HSI Classification Method Based on U-Nets and GCN-CNN in the Background of Artificial Intelligence

Yujin Zou[1,2], Renwang Li[1*]
[1]Faculty of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2]School of Business Intelligence, Zhejiang Institute of Economics and Trade, Hangzhou 310018, China
E-mail: 201810501018@zstu.edu.cn; zjustu@163.com;
*Corresponding author

Keywords: U-shaped network structure, graph convolutional network, multi-scale features, hyperspectral images, image classification

Received: September 1, 2025

Hyperspectral images typically have large-sized features and fuse a large amount of spatial and spectral data, increasing the complexity of feature selection and effective mining. In addition, its dimensional redundancy and feature intersection further exacerbate the interpretability problem of the model, limiting the overall improvement of classification performance. Therefore, this paper proposes an HSI classification model built on U-Nets and a graph convolutional network. This model utilizes multi-scale superpixel segmentation to enhance the flexibility of spatial structure modeling and achieves synchronous extraction of spatial topological relationships between land features from multiple scales through a multi-scale graph convolution architecture. The experiment showed that the proposed model achieved an F1 value of 96.7% on the comprehensive datasets (Indian Pines, Pavia University, Salinas, and GRSS 2013), demonstrating good robustness and generalization ability. Regardless of whether under interference conditions, the average classification entropy and average mutual information between categories of the proposed model were significantly lower than those of comparative models. Under the condition of random loss of some bands, the average classification entropy and average mutual information value between categories of the research model were 0.28 and 0.79, and 0.31 and 0.77 under Gaussian noise interference. The research model has strong discriminative ability in hyperspectral image classification tasks and effectively deals with complex scenes such as noise interference and data loss.

Povzetek: Predlagani model, ki združuje U-Net in grafne konvolucijske mreže, učinkovito izboljša klasifikacijo hiperspektralnih slik ter doseže visoko robustnost, generalizacijo in natančnost tudi ob šumu in izgubi podatkov.

1 Introduction

Hyperspectral Image (HSI) has a spectral coverage range from visible light to near-infrared, with high spectral resolution and strong band continuity. It can reveal land information that cannot be reflected by single-band or multi-band images. However, the high dimensionality of hyperspectral data gives it some special properties that are different from traditional 3D data spaces, making it difficult to process and classify HSI data using conventional methods [1-2]. The spectral correlation of HSI is stronger than the spatial correlation. Therefore, based on the above characteristics, reducing data dimensionality and fusing effective information are necessary for HSI analysis. The advancement of artificial intelligence technology has made Convolutional Neural Networks (CNNs) gradually become the mainstream method for HSI classification due to their excellent nonlinear fitting ability and local Spatial Feature (SF) extraction ability [3-4]. In related research, Ari developed an HSI classification method built on a multi-path CNN and squeeze excitation network. In the 5%WHLK, WHHC, and WHHH training samples, the Overall Accuracy (OA) of this method reached 99.86%, 97.51%, and 97.64% [5]. To simultaneously extract local and global features from HSI, Li et al. designed a parallel dual-branch structure based on CNN and Transformer. In four hyperspectral datasets, the OA of this method reached 99.21%, 99.61%, 92.40%, and 98.17% [6]. Giri et al. proposed an innovative method of using a pre-trained CNN to extract robust SFs from HSI to assist classification. In the Salinas dataset, the OA of this method was 99.12%, the mean precision was 99.40%, and the Kappa coefficient was 0.9901 [7].

U-shaped Network Structure (U-Nets) is a deep learning model specifically designed for medical image segmentation. U-Nets can effectively capture contextual information in images and accurately locate target areas [8-9]. Its powerful feature extraction ability and sensitivity to details provide a new perspective for HSI classification. Deng et al. proposed a U-Nets model that integrates residual structure and a selective convolutional kernel attention mechanism. This model effectively enhanced the SF expression ability while maintaining the integrity of low-resolution and high-resolution spectral information. Meanwhile, it also enhanced the spatial detail representation of the reconstructed HSI, thereby achieving higher quality image super-resolution reconstruction [10]. Subba Reddy et al. put forward an HSI classification method based on U-Nets and the honey badger optimization algorithm. The accuracy of this method was 0.907, the sensitivity was 0.914, and the specificity was 0.904 [11]. In addition, Tang et al. put forth an HSI super-resolution method that combines U-Nets and state space models. This model utilized U-Nets to extract multi-scale SFs and models and predicted time series and spectral dimensions through state space, thereby achieving synergistic improvement of HSI in spatial and spectral resolution [12].

In summary, both CNN and U-Nets models have achieved good results in HSI classification. 3D-CNN models can jointly model in both spectral and spatial dimensions. Their computational complexity is extremely high, especially when dealing with HSI data containing hundreds of bands, which can easily lead to excessive memory consumption and excessively long training time. Meanwhile, 3D convolution kernels are usually fixed in the local receptive domain, making it difficult to effectively capture cross-scale contextual relationships. This leads to a decrease in classification performance in scenarios with high inter-class similarity or blurred spatial boundaries. The traditional Graph Convolutional Network (GCN)-based HSI classification method can capture non-Euclidean spatial relations by using the graph structure. However, its high dependence on the input graph structure makes the model vulnerable to the influence of superpixel partitioning errors and adjacency matrix construction biases, thereby reducing the overall robustness. U-Nets, through their symmetrical Encoder-Decoder Structure (EDS) and skip connections, can enhance the ability to preserve spatial details and to some extent compensate for the insufficient capture of fine-grained information in traditional CNN models. However, U-Nets mainly rely on spatial proximity when modeling the relationships between superpixels, ignoring the non-Euclidean spatial structure of inter-class heterogeneity in HSI, resulting in a decrease in accuracy when dealing with areas with fuzzy boundaries or high inter-class similarity. Therefore, the study proposes a U-Nets model based on Multi-scale Superpixel Segmentation (MSS). On the basis of this model, an HSI classification model based on multi-scale features and GCN-CNN is proposed. By combining the ability of U-Nets to preserve pixel-level spatial details with the modeling ability of GCN for superpixel-level global topological relations, the model can simultaneously capture local texture features and cross-regional spatial semantic relations. Multi-scale scroll integral branches can synchronously extract the spatial topological relationships between ground objects from multiple scales, fully mining the global context information and local texture features in HSIs. Through this new model, the study aims to enhance the robustness and generalization capacity of HSI classification, thereby providing more efficient and accurate technical support for HSI intelligent analysis.

2 Methods and materials

2.1 U-Nets model based on MSS

The U-Nets model offers powerful technical support for HSI classification tasks with its excellent EDS and skip connections. To more effectively integrate spatial structure perception with non-Euclidean spatial modeling capabilities, this study proposes a U-Nets model based on MSS. MSS is a method of layer-by-layer superpixel partitioning of images at different spatial scales, which can extract richer and more hierarchical spatial structural information, thereby providing finer spatial support for subsequent feature learning [13-14]. The overall structure of the proposed model is displayed in Fig.1.

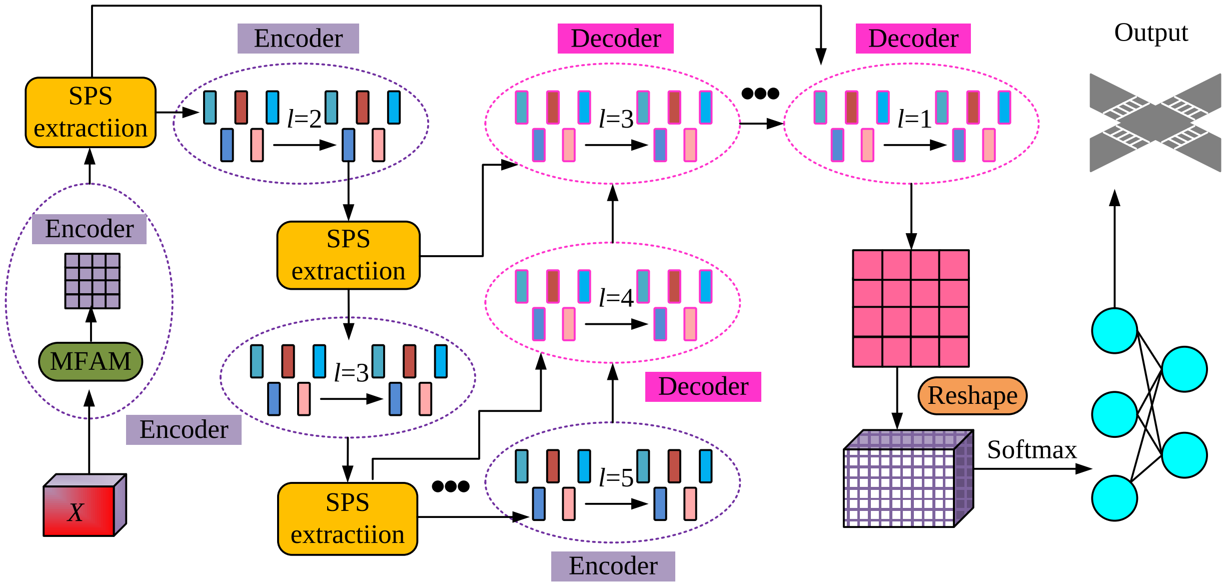


Figure 1: Framework of the U-Nets model based on MSS.

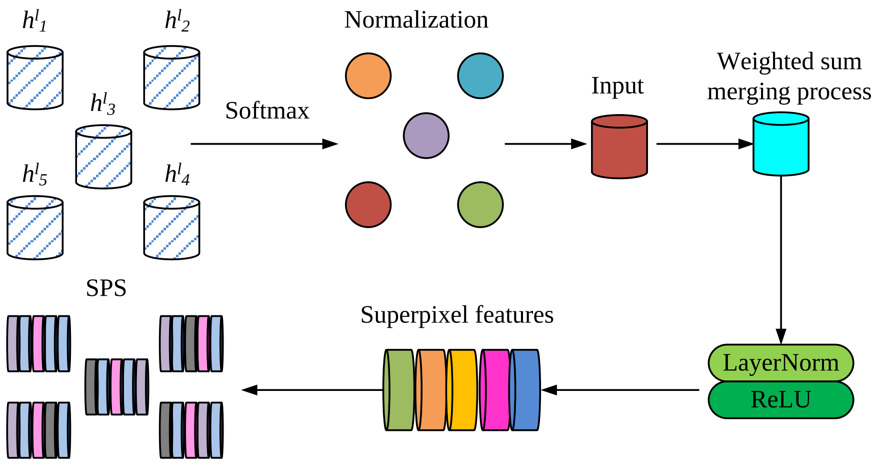


Figure 2: SPS extraction process.

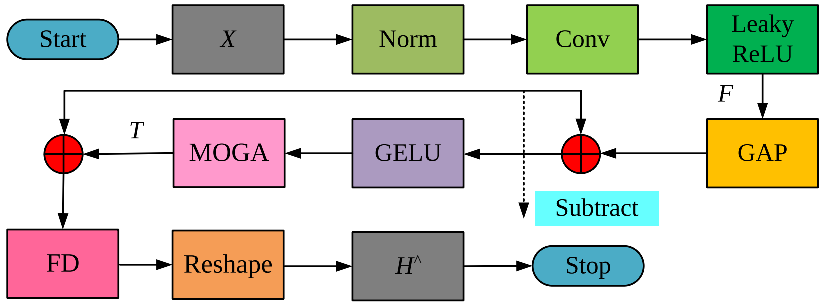


Figure 3: Diagram of the MFAM structure.

In Fig.1, the model includes a Multi-scale Fusion Attention Module (MFAM), a Lightweight Graph Attention Network (LwGAT), and a Superpixel Structure Propagation (SPS). Firstly, the original HSI is input into MFAM, and the response weights of spatial and spectral information are dynamically adjusted through a dual-channel attention mechanism to extract pixel-level fusion features that balance spatial texture and spectral discriminability. Subsequently, this feature is passed as input to the U-Nets encoder section. In each layer of the encoder, this study uses SPS modules to perform superpixel partitioning on image regions and construct a multi-scale image structure. Next, the LwGAT module is utilized to perform graph convolution operations on the constructed graph structure, achieving non-Euclidean space modeling. In the decoder section, the network gradually restores the resolution of the feature map through an upsampling layer and uses skip connections to fuse the features in the encoder with those in the decoder to preserve multi-scale information. After each decoding layer outputs, the feature boundaries and discriminative expressions are further refined through a 5×5 convolution operation. Finally, after the decoding of the last layer is completed, the structural features of the image are mapped back to the original pixel space and input into the Softmax classifier to output a pixel-level classification probability map, achieving accurate determination of the land cover category to which each pixel belongs. The SPS extraction process is shown in Fig.2.

In Fig.2, in traditional graph convolution operations, the features between similar pixels are mainly aggregated through adjacency matrices, and the spectral features of pixels of the same category tend to be consistent [15]. However, HSIs typically contain hundreds of bands, with a significant amount of spectral redundancy and noise, which can easily affect the stability and discriminability of convolution operations [16-17]. Therefore, this study proposes a spectral pixel sequence construction strategy based on superpixels, combined with regional aggregation and an attention mechanism to achieve SPS extraction. The specific process is as follows: Firstly, in the l-layer network, the currently extracted graph structure feature $H^l = \left\{h_1^l, h_2^l, \ldots, h_n^l\right\}$ is matched with the corresponding superpixel structure and divided into several spatially continuous and spectroscopically similar subregions $\left\{S_1^l, S_2^l, \ldots, S_m^l\right\}$. The pixels in each region are regarded as a candidate aggregation unit, and their feature vectors are calculated for similarity with the query vector through dot multiplication to generate attention weights. Subsequently, the attention weights of all candidate pixels are normalized utilizing the Softmax to gain the contribution of each pixel in the current superpixel to its representative features. The specific structure of the MFAM is exhibited in Fig.3.

In Fig.3, MFAM consists of two cascaded submodules, namely Multi-Order Gated Aggregation (MOGA) and Feature Decomposition (FD). Among them, MOGA uses multi-channel parallel Depthwise Separable Convolution (DWConv) to encode multi-scale features. The FD module is responsible for decomposing the fused features, guiding the network to focus on high-order feature interactions, while dynamically suppressing redundant or inefficient information, thereby enhancing the structural and discriminative nature of feature expression. Firstly, the input HSI $X \in R^{h \times w \times c}$ is normalized (Norm) and convolved (Conv) to extract initial pixel level features F . Then, intermediate features are generated using the Leaky ReLU activation function. Subsequently, the feature undergoes Global Average Pooling (GAP) to form channel descriptors, which are then differentiated element by element from the original feature to highlight local spatial variation information. The differential features are converted into an intermediate representation Z through Gaussian Error Linear Unit (GELU) and input into MOGA to achieve interaction and fusion of contextual features at different scales. The calculation formula for F is shown in equation (1).

$$F = \lambda(Conv(Norm(X)) + b) \qquad (1)$$

In equation (1), b is the convolution bias term and λ is the Leaky ReLU. Within MOGA, feature Z is segmented into three sub channels Z_1, Z_2, and Z_3, and local features are extracted through DWConv with different receptive field settings. Subsequently, the three scale features Y_1, Y_2, and Y_3 are concatenated Y, and then transformed using two ordinary convolutions and SiLU function to output the feature response η. This feature is then added element by element to the output T of the previous branch to form the final fused feature representation. Subsequently, the fused features are sent to the FD module for feature decomposition. Finally, the feature is adjusted to a shape compatible with downstream modules of the network through Reshape operation, and output as pixel level feature $\hat{H}$ for subsequent image structure modeling or classification. The formula for channel attention weight α is given by equation (2).

$$\alpha = \sigma(W_2 \cdot \delta(F - F_{gap}) + b_2) \qquad (2)$$

In equation (2), δ is the GELU function, W_2 and b_2 are learnable weights and bias parameters, F_{gap} is the channel global descriptor, and σ represents the Sigmoid function. The calculation of Z is given by equation (3).

$$Z = GELU(F \oplus \tau_s \oplus (F - GAP(F))) \qquad (3)$$

In equation (3), τ_s is a learnable scaling factor, which represents element wise multiplication. The formula for T is shown in equation (4).

$$T = SiLU(Conc(Z) \otimes SiLUConc(Y)) \otimes (\eta \cdot SiLU(Conc(Y))) \qquad (4)$$

The expression for $\hat{H}$ is shown in equation (5).

$$\hat{H} = \mathrm{Re}\,shape(T \oplus F) \qquad (5)$$

2.2 HSI classification model based on multi-scale features and GCN-CNN

The U-Nets model based on MSS can enhance the feature representation ability of the model through MFAM, and improve the graph modeling effect in non-Euclidean space through SPS and LwGAT modules. However, the U-Nets model requires multiple complex graph convolution operations in both the encoder and decoder stages, which can result in significant computational burden and time overhead [18]. To address this issue, this study proposes an HSI classification model based on multi-scale features and GCN-CNN. Fig.4 shows the model’s overall structure.

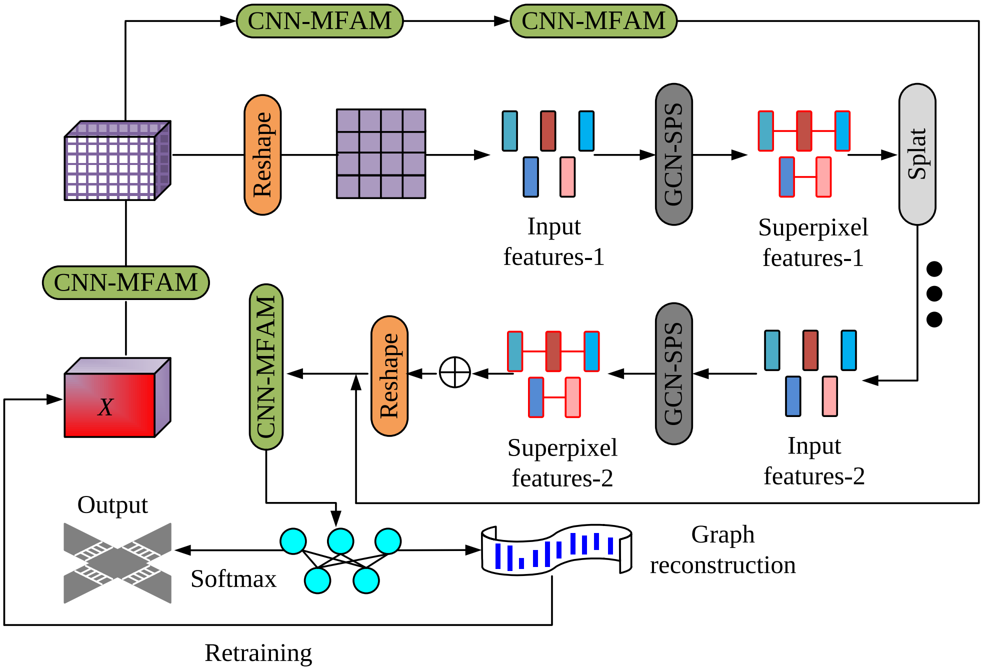


Figure 4: Parallel multi-scale GCN-CNN feature extraction module structure.

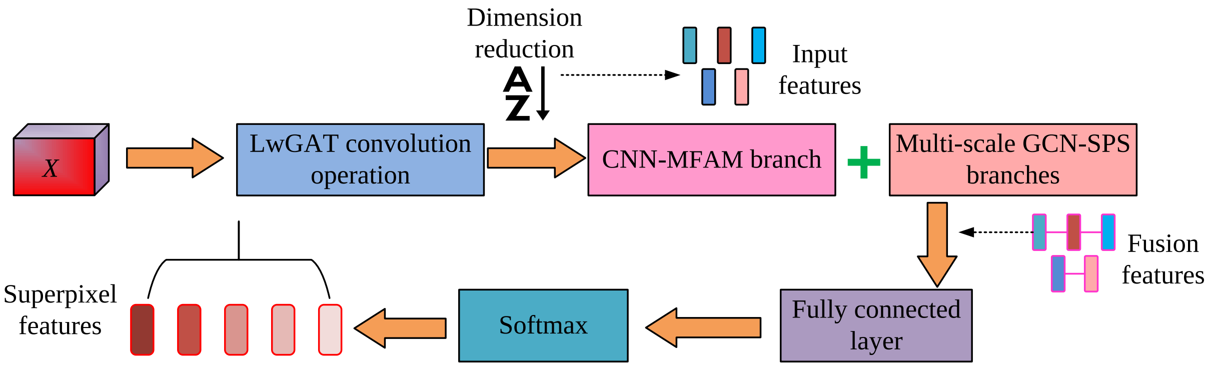


Figure 5: The feature extraction module structure based on the parallel multi-scale GCN-CNN architecture.

In Fig.4, the model mainly consists of a multi-level superpixel GCN-CNN feature extraction module and a graph structure reconstruction module. Among them, the feature extraction module adopts a parallel approach to extract pixel-level features and multi-scale superpixel level features. The graph structure reconstruction module assigns pseudo labels to each superpixel by calculating the feature geometric distance between each superpixel region and the annotated pixels. These pseudo labels, to some extent, reflect the category bias of superpixels, and based on the pseudo label information, further reconstruct the edge weights and connection relationships in the graph structure, thereby obtaining a graph topology structure that is more in line with the true semantic distribution. Finally, the updated graph structure is utilized to retrain the entire network, enhancing the model's ability to distinguish land cover categories in complex scenes. The process of constructing a superpixel map and concatenating multi-scale features is as follows: Firstly, the HSI is input, and superpixel segmentation is performed through Simple Linear Iterative Clustering (SLIC) to obtain N superpixel regions. The node feature vector X_i of each superpixel S_i is defined as the average spectral vector of the pixels within its region, and the calculation formula is shown in equation (6).

$$X_i = \frac{1}{|S_i|_{p\in S_i}} I(p) \qquad (6)$$

In equation (6), $I(p)$ is the spectral vector of pixel p. Subsequently, the adjacency relationship is defined by calculating the spatial centroid and spectral similarity between superpixels, thereby forming a graph structure that can reflect the spatial structure and spectral characteristics. The calculation formula is shown in equation (7).

$$A_{ij} = exp - \frac{\|\mu_i - \mu_j\|^2}{\sigma_s^2} \cdot exp - \frac{\|X_i - X_j\|^2}{\sigma_f^2} \quad (7)$$

In equation (7), A_{ij} represents the adjacency matrix. σ_s and σ_f are the spatial and spectral scale parameters, respectively. X_i and X_j are the spectral features of the superpixel. μ_i and μ_j are the centroid coordinate of the superpixel S_i. Meanwhile, the input image is encoded and decoded through the U-Nets to obtain multi-scale feature maps at different resolutions. On each layer of the feature map, the pixel set corresponding to the superpixel region is mapped as a regional feature, and its semantic information is extracted through average pooling. Then, all the superpixel features are input into the parallel multi-scale GCN-CNN feature extraction module for multi-scale feature concatenation. The flowchart of the parallel multi-scale GCN-CNN feature extraction module based on MSS is shown in Fig.5.

In Fig.5, the module divides the input HSI into 5 layers of superpixel structures with different spatial scales to capture multi-level semantic information from local to global. Firstly, the LwGAT convolution operation is used to perform preliminary dimensionality reduction on the input HSI, to reduce computational costs and preserve key spectral information. Subsequently, the image features are fed into the CNN-MFAM branch and the multi-scale GCN-SPS branch. The former is used to extract local pixel-level features; The latter extracts spatially continuous and spectrally consistent superpixel-level features at multiple segmentation scales, thereby achieving global context modeling. Finally, by fusing features, the confidence information of each pixel is calculated. The confidence level P_i is shown in equation (8).

$$P_i = Soft\max(W_f \cdot F_r + b_f) \qquad (8)$$

In equation (8), W_f and b_f are the weight matrix and bias term, and F_r is the fused pixel feature. The calculation process of GCN is shown in equation (9).

$$H'^{l+1} = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H'^{l}W^{l}) \qquad (9)$$

In equation (9), H'^{l} and H'^{l+1} are the features of nodes in the l-th and $l+1$-th layers of the graph. W^l is the learnable graph convolution weight. $\tilde{A}$ is an adjacency matrix with self-loops. $\tilde{D}$ is the degree matrix corresponding to $\tilde{A}$. The structure of the graph structure reconstruction module that integrates geometric information and label information is shown in Fig.6.

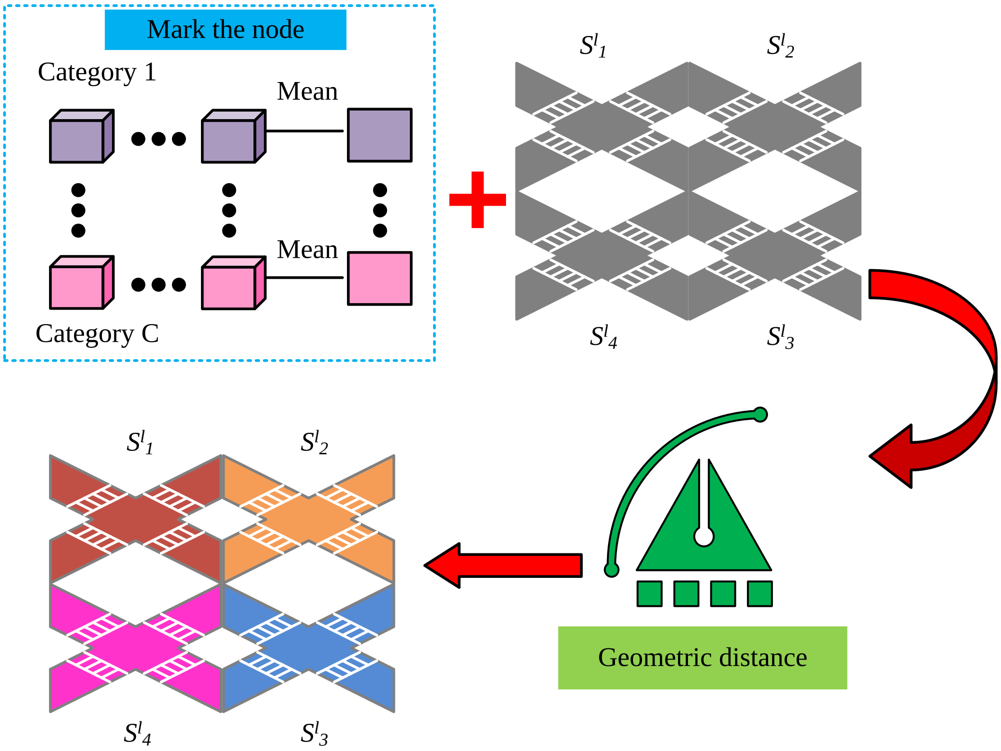


Figure 6: Graph structure reconstruction module structure integrating geometric information and label information.

In Fig.6, the module measures the similarity between two regions by using the information of the marked nodes as an intermediate variable. Firstly, different clustering centers are constructed using the features extracted from the output layer and the matrix. Secondly, a pseudo label is assigned to each superpixel in the l-th layer, enabling weakly supervised guidance of category attributes without the need for complete labeling. After obtaining the pseudo labels, the original graph structure is further reconstructed by combining this information. The new graph structure considers spatial similarity and improves the discriminative power of graph edge connections by introducing geometric distance and pseudo-label consistency as joint constraints. Finally, the reconstructed graph structure is re-injected into the model and subjected to end-to-end retraining using standard cross-entropy loss, effectively improving classification accuracy and model robustness while ensuring the rationality of structural expression. The expression of geometric distance d_{ij} is shown in equation (10).

$$d_{ij} = \left\| F_i - F_j \right\|_2 \tag{10}$$

In equation (10), F_i and F_j are the feature vectors of the i-th superpixel region and the j-th annotated pixel. The construction method of the new adjacency matrix is shown in equation (11).

$$A'_{ij} = \begin{cases} 1 & S^l_i, S^l_j \quad adjacent \\ 0 & else \end{cases} \tag{11}$$

In equation (11), A'_{ij} is the connection strength between the i-th and j-th superpixels in the new image structure. S^l_i and S^l_j are superpixel regions. The cross-entropy loss L is shown in equation (12).

$$L = -\sum_{i \in y_L, c=1}^{C} y_{ic} log P_{ic} \tag{12}$$

In equation (12), y_L is the index set of all labeled pixels, C is the total number of categories, and y_{ic} is the One-hot label of the actual category. P_{ic} is the probability that the model predicts pixel i to belong to category c. Due to the high dimensionality of HSI data, preprocessing should be performed before classification to obtain better parameter estimates and effective information. After preprocessing, dimensionality reduction, and feature extraction, this study trains a classifier using selected samples, determines the discriminant function, and then uses the HSI model for classification.

3 Results

3.1 Performance testing of U-Nets model based on MSS

To verify the performance of the proposed model, a suitable experimental environment is established. Ubuntu 20.04 LTS is an operating system equipped with an Intel Xeon Gold 5218 CPU, NVIDIA RTX 3090 GPU, 64GB of memory, and Pytorch framework. The experiment will have 300 iterations, a learning rate of 0.001, a Dropout ratio of 3, and 5 superpixel layers. Indian Pines, Pavia University, Salinas, and GRSS 2013 publicly available datasets are the data sources. Among them, the Indian Pines dataset is collected from agricultural areas in

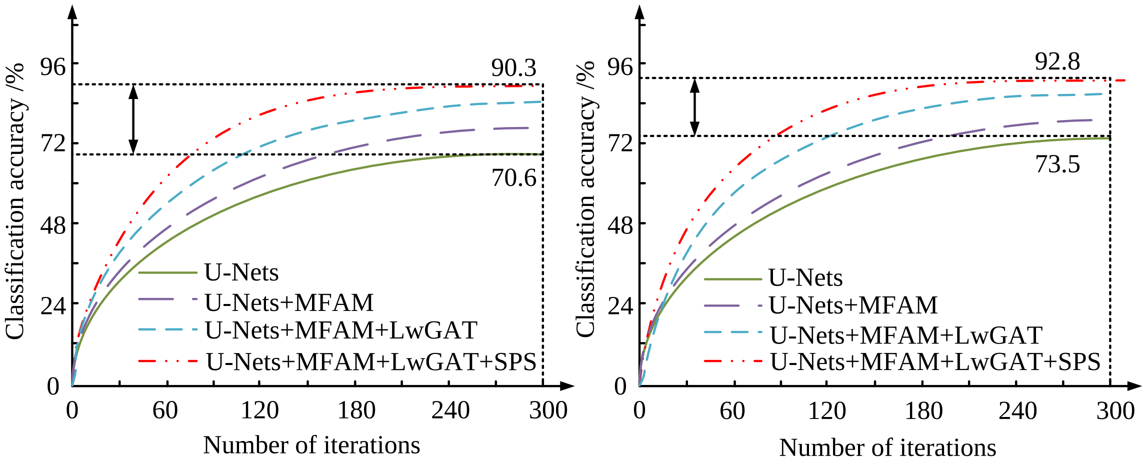
Indiana, USA, and includes 16 land cover categories, mainly different types of crops. The spectral dimension of the image is high, and the spectral similarity between classes is strong, making it difficult to classify. The Pavia University dataset is obtained from aerial photography of urban areas over the University of Pavia in Italy. It includes 9 categories of land features, including roads, buildings, grasslands, etc., with distinct spatial structural characteristics, making it suitable for testing the spatial modeling ability of the model. Salinas is collected in the Salinas Valley agricultural area of California, USA, containing information on 16 types of crops with high spatial resolution and rich spectral information. It is suitable for evaluating the model's ability to classify fine-grained land features. GRSS 2013 is composed of the remote sensing image classification competition dataset provided by IEEE GRSS, which includes multiple urban land cover categories such as buildings, roads, and shadows. It has complex spatial patterns and uneven data distribution, making it highly challenging. During the testing process, each dataset is trained using a five-fold cross-validation to reduce the risk of overfitting and ensure the robustness of the results. The four datasets are divided into the training set and the test set in an 8:2 ratio. Among them, the training set is further divided into a training subset and a validation subset in each cross-validation, which are used for model parameter tuning. Model training and testing are carried out under the same conditions, and CNN, GCN, FCN, and 3D-CNN are used as baseline models for comparison. To verify the generalization ability of the proposed model, experiments are conducted on the comprehensive datasets (Indian Pines, Pavia University, Salinas, and GRSS 2013). Table 1 shows the specific categories and quantities of data.

According to Table 1, firstly, the proposed model is subjected to ablation testing with classification accuracy as the indicator, as shown in Fig.7.

Figs.7 (a) and (b) show the results of studying the model in the training and testing sets. As the iterations increase, the HSI classification accuracy of each module in the research model gradually improves and tends to stabilize in the later stages of training. The U-Nets module performs the worst in both datasets, with a maximum HSI classification accuracy of only 73.5%, indicating significant shortcomings in processing high-dimensional spectral information and complex spatial structures of HSI. After improving the MFAM and LwGAT modules, the HSI classification accuracy of the U-Nets module increases by about 15%. The U-Nets model based on MSS has the best overall performance, with the best performance in HSI reaching 92.8%. The research model can effectively improve the classification precision of HSI by integrating different modules. In addition, the study also conducts a comparative test with average accuracy as the test index. The test results are shown in Figure 8.

Table 1: The specific categories and quantities.

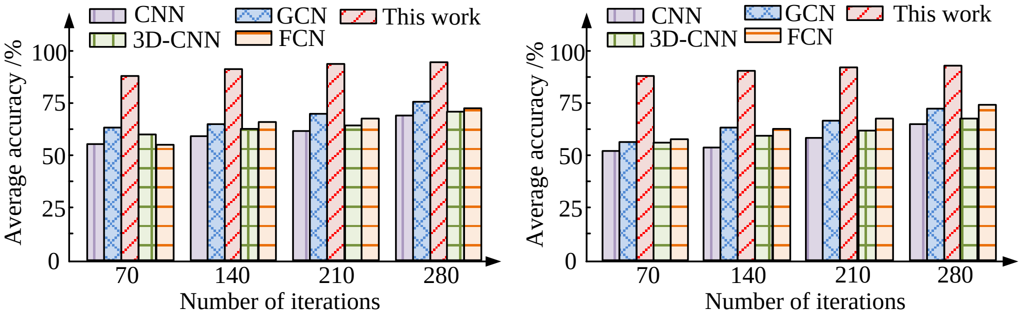
Serial number	Category	Training sample size	Test sample size	Total sample size
1	Meadows	14919	3730	18649
2	Bare Soil	4023	1006	5029
3	Trees	2451	613	3064
4	Asphalt	5304	1327	6631
5	Gravel	1679	420	2099
6	Metal Roofs	1076	269	1345
7	Brick Walls	2945	737	3682
8	Shadows	757	190	947
9	Bitumen	1064	266	1330



(a) Training set

(b) Test set

Figure 7: Ablation test.



(a) Training set

(b) Test set

Figure 8: Average accuracy test results.

Table 2: Test results of ACE and AMI between categories of different models under interference conditions.

Interference type	Interference intensity	Models	ACE (95% CI)	AMI (95% CI)
No interference	/	DUAL-CNN	0.46 (0.44-0.49)	0.62 (0.60-0.64)
		2D-CNN-PCA	0.38 (0.36-0.40)	0.68 (0.66-0.70)
		SSRN	0.34 (0.33-0.36)	0.72 (0.70-0.73)
		Research model	0.21 (0.20-0.23)	0.84 (0.82-0.86)
Gaussian noise	0.03	DUAL-CNN	0.58 (0.56-0.60)	0.55 (0.53-0.57)
		2D-CNN-PCA	0.49 (0.47-0.51)	0.61 (0.59-0.63)
		SSRN	0.44 (0.42-0.46)	0.66 (0.64-0.68)
		Research model	0.31 (0.30-0.33)	0.77 (0.75-0.79)
Band missing	10% band random missing	DUAL-CNN	0.55 (0.53-0.57)	0.57 (0.55-0.59)
		2D-CNN-PCA	0.46 (0.44-0.48)	0.63 (0.61-0.65)
		SSRN	0.42 (0.40-0.44)	0.68 (0.66-0.70)
		Research model	0.28 (0.27-0.30)	0.79 (0.77-0.81)

Fig.8 shows the mean accuracy of different models. The effectiveness of the U-Nets model based on MSS in HSI classification tasks is superior to other baseline models. In Fig.8 (a), CNN exhibits a certain feature learning ability during the training process, with an average accuracy gradually increasing and ultimately stabilizing at 70.6%. Due to its ability to model the structural relationships between pixels, GCN achieves an average accuracy of 75.1% higher than CNN during the training phase. The research model not only has faster convergence speed and stronger stability, but also has an average accuracy of 95.3%. In Fig.8 (b), both CNN and GCN show a certain degree of performance degradation, dropping to 64.2% and 73.9%, indicating that their generalization ability is still insufficient in complex sample distributions. The average accuracy of the research model in the test set is 93.1%, further verifying the effectiveness of the multi-level superpixel structure and graph attention mechanism in modeling the spatial spectral relationship of HSI.

3.2 Simulation testing of HSI classification model

To further validate the performance of the final model (HSI classification model based on multi-scale features and GCN-CNN), Gaussian noise and random loss of some bands are added to the original test set in this study. At present, the most popular and representative model in HSI classification is the 2D-CNN with Principal Component Analysis (2D-CNN-PCA), Dual-Branch CNN (DUAL-CNN), and Spectral Spatial Residual Network (SSRN), which are compared and tested. Under interference conditions, the Average Classification Entropy (ACE) and Average Mutual Information (AMI) test data between different models are listed in Table 2.

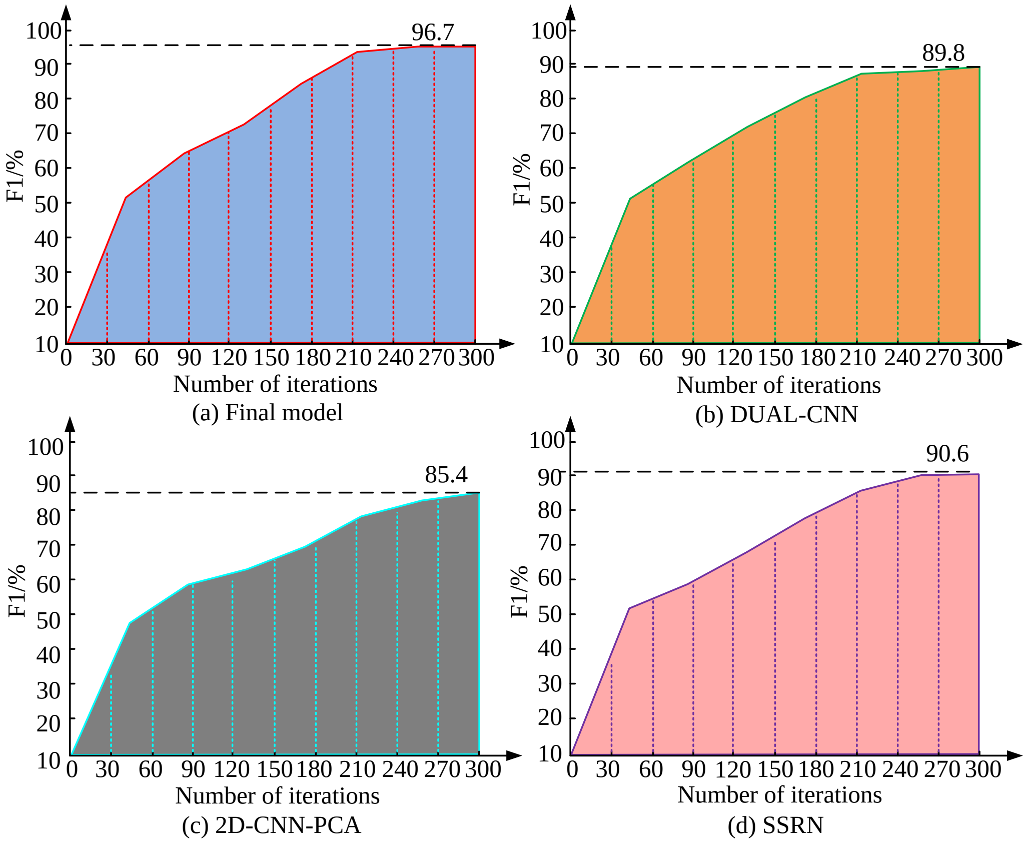
In Table 2, with the addition of noise and band loss interference, the entropy values of each model generally increase and the AMI decrease, reflecting the effect of interference on the model's discriminative ability. The entropy value and AMI variation amplitude of the research model are relatively small, showing better robustness and stability. Moreover, the 95% Confidence Interval (CI) of the research model in the three types of scenarios is generally small, and the numerical intervals basically do not overlap with those of the control models. Under Gaussian noise interference, the ACE and AMI values of the research model are 0.31 and 0.77. Under the interference of random band loss, the ACE and AMI values are 0.28 and 0.79, indicating that the model can better distinguish different categories and avoid category confusion. The multi-scale GCN can adaptively capture the spectral and spatial correlations at different scales, thereby reducing the impact of local noise on the overall prediction. The jump connection of U-Nets helps to restore the key feature regions, and the prediction performance can be maintained even if some bands are missing. For the interpretability, the weight distribution of GCN nodes reveals the degree of attention of the model to different bands and spatial positions. The visualization of the U-Nets intermediate feature map shows the regions that the model focuses on in the spectral image, providing an intuitive basis for understanding the model's decision-making. In practical HSI applications, this model has extensive application value. For instance, in agricultural remote sensing, even if some bands are affected by clouds, fog, or sensor malfunctions, the model can still accurately identify the type and growth status of crops. In urban remote sensing or geological exploration, the multi-scale feature recovery capability enables the model to address spectral deficiency and aliasing issues, enhancing the reliability and stability of classification, target detection, and anomaly identification. Therefore, the model performs well on experimental datasets and has practical application potential across scenarios and datasets.

After analyzing the performance and robustness of the model, the similarities between the model and commonly used control methods in nonlinear and uncertain systems are further explored, revealing its adaptive characteristics and stability mechanism. The adaptive behavior of the HSI classification model based on multi-scale features and GCN-CNN in dealing with noise, band missing, and spectral-spatial complexity is analyzed. It is compared with methods such as Adaptive Fuzzy Control, Output-Feedback Controller, and Robust Neural Adaptive Control, etc. The similarity analysis results are shown in Table 3.

Table 3 shows that the multi-scale GCN-CNN feature extraction module adaptively captures the spectral-SFs at different scales, similar to the mechanism of Adaptive Fuzzy Control that dynamically adjusts the control law when dealing with system uncertainties. The encoder-decoding structure and skip connection of U-Nets are equivalent to the Backstepping Control process of eliminating system uncertainties step by step, minimizing the impact of local disturbances on the overall output. The F1 values predicted by different models for classification are shown in Fig.9.

Table 3: Similarity analysis results.

Control methods	References	Correspondence with the research model	Robustness demonstration
Adaptive fuzzy control	[19]	Adaptive adjustment of multi-scale features is similar to fuzzy rule adaptation	Fast convergence and resistance to external disturbances
Output-Feedback Controller	[20]	Output-feedback mechanism corresponds to graph convolution information fusion in the model	Resistant to input nonlinearities and noise
Robust neural adaptive control	[21]	Adaptive update of neural network weights is similar to GCN node weight adjustment	Robust against multivariable uncertainties
Adaptive backstepping control	[22]	Stepwise feature extraction corresponds to multi-scale GCN-CNN hierarchical feature aggregation	Stable prediction under noise and missing bands
Nonlinear optimal control	[23]	Optimal control concept corresponds to model parameter optimization and training	Improves overall classification accuracy and robustness
High-gain observer-based adaptive fuzzy control	[24]	High-gain observer corresponds to model feature sensitivity adjustment	Sensitive to local anomalies or missing bands
Fuzzy state-feedback control	[25]	State-feedback mechanism corresponds to U-Nets encoder-decoder skip connections	Preserves key information regions, mitigating disturbance effects



(a) Final model

(b) DUAL-CNN

(c) 2D-CNN-PCA

(d) SSRN

Figure 9: Classification prediction F1 values of different models.

Table 4: Comparison of runtime of four models.

Dataset	Models	1st (s)	2nd (s)	3rd (s)	4th (s)	5th (s)	6th (s)	7th (s)
Training set	DUAL-CNN	31.5	31.8	32.2	32.6	33.0	33.5	34.1
	2D-CNN-PCA	28.3	28.5	28.8	29.2	29.5	28.9	30.3
	SSRN	35.9	36.2	36.6	37.1	37.5	38.0	38.5
	Research model	24.9	25.2	25.5	25.8	26.2	26.5	26.8
Test set	DUAL-CNN	13.4	13.6	13.8	14.0	14.3	14.6	15.0
	2D-CNN-PCA	11.2	11.3	11.5	11.7	12.0	12.3	12.6
	SSRN	15.6	15.8	16.0	16.3	16.6	16.8	17.1
	Research model	9.3	9.5	9.7	9.9	10.1	10.3	10.5

Figs.9 (a) to (d) show the comparison curves of F1 values between the research model, DUAL-CNN, 2D-CNN-PCA, and the SSRN model. The F1 values of the research model on public datasets are superior to other models, indicating that it has stronger stability and generalization ability when dealing with typical HSI classification problems such as uneven class distribution, fuzzy boundaries, and high-dimensional redundant features. The F1 values of 2D-CNN-PCA, DUAL-CNN, SSRN, and the research models are 85.4%, 89.8%, 90.6%, and 96.7%. The fundamental reason for the improvement of model performance is that MSS enhances the model's perception ability of spatial regions, enabling it to more accurately identify land boundaries and local patterns. GCN introduces non-Euclidean structural modeling capabilities, which can effectively capture semantic correlations and global dependencies between samples. CNN is good at extracting local spatial contextual information and plays a key role in restoring SF details. Finally, to further validate the computational efficiency of the proposed model in practical applications, a comparative evaluation is conducted on the operational complexity of various models under the same hardware conditions, as listed in Table 4.

In Table 4, as the number of runs grows, the runtime of the four models on both sets shows an increasing trend. In the test set, the runtime of SDUAL-CNN, 2D-CNN-PCA, SSRN, and the research model increases by 2.6 s, 2.3 s, 2.7 s, and 1.9 s. In the training set, the runtime of the four models increases by 1.6 s, 1.4 s, 1.5 s, and 1.2 s. Among them, the overall growth rate of the research model is relatively small, and its effect on the overall performance is relatively small. The computational complexity of the research model is lower, the structural design is more efficient, and it can effectively reduce the consumption of computing resources and processing time.

4 Discussion

To comprehensively verify the advancement and effectiveness of the HSI classification model based on multi-scale features and GCN-CNN, their performance in the existing literature was studied, compared, and analyzed. Firstly, in terms of classification accuracy, existing studies have shown that the classification accuracy rates of the HSI classification method based on 3D-CNN proposed by Atik SO in the datasets of Indian Pines, Salinas, and Pavia University were 92.43%, 95.06%, and 99.00%, respectively [26]. The constructed HSI classification model based on multi-scale features and GCN-CNN had an F1 value of 96.7% on the comprehensive datasets (Indian Pines, Pavia University, Salinas, and GRSS 2013). Under the interference of Gaussian noise, the ACE and AMI values of the research model were 0.31 and 0.77, respectively. Therefore, by combining multi-scale feature extraction with GCNs, the model could effectively capture the spatial dependencies and spectral features between pixels, achieving in-depth modeling of complex spectral information.

In terms of computational complexity, Banerjee A et al. proposed an HSI classification method based on a 3D-2D-1D CNN. The experimental results showed that 3D-CNN and 2D-CNN effectively utilized the spectral space features, while 1D-CNN could perform feature extraction pixel by pixel. The accuracy rate of the 3D-2D-1D CNN model in Indian Pines was 98.87%, and that in Pavia Center was 99.92% [27]. Although the 3D-2D-1D CNN model could extract local spectral or SFs, its computational complexity was extremely high. Especially when dealing with HSI data containing hundreds of bands, it was prone to excessive memory consumption and excessively long training time. The research model maintained the shortest average running time in both the training and testing phases. The average time consumption of the training set was only 25.8 seconds, while that of the test set was 9.9 seconds.

In conclusion, the research model outperforms existing methods in terms of classification accuracy, few-shot category recognition, computational complexity, and spatial structure modeling ability, demonstrating its advancement and effectiveness in the HSI classification task.

5 Conclusion

Due to the high dimensionality of HSI spectral data, directly using traditional methods can bring difficulties and low classification accuracy to land cover classification, as well as the shortcomings of existing methods in handling spectral spatial joint modeling and structural expression capabilities. This study proposed an HSI classification model based on multi-scale features and GCN-CNN. This model constructed a parallel GCN-CNN structure, combined MFAM to enhance spatial spectral feature expression, and utilized a graph structure reconstruction mechanism to introduce geometric and label prior information, effectively improving the discriminative ability of complex land features. In the experiment, under Gaussian noise interference, the ACE of the research model was 0.31, and the AMI was 0.77. Under the condition of random loss in the 10% band, ACE further decreased to 0.28, and AMI increased to 0.79, indicating that the model still has good category discrimination ability and stability in the face of high-dimensional data degradation, and can effectively avoid category ambiguity and confusion. The F1 values of the research model on the public dataset were superior to those of the comparison model, reaching 96.7%. This fully validated its stronger generalization ability and stability in dealing with typical HSI classification problems such as uneven class distribution, fuzzy boundaries, and high-dimensional redundant features. In addition, the research model maintained the shortest average runtime during both training and testing phases, with an average training time of only 25.8 s and an average testing time of 9.9 s. The research model outperforms existing methods in classification accuracy, anti-interference ability, and computational efficiency, demonstrating good practical potential and engineering scalability. However, this study mainly focuses on standard dataset experiments and has not yet covered multi-temporal data and cross-regional HSI. Future work can combine transfer learning and multi-source information fusion to further enhance the generalization ability and application breadth of the model.

6 Funding

The research is supported by the Fundamental Research Funds for the Special Technology Mission Team Project at Zhejiang Institute of Economics and Trade, under the project entitled “Research on Part Recognition Technology Based on Deep Learning” (Grant No. 22KJTPY10).

References

[1] Qiang Zhang, Yaming Zheng, Qiangqiang Yuan, Meiping Song, Haoyang Yu, and Yi Xiao. Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven. IEEE Transactions on Neural Networks and Learning Systems, 35(10):13143-13163, 2023. https://doi.org/10.1109/TNNLS.2023.3278866

[2] Harshula Tulapurkar, Biplab Banerjee, and Krishna Mohan Buddhiraju. Multi-head attention with CNN and wavelet for classification of hyperspectral image. Neural Computing and Applications, 35(10):7595-7609, 2023. https://doi.org/10.1007/s00521-022-08056-w

[3] Ran Ran, Liang-Jian Deng, Tai-Xiang Jiang, Jin-Fan Hu, Jocelyn Chanussot, and Gemine Vivone. GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution. IEEE Transactions on Cybernetics, 53(7):4148-4161, 2023. https://doi.org/10.1109/TCYB.2023.3238200

[4] R. Anand, Bilal Khan, Vinay Kumar Nassa, Digvijay Pandey, Dharmesh Dhabliya, Binay Kumar Pandey, and Pankaj Dadheech. Hybrid convolutional neural network (CNN) for Kennedy Space Center hyperspectral image. Aerospace Systems, 6(1):71-78, 2023. https://doi.org/10.1007/S42401-022-00168-4

[5] Ali Ari. Multipath feature fusion for hyperspectral image classification based on hybrid 3D/2D CNN and squeeze-excitation network. Earth Science Informatics, 16(1):175-191, 2023. https://doi.org/10.1007/s12145-022-00929-x

[6] Zhongwei Li, Wenhao Huang, Leiquan Wang, Ziqi Xin, and Qiao Meng. CNN and Transformer interaction network for hyperspectral image classification. International Journal of Remote Sensing, 44(18):5548-5573, 2023. https://doi.org/10.1080/01431161.2023.2249598

[7] Ram Nivas Giri, Rekh Ram Janghel, Saroj Kumar Pandey, Himanshu Govil, and Anurag Sinha. Enhanced hyperspectral image classification through pretrained CNN model for robust spatial feature extraction. Journal of Optics, 53(3):2287-2300, 2024. https://doi.org/10.1007/s12596-023-01473-7

[8] Arati Paul, and Sanghamita Bhoumik. Classification of hyperspectral imagery using spectrally partitioned HyperUnet. Neural Computing and Applications, 34(3):2073-2082, 2022. https://doi.org/10.1007/s00521-021-06532-3

[9] Bo Peng, Yuxuan Yao, Qunxia Li, Xinyu Li, Guoting Lin, Lin Chen, and Jianjun Lei. Clustering information-constrained 3D U-Net subspace clustering for hyperspectral image. Remote Sensing Letters, 13(11):1131-1141, 2022. https://doi.org/10.1080/2150704X.2022.2132122

[10] Jiawei Deng, and Bin Yang. Hyperspectral and multispectral image fusion via residual selective kernel attention-based U-net. International Journal of Remote Sensing, 45(5):1699-1726, 2024. https://doi.org/10.1080/01431161.2024.2318766

[11] Tatireddy Subba Reddy, V. V. Krishna Reddy, R. Vijaya Kumar Reddy, Chandra Sekhar Kolli, V. Sitharamulu, and Majjaru Chandrababu. SHBO-based U-Net for image segmentation and FSHBO-enabled DBN for classification using hyperspectral image. The Imaging Science Journal, 72(4):479-498, 2024. https://doi.org/10.1080/13682199.2023.2208927

[12] Ting Tang, Weihong Yan, Geli Bai, Xin Pan, and Jiangping Liu. UVMSR: a novel approach to hyperspectral image super-resolution by fusing U-Net and Mamba. International Journal of Remote Sensing, 46(5):2023-2054, 2025. https://doi.org/10.1080/01431161.2024.2443619

[13] Subhashish Nabajja, and Mahendra Kanojia. Choledochal cancer region detection in hyperspectral images using U-Net based models. International Journal of Hybrid Intelligent Systems, 21(2):96-114, 2025. https://doi.org/10.3233/HIS-240024

[14] Hui Zhang, Yixia Pan, Yuan Chen, Hongxu Zhang, Jianhui Xie, Xingchu Gong, Jieqiang Zhua, and Jizhong Yan. Improving the geographical origin classification of Radix glycyrrhizae (licorice) through hyperspectral imaging assisted by U-Net fine structure recognition. Analyst, 149(6):1837-1848, 2024. https://doi.org/10.1039/D3AN02064A

[15] Xuan Tung Nguyen, and Giang Son Tran. Hyperspectral image classification using an encoder-

decoder model with depthwise separable convolution, squeeze and excitation blocks. Earth Science Informatics, 17(1):527-538, 2024. https://doi.org/10.1007/s12145-023-01181-7

[16] Fulin Luo, Xi Chen, Xiuwen Gong, Weiwen Wu, and Tan Guo. Dual-window multiscale transformer for hyperspectral snapshot compressive imaging. Proceedings of the AAAI Conference on Artificial Intelligence, 38(4):3972-3980, 2024. https://doi.org/10.1609/aaai.v38i4.28190

[17] Xueying Li, Zongmin Li, Huimin Qiu, Guangli Hou, and Pingping Fan. An overview of hyperspectral image feature extraction, classification methods and the methods based on small samples. Applied Spectroscopy Reviews, 58(6):367-400, 2023. https://doi.org/10.1080/05704928.2021.1999252

[18] Peter P. Groumpos. A critical historic overview of artificial intelligence: Issues, challenges, opportunities, and threats. Artificial Intelligence and Applications. 1(4):197-213, 2023. https://doi.org/10.47852/bonviewAIA3202689

[19] Abdesselem Boulkroune, Farouk Zouari, and Amina Boubellouta. Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. Journal of Vibration and Control, 2025. https://doi.org/10.1177/10775463251320258

[20] Abdesselem Boulkroune, Sarah Hamel, Farouk Zouari, Abdelkrim Boukabou, and Asier Ibeas. Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities. Mathematical Problems in Engineering, 2017(1):1-12, 2017. https://doi.org/10.1155/2017/8045803.

[21] Mou Chen, Shuzhi Sam Ge, and Bernard Voon Ee How. Robust adaptive neural network control for a class of uncertain MIMO nonlinear systems with input nonlinearities. IEEE Transactions on Neural Networks, 21(5):796-812, 2010. https://doi.org/10.1109/TNN.2010.2042611

[22] Farouk Zouari, Kamel Ben Saad, and Mohamed Benrejeb. Adaptive backstepping control for a class of uncertain single input single output nonlinear systems. 10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13), 2013:1-6, 2013. https://doi.org/10.1109/SSD.2013.6564134

[23] G. Rigatos, M. Abbaszadeh, and B. Sari. Nonlinear optimal control for a gas compressor driven by an induction motor. Results in Control and Optimization, 11:100226, 2023. https://doi.org/10.1016/J.RICO.2023.100226

[24] L. Merazka, F. Zouari, and A. Boulkroune. High-gain observer-based adaptive fuzzy control for a class of multivariable nonlinear systems. 2017 6th International Conference on Systems and Control (ICSC), 2017:96-102, 2017. https://doi.org/10.1109/ICoSC.2017.7958728

[25] L. Merazka, F. Zouari, and A. Boulkroune. Fuzzy state-feedback control of uncertain nonlinear MIMO systems. 2017 6th International Conference on Systems and Control (ICSC), 2017:103-108. https://doi.org/10.1109/ICoSC.2017.7958730

[26] Saziye Ozge Atik. Dual-stream spectral-spatial convolutional neural network for hyperspectral image classification and optimal band selection. Advances in Space Research, 74(5):2025-2041, 2024. https://doi.org/10.1016/j.asr.2024.05.064

[27] Anasua Banerjee, Satyajit Swain, Minakhi Rout, and Mainak Bandyopadhyay. Composite spectral spatial pixel CNN for land-use hyperspectral image classification with hybrid activation function. Multimedia Tools and Applications, 84(12):10527-10550, 2025. https://doi.org/10.1007/s11042-024-19327-0