_1 and α_2 . Then, α_1 and α_2 will balance the spectral signatures and spatial features, enhancing the fusion of different information according to the impact of various features on classification accuracy. The detailed operation is illustrated as follows:

$$FM = \alpha_1 \cdot FM_{spe}^{j+4} + \alpha_2 \cdot FM_{spa}^{j+3}, \quad (\alpha_1 + \alpha_2 = 1) \quad (18)$$

3.3. Spectral-Enhanced GCN Module

To learn the features in the non-grid image data scene, the GCN is employed. It can capture information different from the CNN, thereby enhancing the classification accuracy. The graph is usually defined as $G = (V, E)$, where V represent the set of nodes or vertices, and E represent the set of edges. Assume $v_i \in V$ to indicate a node and $e_{ij} = (v_i, v_j) \in E$ to indicate an edge starting from v_i to v_j . The neighborhood of a node v is characterized by the set $NUM(v) = \{u \in V | (v, u) \in E\}$. The adjacency matrix $A^{n \times n}$ of graph G is defined as follows:

$$A^{n \times n} = \begin{cases} 1, & e_{ij} \in E \\ 0, & e_{ij} \notin E \end{cases} \quad (19)$$

Graph G might possess node attributes denoted as S , where $S \in R^{m \times c}$ is a node feature matrix, and $s_v \in R^c$ represents the feature vector of a node v . Simultaneously, a graph G might possess edge attributes denoted as S_e , where $S_e \in R^{m \times d}$ is an edge feature matrix, and $s_e^{v,u} \in R^d$ denotes the feature vector of an edge (v, u) .

In this part, the spectral-enhanced GCN module is developed to extract and accentuate the features from the spectral channel. Considering the input scalar types required by the GCN, we need to construct an undirected graph $G = (V, E)$ based on each patch from HSI data. Due to the abundance of spectral signatures in each pixel of original HSI, the GCN methods based on the origin HSI data will create a larger graph, resulting in an abundance of computational resource costs [37,57]. To address the above-mentioned issue, inspired by [58], we employ the BSNet to select important spectral signatures from the original HSI patch $FM_{HSI} \in \mathbb{R}^{h \times w \times c}$ in the spectral channel.

$$FM_{BS} = BSNet(FM_{HSI}; \theta) \quad (20)$$

where FM_{BS} is the output of BSNet, and $BSNet()$ function indicates the BSNet. Here, θ is the learnable parameter in BSNet. This selection enhances the performance of GCN for HSI classification by considering the nonlinear interdependencies among different spectral bands.

By stacking multi-layer GCN, it is possible to learn deeper node information from the constructed graph. However, stacking a certain number of GCN layers will result in a decrease in model performance and lower classification accuracy. In this part, to avoid the mentioned issues, the multi-hop adjacency matrix [57] is constructed, in which record nodes are at a distance of d hops from the selected nodes. It can excavate the underlying feature relationships and enlarge the receptive field. The used multi-hop adjacency matrix is constructed in the spectral channel to help the GCN learn the spectral signature in the HSI.

For constructing the multi-hop adjacency matrix, the FM_{HSI} processed by the BSNet is first transformed into feature nodes $M_{BS} \in \mathbb{R}^{hwc}$:

$$M_{BS} = reshape(FM_{BS}) \quad (21)$$

where M_{BS} is the result processed by BSNet, abbreviated as the BS-based feature matrix, and $reshape()$ indicates the reshape operation. Then, the multi-hop adjacency matrix A is constructed based on the M_{BS} .

The spectral-enhanced GCN module, depicted in Figure 4, is employed to learn the feature relationships among multi-hop adjacency matrix through a multi-layer GCN. To learn the features of HSIs from the perspective of GCN, a four-layer cascaded GCN, in which the number of GCN layers is discussed in the Section 5.2, is first implemented in a one-shot strategy. In each GCN layer, a collection of feature nodes from the M_{BS} , denoted as $N_i = \{N_1, N_2, N_3, \dots, N_n\}$, $N \in \mathbb{R}$, where n represents the number of feature nodes, are fed into the GCN layer. And a learnable parameter matrix $W \in \mathbb{R}^{d \times d}$ is employed on each node, resulting in the nodes N_j^{i+1} with comprehensive expressive abilities. Then, the nodes N_{i+1}^j are multiplied by the adjacency matrix A . An exponential linear unit (Elu) activation function is used to accelerate the GCN learning process, which is denoted as:

$$Elu(N_j^{i+1}) = \begin{cases} N_j^{i+1}, & \text{if } N_j^{i+1} > 0 \\ \alpha(\exp(N_j^{i+1})), & \text{if } N_j^{i+1} \leq 0 \end{cases} \quad (22)$$

In the above expression, the hyperparameter α controls the points at which Elu function saturates towards negative values for negative inputs in the GCN layer. Then, the concatenation operation is utilized to concatenate the features from the outputs of the four GCN layers.

$$F^{i+5} = Elu\left(\left(\left(N^i, W^i\right), A\right)\right) \quad (23)$$

$$F^{i+5} = Cat\left(F^{i+1}, F^{i+2}, F^{i+3}, F^{i+4}\right) \quad (24)$$

where $Cat()$ denotes the concatenation operation. F_{i+1} indicates the features yielded from the i -th GCN layer. To mitigate the overfitting problem in the four-layer GCN, which may lead to a decrease in HSI classification performance, the dropout technique is utilized. The parameter p (here, we set the p value to 0.3) in the dropout layer of each GCN layer is used as a threshold value that determines which part of the features in the GCN layer is dropped. To further learn and integrate information from the HSIs, another one-layer GCN is employed after these four GCN layers. Meanwhile, an Elu nonlinearity activation function is used. Finally, to ensure comparability among the features yielded from different GCN layers, the softmax function is applied.

$$F^{i+6} = \frac{\exp(\text{Elu}(((F^{i+5}, W^{i+5}), A)))}{\sum_{k \in N_i} \text{Elu}(((F^{i+5}, W^{i+5}), A))} \tag{25}$$

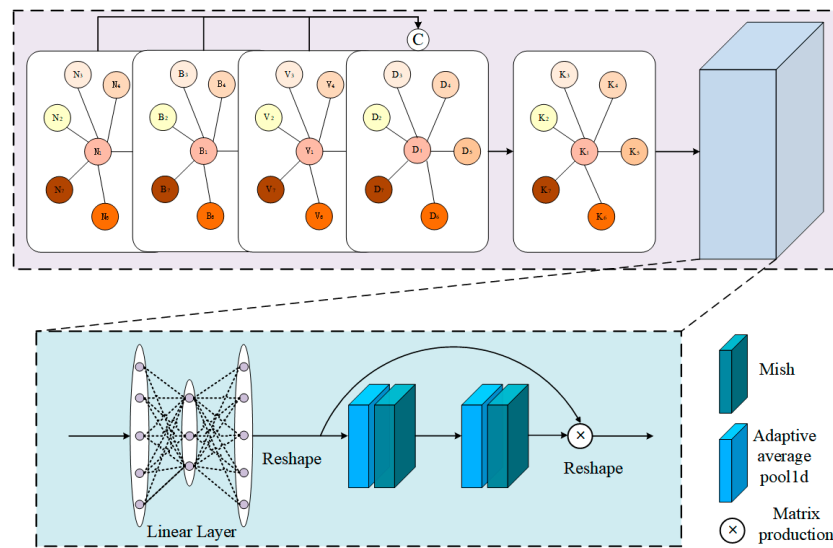


Figure 4. The detailed structure of the spectral-enhanced GCN module.

After obtaining the features extracted from the four-layer cascaded GCN and the one-layer GCN, we further enhance them through a spectral-enhanced method. Specifically, the linear layer is employed to improve the linear expressive skill of the features. Then, we reshape the features to obtain the feature matrix $M_1 \in \mathbb{R}^{c \times hw}$. In order to learn the significant spectral signature, the two cascaded adaptive average pooling layers are used. Additionally, a Mish activation function is included in the adaptive average pooling layer to avoid gradient saturation. Then, the feature matrix $M_2 \in \mathbb{R}^{c \times 1}$ is obtained. Finally, the $M_1 \in \mathbb{R}^{c \times hw}$ is multiplied with the $M_2 \in \mathbb{R}^{c \times 1}$ to obtain the feature matrix $M_{GCN} \in \mathbb{R}^{c \times hw}$ with significant information in the spectral channel. Furthermore, the important values in the spectral channel of the feature matrix are emphasized. And the detailed procedure is shown as follows:

$$M_1 = \text{reshape}(F^{i+6}) \tag{26}$$

$$M_2 = \text{Mish}(\text{AAPool}_2(\text{Mish}(\text{AAPool}_1(M_1)))) \tag{27}$$

$$M_{GCN} = M_1 \otimes M_2 \tag{28}$$

where \otimes in Equation (33) indicates the matrix product operation. M_{GCN} is the result of processing by a spatial-based GCN module, abbreviated as the GCN-based feature matrix.

In this module, the GCN layers are employed to learn the inherent features of nodes, which are different for CNN to extract. Then, the spectral-based method is employed to accentuate significant features in the spectral dimension. The significant spectral signature

can be learned and accentuated, and the HSI classification performance can be enhanced through the spectral-enhanced GCN module.

3.4. Mutual-Cooperative Attention Mechanism

After obtaining the feature matrix M_{GCN} , extracted by spectral-enhanced GCN module, and the feature matrix M_{BS} , extracted by BSNet, we construct a customized mutual-cooperative attention mechanism (MCAM) to align the spectral signature between M_{GCN} and M_{BS} . As shown in Figure 5, the MCAM mainly includes two cross multi-head self-attention mechanisms (CMSM). One improved CMSM, referred to as the M_{BS} to M_{GCN} cross multi-head self-attention block (BG-CMSB), enables the transfer of spectral signatures from M_{BS} into M_{GCN} . Vice versa, the other CMSM, referred to as the M_{GCN} to M_{BS} cross multi-head self-attention block (GB-CMSB), enables the transfer of enhanced spectral signature from M_{GCN} to M_{BS} . Subsequently, the obtained two-feature matrices are merged using an element-wise addition operation. The detailed process is shown as follows:

$$M = A(f_{BG-CMSB}(M_{BS}, M_{GCN}), f_{GB-CMSB}(M_{GCN}, M_{BS})) \tag{29}$$

where $f_{BG-CMSB}()$, $f_{GB-CMSB}()$, and $A()$ represent the BG-CMSB, GB-CMSB, and the element-wise addition, respectively. M indicates the output feature matrix of the mutual-cooperative attention mechanism.

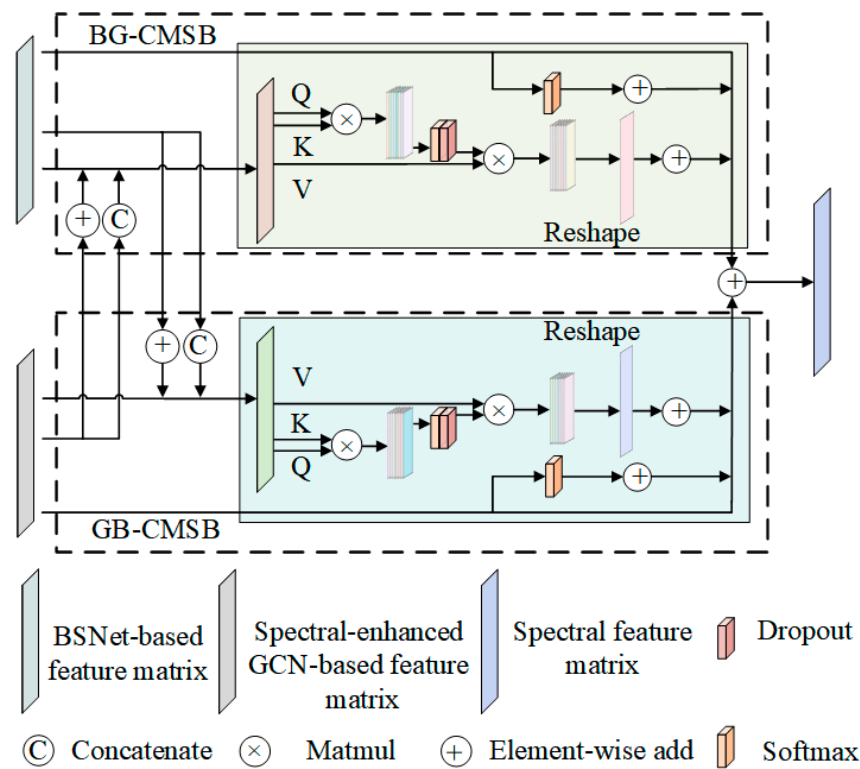


Figure 5. The structure of the mutual-cooperative attention mechanism.

3.4.1. BS-Based Feature Matrix to GCN-Based Feature Matrix Cross Multi-Head Self-Attention Block

The proposal of this block is aimed at improving the transfer and expression of spectral signatures between BS-based feature matrix M_{BS} and GCN-based feature matrix M_{GCN} . Given $M_{BS} \in \mathbb{R}^{hw \times c}$ and $M_{GCN} \in \mathbb{R}^{hw \times c}$ are inputs to BG-CMSB. The BG-CMSB initially combines the M_{BS} and M_{GCN} using the element-wise addition, resulting in the feature matrix M_{BG}^{i-1} . Then, M_{BG}^{i-1} is concatenated with M_{BS} to construct a new feature matrix. Next, the new feature matrix is separately multiplied with two different matrices, thereby linearly constructing the key $K \in \mathbb{R}^{hw \times 2c}$ and value $V \in \mathbb{R}^{hw \times 2c}$ simultaneously. At the same time,

the M_{BS} is used, being multiplied with a weight matrix to linearly construct the query $Q \in \mathbb{R}^{hw \times c}$. Additionally, the Q , K , and V are all projection matrices. Then Q , K , and the scale factor are processed by softmax, subsequently with V , to calculate the cross multi-head self-attention score from M_{BS} to M_{GCN} . Moreover, the scale factor is used to control the gradient of the model during the training process. The detailed overall operation is shown as follows:

$$CA_{BG} = Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_{dim}}}\right) \quad (30)$$

$$d_{dim} = \frac{1}{\sqrt{c}} \quad (31)$$

To obtain stable results, we execute the attention calculation process multiple times in parallel; here, it is executed eight times. Then, we reshape the CA_{BG} , and the linear operation is used to project the reshaped attention score into $M_{BG}^{i,2} \in \mathbb{R}^{hw \times c}$. Subsequently, a softmax operation is performed. To enhance the important features, the residual connection is added on the M_{BS} . Subsequently, element-wise addition is used to concatenate the M_{BS} and the processed result of M_{BS} using softmax as well as $M_{BG}^{i,2}$. The process is shown as follows:

$$M_{BG-CMSB} = M_{BS} + softmax(M_{BS}) + M_{BG}^{i,2} \quad (32)$$

where $M_{BG-CMSB}$ indicates the resulting feature maps from BG-CMSB, and $softmax()$ is the softmax function.

3.4.2. GCN-Based Feature Matrix to BS-Based Feature Matrix Cross Multi-Head Self-Attention Block

This block is devised to transfer and improve the expression of enhanced spectral signature between GCN-based feature matrix M_{GCN} and BS-based feature matrix M_{BS} . Given $M_{BS} \in \mathbb{R}^{hw \times c}$ and $M_{GCN} \in \mathbb{R}^{hw \times c}$ as inputs to GB-CMSB, the GB-CMSB first combines the M_{GCN} and M_{BS} using the element-wise addition, resulting in the feature matrix $M_{GB}^{i,1}$. Then, $M_{GB}^{i,1}$ is concatenated with M_{GCN} to linearly build the key $K \in \mathbb{R}^{hw \times 2c}$ and value $V \in \mathbb{R}^{hw \times 2c}$ simultaneously. At the same time, the M_{GCN} is employed to linearly construct the value $Q \in \mathbb{R}^{hw \times c}$. The following operations are similar to those of the BG-CMSB:

$$M_{GB}^{i,2} = reshape\left(CA_{GB}\left(M_{GB}^{i,1}, M_{GCN}\right)\right) \quad (33)$$

$$M_{GB-CMSB} = M_{GCN} + softmax(M_{GCN}) + M_{GB}^{i,2} \quad (34)$$

where CA_{GB} is the result of the CMSM score from M_{GCN} to M_{BS} , and $reshape()$ indicates the reshape operation.

Then, the obtained $M_{BG-CMSB}$ and $M_{GB-CMSB}$ are concatenated to obtain an output result for the mutual-cooperative attention mechanism. With the help of the mutual-cooperative attention mechanism, integrating comprehensive spectral features from various neural networks has a positive impact on subsequent classification tasks.

3.5. Additive Feature Fusion Based on CNN-Based Subnetworks and GCN-Based Subnetworks

Considering that the features extracted from different network architectures may play different roles in HSI classification, it is important to fuse multi-type features with a proper fusion strategy. For instance, classification performance can be enhanced when different types of features are effectively integrated. Conversely, when different types of features cannot be effectively integrated, the classification performance may be degraded.

In this article, the proposed network includes two subnetworks: one is the CNN-based subnetwork, and the other is the GCN-based subnetwork, which includes a spectral-enhanced module and a mutual-cooperative attention mechanism. Specifically, we first reshape the features extracted from CNN subnetwork, and then a linear layer is utilized to yield the final classification performance of CNN subnetwork. Concurrently, the features

extracted from the GCN subnetwork are reshaped. An adaptive average pooling layer is utilized to produce the final classification performance of the GCN subnetwork. Finally, an additive fusion strategy is employed to achieve the final classification results of PCCGC. Based on the above operations, the CNN-based features can be well integrated with the GCN-based features in an additive strategy for the HSI processing.



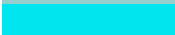






4. Experiments

4.1. Experimental Datasets

In this part of the experiments, five widely used real HSI datasets are utilized to validate the robustness and practicality of our designed model, e.g., the Pavia University dataset, WHU-Hi-Honghu dataset, Houston University dataset, and Indian Pines dataset.

(1) Pavia University (PU) dataset: The PU is collected by using the ROSIS equipment over the University of Pavia, Italy, and its surrounding areas. The spatial area of the PU dataset is 610×340 pixels, with an approximate resolution of 1.3 m per pixel. The PU dataset in our experiment includes 103 spectral bands, spanning a spectrum wavelength from 430 to 860 nm. The quantities of samples for each class utilized in the training, validation, and testing sets are exhibited in Table 1.

Table 1. The landcover classes of the PU, the color of each class, and the number of each class in the training set, validation set, and test set.

Class	Color	Total	Train	Validation	Test
C1		6631	66	66	6499
C2		18,649	186	186	18,277
C3		2099	20	20	2059
C4		3064	30	30	3004
C5		1345	13	13	1319
C6		5029	50	50	4929
C7		1330	13	13	1304
C8		3682	36	36	3610
C9		947	9	9	929
Total		42,776	423	423	41,930

(2) Houston University (Houston) dataset: The Houston dataset is gathered over the campus of University of Houston, Houston, USA. The spatial dimension size of the Houston dataset is 349×1905 pixels, with 144 spectral bands from 380 to 1050 nm, and the spatial resolution of Houston is about 2.5 m per pixel. The Houston dataset contains 15 categories, and the number of samples used for training, validation, and testing are recorded in Table 2.

Table 2. The landcover classes of the Houston, the color of each class, and the number of each class in the training set, validation set, and test set.



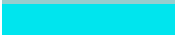


























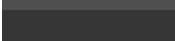

Class	Color	Total	Train	Validation	Test
C1		1251	25	25	1201
C2		1254	25	25	1204
C3		697	13	13	671
C4		1244	24	24	1196
C5		1242	24	24	1194
C6		325	6	6	313
C7		1268	25	25	1218

Table 2. *Cont.*

Class	Color	Total	Train	Validation	Test
C8		1244	24	24	1196
C9		1252	25	25	1202
C10		1227	24	24	1179
C11		1235	24	24	1187
C12		1233	24	24	1185
C13		469	9	9	451
C14		428	8	8	412
C15		660	13	13	634
Total		15,029	293	293	14,443



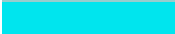










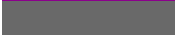


(3) WHU-Hi-Honghu (Honghu) dataset: The Honghu dataset is obtained in Hubei Province, China via imaging sensors mounted on a UAV platform. And, the Honghu dataset has a spatial size of 940×475 pixels, containing 270 spectral bands spanning from 400 to 1000 nm. In our experiments, only 16 categories are selected due to the limitations of the utilized device. The numbers of training, validation, and testing samples for each selected class, as well as the corresponding totals, are listed in Table 3.

Table 3. The landcover classes of the Honghu, the color of each class, and the number of each class in the training set, validation set, and test set.

Class	Color	Total	Train	Validation	Test
C1		3320	33	33	3254
C2		1482	14	14	1454
C3		18,725	187	187	18,351
C4		1792	17	17	1758
C5		14,939	149	149	14,641
C6		5808	58	58	5692
C7		4054	40	40	3974
C8		2375	23	23	2329
C9		939	9	9	921
C10		2584	25	25	2534
C11		3979	39	39	3901
C12		4307	43	43	4221
C13		1002	10	10	982
C14		563	5	5	553
C15		973	9	9	955
C16		2037	20	20	1997
Total		68,879	681	681	67,517

(4) Indian Pines (IP) dataset: The IP scene data is acquired using the Airborne/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines area of northwestern Indiana. The spatial scale of the IP imagery is 145×145 pixels, consisting of 220 spectral bands ranging from 400 nm to 2500 nm. The IP image includes 16 categories, and the number of labelled samples for training, validation, and testing are displayed in Table 4.

Table 4. The landcover classes of the IP, the color of each class, and the number of each class in the training set, validation set, and test set.

Class	Color	Total	Train	Validation	Test
C1		46	2	2	42
C2		1428	71	71	1286
C3		830	41	41	748
C4		237	11	11	215
C5		483	24	24	435
C6		730	36	36	658
C7		28	1	1	26
C8		478	23	23	432
C9		20	1	1	18
C10		972	48	48	876
C11		2455	122	122	2211
C12		593	29	29	535
C13		205	10	10	185
C14		1265	63	63	1139
C15		386	19	19	348
C16		93	4	4	85
Total		10,249	505	505	9239

(5) Xiongan New Area (Xiongan) dataset: The Xiongan (Matiwan Village) scene data is acquired using the Visible and Near-Infrared Imaging Spectrometer over the Xiongan Country, and Baiyangdian Lake. The spatial range of the Xiongan imagery is 3750×1580 pixels, containing 250 spectral bands ranging from 400 to 1000 nm. The Xiongan image includes 19 categories, and the number of labelled samples for training, validation, and testing are listed in Table 5.

Table 5. The landcover classes of the Xiongan, the color of each class, and the number of each class in the training set, validation set, and test set.






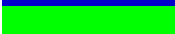









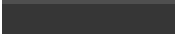
Class	Color	Total	Train	Validation	Test
C1		426,138	4261	4261	417,616
C2		187,425	1874	1874	183,677
C3		124,862	1248	1248	122,366
C4		91,518	915	915	89,688
C5		197,218	1972	1972	193,274
C6		19,663	196	196	19,271
C7		296,538	2965	2965	290,608
C8		276,755	2767	2767	271,221
C9		44,232	442	442	43,348
C10		372,708	3727	3727	365,254
C11		67,210	672	672	65,866
C12		29,763	297	297	29,169
C13		85,547	855	855	83,837
C14		68,885	688	688	67,509
C15		986,139	9861	9861	966,417
C16		7456	74	74	7308

Table 5. Cont.

Class	Color	Total	Train	Validation	Test
C17		27,178	271	271	26,636
C18		6506	65	65	6376
C19		26,140	261	261	25,618
Total		3,341,881	33,411	33,411	3,275,059

4.2. Experimental Setting

4.2.1. The Details of Experiment Implementation

To validate the superior classification performance of our devised method, fourteen different comparative methods, e.g., SSRN [59], DBDA [60], SSGCA [61], PCIA [62], MDB-Net [63], HDDA [64], DBPFA [65], ChebNet [66], GCN [67], MVAHN [68], DGFNet [38], DKDMN [69], FTINet [49], and MRCAG [41] are used for comparison with our method, and the description of each comparative method is shown in Section 4.2.2. To provide a clearer perspective on the classification results for each method, the metric of overall accuracy (OA), average accuracy (AA), and kappa (K), and each class classification accuracies are used to assess their classification performance. All experiments are performed in the same environment, namely, a mini base station equipped with 128 GB of DDR4 RAM as well as $8 \times$ NVIDIA GeForce RTX 2080Ti Graphical Processing Units, with a memory of 11 GB. The software environment used in our experiment includes CUDA Version 11.6, PyTorch 1.10.1, and python 3.8.

To keep a fair comparison environment, we standardized the parameters, optimizer, and architecture of the other fourteen comparison methods to be consistent with the experimental settings of our proposed model. During the training process of our proposed model, the parameter is updated by applying the Adam optimizer. And, the sets {0.0005, 350}, {0.0009, 150}, {0.0007, 120}, and {0.0003, 130} are elected as the learning rate and number of epochs for the proposed model in the PU, Houston, Honghu, and IP data scenes separately, which is discussed in Section 5.3. Moreover, the set {0.0007, 200} is selected as the learning rate and number of epochs for the proposed model in the Xiongan data scenes. The spatial size of 9×9 is employed for the HSI patch cube, and the batch size of 64 is chosen. Early stopping is utilized in the training process of our model. Also, the numbers for the training set, validation set, and the test set for PU, Honghu, Houston, IP, and Xiongan data scenes can be observed in Tables 1–5.

The averaged results and standard deviation of quantitative assessments, in terms of OA, AA, K , and the accuracy values for each class, and qualitative assessments for the fourteen comparative methods and our proposed method on the five HSI datasets are recorded in Tables 6–10, and Figures 6–10, respectively. And, the averaged results and standard deviations for all measurements are derived based on ten repeated experiments. Additionally, the highest values for the three indices and accuracy values for each class are bolded.

Table 9. Classification results of the IP data based on 5% training samples.

Class	SSRN	DBDA	SSGCA	PCIA	MDBNet	HDDA	DBPFA	ChebNet	GCN	MVAHN	DGFNet	FTINet	DKDMN	MRCAG	Ours
C1	0.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	97.55 ± 0.82	64.44 ± 1.79	94.03 ± 1.91	100.00 ± 0.00	100.00 ± 0.00	73.83 ± 4.12	96.87 ± 0.52	100.00 ± 0.00	100.00 ± 0.00	97.63 ± 0.03	0.00 ± 0.00	100.00 ± 0.00
C2	89.93 ± 10.42	94.41 ± 3.03	97.63 ± 1.21	98.62 ± 1.70	47.25 ± 0.74	94.85 ± 0.50	98.16 ± 0.51	71.35 ± 0.77	76.26 ± 0.90	97.61 ± 0.26	85.29 ± 1.08	65.15 ± 1.31	97.77 ± 0.44	60.57 ± 1.75	97.40 ± 0.54
C3	84.88 ± 5.88	94.66 ± 0.20	98.44 ± 0.66	97.11 ± 0.31	46.46 ± 0.81	96.20 ± 0.31	92.39 ± 0.57	67.13 ± 1.10	46.93 ± 4.01	96.63 ± 0.18	89.36 ± 1.93	52.57 ± 1.46	96.50 ± 0.44	61.54 ± 3.65	98.99 ± 0.42
C4	60.00 ± 48.98	98.42 ± 1.36	94.93 ± 2.58	99.10 ± 0.33	56.10 ± 1.98	95.06 ± 1.20	99.03 ± 2.29	63.51 ± 4.08	57.63 ± 4.48	99.48 ± 0.14	85.04 ± 4.29	41.21 ± 3.51	100.00 ± 0.00	59.34 ± 3.46	92.98 ± 0.28
C5	71.86 ± 21.30	96.05 ± 1.03	93.77 ± 1.88	94.40 ± 0.64	63.99 ± 0.97	95.13 ± 0.69	97.62 ± 0.50	94.35 ± 0.61	74.96 ± 1.87	99.74 ± 0.00	96.64 ± 0.20	79.13 ± 2.32	99.02 ± 0.66	83.44 ± 0.88	97.04 ± 0.55
C6	80.70 ± 4.30	99.18 ± 0.12	97.74 ± 0.20	96.80 ± 0.15	76.59 ± 0.45	99.13 ± 0.54	96.68 ± 0.69	77.77 ± 0.48	85.92 ± 1.29	98.59 ± 0.07	98.14 ± 0.21	67.46 ± 0.99	99.95 ± 0.07	75.49 ± 2.63	98.31 ± 0.60
C7	0.00 ± 0.00	72.61 ± 9.26	62.86 ± 22.03	69.96 ± 35.02	50.92 ± 7.44	100.00 ± 0.00	90.00 ± 30.00	0.00 ± 0.00	90.92 ± 14.46	63.15 ± 1.59	81.23 ± 13.36	80.00 ± 28.28	43.58 ± 0.35	0.00 ± 0.00	98.89 ± 2.22
C8	96.65 ± 4.01	93.57 ± 0.60	99.36 ± 1.41	99.84 ± 0.41	78.48 ± 1.56	99.04 ± 0.32	96.09 ± 0.33	89.18 ± 0.35	89.13 ± 0.96	96.68 ± 1.27	97.60 ± 1.36	89.09 ± 0.60	99.85 ± 0.11	88.17 ± 0.82	100.00 ± 0.00
C9	0.00 ± 0.00	100.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	24.72 ± 4.38	96.32 ± 2.41	0.00 ± 0.00	0.00 ± 0.00	53.33 ± 6.67	100.00 ± 0.00	92.31 ± 10.88	16.67 ± 23.57	86.63 ± 0.73	0.00 ± 0.00	73.85 ± 2.31
C10	63.66 ± 4.72	94.64 ± 1.36	93.50 ± 1.49	95.07 ± 0.44	44.05 ± 1.09	93.06 ± 0.59	92.37 ± 4.97	85.09 ± 0.60	73.67 ± 1.22	96.33 ± 0.60	94.22 ± 1.70	72.67 ± 0.27	95.75 ± 0.10	53.86 ± 2.15	96.79 ± 0.36
C11	93.00 ± 2.71	96.20 ± 0.25	98.50 ± 0.89	95.48 ± 0.24	60.26 ± 0.41	96.20 ± 0.29	95.70 ± 0.47	68.66 ± 0.68	77.30 ± 3.00	97.84 ± 0.13	90.66 ± 1.56	67.85 ± 0.57	98.12 ± 0.30	71.57 ± 0.68	98.97 ± 0.42
C12	71.53 ± 12.24	96.95 ± 0.18	93.14 ± 6.70	96.91 ± 0.15	40.18 ± 1.31	95.00 ± 1.08	96.45 ± 0.86	47.91 ± 0.49	51.25 ± 1.98	98.03 ± 0.65	84.41 ± 0.83	48.40 ± 2.91	92.82 ± 0.68	54.63 ± 0.49	95.85 ± 0.72
C13	98.28 ± 2.57	99.39 ± 0.17	93.98 ± 3.18	98.98 ± 0.16	66.73 ± 0.66	98.35 ± 0.28	100.00 ± 0.00	93.07 ± 0.43	73.91 ± 2.10	98.35 ± 0.28	99.61 ± 0.27	89.17 ± 1.46	100.00 ± 0.00	97.57 ± 1.23	100.00 ± 0.00
C14	96.40 ± 3.05	97.92 ± 0.23	99.75 ± 0.10	97.00 ± 0.14	77.88 ± 1.29	96.53 ± 0.21	97.85 ± 0.67	86.36 ± 0.69	87.83 ± 3.69	97.18 ± 0.24	97.05 ± 0.48	94.41 ± 0.33	96.35 ± 0.70	87.82 ± 2.09	97.78 ± 0.43
C15	94.16 ± 7.94	96.39 ± 0.58	90.55 ± 1.31	96.08 ± 0.26	70.61 ± 1.03	93.91 ± 0.50	94.50 ± 1.17	82.25 ± 0.75	70.41 ± 3.42	98.10 ± 0.26	94.20 ± 2.01	70.71 ± 0.94	93.89 ± 0.23	83.21 ± 1.36	98.70 ± 0.13
C16	87.48 ± 9.94	97.51 ± 0.03	98.74 ± 0.04	98.79 ± 0.00	59.59 ± 3.10	94.22 ± 2.83	98.73 ± 0.04	96.63 ± 1.15	53.61 ± 3.20	98.45 ± 0.52	97.54 ± 0.98	100.00 ± 0.00	96.18 ± 0.51	91.96 ± 1.35	96.03 ± 0.68
OA	84.12 ± 3.52	96.03 ± 0.65	96.86 ± 0.97	96.74 ± 0.22	60.16 ± 0.11	95.87 ± 0.14	96.02 ± 0.87	75.38 ± 0.13	74.00 ± 1.78	97.48 ± 0.12	91.80 ± 0.50	70.68 ± 0.38	96.94 ± 0.11	70.88 ± 0.89	97.92 ± 0.12
AA	68.03 ± 4.54	95.49 ± 0.22	88.31 ± 2.00	89.48 ± 2.26	58.02 ± 0.43	96.06 ± 0.33	90.35 ± 2.47	70.20 ± 0.34	71.06 ± 1.26	95.81 ± 0.25	92.71 ± 0.36	70.91 ± 2.66	93.38 ± 0.10	60.58 ± 0.57	96.35 ± 0.13
K × 100	81.82 ± 4.04	95.47 ± 0.74	96.42 ± 1.11	96.27 ± 0.25	54.13 ± 0.14	95.29 ± 0.16	95.45 ± 0.99	71.49 ± 0.16	70.29 ± 2.10	97.12 ± 0.14	90.62 ± 0.58	66.15 ± 0.44	96.51 ± 0.12	66.49 ± 1.05	97.63 ± 0.13

Table 10. Classification results of the Xiong'an data based on 1% training samples.

Class	SSRN	DBDA	SSGCA	PCIA	MDBNet	HDDA	DBPFA	ChebNet	GCN	MVAHN	DGFNet	FTINet	DKDMN	MRCAG	Ours
C1	100.00 ± 0.00	99.78 ± 0.19	99.96 ± 0.03	99.99 ± 0.00	99.98 ± 0.01	99.97 ± 0.03	99.74 ± 0.20	98.54 ± 1.08	99.63 ± 0.12	99.40 ± 0.19	99.95 ± 0.02	99.83 ± 0.14	99.99 ± 0.00	99.83 ± 0.04	99.99 ± 0.00
C2	98.23 ± 2.46	99.59 ± 0.10	99.90 ± 0.08	99.86 ± 0.07	99.83 ± 0.09	99.94 ± 0.03	99.32 ± 0.31	96.38 ± 1.61	98.77 ± 0.44	99.75 ± 0.01	99.70 ± 0.09	99.53 ± 0.09	99.95 ± 0.01	99.56 ± 0.07	99.85 ± 0.12
C3	0.00 ± 0.00	99.81 ± 0.06	97.85 ± 0.54	99.62 ± 0.03	99.73 ± 0.03	99.40 ± 0.59	98.64 ± 0.37	93.36 ± 2.38	98.79 ± 0.42	99.84 ± 0.02	99.34 ± 0.22	99.05 ± 0.36	99.91 ± 0.03	99.49 ± 0.10	99.63 ± 0.03
C4	0.00 ± 0.00	90.32 ± 1.66	90.51 ± 4.25	91.70 ± 0.73	94.56 ± 0.29	84.12 ± 15.56	82.52 ± 6.36	69.78 ± 2.19	86.06 ± 0.71	93.58 ± 1.52	96.09 ± 0.26	90.54 ± 2.97	94.86 ± 0.96	91.83 ± 1.77	96.91 ± 0.18
C5	32.23 ± 9.18	92.43 ± 5.12	88.48 ± 3.61	93.13 ± 0.23	98.12 ± 0.49	94.51 ± 6.08	92.22 ± 2.16	75.28 ± 7.74	91.31 ± 0.75	98.30 ± 0.32	97.93 ± 0.49	93.76 ± 0.40	98.15 ± 0.32	96.03 ± 0.57	98.46 ± 0.22
C6	0.00 ± 0.00	88.32 ± 1.40	98.55 ± 0.30	90.20 ± 0.87	94.88 ± 0.45	96.32 ± 0.69	97.04 ± 1.44	63.51 ± 16.50	75.40 ± 4.80	93.89 ± 0.96	96.48 ± 0.51	87.09 ± 5.12	96.37 ± 1.18	91.04 ± 1.37	96.19 ± 1.09
C7	80.50 ± 6.06	92.18 ± 1.25	84.66 ± 7.52	88.95 ± 2.25	95.23 ± 0.86	90.00 ± 8.54	82.93 ± 1.88	69.82 ± 3.18	86.47 ± 0.90	95.34 ± 0.80	95.16 ± 2.06	90.43 ± 0.31	94.98 ± 0.29	92.43 ± 1.18	96.56 ± 0.29
C8	51.68 ± 40.89	88.89 ± 1.98	93.28 ± 4.18	84.64 ± 0.88	93.49 ± 0.56	93.20 ± 3.32	78.54 ± 3.23	70.64 ± 4.37	83.63 ± 1.34	93.60 ± 0.67	95.79 ± 0.46	90.15 ± 1.37	92.33 ± 0.74	91.14 ± 0.54	94.55 ± 0.21
C9	0.00 ± 0.00	94.33 ± 2.15	86.48 ± 3.62	95.76 ± 0.41	90.91 ± 0.35	96.31 ± 2.77	99.56 ± 0.18	85.51 ± 3.69	91.52 ± 0.48	95.14 ± 0.27	88.31 ± 0.59	88.79 ± 1.25	93.08 ± 0.77	87.02 ± 3.78	95.90 ± 0.58
C10	63.74 ± 14.91	90.53 ± 3.95	97.67 ± 1.24	89.41 ± 1.68	96.34 ± 0.19	94.98 ± 2.19	93.51 ± 1.49	63.64 ± 5.56	88.00 ± 0.60	96.67 ± 0.46	97.93 ± 0.27	92.58 ± 1.31	95.81 ± 0.79	94.37 ± 0.35	98.19 ± 0.34
C11	0.00 ± 0.00	97.30 ± 0.92	96.21 ± 1.36	96.78 ± 0.20	96.33 ± 0.73	96.99 ± 0.78	98.00 ± 1.10	83.84 ± 9.08	93.78 ± 1.40	96.71 ± 0.26	98.00 ± 0.79	91.32 ± 2.73	97.49 ± 0.25	94.13 ± 0.54	98.10 ± 0.71
C12	0.00 ± 0.00	80.38 ± 1.64	58.62 ± 7.80	87.68 ± 2.58	82.00 ± 2.42	80.46 ± 10.32	78.03 ± 4.57	39.88 ± 15.45	79.10 ± 2.46	85.88 ± 1.32	88.53 ± 1.33	71.02 ± 2.00	84.82 ± 0.63	77.96 ± 1.56	93.03 ± 1.00
C13	64.70 ± 45.79	80.04 ± 5.37	73.53 ± 1.02	84.17 ± 1.86	86.61 ± 1.58	76.37 ± 16.98	89.34 ± 1.49	76.51 ± 1.55	79.02 ± 1.02	83.06 ± 0.58	89.18 ± 0.12	78.95 ± 1.65	83.35 ± 0.35	81.30 ± 1.87	88.43 ± 1.09
C14	66.42 ± 46.97	85.55 ± 2.96	66.59 ± 3.75	82.75 ± 2.29	90.89 ± 0.63	90.00 ± 4.76	91.32 ± 4.87	63.62 ± 1.93	72.84 ± 2.67	92.93 ± 0.47	92.14 ± 0.89	86.26 ± 0.33	89.46 ± 0.99	85.11 ± 1.61	94.24 ± 1.09
C15	58.84 ± 3.72	93.69 ± 0.95	84.31 ± 5.34	92.99 ± 1.31	95.74 ± 0.55	94.86 ± 3.08	87.22 ± 2.55	74.14 ± 3.30	88.88 ± 0.40	97.67 ± 0.27	93.26 ± 1.13	93.15 ± 0.08	96.01 ± 0.20	94.51 ± 0.37	97.63 ± 0.54
C16	0.00 ± 0.00	88.50 ± 1.46	78.74 ± 7.44	95.92 ± 1.45	84.86 ± 1.69	88.46 ± 20.3	95.69 ± 0.58	74.06 ± 8.96	71.91 ± 3.61	84.28 ± 1.52	88.12 ± 2.40	77.68 ± 2.72	93.04 ± 0.59	85.89 ± 1.22	88.42 ± 0.83
C17	0.00 ± 0.00	93.56 ± 1.38	84.63 ± 9.40	95.75 ± 1.84	96.16 ± 0.45	98.27 ± 0.31	98.47 ± 0.56	0.00 ± 0.00	86.94 ± 4.19	98.03 ± 0.14	94.76 ± 3.07	96.41 ± 0.31	96.41 ± 0.34	98.82 ± 0.23	98.82 ± 0.23
C18	0.00 ± 0.00	76.80 ± 10.70	33.16 ± 46.90	95.55 ± 2.38	85.99 ± 3.50	73.69 ± 22.98	73.94 ± 9.50	0.00 ± 0.00	71.11 ± 3.12	89.63 ± 1.04	93.83 ± 0.34	74.54 ± 1.31	90.70 ± 1.33	86.79 ± 2.12	92.54 ± 1.17
C19	100.00 ± 0.00	98.88 ± 0.58	99.27 ± 0.47	98.53 ± 0.18	98.83 ± 0.32	99.05 ± 0.69	96.02 ± 2.76	0.00 ± 0.00	96.92 ± 1.04	98.93 ± 0.28	99.86 ± 0.03	96.57 ± 2.24	98.67 ± 0.09	97.88 ± 0.83	98.78 ± 0.47
OA	62.40 ± 2.47	93.38 ± 1.47	89.00 ± 2.59	92.65 ± 0.39	96.09 ± 0.06	94.04 ± 4.22	89.81 ± 1.19	77.14 ± 2.75	89.90 ± 0.29	96.66 ± 0.11	95.89 ± 0.54	93.14 ± 0.39	96.02 ± 0.11	94.39 ± 0.20	97.49 ± 0.15
AA	37.70 ± 5.49	91.10 ± 1.56	84.86 ± 3.50	92.81 ± 0.24	93.71 ± 0.43	91.94 ± 4.87	91.16 ± 1.02	63.08 ± 3.14	86.32 ± 0.91	94.36 ± 0.18	95.14 ± 0.08	89.26 ± 1.19	94.49 ± 0.09	91.72 ± 0.07	96.12 ± 0.24
K × 100	54.16 ± 2.56	92.28 ± 1.73	87.04 ± 3.13	91.42 ± 0.47	95.44 ± 0.08	93.07 ± 4.90	88.02 ± 1.43	72.89 ± 3.39	88.20 ± 0.34	96.11 ± 0.13	95.19 ± 0.64	92.00 ± 0.45	95.37 ± 0.12	93.47 ± 0.24	97.07 ± 0.18

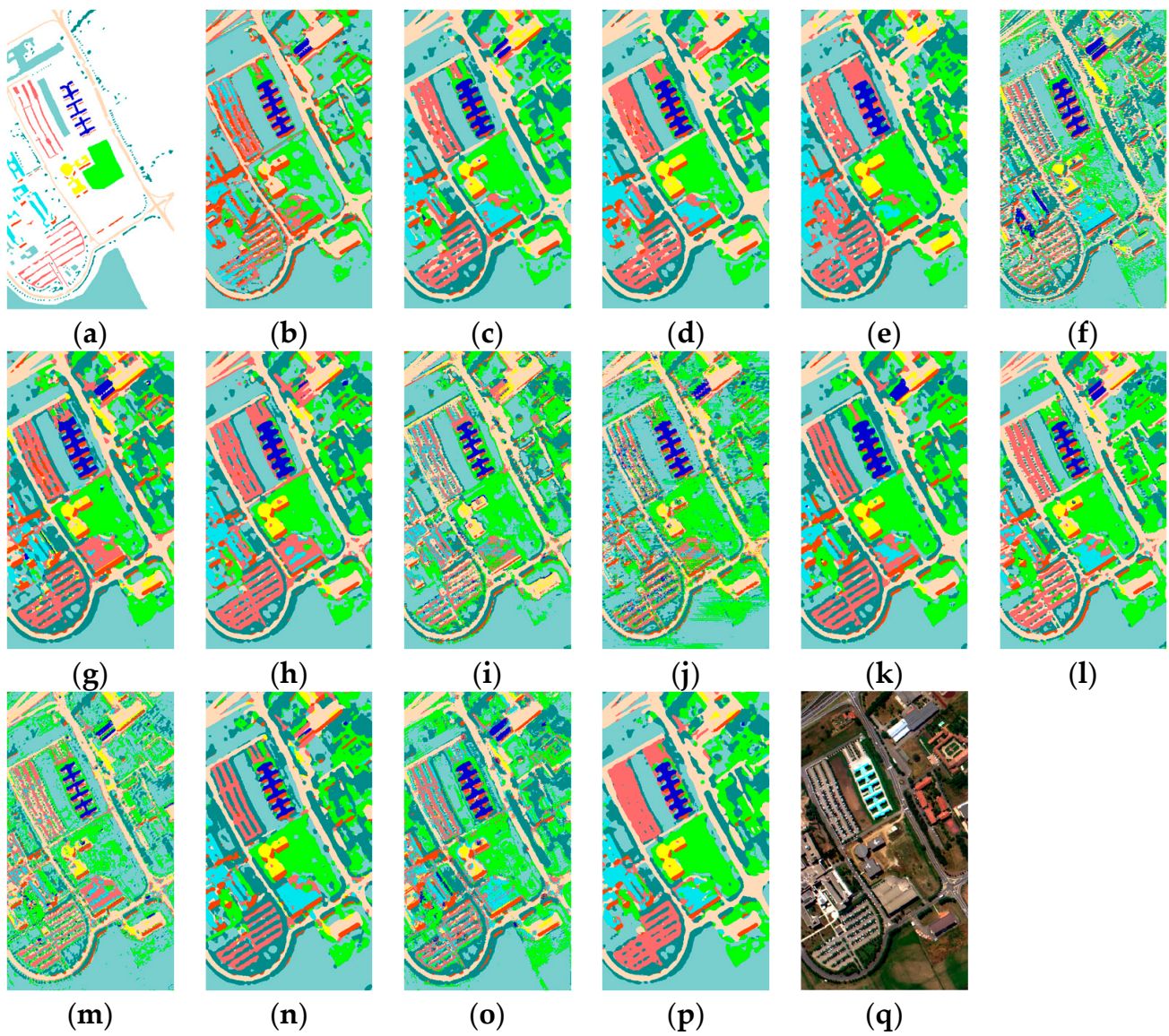


Figure 6. Full-pixel classification maps for the PU data scene. (a) Ground-truth; (b) SSRN; (c) DBDA; (d) SSGCA; (e) PCIA; (f) MDBNet; (g) HDDA; (h) DBPFA; (i) ChebNet; (j) GCN; (k) MVAHN; (l) DGFNet; (m) FTINet; (n) DKDMN; (o) MRCAG; (p) Ours; (q) False-color image.

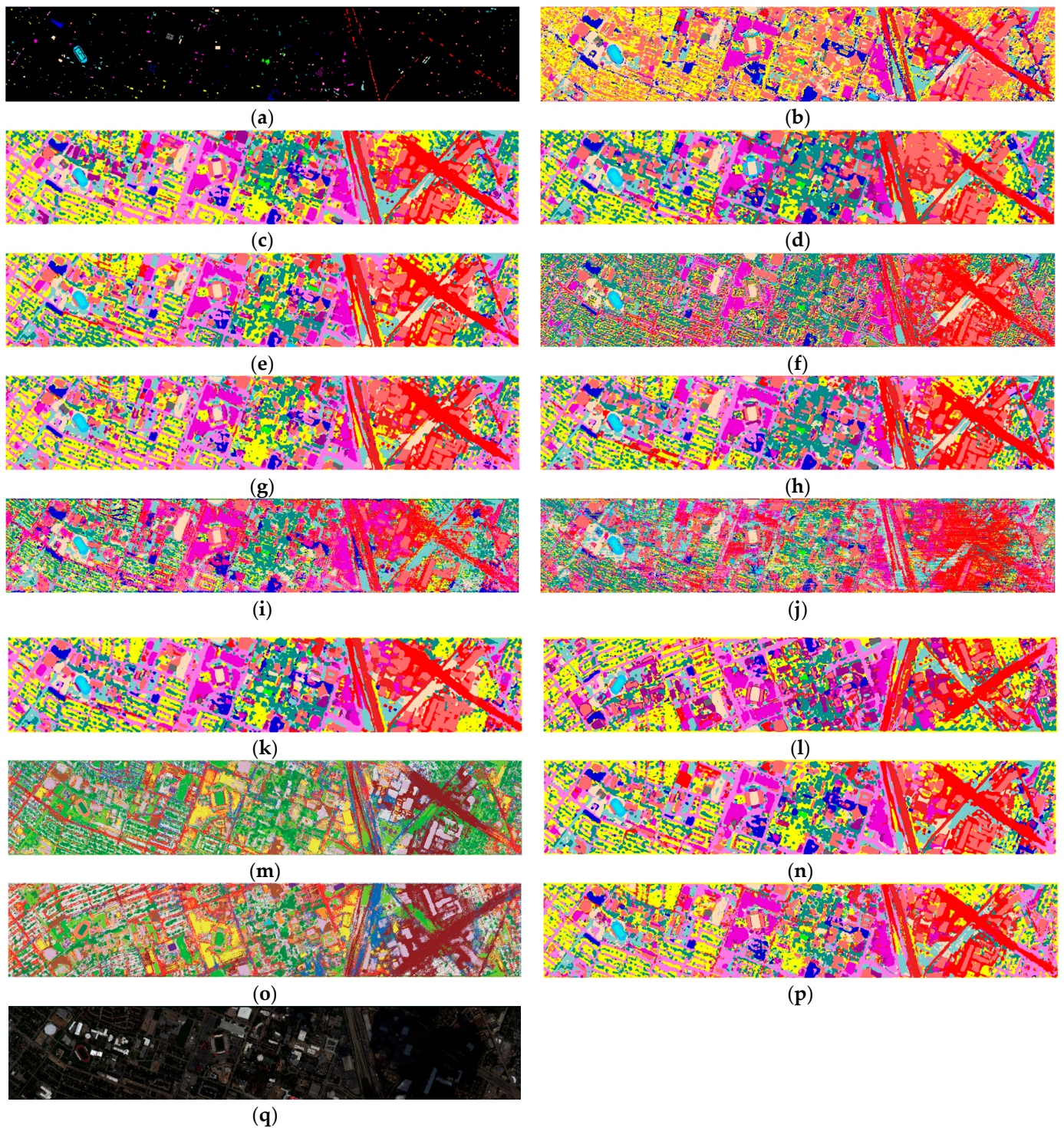


Figure 7. Full-pixel classification maps for the Houston data scene. (a) Ground-truth; (b) SSRN; (c) DBDA; (d) SSGCA; (e) PCIA; (f) MDBNet; (g) HDDA; (h) DBPFA; (i) ChebNet; (j) GCN; (k) MVAHN; (l) DGFNet; (m) FTINet; (n) DKDMN; (o) MRCAG; (p) Ours; (q) False-color image.

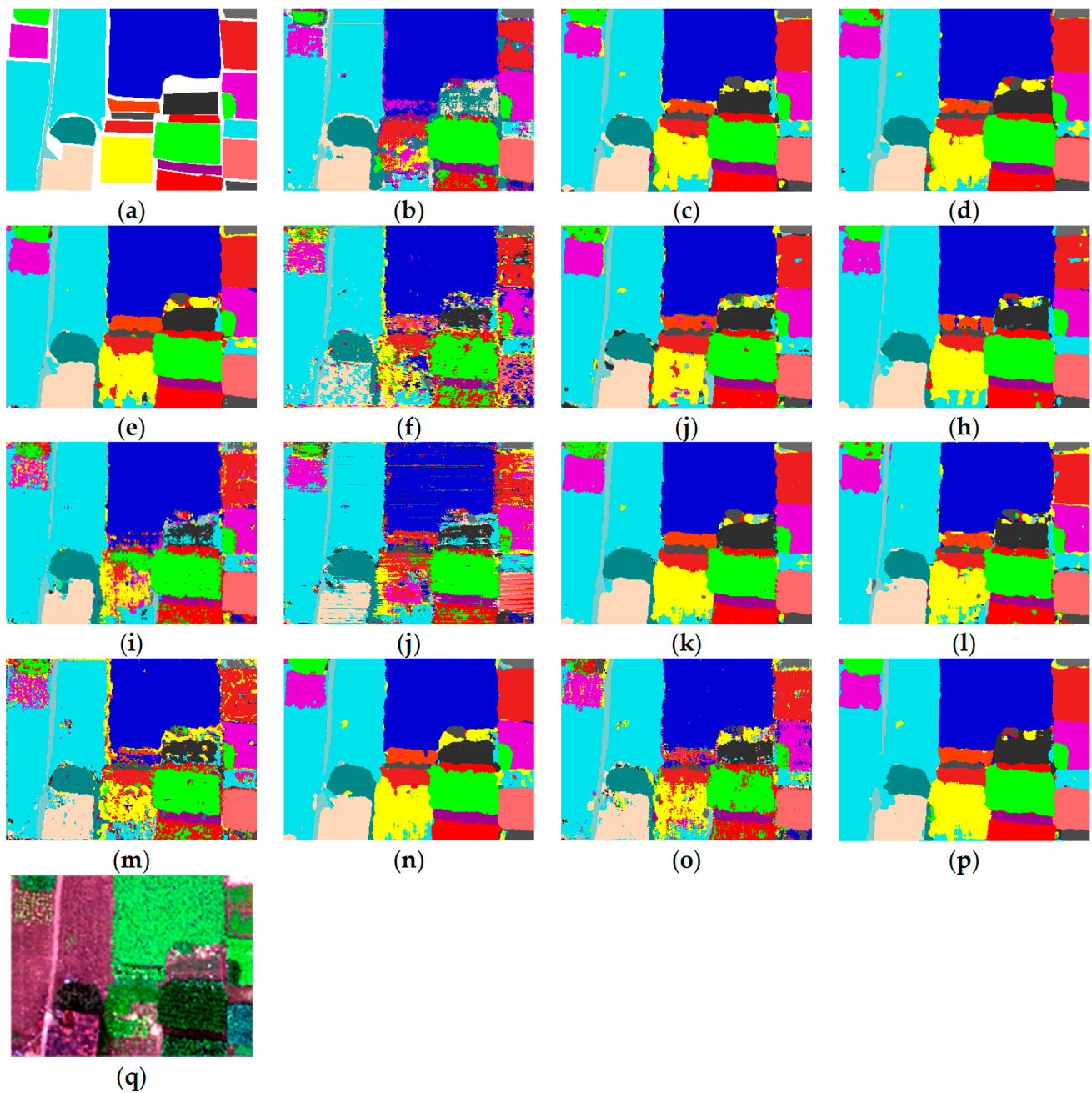


Figure 8. Full-pixel classification maps for the Houghu data scene. (a) Ground-truth; (b) SSRN; (c) DBDA; (d) SSGCA; (e) PCIA; (f) MDBNet; (g) HDDA; (h) DBPFA; (i) ChebNet; (j) GCN; (k) MVAHN; (l) DGFNet; (m) FTINet; (n) DKDMN; (o) MRCAG; (p) Ours; (q) False-color image.

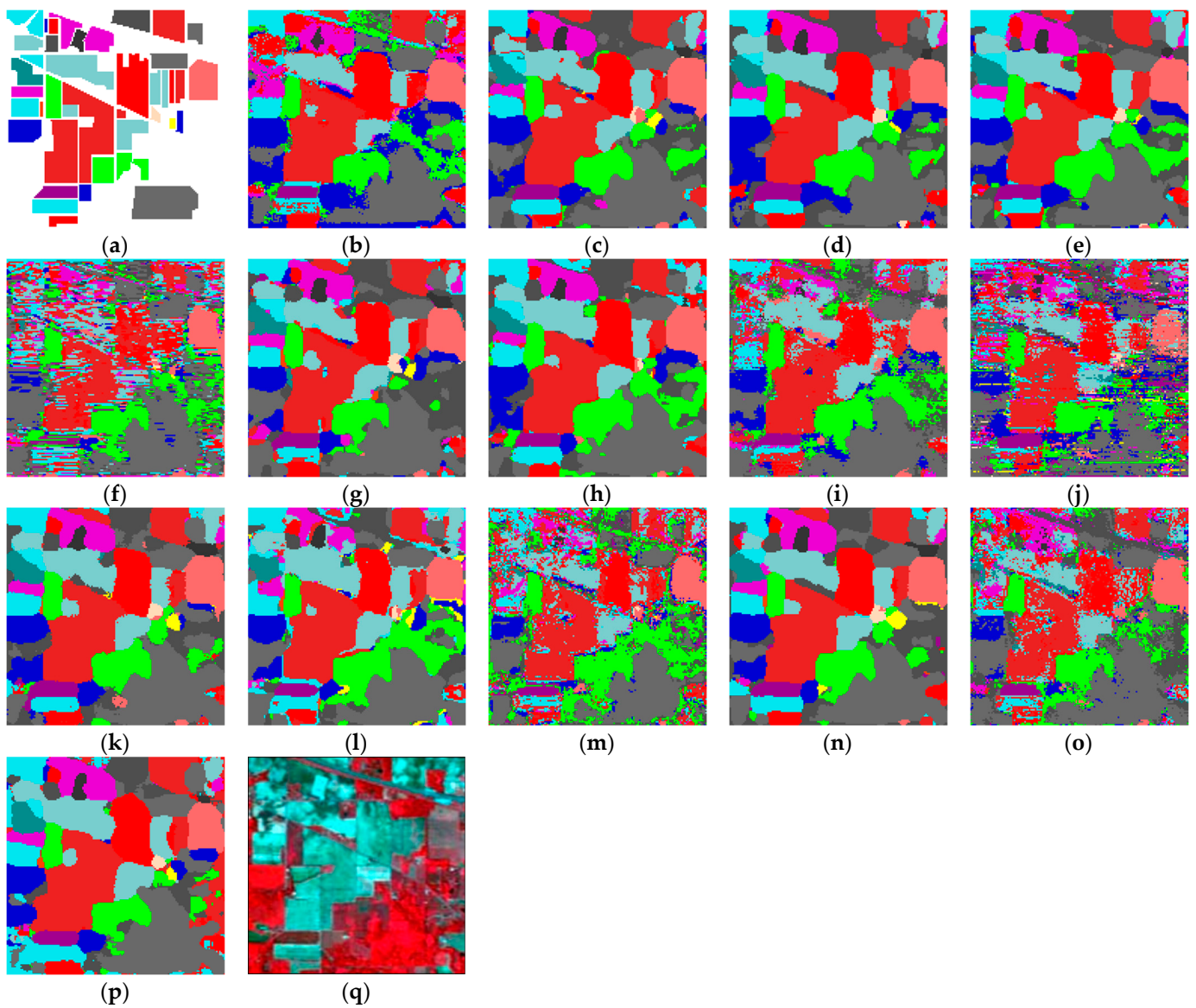


Figure 9. Full-pixel classification maps for the IP data scene. (a) Ground-truth; (b) SSRN; (c) DBDA; (d) SSGCA; (e) PCIA; (f) MDBNet; (g) HDDA; (h) DBPFA; (i) ChebNet; (j) GCN; (k) MVAHN; (l) DGFNet; (m) FTINet; (n) DKDMN; (o) MRCAG; (p) Ours; (q) False-color image.

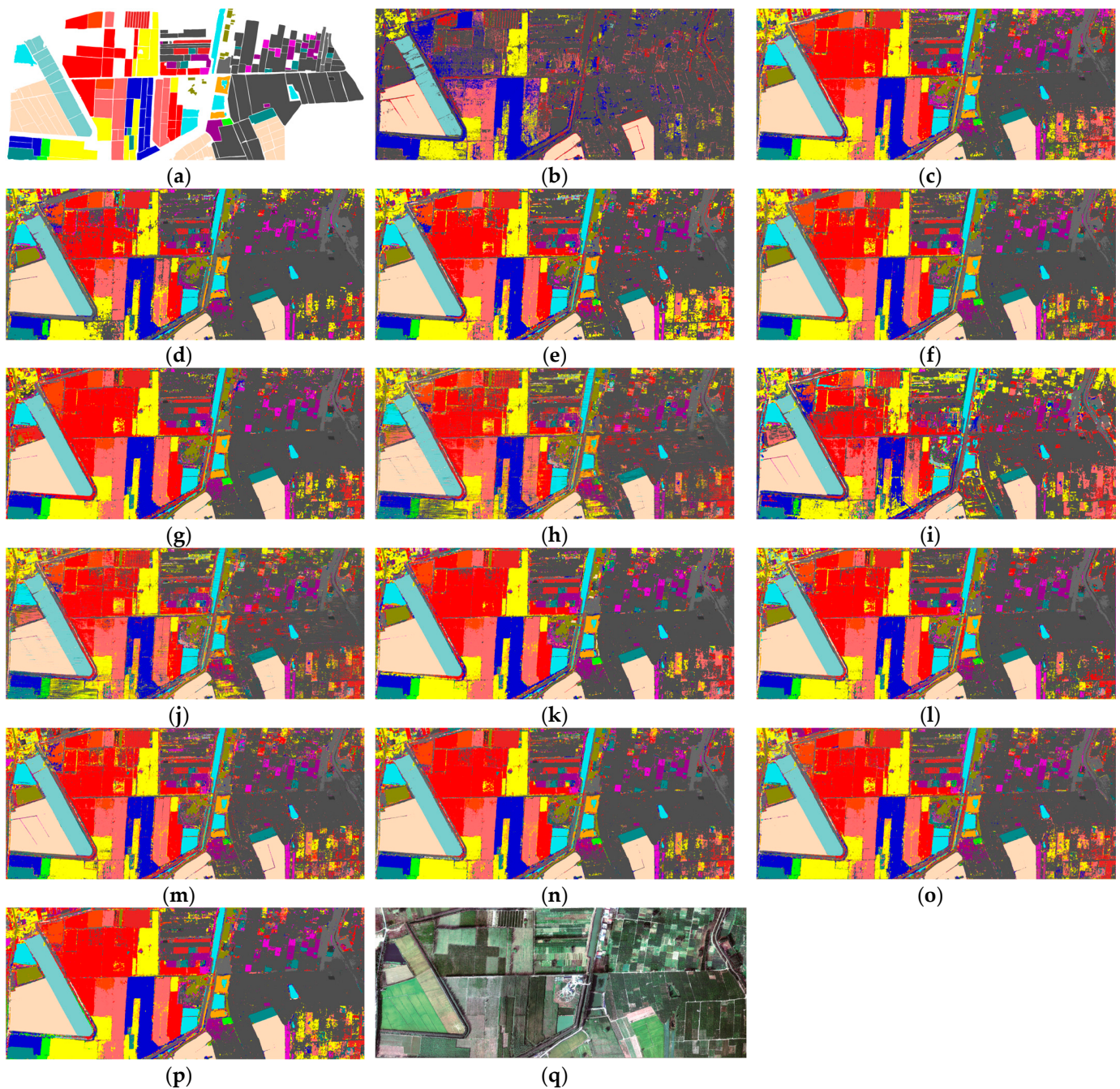


Figure 10. Full-pixel classification maps for the IP data scene. (a) Ground-truth; (b) SSRN; (c) DBDA; (d) SSGCA; (e) PCIA; (f) MDBNet; (g) HDDA; (h) DBPFA; (i) ChebNet; (j) GCN; (k) MVAHN; (l) DGFNet; (m) FTINet; (n) DKDMN; (o) MRCAG; (p) Ours; (q) False-color image.

4.2.2. The Fourteen State-of-the-Art Comparison Methods

(1) SSRN: The SSRN adopts spectral and spatial residual modules as its backbone and combines them in a consecutive manner to address the accuracy decreasing problem. It first extracts spectral signatures and then extracts the spatial features for pixel-wise HSI classification. Additionally, batch normalization is used in each 3D convolutional layer to regulate the feature extraction process.

(2) DBDA: The DBDA has spectral and spatial branches, with dense spectral block and channel attention mechanisms included in the spectral branch for extracting and refining spectral features, and a spatial attention block and a dense spatial block included in the

spatial branch for learning and optimizing spatial information. Then, a concatenation operation is utilized to fuse spectral and spatial information.

(3) SSGCA: The SSGCA first uses a spectral–spatial module to excavate spectral and spatial features separately. Then, a channel global context attention mechanism is developed to enhance the significance of the extracted spectral signatures, and a position global context attention mechanism is devised to enhance the importance of the extracted spatial features.

(4) PCIA: The PCIA is a dual-branch model, which first uses spectral and spatial pyramidal blocks to efficiently learn spectral and spatial information. Then, a novel iterative attention, namely, a new expectation-maximization attention, is employed to refine the learned spectral and spatial information. Finally, the refined spectral-spatial information is conveyed to the fully connected layer for the final classification outcomes.

(5) MDBNet: The MDBNet uses PCA to operate on the original dataset, yielding the processed dataset. The processed dataset is then processed by the multiscale spectral–spatial feature extraction module to extract the multiscale spectral and spatial features. Then, a dual-branch information fusion block consisting of residual connections and dense connections is used to learn discriminant features. Finally, a new shuffle attention is proposed to adaptively weigh the spectral and spatial features, resulting in improved classification accuracy.

(6) HDDA: The HDDA architecture features a novel hybrid dense module and dual attention mechanisms. It utilizes a stacked autoencoder to decrease the number of channels in the HSI. And then, a hybrid 2D-3D CNN module is employed to extract the spectral and spatial information. The channel and spatial attention mechanism is designed to refine the extracted spectral and spatial features separately. Additionally, a dropout layer and batch normalization are employed separately to mitigate overfitting and enhance computational efficiency.

(7) DBPFA: The DBPFA mainly consists of dual-branches and an improved attention mechanism for HSI classification. It includes a spectral feature extraction branch for extracting spectral signatures and a spatial feature excavating branch for extracting spatial features. It also includes a polarized fully attention for learning context feature information.

(8) ChebNet: The Spectral GCN is a graph convolutional operation with fast local convolutional filters, where the filter is approximated by a K -order Chebyshev polynomial. And, it is applicable to any graph structure. In this article, a filter with a 1-order Chebyshev polynomial is used.

(9) GCN: The GCN is an efficient tool based on convolutional neural networks and constructed by an approximation of localized first-order spectral graph convolutions, which can be directly operate on graphs. The GCN is also a linear layer model that can learn the representation of both graph edges relations and node information. In this article, a five-layer GCN is employed.

(10) MVAHN: The MVAHN is a new hybrid vision architecture-based model, which first utilizes CNN to extract the spectral signatures and spatial features from HSIs. Next, the generated features are divided into two components; one is for the GCN module, and another is for the transformer module. Finally, a residual learning block is used to fuse the extracted features.

(11) MRCAG: The MRCAG model mainly has three components: a developed multi-scale random shape convolution part for learning convolution-based multiscale features, where the convolution kernels used are randomized. A designed adaptive graph convolution part for learning graph-based features, where the weights for neighborhood nodes are learned adaptively. A local feature processing part is designed to exploit CNN-based features and GCN-based features, enhancing the feature representation.

(12) FTINet: The FTINet method consists of three stages: First, multiple stacked Ciformers are used to learn the dynamic and static spatial contextual information of the data. Then, concatenated FTCUs are employed to learn the spectral and topological features of the processed data. Meanwhile, the edges of the graph are learned for information

aggregation and propagation. Finally, the learned features and information are fed into the classification stage to produce the classification results.

(13) DKDMN: The DKDMN method is a hybrid neural network architecture. It first employs the proposed multi-scale spectral signature extraction module for spectral signature extraction. Then, the extracted multiscale spectral signature is combined with positional embedding for Transformer preprocessing. Then, the signatures are fed into the designed module for comprehensive spectral signature learning. This module is composed of multiple CNN-Transformer blocks and a residual GCN. To better achieve the final classification results, the learned spectral signature is combined with the features extracted by the diffusion model.

(14) DGFNet: The DGFNet is a dual-branch GNN fusion network, which includes a spatial-based branch and a spectral-based branch. It takes HSI subcubes as input data. The spatial branch first employs a Graph Attention Network to learn the intrinsic relationships within the input data. Then, it develops a local guidance module to learn significant features. The spectral branch employs weights for different spectral bands to obtain spectral features. Finally, a linear layer is used to fuse the spatial features and spectral features.

4.3. Experimental Results

In this section, for the five widely used data scenes, the training set, validation set, and test set are composed of randomly selected labeled samples from each category on the five HSI data scenes, which are recorded in Tables 1–5 separately. Tables 6–10 show the accuracy of OA, AA, K , and each class of the fourteen competitive methods as well as our proposed model on the five widely used data scenes. Figures 6–10 show the ground truth map, full-pixel classification maps, and False-color images on the PU, Houston, Honghu, and IP data scenes. The detailed discussion of classification results on the five used data scenes is shown as follows:

(1) Classification results on the PU dataset: In the PU dataset, as recorded in Table 6, in terms of the value of OA, our proposed method surpasses SSRN, DBDA, SSGCA, PCIA, MDBNet, HDDA, DBPFA, ChebNet, GCN, and MVAHN by about 13.33%, 2.76%, 0.76%, 1.30%, 15.63%, 2.89%, 1.38%, 10.93%, 9.88%, and 0.50%, respectively, demonstrating the superior classification performance of our method. The methods based on CNN, namely SSRN, DBDA, SSGCA, PCIA, MDBNet, HDDA, and DBPFA, all achieved excellent classification performance in terms of OA, AA, K , and accuracy values for each class, except for SSRN and MDBNet. This can be attributed to CNN being one of the powerful data fitting tools of deep learning. The GCN-based methods, namely ChebNet and GCN, exhibit lower classification performance, except for MVAHN. This can be ascribed to the fact that GCN-based methods only consider features from a single perspective. To some extent, the pure GCN-based method has limited feature extraction performance compared to the spectral and spatial CNN-based method. The hybrid vision architectures, namely our proposed method and MVAHN, which are based on CNN and GCN architecture, both achieve outstanding classification performance on the PU dataset. Specifically, our proposed method achieve the highest OA, AA, and K values among the other fourteen comparative methods on the PU dataset. Furthermore, from Figure 6, the classification map yielded by our proposed method is not only clearer than those from the other fourteen comparative experiments, but also has smoother land cover edges. Conversely, the classification maps yielded by MDBNet and GCN show a lot of salt-and-pepper scenarios.

(2) Classification results on the Houston dataset: For the Houston dataset, our proposed method achieves the highest records in terms of OA, AA, and K compared to the fourteen comparative methods, as documented in Table 7. The OA value of our proposed method is approximately 0.47% higher than that of MVAHA, about 0.33% higher than SSGCA, a CNN-based method with higher OA among CNN-based methods, and about 18.53% higher than ChebNet, a GCN-based method with higher OA among GCN-based methods.

This is due to the fact that our proposed method is designed by combining the CNN and GCN networks, which can excavate the multiscale spectral-spatial features and learn

the pixel-wise spectral signature among the graphs. From Figure 7, it can be observed that our proposed method exhibits a classification map that closely resembles the ground-truth feature map in comparison to the other fourteen classification methods.

(3) Classification results on the Honghu dataset: On the Honghu dataset, as seen in Figure 8a, the terrain regions of same land covers are more concentrated, so this dataset is more conducive to being distinguished. From Table 8, our proposed method shown higher values of OA, AA, and K compared with the fourteen comparative methods. The MVAHN achieve the second classification performance. The reason for this can be attributed to the method combining the CNN and GCN, which enables the extraction of different types of features. The CNN-based methods, namely PCIA and DBPFA, all get good classification performance, but worse than hybrid CNN and GCN methods. It is also shown that the features extracted by CNN-based methods are less expressive than the hybrid CNN and GCN methods. Furthermore, the classification performance of categories of C1, C4, C5, C12, and C15 that are obtained by our proposed method is higher than the other fourteen comparative methods, which is shown the in better feature extraction ability of our proposed method. At the same time, from Figure 8, the classification map obtained by our proposed method has less salt-and-pepper noise and more similarity to the ground truth map.

(4) Classification results on the IP dataset: To validate the classification performance under limited training samples, the IP data scene is used. From Table 9, it can be seen that our proposed model gets the best classification accuracy on OA, AA, and K . Meanwhile, our proposed model gets 100% classification accuracy on the classes of C1, C8, and C13, which demonstrates the effectiveness of our devised model under the IP data scene. Especially under the condition of limited sample quantities from classes C1, C7, C9, and C16, our proposed model achieves good accuracy in individual class classification. The SSRN and ChebNet receive 0% classification accuracy on C7 and C9, respectively. However, both the MVAHN and our proposed model achieve better classification accuracy on C7 and C9 compared to SSRN and ChebNet. This shows that the model combining CNN with GCN has better feature extraction performance than a model only based on CNN or GCN architecture. In Figure 9, the classification map generated by our proposed model shows clear boundaries between different classes. Compared to other comparative methods, MDBNet gets a worse OA value, and its classification map exhibits a higher occurrence of the salt-and-pepper noise phenomenon.

(5) Classification results on the Xiongan dataset: To further validate the superior classification performance of our designed method, we use the Xiongan dataset. From Table 10, it can be observed that our method achieves the best OA, AA, and k compared to other methods. Conversely, the SSRN method, which is based on CNN, has lower classification results, especially for the AA value, compared to other comparative methods. The classification accuracy of C3, C4, C6, C9, C11, C12, C16, C17, and C18 produced by SSRN is 0%, which contributes to the lower AA value. MVAHN, a hybrid architecture that combines CNN and GCN models, yields comparable results but has slightly lower classification performance than our method. For GCN-based comparative methods, such as ChebNet and GCN, the classification results are relatively lower compared to other methods. Other comparative methods, such as DGFNet, FTINet, and MRCAG exhibit acceptable classification results. From Figure 10, the full-pixel classification map of SSRN displays unclear edges between different classes, resulting in a poor classification map. Our method shows a classification map that is more similar to the ground-truth map, demonstrating superior classification performance.

5. Discussion

5.1. The Importance of an Adaptive Feature-Weighted Strategy in Feature Fusion

In the CNN subnetwork of our proposed method, the extracted spectral and spatial information plays unequally significant roles in the classification process. The spectral signature learning weight α_1 and spatial features learning weight α_2 are employed to

show the importance of spectral and spatial features in the training accuracy and validation accuracy of the four used HSI datasets. As shown in Figure 11a, it can be observed that for the PU data scene, the value of the spectral learning weight α_1 is higher than that of the spatial weight α_2 , and the difference between them becomes increasingly larger as the epochs increase, while both the training accuracy and validation accuracy are increasing. Therefore, when the proposed method achieves higher training and validation accuracy, it demonstrates that the spectral learning weight α_1 and the spatial learning weight α_2 play different roles in classification, which shows the appropriateness of the used adaptive feature-weighted fusion strategy. The extracted spectral signatures have a relatively larger proportion than spatial features in classification. From Figure 11a, when the value of α_1 is 0.5329 and the value of α_2 is 0.4671, the proposed model gets the best performance in training and validation accuracy on the PU dataset.

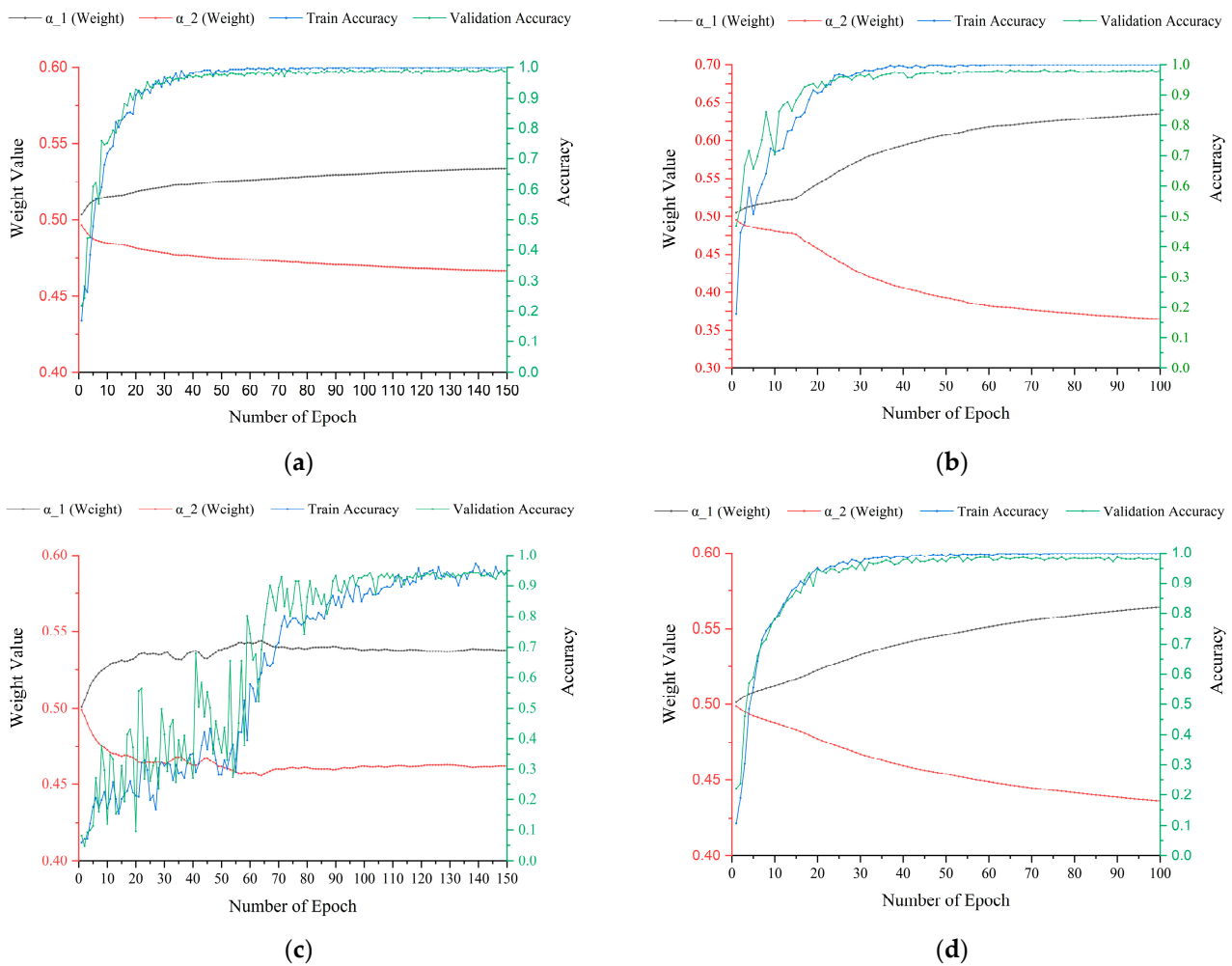


Figure 11. The train accuracy and validation accuracy under the influence of weight α_1 and weight α_2 on (a) PU; (b) Honghu; (c) Houston; and (d) IP data scenes.

Additionally, in Figure 11b–d, feature learning weights α_1 and α_2 are analyzed respectively on Honghu, Houston, and IP data scene. In these three figures, phenomena similar to those in Figure 11a can be observed. Especially in the Honghu data scene, when the value of α_1 is 0.6279 and the value of α_2 is 0.3721, the training accuracy reaches a value of 100%, and the validation accuracy reaches a value of 98.09%.

5.2. The Value of n in n -Layer GCN of the Spectral-Enhanced GCN Module

To learn the features from the perspective of GCN as well as to learn the relations in the spectral feature matrix, an n -layer GCN is used. In general, GCN-based architectures can learn deep node features by using more stacked GCN layers. However, when the GCN layers reach a certain level, it will lead to a drop in the performance of the spectral-enhanced GCN module. Additionally, using the spectral-enhanced GCN module with fewer GCN layers results in inadequate capture of deep information. To ensure that the values of n are reasonable, an experiment is conducted. To be specific, we implement experiments with different n -layer ($n = 1, 2, 3, 4, 5$) GCNs in spectral-enhanced GCN modules on the four datasets used. The purpose is to determine the optimal value of n to achieve the best performance of our model. From Figure 12, when the value of n is 4, the OA of our model gets the best classification accuracy on four used data scenes separately. Therefore, a four-layer GCN is utilized in the spectral-enhanced GCN module, which is beneficial for spectral signature learning.

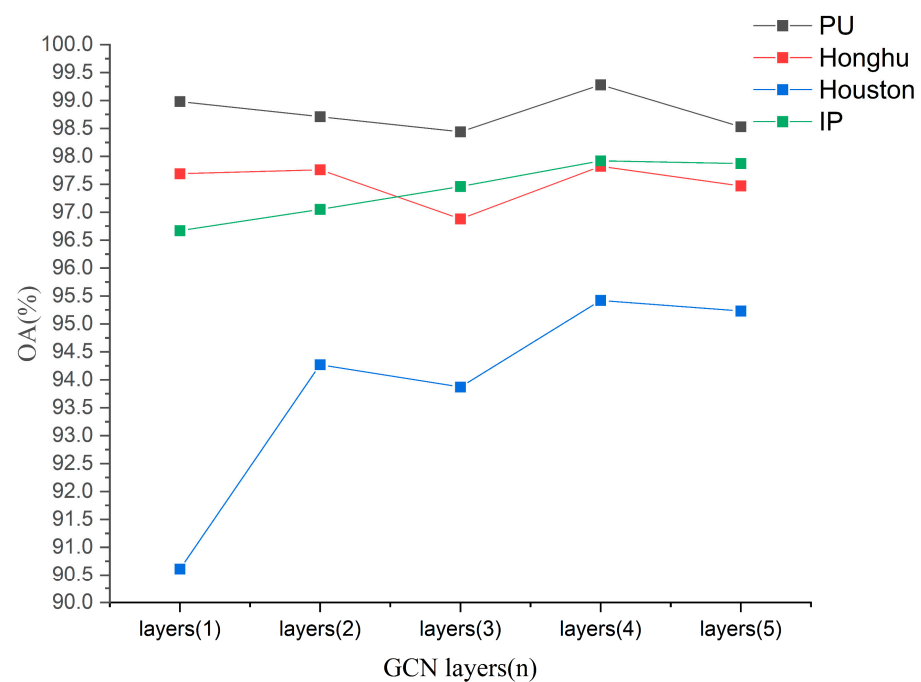


Figure 12. The OA of our proposed model under different GCN layers on PU, Honghu, Houston, and IP data scenes.

5.3. The Learning Rate under Different Epoch Numbers

The learning rate, a vital hyperparameter, plays a great role in training a deep learning-based model. And, it has a significant influence on both the convergence and the final classification performance of our model. Additionally, the number of epochs can affect the convergence speed of the model as well as its training time. So, we evaluate the impact of various learning rates on the classification accuracy of our proposed model under the different number of epochs on the four used data scenes, and the results are shown in Figure 13. To analyze the impact of learning rates and various epochs on the proposed model, the learning rate for the PU, Honghu, Houston, IP, and Xiongan are selected from the sets $\{0.007, 3 \times 10^{-4}, 5 \times 10^{-4}, 7 \times 10^{-4}, 9 \times 10^{-4}\}$, $\{0.003, 0.007, 3 \times 10^{-4}, 7 \times 10^{-4}, 9 \times 10^{-4}\}$, $\{0.003, 0.005, 3 \times 10^{-4}, 5 \times 10^{-4}, 7 \times 10^{-4}, 9 \times 10^{-4}\}$, and $\{0.01, 0.005, 3 \times 10^{-4}, 5 \times 10^{-4}, 7 \times 10^{-4}\}$. Simultaneously, the learning rate settings for PU, Honghu, Houston, and IP are tested under various epochs sets $\{200, 250, 300, 350, 400\}$, $\{100, 120, 150, 180, 200\}$, $\{100, 120, 150, 180, 200\}$, and $\{130, 150, 200, 250, 300\}$ respectively. From Figure 13, it is observed that when the set of learning rate and epochs $\{5 \times 10^{-4}, 350\}$, $\{7 \times 10^{-4},$

120}, $\{9 \times 10^{-4}, 150\}$, and $\{3 \times 10^{-4}, 130\}$ are selected for the PU, Honghu, Houston, IP, and Xiongan data scenarios separately, our proposed model gets the best classification performance in each of these data scenes.

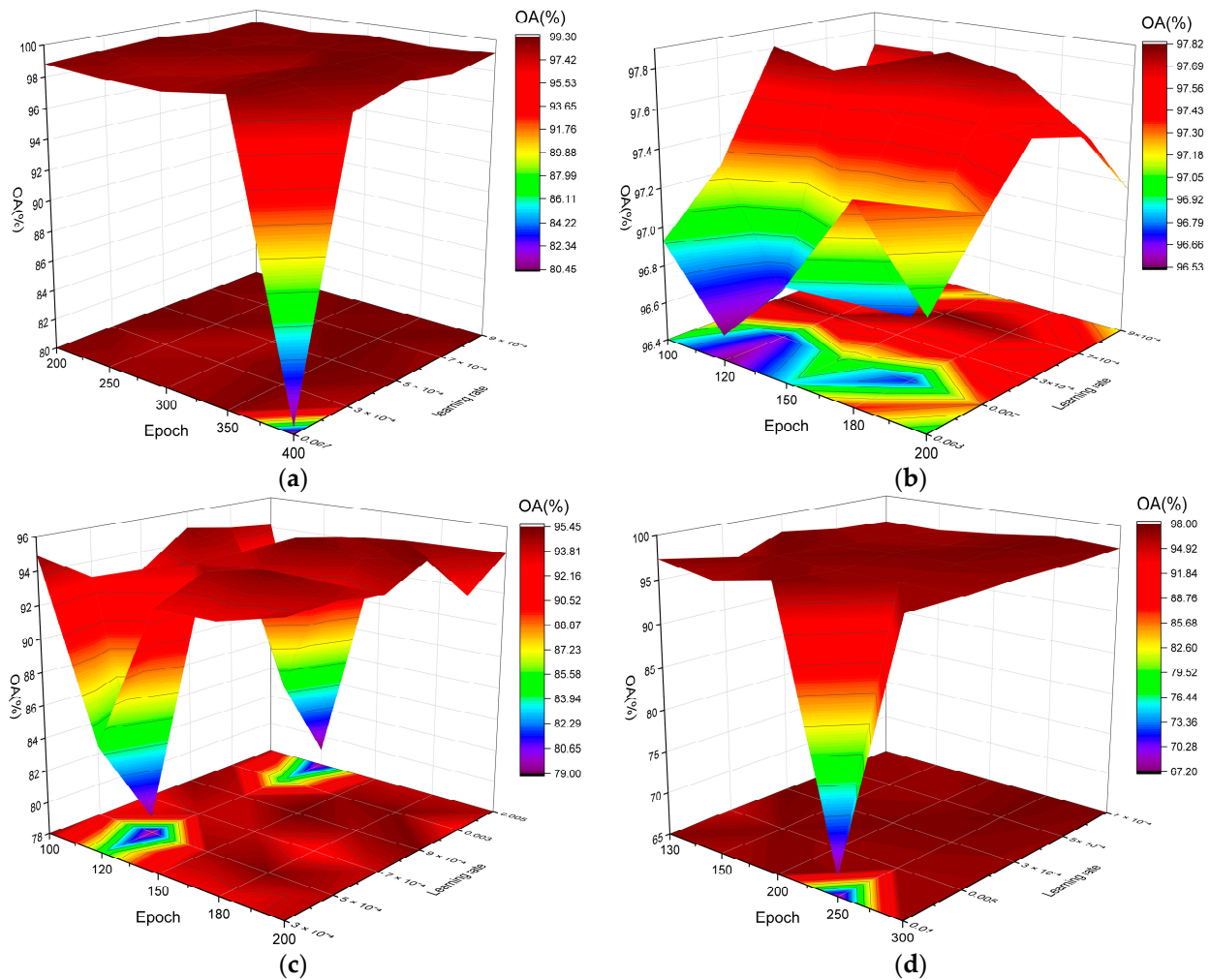


Figure 13. The OA of learning rate of our method under different epochs on (a) PU; (b) Honghu; (c) Houston; (d) IP data scenes.

5.4. Impact of Different Training Samples on the Classification Result

For our proposed model and other comparative methods, as deep learning-based models, the training sample is an important factor that influences the classification performance.

To analyze the classification performance of our proposed model and the other fourteen comparison algorithms under different limited training samples, $\{0.5\%, 1\%, 2\%, 3\%, 4\%\}$, $\{1\%, 2\%, 3\%, 4\%, 5\%\}$, $\{2\%, 3\%, 4\%, 5\%, 6\%\}$, and $\{5\%, 6\%, 7\%, 8\%, 9\%\}$ training samples sets are separately selected to train our proposed model and fourteen different comparison algorithms and to evaluate their classification performances on the PU, Honghu, Houston, and IP data scenes separately. From Figure 14, it is clear that our proposed model exhibits the best classification accuracy under different training samples, especially with limited training samples.

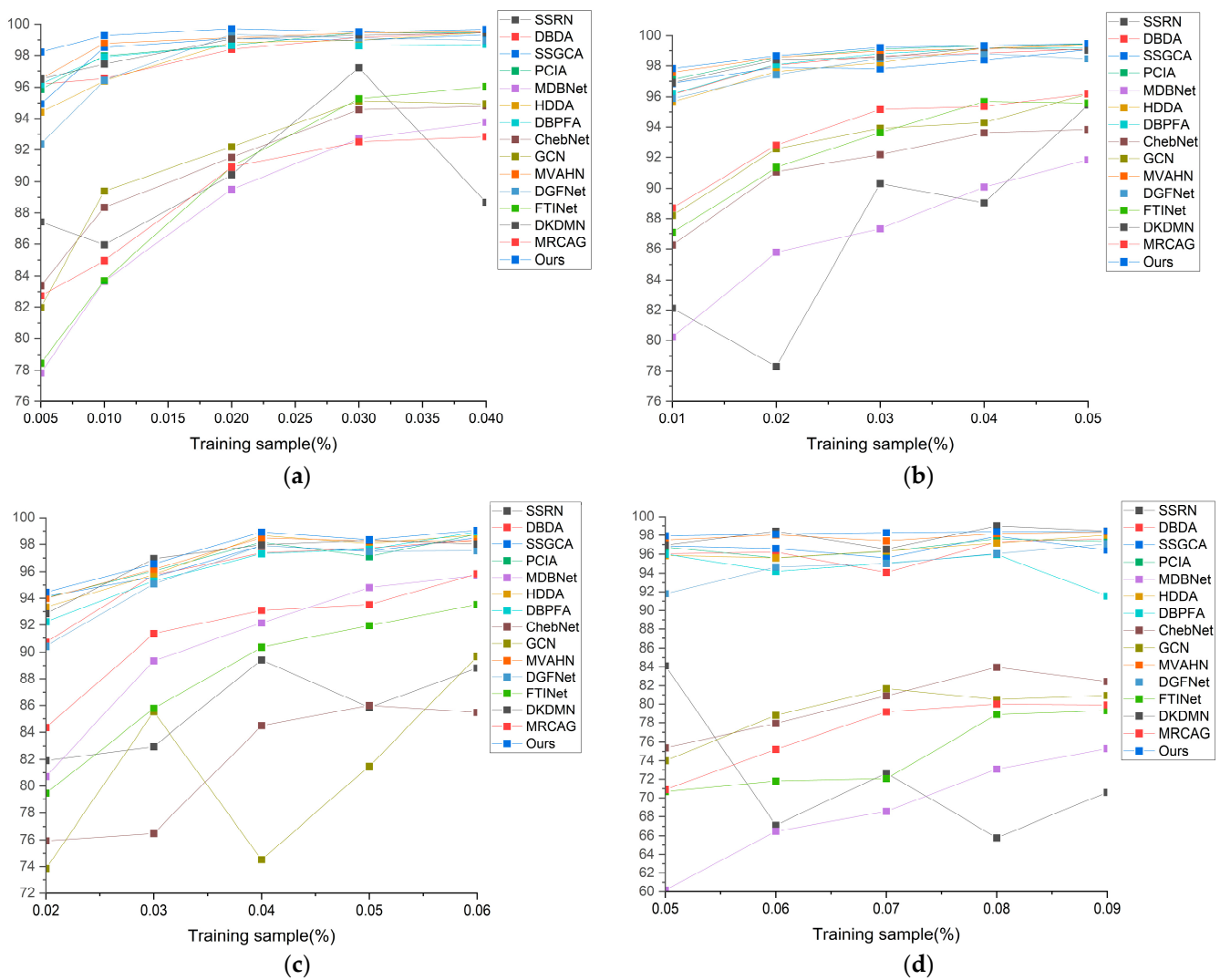


Figure 14. The OA of different classification methods under different training samples on (a) PU; (b) Honghu; (c) Houston; (d) IP data scenes.

5.5. Visual Results about Different Methods

In this section, to intuitively visualize the classification performance of fourteen comparative methods and our devised model, a t-distributed stochastic neighbor embedding (T-SNE) technique is employed. And, we take IP data scene as an example. From the Figure 15e, it can be observed that the classification map of MDBNet shows different class mix-ups together, causing a phenomenon of confusion. This is consistent with the result from Table 9, that the OA of MDBNet is lower than other comparative methods. Additionally, the T-SNE-based classification maps of DBPFA, MVAHN, and our proposed model demonstrate a clearer inter-class separation phenomenon. Furthermore, the intra-class separation phenomenon of the T-SNE-based classification maps produced by our model is the best compared to DBPFA and MVAHN, demonstrating the superior classification performance of our method. Meanwhile, as depicted in Figure 15f,j,k, the T-SNE-based classification map generated by the hybrid GCN and CNN model exhibits a better inter-class separation phenomenon compared to the T-SNE-based classification maps generated by models based solely on GCN or CNN, which shows the benefit of models that combine the CNN and GCN for feature extraction. From Figure 15l, the T-SNE-based classification map of our proposed model, which shows a better inter-class and intra-class clustering scenario, indicates better classification performance compared to other comparative methods.

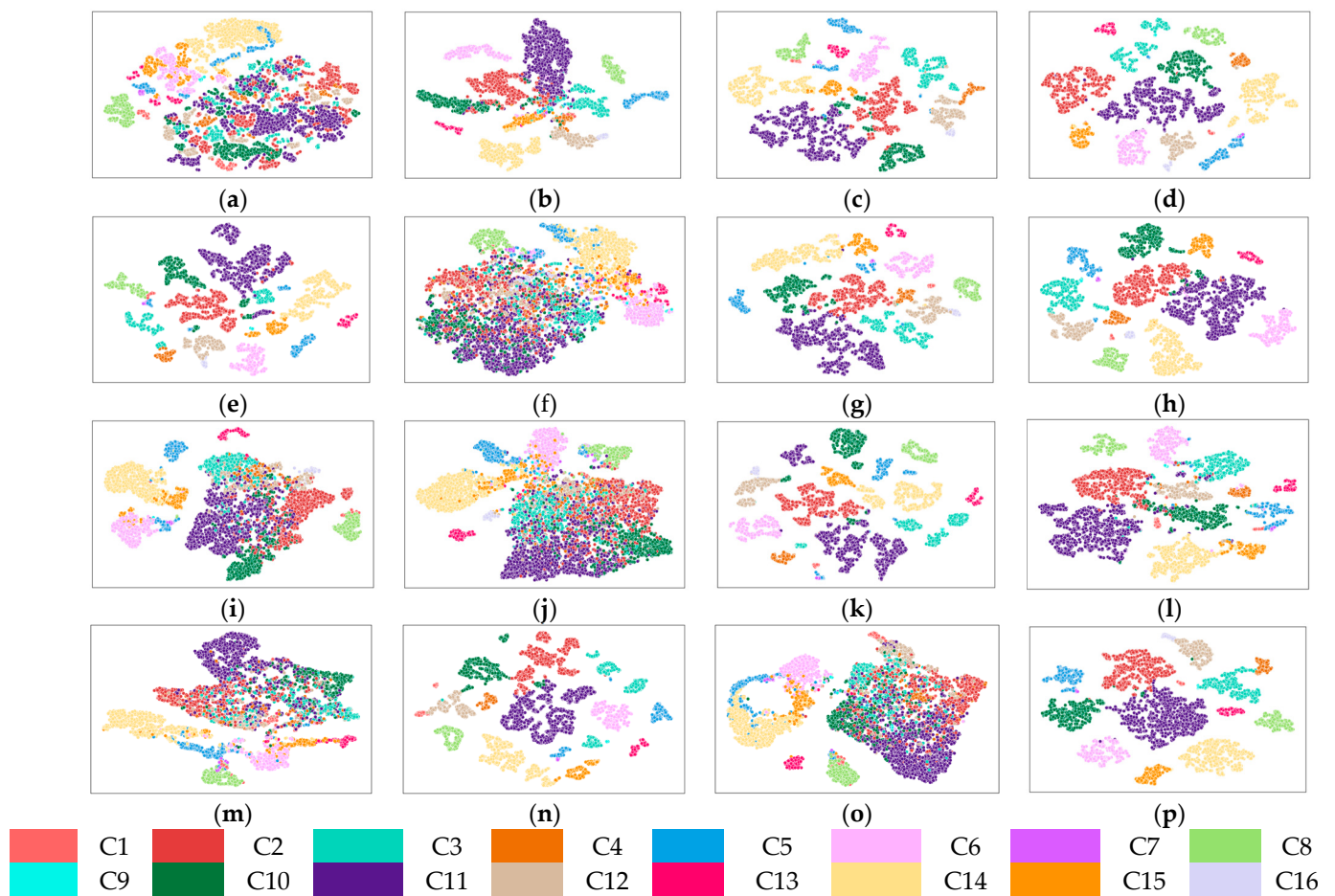


Figure 15. Feature visualization map of ten comparative methods and our proposed method on IP data scene. (a) Origin data; (b) SSRN; (c) DBDA; (d) SSGCA; (e) PCIA; (f) MDBNet; (g) HDDA; (h) DBPFA; (i) ChebNet; (j) GCN; (k) MVAHN; (l) DGFNet; (m) FTINet; (n) DKDMN; (o) MRCAG; (p) ours.

5.6. Ablation Experiment

As described in the Methods Section, our proposed model mainly includes a CNN-based subnetwork and a GCN-based subnetwork. And in each subnetwork, the spectral and spatial pyramid hybrid convolution block, adaptive feature-weighted fusion strategy, spectral-enhanced GCN module, and mutual-cooperative attention mechanism are crucial for our proposed model. In this section, some ablation experiments are performed to verify the effectiveness of various designed modules in our proposed model. The environment settings of ablation experiments performed on PU, Honghu, Houston, and IP image data scenes are described in Section 4.2. And, we take the value of OA as an evaluation indicator to show the effectiveness of various devised modules, which will compare the complete model with the model lacking corresponding modules to show the effectiveness of each corresponding module. In detail, from Figure 16, the OA of model_0 is higher than that of model_1, model_2, model_3, model_4, model_5, model_6, model_7, and model_8, which shows the reasonability and the superlative classification accuracy of our proposed model.

To demonstrate the validity of the designed spectral pyramid hybrid convolution block and spatial pyramid hybrid convolution block, we individually eliminate the spatial pyramid hybrid convolution block and spectral pyramid hybrid convolution block from the complete model. As demonstrated in Figure 16, the OAs of model_8 and model_7 are both lower than that of model_0, which indicates the importance of the spectral pyramid hybrid block and spatial pyramid hybrid convolution block in multiscale feature extraction.

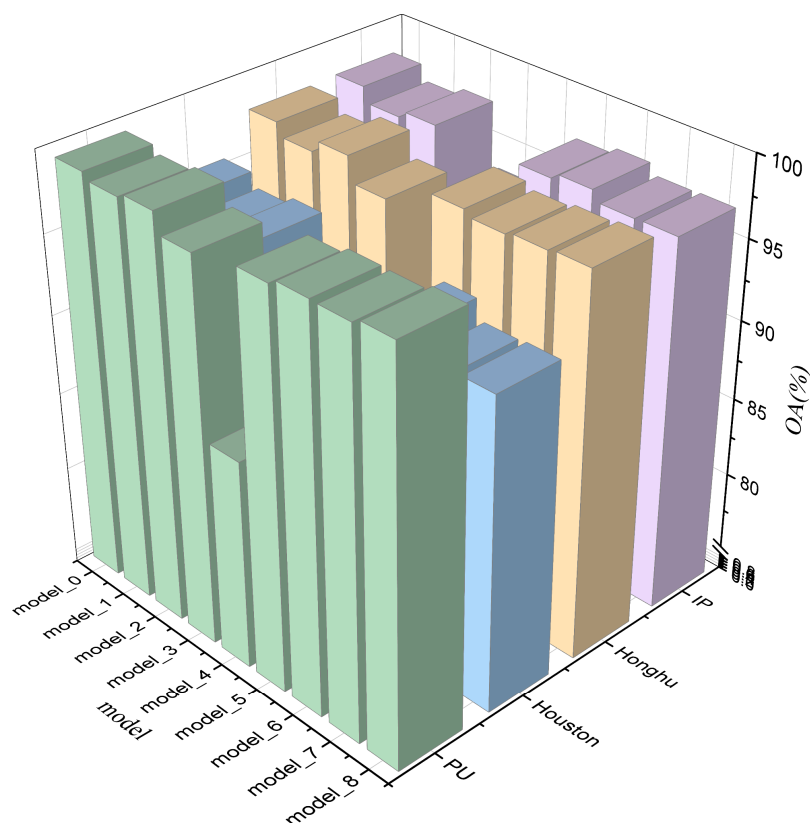


Figure 16. Ablation experiments of our proposed model on PU, Houston, Honghu, and IP data scenes: model_0: complete model; model_1: model without mutual-cooperative mechanism; model_2: model with mutual-cooperative attention mechanism that includes GCN-based spectral signature; model_3: model without spectral-enhanced GCN module; model_4: the model that only includes GCN-based subnetwork; model_5: model that only includes CNN-based subnetwork; model_6: model without adaptive feature-weighted fusion strategy; model_7: model without spectral pyramid hybrid convolution block; model_8: model without spatial pyramid hybrid convolution block.

To verify the effectiveness of the devised GCN-based subnetwork, we remove the GCN-based subnetwork from the proposed model. The OA of model_5 is lower than that of model_0, which shows the effectiveness of the GCN-based subnetwork.

To demonstrate the robustness of the CNN-based subnetwork to the complete model, we remove the CNN-based subnetwork; as shown in Figure 16, the OA of model_4 is lower than that of model_0 on the four widely used HSI data scenes. Meanwhile, the OA of model_4 is lower than others on the model axis; especially on the IP data scene, the OA of model_4 is much lower than model_0. And, the limited feature extraction ability of the GCN-based subnetwork is also shown.

To show the importance of the mutual-cooperative attention mechanism, we remove the mutual-cooperative attention mechanism from the proposed model; as shown in Figure 16, the OA of model_1 is lower than the OA of model_0, which indicates the importance of the mutual-cooperative attention mechanism. Meanwhile, the OAs of model_2 and model_3 are both lower than the OA of model_0, which shows the significance of the mutual-cooperative attention mechanism without BNet-based spectral features and GCN-based spectral signatures separately. And, the phenomenon that the OA of model_3 is lower than the OA of model_0 also demonstrates the significance of spectral-enhanced GCN modules.

5.7. The Visualization of the Spectral-Enhanced GCN Module

To validate the effectiveness of the proposed spectral-enhanced GCN modules, the heatmaps of features before and after using the spectral-enhanced GCN module are shown in Figure 17. Taking the PU data scene as an example, we randomly chose a 9×11 pixel region from the spectral matrix to show the features contained in the spectral matrix. From Figure 17a, it can be seen that the image has lighter color, especially in the range of lighter green, and each pixel does not show significant differences. It shows that the features within a certain region before using the spectral-enhanced GCN module have weak differences. Conversely, from Figure 17b, it can be seen that the heatmap displays darker colors, with different darker shades among pixels for different classes. According to Figure 17b,d,f,g,h, the features within the feature matrix are accentuated, which demonstrates the effectiveness of the proposed spectral-enhanced GCN module. On the Houston, Honghu, and IP datasets, Figure 17c,e,g shows that the feature matrices display relatively lighter colors. In contrast, Figure 17d,f,h present relatively darker feature maps, which are processed by the spectral-enhanced GCN module. These heatmaps exhibit more pronounced differences between pixels of different classes, highlighting the significant features in the feature map.

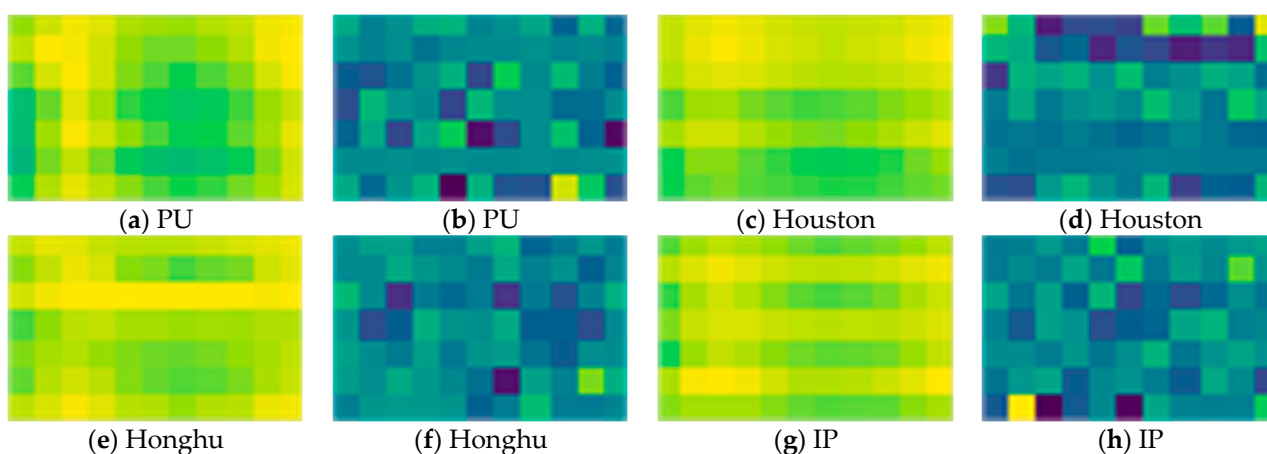


Figure 17. The features before the spectral-enhanced GCN module: (a,c,e,g), the features after the spectral-enhanced GCN module: (b,d,f,h).

5.8. Training Times

In this subsection, the time consumed in experiments is discussed to compare the efficiency of our proposed method on each dataset used. Tables 11–14 show the detailed training and testing times for each comparative method and our method. Taking the PU data scene as an example, Table 11 shows that the method called HDDA has the highest training time. HDDA also has the highest testing times compared to other methods. The GCN comparative method has the lowest training and testing times. SSRN, DBDA, and MDBNet have much higher training and testing times than our proposed method. For the other datasets used in our experiment section, our method exhibits similar training and testing times compared to other comparative methods. Although our proposed method does not have the lowest training and testing times, its time efficiency is acceptable compared to other comparative methods.

Table 11. Training and testing times of different comparative methods and our method on the PU data.

	SSRN	DBDA	SSGCA	PCIA	MDBNet	HDDA	DBPFA	ChebNet	GCN	MVAHN	DGFNet	FTINet	DKDMN	MRCAG	Ours
Train Time (s)	23.79	42.37	31.39	26.77	51.72	211.94	50.55	12.86	13.25	48.39	92.50	69.94	84.06	23.00	50.12
Test Time (s)	3.88	7.20	4.19	5.42	12.50	21.42	7.93	2.30	2.32	8.16	28.28	9.67	12.09	18.78	7.15

Table 12. Training and testing times of different comparative methods and our method on the Houston data.

	SSRN	DBDA	SSGCA	PCIA	MDBNet	HDDA	DBPFA	ChebNet	GCN	MVAHN	DGFNet	FTINet	DKDMN	MRCAG	Ours
Train Time (s)	177.35	10.30	10.00	10.15	14.80	22.77	11.42	4.14	3.81	16.33	46.42	78.58	53.41	23.46	17.79
Test Time (s)	0.64	0.84	0.87	1.10	1.51	2.36	1.03	0.46	0.44	1.51	8.23	3.51	3.17	5.84	1.56

Table 13. Training and testing times of different comparative methods and our method on the Honghu data.

	SSRN	DBDA	SSGCA	PCIA	MDBNet	HDDA	DBPFA	ChebNet	GCN	MVAHN	DGFNet	FTINet	DKDMN	MRCAG	Ours
Train Time (s)	793.08	206.92	194.94	215.90	252.04	1026.96	190.63	7.90	10.25	51.13	110.74	102.42	156.44	61.60	362.80
Test Time (s)	14.94	20.98	17.51	24.08	66.60	86.48	21.48	3.16	3.26	13.76	51.80	16.97	30.97	31.92	44.05

Table 14. Training and testing times of different comparative methods and our method on the IP data.

	SSRN	DBDA	SSGCA	PCIA	MDBNet	HDDA	DBPFA	ChebNet	GCN	MVAHN	DGFNet	FTINet	DKDMN	MRCAG	Ours
Train Time (s)	798.44	152.24	86.51	111.95	127.58	442.71	81.32	14.04	10.91	51.09	86.10	98.10	90.18	34.22	114.91
Test Time (s)	1.64	3.18	1.88	2.27	5.64	8.87	2.40	1.08	0.69	4.71	6.41	2.36	2.46	4.26	3.79

6. Conclusions

In this article, we propose a novel PCCGC method that combines CNN and GCN for HSI classification. It contains two parallel subnetworks, namely, a CNN-based subnetwork and a GCN-based subnetwork. Specifically, in the CNN subnetwork, a SpePRCM is employed to extract multiscale spectral signatures. Meanwhile, SpaPCCM is used to extract multiscale spatial features. Furthermore, an adaptive feature-weighted fusion strategy is employed to adaptively fuse multiscale spectral and spatial features based on their respective weights. Based on the above, the CNN subnetwork can enhance the robustness of the proposed model in classifying HSIs. In the GCN subnetwork, a BSNet is first used to learn the spectral signatures in the origin HSI using nonlinear inter-band dependencies. Then, the spectral-enhanced GCN module is employed to learn and accentuate the important features in the spectral channel. Subsequently, a mutual-cooperative attention mechanism is constructed that can align the spectral signatures between BSNet-based matrix with spectral-enhanced GCN-based matrix for spectral signature integration. Finally, the additive fusion strategy is utilized to fusion the features extracted from GCN-based and CNN-based subnetworks. The effectiveness and robustness of our designed model are demonstrated by quantitative and qualitative experiments. In addition, to verify the superior performance of our model, numerous parametric analyses and ablation experiments are conducted.

However, the spectral-enhanced module used in GCN-based subnetwork only learns the significant features in the spectral channel. In the future, the designed GNN will be employed to extract the features from the spectral and spatial channels simultaneously. And, a fusion-based mechanism will be employed to exquisitely combine the CNN and GCN model.

Author Contributions: Conceptualization, H.P. and H.Y.; methodology, H.P., H.Y. and H.G.; software, H.P., H.Y. and H.G.; validation, H.P., H.Y. and H.G.; formal analysis, H.P. and H.Y.; investigation, H.P., H.Y. and H.G.; resources, H.P., L.W. and C.S.; data curation, H.P. and H.G.; writing—original draft preparation, H.P. and H.Y.; writing—review and editing, H.P. and H.Y.; visualization, H.P., H.Y. and H.G.; supervision, H.P. and H.Y.; project administration, H.P., L.W. and C.S.; funding acquisition, H.P. and C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Heilongjiang Provincial Natural Science Foundation of China, grant number LH2023F050, the Fundamental Research Funds in Heilongjiang Provincial Universities under Grant, grant number 145309208, the National Natural Science Foundation of China, grant number 42271409, and the Heilongjiang Provincial Higher Education Teaching and Reform Project, grant number SJGZ20220112.

Data Availability Statement: Data available in a publicly accessible repository: Pavia University dataset and Indian Pines (http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, accessed on 24 June 2024); Houston 2013 dataset (<https://www.grss-ieee.org/resources/tutorials-documents/>, accessed on 24 June 2024), WHU-Hi-Honghu (http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm, accessed on 24 June 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lorenz, S.; Salehi, S.; Kirsch, M.; Zimmermann, R.; Unger, G.; Vest Sørensen, E.; Gloaguen, R. Radiometric Correction and 3D Integration of Long-Range Ground-Based Hyperspectral Imagery for Mineral Exploration of Vertical Outcrops. *Remote Sens.* **2018**, *10*, 176. [CrossRef]
2. Rajabi, R.; Zehtabian, A.; Singh, K.D.; Tabatabaenejad, A.; Ghamisi, P.; Homayouni, S. Editorial: Hyperspectral Imaging in Environmental Monitoring and Analysis. *Front. Environ. Sci.* **2024**, *11*, 1353447. [CrossRef]
3. Kütük, M.; Geneci, İ.; Bilge Özdemir, O.; Koz, A.; Esentürk, O.; Yardımcı Çetin, Y.; Alatan, A.A. Ground-Based Hyperspectral Image Surveillance System for Explosive Detection: Methods, Experiments, and Comparisons. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 8747–8763. [CrossRef]
4. Waczak, J.; Aker, A.; Wijeratne, L.O.H.; Talebi, S.; Fernando, A.; Dewage, P.M.H.; Iqbal, M.; Lary, M.; Schaefer, D.; Lary, D.J. Characterizing Water Composition with an Autonomous Robotic Team Employing Comprehensive In Situ Sensing, Hyperspectral Imaging, Machine Learning, and Conformal Prediction. *Remote Sens.* **2024**, *16*, 996. [CrossRef]

5. Gallacher, C.; Benz, S.; Boehnke, D.; Jehling, M. A Collaborative Approach for the Identification of Thermal Hot-Spots: From Remote Sensing Data to Urban Planning Interventions. *AGILE GIScience Ser.* **2024**, *5*, 23. [[CrossRef](#)]
6. Ge, H.; Wang, L.; Liu, M.; Zhao, X.; Zhu, Y.; Pan, H.; Liu, Y. Pyramidal Multiscale Convolutional Network with Polarized Self-Attention for Pixel-Wise Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5504018. [[CrossRef](#)]
7. Jia, X.; Kuo, B.-C.; Crawford, M.M. Feature Mining for Hyperspectral Image Classification. *Proc. IEEE* **2013**, *101*, 676–697. [[CrossRef](#)]
8. Liu, Q.; Xiao, L.; Yang, J.; Wei, Z. CNN-Enhanced Graph Convolutional Network with Pixel- and Superpixel-Level Feature Fusion for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8657–8671. [[CrossRef](#)]
9. Liu, Q.; Dong, Y.; Zhang, Y.; Luo, H. A Fast Dynamic Graph Convolutional Network and CNN Parallel Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5530215. [[CrossRef](#)]
10. Zhao, C.; Wan, X.; Zhao, G.; Cui, B.; Liu, W.; Qi, B. Spectral-Spatial Classification of Hyperspectral Imagery Based on Stacked Sparse Autoencoder and Random Forest. *Eur. J. Remote Sens.* **2017**, *50*, 47–63. [[CrossRef](#)]
11. Guo, Y.; Cao, H.; Han, S.; Sun, Y.; Bai, Y. Spectral–Spatial Hyperspectral Image Classification With K-Nearest Neighbor and Guided Filter. *IEEE Access* **2018**, *6*, 18582–18591. [[CrossRef](#)]
12. Chen, Y.-N.; Thaipisutikul, T.; Han, C.-C.; Liu, T.-J.; Fan, K.-C. Feature Line Embedding Based on Support Vector Machine for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 130. [[CrossRef](#)]
13. Sun, W.; Yang, G.; Peng, J.; Du, Q. Lateral-Slice Sparse Tensor Robust Principal Component Analysis for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 107–111. [[CrossRef](#)]
14. Luo, H.; Tang, Y.Y.; Biuk-Aghai, R.P.; Yang, X.; Yang, L.; Wang, Y. Wavelet-Based Extended Morphological Profile and Deep Autoencoder for Hyperspectral Image Classification. *Int. J. Wavelets Multiresolution Inf. Process.* **2018**, *16*, 1850016. [[CrossRef](#)]
15. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–Spatial Classification of Hyperspectral Images With a Superpixel-Based Discriminative Sparse Model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4186–4201. [[CrossRef](#)]
16. Wang, T.; Wang, G.; Tan, K.E.; Tan, D. Spectral Pyramid Graph Attention Network for Hyperspectral Image Classification. *arXiv* **2020**, arXiv:2001.07108.
17. Zhang, X.; Chen, S.; Zhu, P.; Tang, X.; Feng, J.; Jiao, L. Spatial Pooling Graph Convolutional Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [[CrossRef](#)]
18. Liu, M.; Pan, H.; Ge, H.; Wang, L. MS3Net: Multiscale Stratified-Split Symmetric Network with Quadra-View Attention for Hyperspectral Image Classification. *Signal Process.* **2023**, *212*, 109153. [[CrossRef](#)]
19. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *J. Sens.* **2015**, *2015*, 258619. [[CrossRef](#)]
20. Hang, R.; Liu, Q.; Hong, D.; Ghamisi, P. Cascaded Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5384–5394. [[CrossRef](#)]
21. Yang, A.; Li, M.; Ding, Y.; Hong, D.; Lv, Y.; He, Y. GTFN: GCN and Transformer Fusion Network With Spatial-Spectral Features for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 6600115. [[CrossRef](#)]
22. Yao, J.; Zhang, B.; Li, C.; Hong, D.; Chanussot, J. Extended Vision Transformer (ExViT) for Land Use and Land Cover Classification: A Multimodal Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5514415. [[CrossRef](#)]
23. Mei, S.; Song, C.; Ma, M.; Xu, F. Hyperspectral Image Classification Using Group-Aware Hierarchical Transformer. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5539014. [[CrossRef](#)]
24. Wu, K.; Fan, J.; Ye, P.; Zhu, M. Hyperspectral Image Classification Using Spectral–Spatial Token Enhanced Transformer With Hash-Based Positional Embedding. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5507016. [[CrossRef](#)]
25. Miftahushudur, T.; Grieve, B.; Yin, H. Permuted KPCA and SMOTE to Guide GAN-Based Oversampling for Imbalanced HSI Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 489–505. [[CrossRef](#)]
26. Xu, M.; Sun, J.; Yao, K.; Cai, Q.; Shen, J.; Tian, Y.; Zhou, X. Developing Deep Learning Based Regression Approaches for Prediction of Firmness and pH in Kyoho Grape Using Vis/NIR Hyperspectral Imaging. *Infrared Phys. Technol.* **2022**, *120*, 104003. [[CrossRef](#)]
27. Yang, J.; Zhao, Y.-Q.; Chan, J.C.-W. Learning and Transferring Deep Joint Spectral–Spatial Features for Hyperspectral Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4729–4742. [[CrossRef](#)]
28. Qi, W.; Zhang, X.; Wang, N.; Zhang, M.; Cen, Y. A Spectral-Spatial Cascaded 3D Convolutional Neural Network with a Convolutional Long Short-Term Memory Network for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 2363. [[CrossRef](#)]
29. Fu, C.; Du, B.; Zhang, L. ReSC-Net: Hyperspectral Image Classification Based on Attention-Enhanced Residual Module and Spatial-Channel Attention. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5518615. [[CrossRef](#)]
30. Li, Y.; Yang, X.; Tang, D.; Zhou, Z. RDTN: Residual Densely Transformer Network for Hyperspectral Image Classification. *Expert Syst. Appl.* **2024**, *250*, 123939. [[CrossRef](#)]
31. Pan, H.; Zhao, X.; Ge, H.; Liu, M.; Shi, C. Hyperspectral Image Classification Based on Multiscale Hybrid Networks and Attention Mechanisms. *Remote Sens.* **2023**, *15*, 2720. [[CrossRef](#)]
32. Ding, Y.; Chong, Y.; Pan, S.; Zheng, C.-H. Class-Imbalanced Graph Convolution Smoothing for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5510618. [[CrossRef](#)]
33. Wan, S.; Gong, C.; Zhong, P.; Du, B.; Zhang, L.; Yang, J. Multi-scale Dynamic Graph Convolutional Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3162–3177. [[CrossRef](#)]

34. Zhang, Y.; Miao, R.; Dong, Y.; Du, B. Multiorder Graph Convolutional Network With Channel Attention for Hyperspectral Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1523–1534. [[CrossRef](#)]
35. Dong, Y.; Liu, Q.; Du, B.; Zhang, L. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2022**, *31*, 1559–1572. [[CrossRef](#)]
36. Guan, R.; Li, Z.; Tu, W.; Wang, J.; Liu, Y.; Li, X.; Tang, C.; Feng, R. Contrastive Multiview Subspace Clustering of Hyperspectral Images Based on Graph Convolutional Networks. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5510514. [[CrossRef](#)]
37. Zhou, H.; Luo, F.; Zhuang, H.; Weng, Z.; Gong, X.; Lin, Z. Attention Multihop Graph and Multiscale Convolutional Fusion Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5508614. [[CrossRef](#)]
38. Li, W.; Liu, Q.; Fan, S.; Xu, C.; Bai, H. Dual-Stream GNN Fusion Network for Hyperspectral Classification. *Appl. Intell.* **2023**, *53*, 26542–26567. [[CrossRef](#)]
39. Jia, S.; Jiang, S.; Zhang, S.; Xu, M.; Jia, X. Graph-in-Graph Convolutional Network for Hyperspectral Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 1157–1171. [[CrossRef](#)]
40. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph Convolutional Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5966–5978. [[CrossRef](#)]
41. Gao, H.; Sheng, R.; Chen, Z.; Liu, H.; Xu, S.; Zhang, B. Multiscale Random-Shape Convolution and Adaptive Graph Convolution Fusion Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516017. [[CrossRef](#)]
42. Zhang, Z.; Cai, Y.; Liu, X.; Zhang, M.; Meng, Y. An Efficient Graph Convolutional RVFL Network for Hyperspectral Image Classification. *Remote Sens.* **2023**, *16*, 37. [[CrossRef](#)]
43. Liu, X.; Liu, S.; Chen, W.; Qu, S. HDECGCN: A Heterogeneous Dual Enhanced Network Based on Hybrid CNNs Joint Multiscale Dynamic GCNs for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5515717. [[CrossRef](#)]
44. Shi, C.; Yue, S.; Wu, H.; Zhu, F.; Wang, L. A Multihop Graph Rectify Attention and Spectral Overlap Grouping Convolutional Fusion Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5520517. [[CrossRef](#)]
45. K Ghotekar, R.; Shaw, K.; Rout, M. Deep Feature Segmentation Model Driven by Hybrid Convolution Network for Hyper Spectral Image Classification. *Int. J. Comput. Digit. Syst.* **2024**, *15*, 719–738. [[CrossRef](#)] [[PubMed](#)]
46. Lu, Y.; Mei, S.; Xu, F.; Ma, M.; Wang, X. Separable Deep Graph Convolutional Network Integrated With CNN and Prototype Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516216. [[CrossRef](#)]
47. Li, H.; Xiong, X.; Liu, C.; Ma, Y.; Zeng, S.; Li, Y. SFFNet: Staged Feature Fusion Network of Connecting Convolutional Neural Networks and Graph Convolutional Neural Networks for Hyperspectral Image Classification. *Appl. Sci.* **2024**, *14*, 2327. [[CrossRef](#)]
48. El Abady, N.F.; Zayed, H.H.; Taha, M. An Efficient Technique for Detecting Document Forgery in Hyperspectral Document Images. *Alex. Eng. J.* **2023**, *85*, 207–217. [[CrossRef](#)]
49. Fan, S.; Liu, Q.; Li, W.; Bai, H. A Frequency and Topology Interaction Network for Hyperspectral Image Classification. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108234. [[CrossRef](#)]
50. Feng, H.; Wang, Y.; Chen, C.; Xu, D.; Zhao, Z.; Zhao, T. Hyperspectral Image Classification Framework Based on Multichannel Graph Convolutional Networks and Class-Guided Attention Mechanism. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5515115. [[CrossRef](#)]
51. Jiang, M.; Su, Y.; Gao, L.; Plaza, A.; Zhao, X.-L.; Sun, X.; Liu, G. GraphGST: Graph Generative Structure-Aware Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5504016. [[CrossRef](#)]
52. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *32*, 4–24. [[CrossRef](#)] [[PubMed](#)]
53. Yu, W.; Wan, S.; Li, G.; Yang, J.; Gong, C. Hyperspectral Image Classification with Contrastive Graph Convolutional Network. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5503015. [[CrossRef](#)]
54. Shi, C.; Wu, H.; Wang, L. CEGAT: A CNN and Enhanced-GAT Based on Key Sample Selection Strategy for Hyperspectral Image Classification. *Neural Netw.* **2023**, *168*, 105–122. [[CrossRef](#)] [[PubMed](#)]
55. Wang, Q.; Huang, J.; Meng, Y.; Shen, T. DF2Net: Differential Feature Fusion Network for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 10660–10673. [[CrossRef](#)]
56. Wang, J.; Li, J.; Shi, Y.; Lai, J.; Tan, X. AM³Net: Adaptive Mutual-Learning-Based Multimodal Data Fusion Network. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 5411–5426. [[CrossRef](#)]
57. Xue, H.; Sun, X.-K.; Sun, W.-X. Multi-Hop Hierarchical Graph Neural Networks. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; IEEE: Busan, Republic of Korea, 2020; pp. 82–89.
58. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An End-to-End Framework for Band Selection of Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1969–1984. [[CrossRef](#)]
59. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 847–858. [[CrossRef](#)]
60. Li, R.; Zheng, S.; Duan, C.; Yang, Y.; Wang, X. Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network. *Remote Sens.* **2020**, *12*, 582. [[CrossRef](#)]
61. Li, Z.; Cui, X.; Wang, L.; Zhang, H.; Zhu, X.; Zhang, Y. Spectral and Spatial Global Context Attention for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 771. [[CrossRef](#)]

62. Shi, H.; Cao, G.; Ge, Z.; Zhang, Y.; Fu, P. Double-Branch Network with Pyramidal Convolution and Iterative Attention for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 1403. [[CrossRef](#)]
63. Gao, H.; Zhang, Y.; Chen, Z.; Li, C. A Multiscale Dual-Branch Feature Fusion and Attention Network for Hyperspectral Images Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8180–8192. [[CrossRef](#)]
64. Zhao, J.; Hu, L.; Dong, Y.; Huang, L. Hybrid Dense Network with Dual Attention for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 4921. [[CrossRef](#)]
65. Ge, H.; Wang, L.; Liu, M.; Zhu, Y.; Zhao, X.; Pan, H.; Liu, Y. Two-Branch Convolutional Neural Network with Polarized Full Attention for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 848. [[CrossRef](#)]
66. Defferrard, M.; Bresson, X.; Vandergheynst, P. *Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering*; EPFL: Lausanne, Switzerland, 2017.
67. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2016**, arXiv:1609.02907.
68. Zhao, F.; Zhang, J.; Meng, Z.; Liu, H.; Chang, Z.; Fan, J. Multiple Vision Architectures-Based Hybrid Network for Hyperspectral Image Classification. *Expert Syst. Appl.* **2023**, *234*, 121032. [[CrossRef](#)]
69. Zhang, J.; Zhao, F.; Liu, H.; Yu, J. Data and Knowledge-Driven Deep Multiview Fusion Network Based on Diffusion Model for Hyperspectral Image Classification. *Expert Syst. Appl.* **2024**, *249*, 123796. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.