



Hyperspectral Image Classification Based on 3D–2D Hybrid Convolution and Graph Attention Mechanism

Hui Zhang^{1,2} · Kaiping Tu¹ · Huanhuan Lv^{1,2} · Ruiqin Wang^{1,2}

Accepted: 23 February 2024
© The Author(s) 2024

Abstract

Convolutional neural networks and graph convolutional neural networks are two classical deep learning models that have been widely used in hyperspectral image classification tasks with remarkable achievements. However, hyperspectral image classification models based on graph convolutional neural networks using only shallow spectral or spatial features are insufficient to provide reliable similarity measures for constructing graph structures, limiting their classification performance. To address this problem, we propose a new end-to-end hyperspectral image classification model combining 3D–2D hybrid convolution and a graph attention mechanism (3D–2D-GAT). The model utilizes the collaborative work of hybrid convolutional feature extraction module and GAT module to improve classification accuracy. First, a 3D–2D hybrid convolutional network is constructed and used to quickly extract the discriminant deep spatial-spectral features of various ground objects in hyperspectral image. Then, the graph is built based on deep spatial-spectral features to enhance the feature representation ability. Finally, a network of graph attention mechanism is adopted to learn long-range spatial relationship and distinguish the intra-class variation and inter-class similarity among different samples. The experimental results on three datasets, Indian Pine, the University of Pavia and Salinas Valley show that the proposed method can achieve higher classification accuracy compared with other advanced methods.

Keywords Hyperspectral image classification · Convolutional neural network · Graph convolutional neural network · Attention mechanism

1 Introduction

Hyperspectral image classification (HSIC) aims to assign a unique category identity to each element in the image, which is a key technology for the intelligent interpretation of the

✉ Huanhuan Lv
03081@zjhu.edu.cn

¹ Huzhou University, Huzhou 313000, China

² Zhejiang Province Key Laboratory of Smart Management and Application of Modern Agricultural Resources, Huzhou 313000, China

hyperspectral image (HSI) and has been widely used in many fields such as urban development planning [1, 2], agricultural land use [3–5], military target detection [6, 7] and medical pathological diagnosis [8, 9]. However, the problems of low spatial resolution, high spectral dimensionality and lack of labelled samples in HSI pose great challenges to the classification task [10–12]. In the early days, researchers proposed a series of feature extraction methods such as principal component analysis [13, 14], independent component analysis [15–17], and linear discriminant analysis [17, 18], and combined them with machine learning classifiers such as support vector machines [19, 20], random forests [21, 22], and Gaussian mixture model [23, 24] to classify the HSI. These methods can effectively alleviate the Hughes phenomenon [25] that classification accuracy decreases with increasing spectral dimension, but because only spectral features are considered and based on manual design, the classification accuracy and applicability are not ideal. Therefore, it is important to design a HSIC method with high accuracy and applicability.

Recently, with the improvement of computer computing power, deep learning models represented by convolutional neural network (CNN) have been widely used in HSIC [26, 27]. Yue et al. [28] used a two-dimensional convolutional neural network (2D-CNN) to automatically extract deep features. However, the HSI is a three-dimensional cube, and the use of 2D convolution can only extract the spatial features and cannot fully utilize the spectral features. Another HSIC model based on three-dimensional convolution neural network (3D-CNN) is designed by Chen et al. [29], which introduced 3D convolution to directly extract the spatial-spectral features and achieved better classification results. Zhong et al. [30] employed a 3D residual network, using the idea of residual learning to effectively alleviate the problem of model degradation caused by deepening the network layers. Sun et al. [31] utilized spatial and spectral attention mechanisms to enhance the expression of the HSI spatial-spectral features and improved the classification accuracy. Although the introduction of 3D convolution allows the extraction of both rich spatial and spectral features, it also leads to an increase in the number of parameters of the network model. To address this problem, Roy et al. [32] developed a hybrid model (HybridSN) of 2D convolution and 3D convolution, which can obtain higher classification accuracy while reducing parameters. The above CNN-based methods have made some progress, but they only perform convolutional operations on the regular regions of the image and cannot fully express the local features of irregular regions, leading to higher misclassification in the case of small samples.

Graph Convolutional Neural Network (GCN) is a deep learning network model that performs convolutional operations on graph-structured data based on graph theory, which can make full use of the spatial structure and neighbourhood information among nodes to effectively learn the features of irregular data, and has been widely used in the field of HSIC. Qin et al. [33] used a semi-supervised spectral-spatial graph convolution network, which effectively utilizes the spatial neighbourhood information of images. Ding et al. [34] proposed a global consistent graph convolution HSIC model (GCGCN), which achieves global feature smoothing for all in-class samples by combining an adaptive global high-order graph structure with a two-layer network. Wan et al. [35] introduced a multi-scale dynamic graph convolution network, which fully exploits the spatial information using multi-scale graph convolution and obtains a better feature representation. Shi et al. [36] used convolution to extract pixel-level features, combined with the results of super-pixel segmentation to obtain a graph structure with stronger expressiveness, and then combined with GCN and Transformer encoder structure to optimize features and obtain good classification results.

The HSIC models based on graph convolution can effectively improve the classification accuracy, but there is the problem that the weights between nodes cannot be changed in the process of building the graph, which limits the representation capability of the model.

Veličković et al. [37] constructed a graph attention mechanism network model, which improves the representational power of the model in operations by introducing an attention mechanism to assign weights to different nodes in the neighbourhood. Dong et al. [38] constructed a graph structure based on hyperpixels and extracted hyperpixel-level features of the image through a graph attention network, and then achieved classification by weighted fusion with pixel-level features extracted by a CNN based on spectral and spatial attention. Sha et al. [39] applied a graph attention mechanism network to the HSIC and represented the relationship between neighbouring nodes adaptively, but did not fully use the deep spatial-spectral features of HSI when constructing the graph structure. An HSIC model (S2RGANet) using a spatial-spectral residual graph attention mechanism was proposed by Xu et al. [40]. This model effectively improves the classification accuracy by constructing a deep convolutional residual module while introducing a network of graph attention mechanisms to obtain more important spatial information, but the training time of this model is long. Therefore, how to combine the advantages of CNN and graph attention mechanism networks to design an end-to-end network model that can effectively reduce the network training time while improving the model accuracy is a problem that needs to be solved.

To address the above problems, this paper proposes an HSIC model (3D-2D-GAT) based on a combination of 3D-2D hybrid convolution and a graph attention mechanism network. The main contributions can be summarized as follows.

- Combining the respective advantages of 2D convolution and 3D convolution, a feature extraction module based on hybrid convolution is constructed, which can quickly extract the discriminant deep spatial-spectral features of various ground objects in HSI.
- The samples are randomly selected in the global range and the adjacency matrix is constructed using their deep spatial-spectral features, so that the graph structure can better express the spectral and spatial connections between different nodes.
- GAT is used to automatically adjust the weights between different nodes and learn the long-distance spatial relationship between hyperspectral data, which can improve the model's ability to distinguish the intra-class variation and inter-class similarity among different samples.
- A new end-to-end HSIC model is proposed, which utilizes the collaborative work of hybrid convolutional feature extraction module and GAT module to improve classification accuracy. Experimental analysis is performed on three HSI datasets, and the experimental results confirm the superiority of our method.

2 Proposed Method

2.1 Model Overview

The overall flow of our proposed method is shown in Fig. 1. The method consists of three main parts: feature extraction based on hybrid convolution, graph structure construction using K-Nearest Neighbour (KNN), and classification by introducing Graph Attention Network (GAT). Firstly, the spatial neighbourhood information of the original HSI is extracted and deep spatial-spectral features are extracted by constructing a hybrid convolutional network. Then, KNN is adopted to construct a graph structure for the extracted deep spatial-spectral features. Finally, the features of the graph nodes are extracted using GAT and classified by a softmax classifier to obtain the classification results of the image.

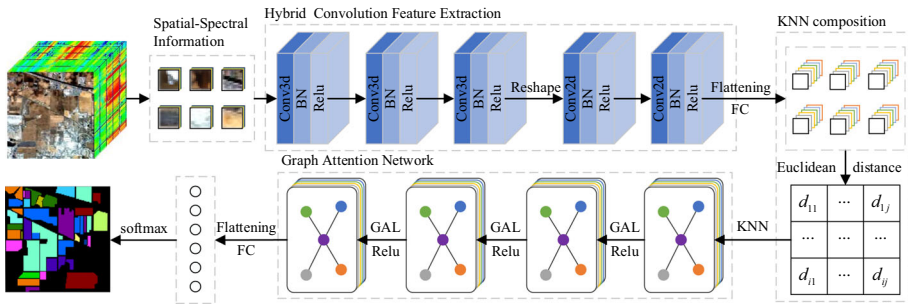


Fig. 1 Overall flow of the proposed method

2.2 Hybrid Convolutional Feature Extraction Network

2D-CNN uses two-dimensional convolutional kernels to sequentially extract local features of the HSI by window sliding, but this approach can only perform the extraction of spatial features and cannot obtain the spectral features of images. 3D-CNN directly uses the original image as the network input without complicated preprocessing and can extract both spatial and spectral features, but has the problem of high computational complexity. To combine the advantages of the above two methods, a hybrid feature convolutional extraction network is proposed in this paper as shown in Fig. 1. It mainly includes three 3D convolutional layers, two 2D convolutional layers, and one fully connected layer.

The input data blocks are first subjected to 3D convolutional operations to simultaneously learn the joint spatial-spectral features of HSI. The parameters of the first 3D convolutional layer is set to (8, (3, 3, 3)), the second is set to (16, (3, 3, 3)), and the third is set to (32, (3, 3, 3)). The step size of all three 3D convolution layers is set to (1, 1, and 1). Then, the extracted features after 3D convolution is calculated by Reshape function. However, with the increase of the depth of the network, there will be a gradual loss of detailed features and an increase in the amount of computation. Therefore, to further enhance the local spatial feature of HSI, 2D convolution is used to extract spatial feature based on 3D convolution. The parameters of the first 2D convolutional layer is set to (64, (3, 3)), and the second is set to (128, (3, 3)). The step size of all two 2D convolution layers is set to (1, 1). Finally, the features are fed into the fully connected network layer to obtain a set of feature maps stacked by 256 channel features.

2.3 Construction of Graph Structure Based on KNN

GAT is a multilayer neural network that can only process data with a graph structure. The feature graph is represented as X . Each pixel location on X is defined as a node, and its feature vector along the channel direction is the initial feature of the node, denoted as $X \in R^{N \times C}$, where N is the number of nodes and C is the feature dimension.

The Euclidean distance d_{ij} between any two nodes can be expressed as:

$$d_{ij} = \left(\sum_{c=1}^C |x_{ic} - x_{jc}|^2 \right)^{\frac{1}{2}} \tag{1}$$

The distance matrix $D = [d_{ij}] \in R^{N \times N}$ is obtained according to the above equation. Then, KNN is used to arrange D in order from largest to smallest, from which the K nearest nodes to each node are selected as its neighbour nodes. The edges are established between these nodes, from which the adjacency matrix $A \in R^{N \times N}$ of the graph can be represented as:

$$A_{ij} = \begin{cases} 1, & X_i \in KNN(X_j) \text{ or } X_j \in KNN(X_i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

2.4 Graph Attention Network

GAT [36–38] is a variant of Graph Neural Network (GNN) and mainly consists of a graph attention layer (GAL). GAT iteratively updates the representation of each node by aggregating the representations of neighbouring nodes using a multi-head attention network mechanism, thus enabling the adaptive assignment of weights to different neighbouring nodes.

The input to GAL is a set of vector $V = (v_1, v_2, \dots, v_N)$, $v_i \in R^F$ where N represents the number of nodes and F is the feature dimension of each vector. The output of GAL is a new set of vector $V' = (v'_1, v'_2, \dots, v'_N)$, $v'_i \in R^{F'}$ where F' represents a feature dimension different from F .

To convert the input features into higher-level features with sufficient expressiveness, the features of each node are parameterized using a weight matrix. The computation of the attention coefficient between node i and node j is expressed as [41]:

$$e_{ij} = \text{LeakyReLU} \left[(\vec{a})^T (W v_i \parallel W v_j) \right] \tag{3}$$

where $W \in R^{F' \times F}$ is a weight matrix and $\vec{a} \in R^{2F'}$ is a single-layer feed-forward neural network. $(\cdot)^T$ denotes the transposition operation and \parallel is the concatenation operation. LeakyReLU is a nonlinear activation function to express the attention coefficients between node i and node j . The softmax function is used to normalize the attention coefficient e_{ij} , and the final attention coefficient α_{ij} between node i and node j can be computed as:

$$\alpha_{ij} = \text{softmax}(e_{ij}) \tag{4}$$

Next, the aggregated feature v' of node j is calculated based on the parameterized features and attention coefficient, and the multi-head attention mechanism allows the model to learn more stable features. The results are computed independently H times and the results of each computation are stitched together as the final aggregated features of node i , which is calculated by:

$$v'_i = \parallel_{h=1}^H \sigma \left(\sum_{j \in N_i} \alpha_{ij}^h W^h v_j \right) \tag{5}$$

where $\sigma(\cdot)$ represents the ReLU activation function [42, 43]. α_{ij}^h is the normalized attention coefficient calculated from the h th attention mechanism and W^h is the corresponding weight matrix.

To maintain high computational efficiency, the number of graph attention layers is set to 3. The overall computational procedure for GAT in this paper is:

$$V_{out} = GAL(GAL(GAL(V_{in}, A), A), A) \tag{6}$$

where V_{in} and V_{out} denote the input and output of GAT, respectively. A is the adjacency matrix of the graph constructed using KNN.

2.5 The Classification Module

The new node features extracted by GAT are input to the fully connected layer to obtain global features F_i . Then, they are fed into softmax for classification, which can be represented as:

$$y = \text{softmax}(W_y F_i + b_y) \tag{7}$$

where W_y is the vector of parameters that can be learned. We use the cross-entropy loss to calculate the loss between the predicted value of the model and the label, which can be expressed as:

$$Loss = -\frac{1}{N} \times \sum_{i=1}^N (y_i \log(y'_i) + (1 - y_i) \log(1 - y'_i)) \tag{8}$$

where N denotes the number of samples. y_i represents the predicted outcome and y'_i represents the true value (Figs. 2, 3).

The overall process of GAT in HSI classification is shown in detail in Algorithm 1 as follows

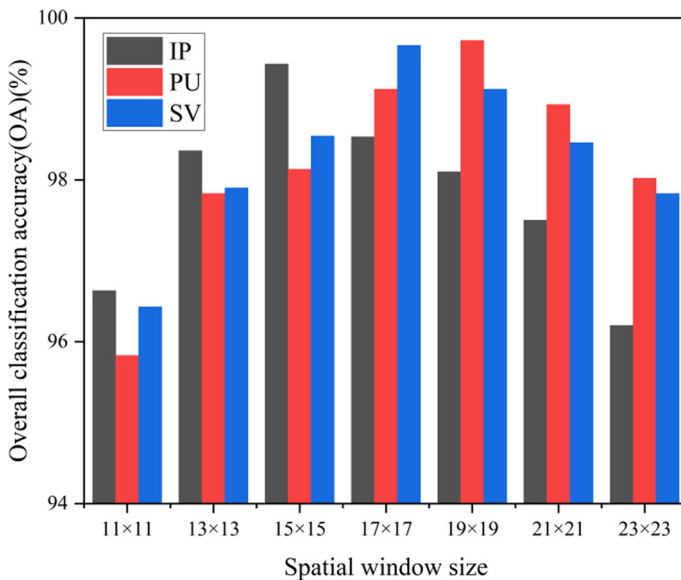


Fig. 2 Classification results of different spatial window size in IP, PU and SV datasets

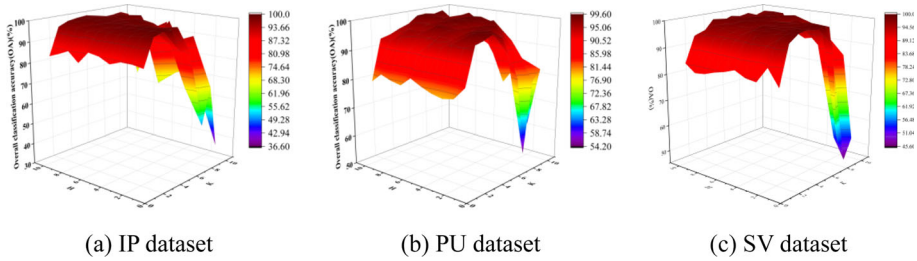


Fig. 3 Classification results of different K and H

Algorithm 1 GAT in HSI classification

Input: Classified HSI X ; heads of the multi-head attention mechanism H ; number of neighbor nodes in the KNN K .

Output: Pixel label

- 1: The spatial-spectral features are extracted by hybrid convolutional feature extraction network;
- 2: The Euclidean distance d_{ij} between any two nodes is computed by Eq. (1) and the adjacency matrix A is constructed by Eq. (2).
- 3: For $h = 1$ to H do
- 4: The attention coefficient of node i and node j is calculated by Eq. (3) and Eq. (4);
- 5: End
- 6: The multi-head attention features are fused by Eq. (5);
- 7: The output feature is calculated by Eq. (6);
- 8: The loss is calculated by Eq. (7) and the weight matrices are updated by Adam gradient descent;
- 9: Predict label for each pixel.

3 Experiment and Result Analysis

3.1 Dataset Description

We used three HSI datasets of Indian Pines (IP), Pavia University (PU) and Salinas Valley (SV) for performance evaluation.

- IP dataset was acquired from the airborne visible/Infrared Imaging Spectrometer (AVIRIS) in Indiana, USA. The image size is 145×145 pixels and the spectral coverage is 400 to 2500 nm, which contains 10,249 pixels and a spatial resolution of 20 m. After removing the bands affected by noise, the remaining 200 bands can be used for classification. The IP dataset has 10,249 samples with 16 classes. The false-colour composite image (30, 60, and 90) and the corresponding ground-truth map are demonstrated in Fig. 4a and b.
- PU dataset was acquired by the Reflection Optics System Imaging Spectrometer (ROSIS) at the University of Pavia. The image size is 610×340 pixels and the spectral coverage is 430 to 860 nm, which contains 42,776 pixels and a spatial resolution of 1.3 m. After removing the bands affected by noise, the remaining 103 bands can be used for classification. The PU dataset has 42,776 samples and 9 classes. The false-colour composite image (20, 60, and 80) and the corresponding ground-truth map are demonstrated in Fig. 5a and b.
- SV dataset was acquired by AVIRIS sensors on the Sarinas Valley, California. The image size is 512×217 pixels and the spectral coverage is 400 to 2500 nm, which contains 54,129 pixels and a spatial resolution of 3.7 m. After removing the bands affected by noise, the remaining 204 bands can be used for classification. The SV dataset has 54,129 samples and 16 classes. The false-colour composite image (30, 60, and 120) and the corresponding ground-truth map are demonstrated in Fig. 6a and b.

Our experimental environment is built under Windows 10 using the Python language and the open-source deep learning framework TensorFlow. The hardware environment is an Intel Core i9-12900KF, 32 GB RAM, and an Nvidia GeForce GTX3060 Ti graphics card.

We use Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient as classification accuracy evaluation metrics to measure the performance of the classification. The OA, AA and Kappa coefficient can be defined as [44–46]:

$$OA = \frac{\sum p_i}{\sum t_i} \quad (9)$$

$$AA = \frac{\sum \frac{p_i}{t_i}}{N} \quad (10)$$

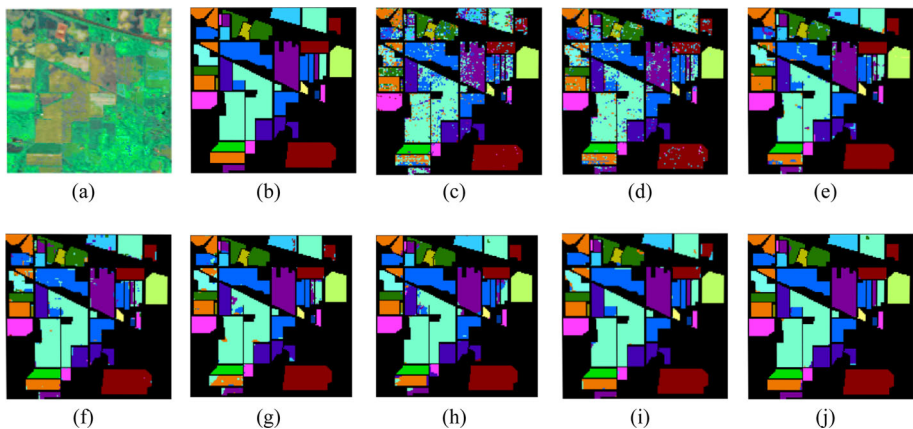


Fig. 4 Classification maps for IP dataset. **a** False-colour composite image (30, 60, 90). **b** Ground truth. **c** 2D-CNN. **d** 3D-CNN. **e** HybridSN. **f** GCGCN. **g** G2T. **h** WFCG. **i** S2RGANet. **j** 3D-2D-GAT

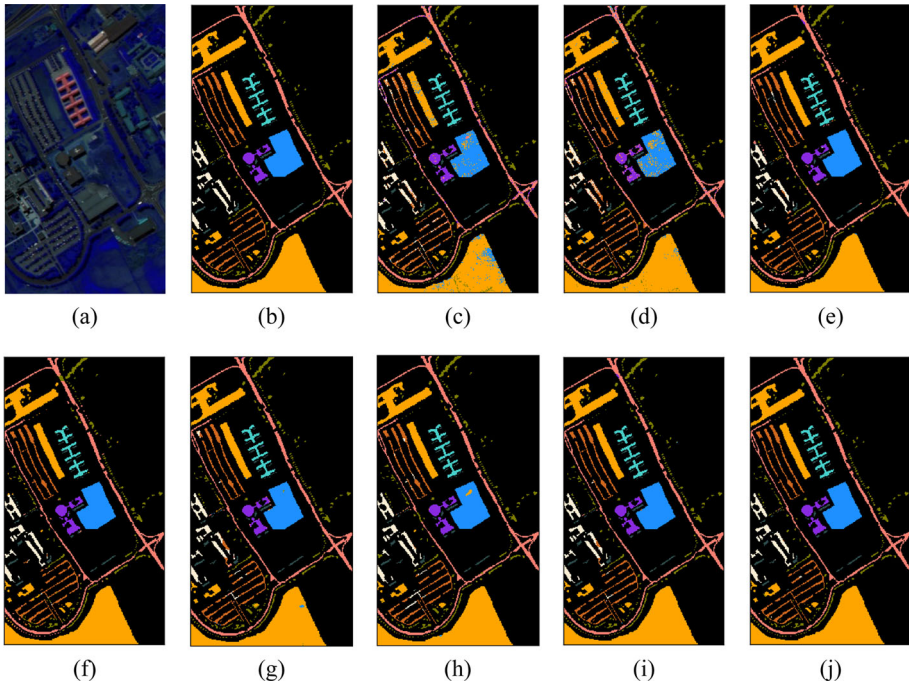


Fig. 5 Classification maps for PU dataset. **a** False-colour composite image (20, 60, 80). **b** Ground truth. **c** 2D-CNN. **d** 3D-CNN. **e** HybridSN. **f** GCGCN. **g** G2T. **h** WFCG. **i** S2RGANet. **j** 3D-2D-GAT

$$\text{Kappa} = \frac{\text{OA} - \frac{\sum p_i \times t_i}{n}}{1 - \frac{\sum p_i \times t_i}{n}} \tag{11}$$

where p_i is the number of correctly classified samples of the i th class, t_i is the number of samples of the i th class in the ground-truth data, and N is the number of classes.

To verify the classification effectiveness of the proposed method, 5% samples of each class of ground objects were randomly selected as the training set in the IP dataset, and the remaining samples were used as the test set. 1% samples of each class of ground objects were randomly selected as the training set in the UP and SV dataset, and the remaining samples were used as the test set. Samples were randomly selected 10 times for all experiments and the average results were obtained.

4 Parameters Setting

The proposed model uses Adam [25] as the optimizer in the training process. The training epoch is set to 300, the batch size is set to 16, and the learning rate is 0.0001. The parameters of the model mainly include the spatial neighbourhood window size, the number of neighbour nodes in KNN and the number of heads of the multi-headed attention mechanism in GAT, and we have experimented and analysed the effects of them as follows.

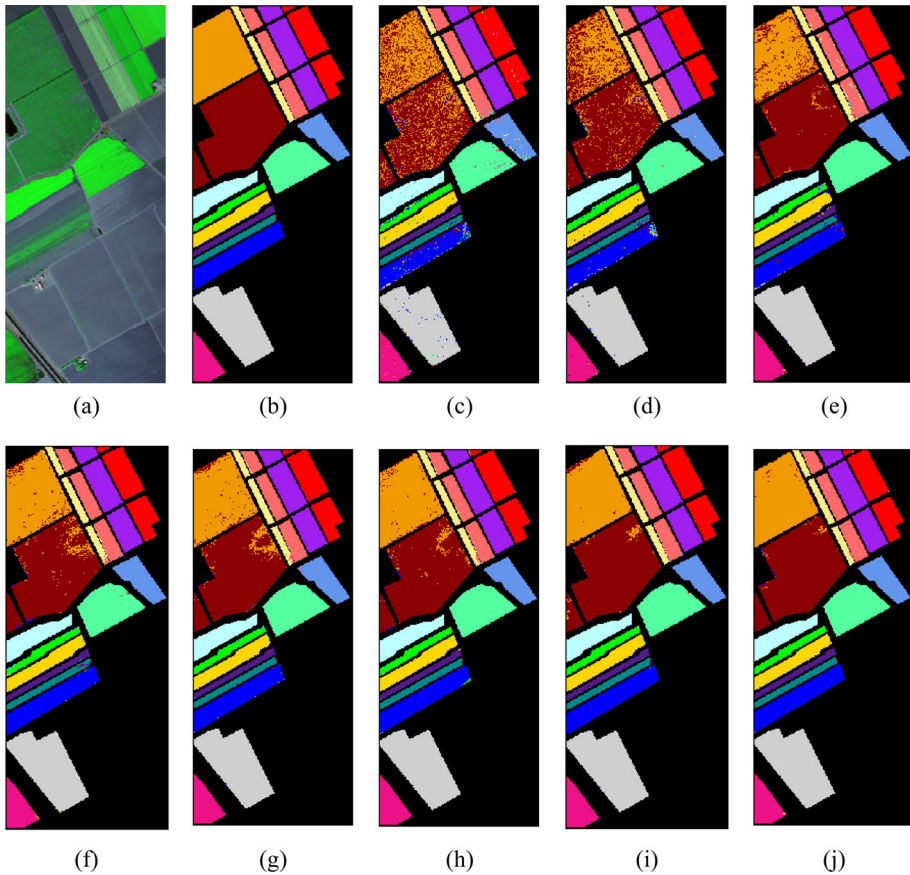


Fig. 6 Classification maps for SV dataset. **a** False-colour composite image (30,60,120). **b** Ground truth. **c** 2D-CNN. **d** 3D-CNN. **e** HybridSN. **f** GCGCN. **g** G2T. **h** WFCG. **i** S2RGANet. **j** 3D-2D-GAT

4.1 Influence of Spatial Window Size

The difference in spatial neighbourhood information contained in different window sizes affects the classification results of the model, so experiments are needed to find the optimal window size on IP, PU and SV datasets. The experimental results are shown in Fig. 2. From the figure, it can be seen that as the window size increases, the spatial information that the model can utilize gradually increases and the classification accuracy also increases. And when the window size is too large, the classification accuracy of the model appears to decrease instead due to the increase of redundant spatial information. Therefore, the input size of the IP data set is set to 15×15 , and the input window size of the PU data set is set to 19×19 , and the input window size of the SV data set is set to 17×17 .

4.1.1 Influence of Different K and H

During the training of the model, the number of neighbour nodes in the KNN algorithm and the number of heads of the multi-headed attention mechanism in the graph attention

mechanism determine the learning efficiency of the model. Therefore, we investigated the effects of them on the model performance on the three datasets.

The optimal values of the number of neighbour nodes K and the number of attention heads H are explored experimentally between 1 and 10 on the IP, PU and SV datasets, and the experimental results are shown in Fig. 3.

From Fig. 3, it can be seen that the experimental results are poor when the value of K is taken low, and the performance of the model is improved as the value of K increases, while too high a value of K will cause the graph structure to fail to accurately express the spatial connections between samples, thus affecting the accuracy of the model. We observe that the change in the number of attention heads H for the same K value does not have a significant effect on the performance of the model, but the model is more likely to converge with higher accuracy when H takes smaller values within a limited number of training batches. Therefore, based on the experimental results, K is set to 4 and H is set to 2 for the IP dataset. For the PU dataset, K is set to 6 and H is set to 2. For the SV dataset, K is set to 6 and H is set to 1.

4.2 Ablation Experiments and Analysis

In order to verify the effectiveness of the hybrid convolutional structure and GAT in the proposed model, the whole network model is divided into three components: 3D convolutional layer, 2D convolutional layer and GAT, and the ablation study are put forward on the three datasets. The experimental results of the ablation experiments are shown in Table 1. As can be seen from the table, the OA obtained on the IP dataset using only 3DCNN is 85.30%, the OA improves by 7.93% when using the hybrid convolutional structure, and by 13.89% when using both the hybrid convolutional structure and GAT. The OA on PU dataset is 89.93% when using only 3DCNN, improved by 6.74% when using the hybrid convolutional structure, and improved by 9.80% when using both the hybrid convolutional structure and GAT. The OA on the SV dataset is 88.67% when using only 3DCNN, the OA is improved by 6.68% when using the hybrid convolutional structure and the OA is improved by 10.76% when both the hybrid convolutional structure and GAT are used. The above experimental results show that using hybrid convolutional structure and GAT can effectively improve the classification accuracy of the model.

Table 1 Ablation experiments results on IP, PU and SV dataset

	3DCNN	2DCNN	GAT	OA(%)
IP	✓			85.30
	✓	✓		94.23
PU	✓	✓	✓	99.19
	✓			89.93
SV	✓	✓		96.67
	✓	✓	✓	99.73
	✓			88.67
	✓	✓		95.35
	✓	✓	✓	99.43

4.3 Comparative Experiments and Analysis

To verify the effectiveness of our proposed method, 2D-CNN [28], 3D-CNN [30], HybridSN [32], GCGCN [34], G2T [36], WFCG [38] and S2RGANet [39] were selected for comparison experiments. To ensure the fairness of the experiments, the comparison methods were configured according to the optimal parameters, and all methods used the same number of samples for both the training and testing sets. Tables 2, 3 and 4 lists the comparison of classification results of different methods on the IP, PU and SV datasets. From these tables, the following conclusions can be drawn:

1. 3D-2D-GAT has the best classification results compared with other comparison methods. Compared to 2D-CNN, 3D-CNN, HybridSN, GCN, G2T, WFCG and S2RGANet, On the IP dataset, the OA of 3D-2D-GAT is improved by 19.3, 13.65, 5.87, 4.06, 3.01, 2.8, and 0.89%, respectively. AA by 19.08, 12.14, 6.75, 3.26, 1.92, 2.03, and 0.62%, respectively, and Kappa improved by 22.07, 15.57, 6.69, 4.63, 3.43, 3.19, and 1.01%, respectively; On the PU dataset, the OA of 3D-2D-GAT improved by 9.04, 7.42, 3.87, 2.35, 1.91, 1.62, and 0.57%, respectively. AA improved by 12.03, 8.36, 5.51, 2.95, 2.83, 2.42, and 1.22%, respectively, and Kappa improved by 11.96, 9.89, 5.13, 3.13, 2.55, 2.16, and 0.78%, respectively. On the SV dataset, the OA of 3D-2D-GAT improved by 13.09, 9.88, 3.93, 2.68, 2.17, 1.58, and 0.65%, respectively. AA improved by 8.1, 5.4, 2.54, 2.28, 1.34, 1.28, and 0.44%, respectively, and Kappa improved by 14.58, 11.05, 4.39, 2.99, 2.42, 1.76, and 0.73%, respectively.
2. Compared with the methods using only 2D-CNN or 3D-CNN, HybridSN shows a greater improvement in classification accuracy on all datasets, indicating that the network can more adequately extract the spatial-spectral features in HSI when it employs both 3D convolution and 2D convolution.
3. GCGCN, WFCG, G2T, S2RGANet and 3D-2D-GAT all use graph structures for feature analysis, and their classification accuracies are higher than those of several other methods using ordinary convolution. This is because the use of graph structure can better learn the intra-class variation and inter-class similarity in hyperspectral data.
4. Compared with GCGCN, the classification accuracies of S2RGANet and 3D-2D-GAT using the graph attention network are significantly improved, indicating that the graph attention network has the feature of dynamically changing the weights between nodes, which is conducive to improving the expressiveness of the networks.
5. G2T adopts a new Graph Guided Transformer structure, but its accuracy is still not as good as that of S2RGANet and 3D-2D-GAT which use graph attention network. Therefore, for graph structure learning, graph attention network is still a structure with strong learning ability.
6. Compared to S2RGANet and 3D-2D-GAT, which use pixels to construct the graph structure, GCGCN, WFCG and G2T all use hyperpixels to construct the graph, which reduces the complexity of the image processing, but is prone to lose the details, which is not conducive to the improvement of the accuracy of the model, and therefore the classification results of these models are poor.
7. S2RGANet introduces only spectral features to construct the graph structure, ignoring important spatial features, and the 3D-2D-GAT model employs deep spatial-spectral features extracted by a hybrid convolutional network to construct the graph, making full use of spatial information and effectively improving the classification accuracy.

Figures 4, 5 and 6 show the classification results of different methods on the three datasets. As can be seen from the figures, the classification results of different methods on the IP, PU

Table 2 Comparison of classification results of different methods on IP dataset (%)

Class	2D-CNN	3D-CNN	HybridSN	GCGCN	G2T	WFCG	S2RGANet	3D-2D-GAT
1	89.47	93.33	75.47	100	100	100	100	100
2	70.19	78.02	89.48	93.77	97.36	93.39	97.91	99.61
3	68.38	80.08	90.69	93.73	93.47	97.87	94.83	98.65
4	59.63	63.49	77.14	92.81	93.36	98.15	98.93	100
5	92.11	90.34	97.64	97.12	99.77	98.52	100	99.32
6	88.42	94.59	94	97.58	99.69	89.67	99.55	97.32
7	91.67	100	88.46	95.83	100	100	100	100
8	93.07	95.74	99.76	100	98.62	100	99.31	99.77
9	41.67	83.33	100	100	100	100	100	100
10	76.04	82.36	92.32	93.87	95.7	94.43	99.31	97.87
11	77.32	85.39	93.47	94.8	93.25	97.09	98.34	99.91
12	79.42	86.56	90.13	88.38	96.69	98.27	99.42	97.76
13	93.68	98.35	100	100	100	99.46	100	100
14	93.07	94.65	99.48	99.21	98.52	99.13	99.46	100
15	76	70.06	92.75	89.53	94.02	97.18	91.87	98.58
16	93.24	98.31	100	100	97.65	93.1	100	100
OA(%)	79.89	85.54	93.32	95.13	96.18	96.39	98.3	99.19
AA(%)	80.22	87.16	92.55	96.04	97.38	97.27	98.68	99.3
Kappat(%)	77	83.5	92.38	94.44	95.64	95.88	98.06	99.07

Table 3 Comparison of classification results of different methods on PU dataset (%)

Class	2D-CNN	3D-CNN	HybridSN	GCGCN	G2T	WFCG	S2RGANet	3D-2D-GAT
1	95.99	93.56	90.73	95.56	98.24	99.49	99.75	99.72
2	97.01	94.53	99.4	98.42	99.51	99.11	99.92	99.99
3	76.21	80.73	94.02	97.47	90.39	86.93	95.83	98.59
4	89.25	96.01	89.3	91.6	93.91	92.63	98.54	98.13
5	98.76	98.16	98.17	99.53	100	99.29	100	100
6	78.84	91.85	99.8	100	98.68	99.75	99.89	100
7	65.1	84.24	97.18	97.35	99.25	100	96.17	100
8	86.76	81.94	88.28	96.18	93.38	98.78	95.92	99.77
9	100	100	89.75	93.6	97.37	98.5	99.22	100
OA(%)	90.69	92.31	95.86	97.38	97.82	98.11	99.16	99.73
AA(%)	87.55	91.22	94.07	96.63	96.75	97.16	98.36	99.58
Kappat(%)	87.68	89.75	94.51	96.51	97.09	97.48	98.86	99.64

Table 4 Comparison of classification results of different methods on SV dataset (%)

Class	2D-CNN	3D-CNN	HybridSN	GCGCN	G2T	WFCG	S2RGANet	3D-2D-GAT
1	98.51	98.28	100	100	100	100	100	100
2	96.32	98.19	99.29	99.64	99.91	99.73	100	99.85
3	93.42	92.18	95.02	97.99	98.53	99.21	98.76	99.66
4	98.31	97.85	88.41	95.54	96.58	94.75	97	98.05
5	96.32	98.69	97.91	96.86	97.47	97.85	99.32	98.84
6	99.21	99.16	98.63	100.00	99.97	99.92	99.97	99.97
7	97.45	99.13	96.9	99.19	99.19	99.56	99.78	99.78
8	74.11	73.99	88.06	95.25	95.24	96.65	97.75	99.17
9	98.25	98.39	99.46	99.93	99.86	99.87	99.98	99.93
10	85.78	94.58	96.89	97.38	98.57	98.56	99.45	99.29
11	86.44	94.92	97.8	95.62	97.24	95.59	98.83	99.37
12	97.13	96.96	98.12	98.14	99.00	98.78	99.47	99.88
13	95.96	91.23	99.74	94.94	100.00	99.12	99.88	100
14	93.7	97.69	98.81	95.74	98.63	98.43	98.64	99.89
15	60.47	75.4	96.56	90.28	91.31	93.77	96.51	98.68
16	91.31	99.36	100.00	99.31	99.44	100.00	100.00	100.00
OA(%)	86.34	89.55	95.5	96.75	97.26	97.85	98.78	99.43
AA(%)	91.42	94.12	96.98	97.24	98.18	98.24	99.08	99.52
Kappa(%)	84.79	88.32	94.98	96.38	96.95	97.61	98.64	99.37

and SV dataset demonstrate that the classification results of 3D-CNN and 2D-CNN have serious salt and pepper phenomenon, and the overall classification effect is poor, and the difference with the ground truth map is relatively large. HybridSN and GCGCN have limited feature learning ability, and many misclassified samples in the classification result maps affect the overall effect. WFCG and S2RGANet have good feature learning ability by combining CNN and GAT, which obtained relatively smooth classification result maps. 3D-2D-GAT has the least number of misclassified features in the classification result map, and the overall classification effect is smoother with only a few noise points, which is closer to the ground truth map.

4.4 Computational Efficiency and Time Consumption

To compare the computational efficiency between different methods, experiments were performed using IP, PU and SV datasets, and the results are shown in Table 5.

Through the statistics of time consumption, it can be found that 3D-CNN requires longer training time than the 2D-CNN, which is due to the fact that 3D-CNN has many parameters. While the structure adopted by HybridSN combines 3D-CNN and 2D-CNN, which can effectively reduce the parameters of the network, and improve the computational efficiency of the network. Compared with the convolution-based models, GCGCN does not use the convolution layers, so it has an advantage in running speed. S2RGANet uses the spectral features extracted by 3D-CNN and leaves the task of extracting spatial features to GAT, which increases the difficulty of model operations. Since WFCG and G2T use the hyperpixel to construct graphs, they are computationally more efficient than S2RGANet and 3D-2D-GAT, which construct graphs based on pixels. In contrast, 3D-2D-GAT first uses 3D-CNN and 2D-CNN to extract the deep spatial-spectral features and then performs the subsequent composition and operations, which reduces the computational cost of the graph attention module and improves the operation speed of the model. Although the running time of 3D-2D-GAT is slightly higher than some comparison methods, it is still a competitive model considering its stable performance and fast test time.

4.5 Influence of Small Samples

To verify the effect of changing the number of training samples on the performance of different classification methods under small sample conditions, experiments were conducted using the IP, PU and SV datasets. 5, 10, 15, 20 and 25 samples were randomly selected from each category of ground objects as training samples, and the experimental results are shown in Fig. 7. As it can be observed from the changing trend of the OAs of different methods under different training sets, increasing the number of training samples has a relatively positive effect on OA of all classification methods. The proposed 3D-2D-GAT method has more robustness and adaptability with better classification accuracy than the six compared methods with limited training samples.

5 Conclusion

In this paper, a new end-to-end HSIC method is proposed to improve the performance of the HSI classification model by introducing three key techniques, namely, deep spatial

Table 5 Comparison of computational efficiency of different methods on IP, PU and SV datasets

dataset	Time (s)	2D-CNN	3D-CNN	HybridSN	GCGCN	G2T	WFCG	S2RGANet	3D-2D-GAT
IP	Train	15.7	51.98	42.10	12.06	24.36	37.23	80.85	48.53
	Test	1.84	6.09	3.12	2.52	2.10	2.76	6.92	4.87
PU	Train	25.11	117.13	83.65	20.30	59.73	94.53	163.36	103.20
	Test	3.59	19.72	12.02	4.08	8.67	13.73	21.86	14.46
SV	Train	29.63	139.58	107.39	23.57	75.48	113.94	225.83	141.63
	Test	4.12	22.83	16.14	4.65	11.23	15.76	31.54	21.21

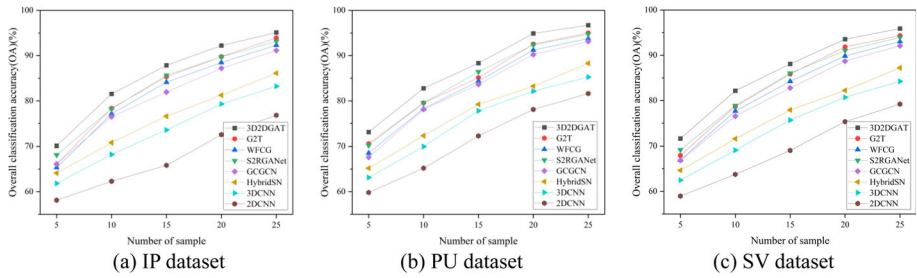


Fig. 7 The classification performance of each method with different number of training samples on IP, PU and SV datasets

and spectral feature extraction, KNN-based graph structure construction and graph attention mechanism. The model uses a 2D–3D hybrid convolutional network to extract deep spatial-spectral features from HSI. In KNN-based graph structure construction, the model utilizes deep spatial-spectral features to improve the representation of the graph structure and performs a transformation between the two modules. We also use GAT to learn long-range spatial links between data and use the extracted spatial features for classification. Our proposed model has some advantages over existing HSIC methods.

In our future research work, we will introduce a semi-supervised method based on this model to further improve the classification accuracy of HSI under small sample conditions.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant 62277016, by the General Scientific Research Project of Zhejiang Education Department under Grant Y202248546, and by Huzhou Science and Technology Bureau under Grants 2023GZ29 and 2023YZ55.

Author contributions Hui Zhang conceptualized and designed the algorithm, prepared the original manuscript draft. Kaiping Tu built the model, verified and analyzed it experimentally, visualized experimental results. Huanhuan Lv assisted with manuscript writing and revisions, supervised the project, provided strategic direction in algorithm development and testing, and conducted a thorough review and final approval of the manuscript prior to submission. Ruiqin Wang contributed to algorithm improvements, and critically revised the manuscript for important intellectual content. All authors reviewed the manuscript.

Declarations

Conflict of interest The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Feng X, Shao Z, Huang X, He L, Lv X, Zhuang Q (2022) Integrating Zhuhai-1 hyperspectral imagery with Sentinel-2 multispectral imagery to improve high-resolution impervious surface area mapping. *IEEE J Sel Top Appl Earth Obs Remote Sens* 15:2410–2424. <https://doi.org/10.1109/JSTARS.2022.3157755>

2. Yang Z, Gong C, Ji T, Hu Y, Li L (2022) Water quality retrieval from ZY1-02D hyperspectral imagery in urban water bodies and comparison with sentinel-2. *Remote Sens* 14(19):5029. <https://doi.org/10.3390/rs14195029>
3. Agilandeewari L, Prabukumar M, Radhesyam V, Phaneendra KLB, Farhan A (2022) Crop classification for agricultural applications in hyperspectral remote sensing images. *Appl Sci* 12(3):1670. <https://doi.org/10.3390/app12031670>
4. Chu X, Miao P, Zhang K, Wei H, Fu H, Liu H, Jiang H, Ma Z (2022) Green Banana maturity classification and quality evaluation using hyperspectral imaging. *Agriculture* 12(4):530. <https://doi.org/10.3390/agriculture12040530>
5. Riefole C, Belmonte A, Quarto R, Quarto F, Ruggieri S, Castrignanò A (2022) Potential of GPR data fusion with hyperspectral data for precision agriculture of the future. *Comput Electron Agric* 199:107109. <https://doi.org/10.1016/j.compag.2022.107109>
6. Li Q, Li J, Li T, Li Z, Zhang P (2023) Spectral-spatial depth-based framework for hyperspectral underwater target detection. *IEEE Trans Geosci Remote Sens* 61:4204615. <https://doi.org/10.1109/TGRS.2023.3275147>
7. Lu H, Bai X, Wang Z, Guo Y, Zhang L, Weng X, Xie J, Liang D, Deng L (2023) Hyperspectral camouflage coating using Palygorskite to simulate water absorption of healthy green leaves. *Mater Sci Semicond Process* 156:107293. <https://doi.org/10.1016/j.mssp.2022.107293>
8. Alopogianni E, Ichimura T, Hamada M, Ishikawa M, Murakami T, Sasaki A, Nakamura K, Kobayashi N, Obi T (2022) Hyperspectral imaging for tumor segmentation on pigmented skin lesions. *J Biomed Opt* 27(10):106007. <https://doi.org/10.1117/1.JBO.27.10.106007>
9. Witteveen M, Sterenborg HJ, van Leeuwen TG, Aalders MC, Ruers TJ, Post AL (2022) Comparison of preprocessing techniques to reduce nontissue-related variations in hyperspectral reflectance imaging. *J Biomed Opt* 27(10):106003. <https://doi.org/10.1117/1.JBO.27.10.106003>
10. Chen W, Zheng X, Lu X (2021) Hyperspectral image super-resolution with self-supervised spectral-spatial residual network. *Remote Sens* 13(7):1260. <https://doi.org/10.3390/rs13071260>
11. Wang Z, Chen B, Lu R, Zhang H, Liu H, Varshney PK (2020) FusionNet: An unsupervised convolutional variational network for hyperspectral and multispectral image fusion. *IEEE Trans Image Process* 29:7565–7577. <https://doi.org/10.1109/TIP.2020.3004261>
12. Karaca AC (2021) Spatial aware probabilistic multi-kernel collaborative representation for hyperspectral image classification using few labelled samples. *Int J Remote Sens* 42(3):839–864. <https://doi.org/10.1080/01431161.2020.1823516>
13. Li L, Ge H, Gao J, Zhang Y, Tong Y, Sun J (2020) A novel geometric mean feature space discriminant analysis method for hyperspectral image feature extraction. *Neural Process Lett* 51(1):515–542. <https://doi.org/10.1007/s11063-019-10101-0>
14. Wang Y, Li T, Chen L, Yu Y, Zhao Y, Zhou J (2021) Tensor-based robust principal component analysis with locality preserving graph and frontal slice sparsity for hyperspectral image classification. *IEEE Trans Geosci Remote Sens*. <https://doi.org/10.1109/TGRS.2021.3093582>
15. Hashemi-Nasab FS, Parastar H (2022) Vis-NIR hyperspectral imaging coupled with independent component analysis for saffron authentication. *Food Chem* 393:133450. <https://doi.org/10.1016/j.foodchem.2022.133450>
16. Lupu D, Necoara I, Garrett JL, Johansen TA (2022) Stochastic higher-order independent component analysis for hyperspectral dimensionality reduction. *IEEE Trans Comput Imaging* 8:1184–1194. <https://doi.org/10.1109/TCI.2022.3230584>
17. Jayaprakash C, Damodaran BB, Viswanathan S, Soman KP (2020) Randomized independent component analysis and linear discriminant analysis dimensionality reduction methods for hyperspectral image classification. *J Appl Remote Sens* 14(3):036507–036507. <https://doi.org/10.1117/1.JRS.14.036507>
18. Li L, Gao J, Ge H, Zhang Y, Yang J (2022) An effective feature extraction approach based on spectral-Gabor space discriminant analysis for hyperspectral image. *Neural Process Lett* 54(2):909–959. <https://doi.org/10.1007/s11063-021-10665-w>
19. Liu G, Wang L, Liu D, Fei L, Yang J (2022) Hyperspectral image classification based on non-parallel support vector machine. *Remote Sens* 14(10):2447. <https://doi.org/10.3390/rs14102447>
20. Qureshi AS, Roos T (2022) Transfer learning with ensembles of deep neural networks for skin cancer detection in imbalanced data sets. *Neural Process Lett*. <https://doi.org/10.1007/s11063-022-11049-4>
21. Huang L, Liu Y, Huang W, Dong Y, Ma H, Wu K, Guo A (2022) Combining random forest and XGBoost methods in detecting early and mid-term winter wheat stripe rust using canopy level hyperspectral measurements. *Agriculture* 12(1):74. <https://doi.org/10.3390/agriculture12010074>
22. Tong F, Zhang Y (2022) Spectral-spatial and cascaded multilayer random forests for tree species classification in airborne hyperspectral images. *IEEE Trans Geosci Remote Sens* 60:1–11. <https://doi.org/10.1109/TGRS.2022.3177935>

23. Park J-J, Park K, Foucher P-Y, Kim T-S, Lee M (2023) Estimation of hazardous and noxious substance (toluene) thickness using hyperspectral remote sensing. *Front Environ Sci* 11:1130585. <https://doi.org/10.3389/fenvs.2023.1130585>
24. Zhu C, Ding J, Zhang Z, Wang J, Wang Z, Chen X, Wang J (2022) SPAD monitoring of saline vegetation based on Gaussian mixture model and UAV hyperspectral image feature classification. *Comput Electron Agric* 200:107236. <https://doi.org/10.1016/j.c-ompag.2022.107236>
25. Mohan A, Venkatesan M (2020) HybridCNN based hyperspectral image classification using multi-scale spatio-spectral features. *Infrared Phys Technol* 108:103326. <https://doi.org/10.1016/j.infrared.2020.103326>
26. Kutluk S, Kayabol K, Akan A (2021) A new CNN training approach with application to hyperspectral image classification. *Digit Signal Process* 113:103016. <https://doi.org/10.1016/j.dsp.2021.103016>
27. Vaddi R, Manoharan P (2020) Hyperspectral image classification using CNN with spectral and spatial features integration. *Infrared Phys Technol* 107:103296. <https://doi.org/10.1016/j.infrared.2020.103296>
28. Yue J, Zhao W, Mao S, Liu H (2015) Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sens Lett* 6(6):468–477. <https://doi.org/10.1080/2150704X.2015.1047045>
29. Chen Y, Jiang H, Li C, Jia X, Ghamisi P (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans Geosci Remote Sens* 54(10):6232–6251. <https://doi.org/10.1109/TGRS.2016.2584107>
30. Zhong Z, Li J, Luo Z, Chapman M (2017) Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans Geosci Remote Sens* 56(2):847–858. <https://doi.org/10.1109/TGRS.2017.2755542>
31. Sun H, Zheng X, Lu X, Wu S (2019) Spectral–spatial attention network for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 58(5):3232–3245. <https://doi.org/10.1109/TGRS.2019.2951160>
32. Roy SK, Krishna G, Dubey SR, Chaudhuri BB (2019) HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 17(2):277–281. <https://doi.org/10.1109/LGRS.2019.2918719>
33. Qin A, Shang Z, Tian J, Wang Y, Zhang T, Tang YY (2018) Spectral–spatial graph convolutional networks for semisupervised hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 16(2):241–245. <https://doi.org/10.1109/LGRS.2018.2869563>
34. Ding Y, Guo Y, Chong Y, Pan S, Feng J (2021) Global consistent graph convolutional network for hyperspectral image classification. *IEEE Trans Instrum Meas* 70:1–16. <https://doi.org/10.1109/TIM.2021.3056750>
35. Wan S, Gong C, Zhong P, Du B, Zhang L, Yang J (2019) Multiscale dynamic graph convolutional network for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 58(5):3162–3177. <https://doi.org/10.1109/TGRS.2019.2949180>
36. Shi C, Liao Q, Li X, Zhao L, Li W (2023) Graph Guided Transformer: An Image-Based Global Learning Framework for Hyperspectral Image Classification. *IEEE Geosci Remote Sens Lett* 20:5512505. <https://doi.org/10.1109/LGRS.2023.3316732>
37. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. *stat* 1050(20):1048550. <https://doi.org/10.48550/arXiv.1710.10903>
38. Dong Y, Liu Q, Du B, Zhang L (2022) Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification. *IEEE Trans Image Process* 31:1559–1572. <https://doi.org/10.1109/TIP.2022.3144017>
39. Sha A, Wang B, Wu X, Zhang L (2020) Semisupervised classification for hyperspectral images using graph attention networks. *IEEE Geosci Remote Sens Lett* 18(1):157–161. <https://doi.org/10.1109/LGRS.2020.2966239>
40. Xu K, Zhao Y, Zhang L, Gao C, Huang H (2021) Spectral–spatial residual graph attention network for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 19:1–5. <https://doi.org/10.1109/LGRS.2021.3111985>
41. Niu D, Yu M, Sun L, Gao T, Wang K (2022) Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Appl Energy* 313:118801. <https://doi.org/10.1016/j.apenergy.2022.118801>
42. Kim Y, Ohn I, Kim D (2021) Fast convergence rates of deep neural networks for classification. *Neural Netw* 138:179–197. <https://doi.org/10.1016/j.neunet.2021.02.012>
43. Bodyanskiy Y, Antonenko T (2021) Deep neural network based on generalized neo-fuzzy neurons and its learning based on backpropagation. *Artif Intell* 26(1):32–41. <https://doi.org/10.15407/jai2021.01.032>
44. Zhao Q, Jia S, Li Y (2021) Hyperspectral remote sensing image classification based on tighter random projection with minimal intra-class variance algorithm. *Pattern Recognit* 111:107635. <https://doi.org/10.1016/j.patcog.2020.107635>

45. Hamidi M, Safari A, Homayouni S (2021) An auto-encoder based classifier for crop mapping from multitemporal multispectral imagery. *Int J Remote Sens* 42(3):986–1016. <https://doi.org/10.1080/01431161.2020.1820619>
46. Zhao X, Yang Y, Duan F, Zhang M, Jiang G, Yan X, Cao S, Zhao W (2022) Identification of construction and demolition waste based on change detection and deep learning. *Int J Remote Sens* 43(6):2012–2028. <https://doi.org/10.1080/01431161.2022.2054296>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.