Supervised Hyperspectral Remote Sensing Image Classification Using Graph Attention Networks with Hybrid Pooling

Hongliang Liu
School of Education, Luoyang Vocational College of Culture and Tourism, Luoyang 471000, China
E-mail: lywllhl@163.com

Keywords: graph neural network, remote sensing images, classification of ground objects

Received: June 17, 2025

Traditional remote sensing image surveys rely on field observations, which have shortcomings such as low efficiency and limited coverage, making it difficult to meet the needs of urban planning and management. In response to this, this article conducts in-depth research on remote sensing image classification methods based on graph neural networks (GNN). This study adopts a joint framework of GAT (Graph Attention Network)+SLC (Simple Linear Iterative Clustering)+Hybrid Pooling. Use SLIC algorithm to perform superpixel segmentation on hyperspectral data and construct a graph structure that integrates spectral spatial features. Utilizing GAT's attention mechanism to enhance feature interaction among neighboring nodes, combined with a hybrid pooling strategy to balance feature dimensionality reduction and key information preservation. The experiment was conducted on the public datasets WHU-RS19 and UC Merced, using traditional CNN, random forest, and minimum distance methods as baseline models. The significance of the differences was verified through independent sample t-test. By expanding the sample coverage, optimizing the model regularization strategy, and continuously iterating training, the generalization performance of the model has been significantly improved. Therefore, from the comparison of classification performance, graph NN classification is superior to minimum distance classification. This method utilizes perspective remote sensing data to achieve accurate ground target classification and dynamic change information extraction, providing reliable data support for urban land survey and management.

Povzetek: Za klasifikacijo hiperspektralnih posnetkov je predlagan GAT+SLIC+Hybrid Pooling, kjer superpikslovna segmentacija zgradi graf s spektralno-prostorskimi značilkami, GAT poudari medvozliščne odnose, hibridno prilagajanje pa ohrani ključne informacije.

1 Introduction

The rational planning and layout of the city, the research on the future development of the city, and the construction of various urban projects all need geographic information data as the support and theoretical basis [1]. The traditional remote sensing survey is carried out by means of on-site survey or observation survey, which is inefficient, and is difficult to update the survey data obtained. Therefore, the urban survey method to obtain data in this way cannot meet the needs of urban planning management [2]. However, remote sensing technology is a technology that uses satellites to observe objects or forms on the earth's surface to obtain a variety of remote sensing image information data. With the development of economy, industrial development requires more and more strict remote sensing image data, which requires not only high quality and large amount of information, but also large quantity and wide coverage. Remote sensing equipment receives and records information, and obtains clearer remote sensing image data containing more information by processing echo information [3]. The data contains a lot of valuable information, which needs to be fully exploited and utilized. Due to frequent interference during remote sensing image acquisition, image classification using a single model may not match the actual feature types, such as bare land, dry land and wetland, which are relatively difficult to distinguish [4]. Therefore, classified calibration maps can only be stored in the database after frequent editing and modification. Object oriented remote sensing image classification technology will solve the above problems to the greatest extent, and provide geographic information with strong usability and good real-time.

The purity of remote sensing image pixels is also a key issue. In actual remote sensing images, pixels are often not pure, meaning that one pixel may contain information about multiple geographic things. This is due to the limited spatial resolution of remote sensing images, which results in one pixel covering partial areas of multiple geographical features. For example, in high-resolution remote sensing images, a pixel may contain information about both buildings and surrounding

vegetation. This brings great difficulties to pixel-based classification methods, which can easily lead to classification errors. The general description of this problem is as follows: Given a remote sensing image of an area, it is preprocessed to meet the recognition requirements. First, the clustering of image pixels is obtained by region segmentation, or further accurate segmentation is performed. Then, the characteristics of regional targets are used for classification and recognition. [5]. Many types of surface objects mainly include roads, buildings, water bodies, vegetation and bare land. It is of great theoretical significance to classify remote sensing images, extract and analyze this information based on graph neural network [6]. Therefore, hyperspectral remote sensing data can be used to easily and quickly identify and monitor various buildings, urban roads, urban green spaces, and other facilities, and extract their dynamic change information, providing good data support for urban land surveys, future development research, and other urban management work.

In response to the low efficiency, limited coverage, and weak generalization ability of traditional remote sensing image survey methods, this study constructs a joint framework of mixed pooling. Constructing a graph structure that integrates spectral and spatial features through superpixel segmentation, utilizing attention mechanisms to enhance node feature interaction and balance feature dimensionality reduction with key information preservation. Use independent sample t-test to verify the superiority of this method in improving the accuracy of complex scene classification and enhancing noise adaptability and generalization ability. At the same time, clarify the application bottlenecks of the model in large-scale sample processing and dynamic scenarios, providing a basis for subsequent optimization. Ultimately achieving accurate ground target classification and dynamic change information extraction, providing reliable technical support for urban land survey and management.

Innovation content:

(1) This article proposes a graph neural network model suitable for remote sensing image classification. The algorithm achieves a balance between efficient feature extraction and noise resistance through a lightweight network architecture design.

(2) This model captures spatial correlation features between pixels in remote sensing images through graph convolution operation, effectively solving the problem of decreased classification accuracy in imbalanced sample scenes.

2 Related work

Yue et al. reconstructed an enhanced U-net structure generative adversarial network using remote sensing image super-resolution. Research has shown that classification accuracy can be significantly improved based on discovered knowledge [7]. Shi et al. conducted research on remote sensing scene image classification based on self compensating convolutional neural networks. The purpose is to assign a class label to each pixel based on its own attributes and its correlation with adjacent pixels, and collect all pixels with the same features to identify the category to which the pixel belongs [8]. Chen et al. analyzed hyperspectral remote sensing image classification based on dense residual 3D convolutional neural network. By utilizing texture, spectrum, shape, and background information to extract informal residential areas from images, high classification accuracy has been achieved [9]. Liu et al. conducted high-resolution remote sensing image super-resolution reconstruction using convolutional neural networks. The neural network model developed based on multi-scale geometric analysis provides a new approach and method for accurate and fast classification of remote sensing images [10]. Huo et al. combined support vector machines to solve the problem of high classification accuracy, but the computation time was long. This method is used for classifying land features, improving classification efficiency and ensuring that classification accuracy remains basically unchanged, demonstrating the advantages of combining the two algorithms [11]. Salvetti et al. demonstrated strong robustness to noisy data and outliers due to the use of sample random extraction and feature random selection in the random forest algorithm [12]. Wang et al.'s remote sensing image feature extraction methods divide complex data into real parts and corresponding imaginary parts, and process the real and imaginary parts to obtain the final classification features. Due to these feature extraction methods only extracting amplitude information from remote sensing images without effectively utilizing phase information, there is a significant loss of information during the feature extraction process [13]. Hu et al. analyzed a remote sensing image super-resolution reconstruction model based on generative adversarial networks. Object oriented classification is an image classification technique based on mathematics, statistics, logic, and inspiration. The development of object-oriented classification can further promote the development of image analysis technology [14]. In response to the confusion caused by traditional classification methods, Huan et al. used an asymmetric multi-scale super-resolution network for remote sensing image reconstruction. The method of hierarchical extraction of remote sensing thematic information was discussed, which involves using decision trees to establish corresponding feature databases based on the different characteristics of different land features, and finally integrating and merging the thematic layers into component class diagrams [15]. Hua et al. put forward that deep NN has a strong feature expression ability, which makes it have great advantages in processing remote sensing images with rich information [16]. Lu and Cheng [17] applied an improved gene expression

programming algorithm based on grouping strategy (bs gep) to remote sensing image classification problems. This avoids the local convergence caused by the destruction of population diversity in traditional gene expression programming methods. Ye et al. [18] proposed a lightweight model that combines VGG-16 and U-Net networks. By combining two convolutional neural networks, they classified the scenes of remote sensing images. Minimize the memory usage of the model while ensuring its accuracy. Wang and Li [19] proposed a multimedia image edge recognition method based on intelligent monitoring robots. This method mainly focuses on the similarity features of pixels and edge detection points between adjacent regions of multimedia images. And use iterative methods to weight adjacent regions of the multimedia image, making the properties of the multimedia image different from traditional images before. The improvement of its viewing mode is loved by the audience, who can experience an immersive state while watching movies and become the subject of the image. Petrovska et al. [20] used transfer learning from pre trained deep convolutional neural networks (CNNs) in remote sensing image classification. The neural network they used in their research was a high-dimensional previously trained CNN on the ImageNet dataset. Transfer learning can be achieved through feature extraction or fine-tuning. Liang et al. [21] introduced a novel dual stream architecture that combines global visual features and object-based positional features to enhance feature representation capabilities. They are the first people to study the dependency relationships between objects in remote sensing scene classification tasks.

Table 1: Research review

Method	Dataset	Accuracy (%)	Kappa / Other indicators	Restriction
Inductive Learning Space Knowledge Discovery	Land classification dataset	82.3	0.792	Relying on prior knowledge to induce quality, limited adaptability to complex scenarios
Multi feature extraction (texture/spectrum, etc.)	Informal Residential Area Dataset	88.7	0.843	Without considering the redundancy between features, the accuracy of complex terrain boundary classification is insufficient
Support Vector Machine Optimization	Ground target classification dataset	90.2	0.863	Long computation time and low training efficiency on large-scale datasets
Real/imaginary part feature extraction	Remote sensing image dataset	79.5	0.768(Information loss rate 18.3%)	Weak generalization ability in complex scene classification
Hierarchical Decision Tree Classification	Remote sensing thematic information dataset	84.7	0.805	The construction of feature databases is cumbersome
Deep Neural Network	High information remote sensing dataset	91.3	0.876	High risk of overfitting in small sample scenarios

As shown in Table 1. In existing methods, even deep neural networks with good performance have an accuracy rate of only 91.3%. The proposed GNN method naturally adapts to the grid characteristics of hyperspectral data through graph structure, and can simultaneously capture the spectral features and spatial neighborhood

dependencies of pixels, solving the problem of spectral spatial feature segmentation in traditional methods. Its attention mechanism can dynamically adjust neighborhood weights, overcome spatial heterogeneity interference, and the hybrid pooling strategy can preserve key information in dimensionality reduction. Compared to deep neural networks, it is more suitable for small sample scenarios and compensates for the shortcomings of existing SOTA in terms of association modeling accuracy and scene adaptability.

3 Research method

To solve the problem of remote sensing image classification, this study constructed a dedicated model based on Graph Attention Network (GAT), and optimized the graph construction and network layer design in a targeted manner. This study adopts a graph topology structure based on superpixel regions and uses the Simple Linear Iterative Clustering (SLIC) algorithm to perform superpixel segmentation on remote sensing images. Set the number of segmentations superpixels to 800-1200 and the compactness coefficient to 15. Through this step, the image is divided into local regions with similar spectral and texture features, which serve as the basic units of the image structure. This study adopts an adaptive learning rate adjustment strategy based on the rate of change of Batch Normalization (BN) statistic. This mechanism improves the convergence speed of the model on the WHU-RS19 dataset, and ultimately improves the classification accuracy compared to fixed learning rates. To extract local texture features of pixels, CNN can be used to preprocess the image first, and the CNN feature map can be used as part of the node features instead of directly using CNN's convolution operation. At this stage, CNN is only used for feature extraction, and the subsequent graph construction and GNN operations are independent core processes. Therefore, it is necessary for our algorithm to have high effectiveness and robustness to meet the requirements of classification accuracy in practical applications. The NN is composed of these very simple neurons layer by layer, and the output of the previous neuron can become the input of the next neuron. Through the above analysis, a ground object classification model remote sensing image based on graph NN is constructed, as shown in Figure 1.

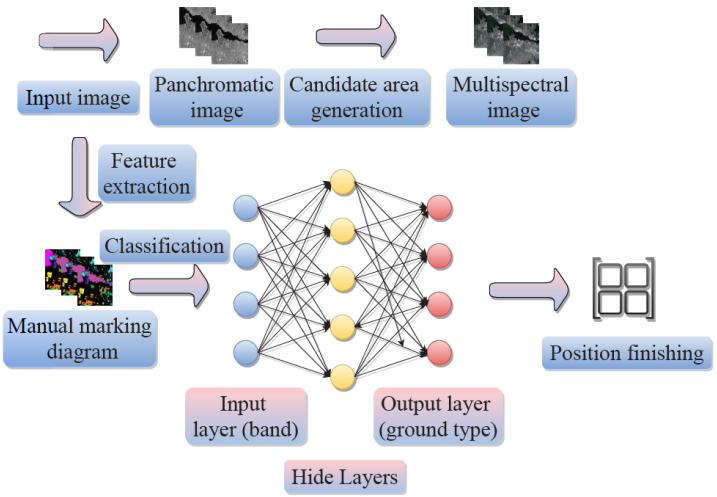

Figure 1: Classification model of remote sensing images based on graph NN

Dynamically adjust the learning rate based on changes in the distribution of network layer data. When the data distribution changes significantly, reduce the learning rate appropriately to avoid network instability caused by excessive parameter updates; When the data distribution changes slightly, the learning rate can be appropriately increased to accelerate the training speed. Implicit layer plays a role in extracting features from input in the network, which directly affects the processing ability, learning rate and generalization ability of the network. According to different problems, choose the corresponding number of hidden layers. Generally, one or two hidden layers are selected, but rarely more than three layers are used. Because too many hidden layers will increase the calculation amount exponentially, under normal circumstances, the number of hidden layer nodes should be increased first, and then the number of hidden layers should be increased according to the actual situation. As the graph NN is a layer-by-layer calculation model, firstly, calculate the mean μ and variance σ_B^2 of ground feature classification according to the following formulas, as shown in formulas (1) and (2).

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x \tag{1}$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m} x_i$$

(2)

m in indicates the number of features in the classification of ground objects, and x_i indicates the i feature.

Then, standardize each feature x_i in the ground feature classification, as shown in formula (3):

$$x_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2}}$$

(3)

Where x_i represents the characteristics after treatment.

Finally, the feature x_i is post-processed, where γ is the scaling parameter, β is the offset parameter, and y_i represents the post-processed feature. As shown in formula (4):

$$y_i = \gamma x_i + \beta$$

(4)

In the process of training the graph NN model, the μ_B and σ_B^2 of each ground feature classification are recorded, and the unbiased estimators of μ_B and σ_B^2 are used for the prediction, as shown in Formula (5) and Formula (6):

$$E(x) = E_B(\mu_B)$$

(5)

x is the node feature of remote sensing image conversion. After the remote sensing image is transformed into a graph, if the node feature vector is $X \in R^d$, then x in the formula represents the input feature vector of the graph node, carrying multidimensional observation information of the land cover. $E(x)$ reflects the "typical value" of land features in statistical sense (such as the average level of spectral features of a certain type of land). E_β represents the expected feature calculation for a single node/local neighborhood. E_B represents the expected calculation of graph nodes for the entire training batch.

$$Var(x) = \frac{m}{m-1} E_B$$

(6)

Essentially, the random deactivation of graph NN is to select different substructure parameters for training in the iterative process, and the final model is equivalent to the combination result of multiple substructure models. It is beneficial to balance the deep network with multiple hidden layers, each of which includes dozens or even hundreds of convolution kernels. Compared with the traditional classification methods, the graph NN classification method is easier to obtain higher precision classification results, especially for complex types of land cover classification, and its advantages can be better reflected. When the dimension of remote sensing image is 3, the dimension of convolution kernel W is also 3, and its calculation can be expressed as formula (7).

$$a = \delta(a * W + b)$$

(7)

Where a is the input image; W is the first layer convolution kernel parameter matrix; b is the convolution kernel offset value; δ is the activation function.

According to the rule of forward propagation, the multiplication and addition of the parameter matrix of the second convolution kernel W^1 and b^1 should be carried out on the basis of the calculation results of the first layer, and the calculation can be expressed as formula (8).

$$a^1 = \delta(a^2 * W^1 + b^1)$$

(8)

Where a^1 corresponds to the input layer; a^2 corresponds to the output layer; δ corresponding activation function; W^1 convolution kernel parameters; b^1 corresponds to the convolution kernel offset value. The higher the accuracy of a^1, the better the learning ability of the model.

$$a^1 = softMax\left(a^1 * W^1 + b^1\right) \tag{9}$$

The a^1 on the left side of the formula represents the output result calculated by the softMax function, which is usually a probability distribution used to represent the probability of samples belonging to various categories in classification tasks. Different W^1 will cause the model to learn different feature mapping relationships, determining how input features affect the final output. SoftMax: It is an activation function that can convert the result of linear transformation (calculated as $a^1 * W^1 + b^1$ into a probability distribution form, so that the output result is in the interval of [0,1], and the sum of probabilities in each dimension is 1.

The $loss$ function commonly used in the classification of remote sensing images is a multi-class cross entropy function, and its content is shown in formula (10).

$$loss = -\sum_j t_{i,j} \log \tag{10}$$

Where $t_{i,j}$ represents the target value.

Then, the upper layer parameters W and b are solved according to Formula (10), and is shown in Formula (11) and Formula (12).

$$a^{l-1} = \frac{\partial E}{\partial a} \frac{\partial a^1}{\partial a^{l-1}} \tag{11}$$

$$\frac{\partial E}{\partial u} = \frac{\partial E}{\partial W^l} + \frac{\partial E}{\partial b^l} \tag{12}$$

Where i and j correspond to convolution kernel positions; u and v correspond to multi-dimensional offset values. This study introduces the Hybrid Graph Pooling mechanism into the GNN architecture and integrates it into the node aggregation process. In the node feature aggregation stage, perform element by element maximum operation on the neighborhood features of each node. The following is a pseudocode block of the hybrid pooling mechanism, which clearly presents its implementation process and operational logic in the node aggregation stage. Graph Attention Network (GAT) learns the feature interaction weights between nodes through attention mechanism. For node v_i, its neighboring node set is $N(v_i) = \{v_i \left| A_{i,j} = 1\right\}$. Perform SLIC segmentation on hyperspectral image I, output superpixel label image $L \in N^{H \times W}$ (each pixel corresponds to a superpixel index), and calculate superpixel feature matrix X.Based on the superpixel label graph L, calculate the adjacency matrix A and construct the graph $G = (V, \varepsilon, X, A)$, where nodes $V = S$ and edges $\varepsilon = \{(i, j) \left| A_{ij} = 1\right\}$.

4 Analysis and discussion of results

This study used the WHU-RS19 remote sensing dataset as the main experimental data. This dataset was released by Wuhan University and contains 19 typical land features. The specific sources were obtained through the Gaofen-2 satellite, covering the central region of China. The imaging time was from 2018 to 2020, and it includes four bands: RGB and near-infrared. The spatial resolution of hyperspectral bands is 10m, the resolution of panchromatic bands is 2m, and the final image resolution after band fusion is 2.5m/pixel. The original dataset contains 4800 remote sensing images, each with a size of 600 × 600 pixels and covering a total area of approximately 1800km ². The selection of training samples should not only ensure the typicality and representativeness of samples, but also ensure the convergence of graph NN training. On the basis of the investigation, a total of 5,143 pixels of training samples were selected from the images, which is shown in Table 2.

Table 2: Training sample statistics of pixels

Classified ground objects	Water body	Lawn	Woodland	Building	Desert
Number of pixels	890	1032	1411	1228	467

Based on the classification results of 5 types of land features (water bodies, grasslands, forests, buildings, deserts) in the line chart, Table 3 shows the actual sample size. the confusion matrix is organized as follows (assuming a total of 274 test samples, the accuracy is calculated according to the horizontal axis of the line chart "Number of test samples", and approximate values are filled in):

Table 3: Actual sample size

Actual Category * * Predicted Category**	(Water body)	(Grassland)	(Woodland)	(Building)	(Sand)	Actual sample size
Water body	45	5	3	2	0	55
Grassland	3	50	5	2	0	60
Woodland	2	4	55	3	1	65
Building (Building)	1	3	2	45	1	52
Desert (Sand)	0	1	1	2	38	42
Predict sample size	51	63	66	54	40	274

Figure 2 - Figure 3 compares the overall classification accuracy and Kappa coefficient of graph NN and minimum distance method under different subject numbers, where the size of visual dictionary is 483.

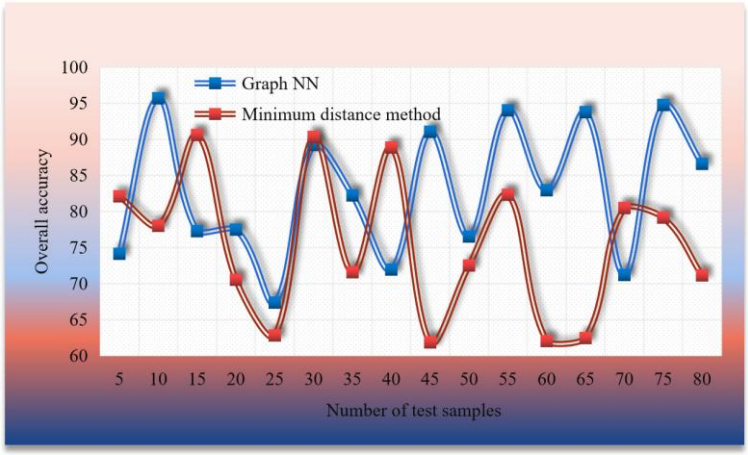

Figure 2: Comparison of classification accuracy

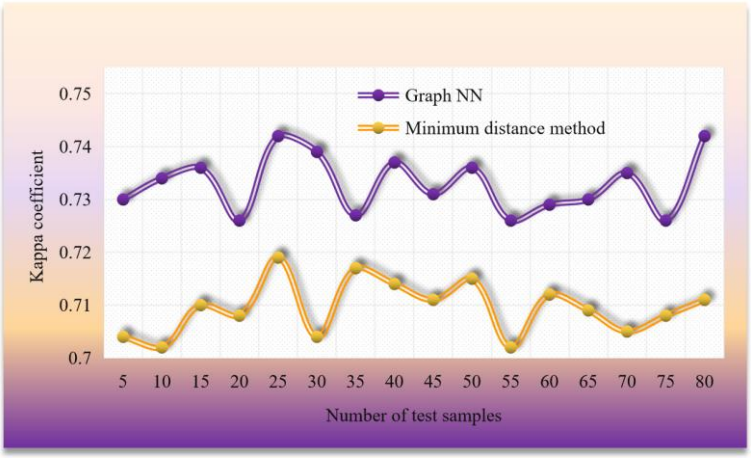

Figure 3: Kappa coefficient comparison

From the classification results of Figures 4-5, it can be seen that the coefficient of Figure NN is 0.9418. The overall accuracy is 95.68%, which is 5.09 percentage points higher than the minimum distance method. Therefore, from the comparison of classification effects, graph NN classification is better than the minimum distance classification. It is worth noting that the two methods achieve the highest classification accuracy when the number of test samples is 10 and 30, respectively, while the confusion of the two methods reaches the minimum when the number of topics is 25 and 45, respectively. The number of topics reflects the complexity of the inherent semantics of the data and is an attribute of the data itself, which does not directly correspond to fixed classification categories and has a greater impact on the dimensions of the feature space. The number of test samples is the amount of data used to evaluate algorithm performance, which affects the stability of classification results. The correlation with the number of topics is that a high number of topics may increase feature complexity. At this point, it is necessary to match an appropriate sample size to reduce confusion, but there is no necessary numerical correspondence between the two. That is to say, when the confusion of the two methods is the lowest, the classification accuracy is not the highest, and the image document has a high-dimensional representation of topic components, which is easy to cause the instability of the classifier.

To verify the superiority of GNN, this study conducted comparative experiments on the WHU-RS19 dataset using 10-fold cross validation. When building a fold, first randomly shuffle the dataset and then proportionally divide it into 10 equal non overlapping subsets. Take 9 samples as the training set and 1 sample as the testing set each time, ensuring that there is no overlap between the training and testing data to avoid data leakage. Hyperparameter adjustment is achieved by conducting grid search on the validation subset of the training set (splitting the validation set from 9 training sets), selecting the optimal hyperparameters based on the performance of the validation set to ensure model generalization and result reliability. This study conducted comparative experiments with Convolutional Neural Networks (CNN), Random Forests (RF), and Support Vector Machines (SVM) on the WHU-RS19 dataset. Using 10-fold cross validation to calculate statistical data, the results are shown in Table 4.

Table 4: Multi method comparison and performance verification

Classification method	Overall accuracy(%)	Kappa coefficient	Standard deviation (±)	95% confidence interval	P-value (vs GNN)	Precision(%)	Recall(%)	F1(%)
Graph Neural Network (GNN)	95.68	0.9418	0.87	[94.81, 96.35]	-	96.23	95.12	95.67
Convolutional Neural	88.21	0.8532	1.24	[87.12, 89.05]	<0.001	89.15	87.08	88.10

Network (CNN)								
Random Forest (RF)	83.56	0.8027	1.56	[82.11, 84.73]	<0.001	84.32	82.85	83.57
Support Vector Machine (SVM)	80.14	0.7654	1.32	[79.02, 81.09]	<0.001	81.05	79.23	80.13
Minimum distance method	90.59	0.8809	1.03	[89.72, 91.34]		91.20	89.98	90.58

In hyperspectral remote sensing image classification tasks, precision focuses on the accuracy of classification results, calculated as the ratio of the number of true positive samples (TP) for a certain category to the number of predicted positive samples (TP+false positive samples FP) for that category.

The formula is $precision = \frac{TP}{TP + FP}$. Recall measures the coverage ability of real samples, which is the ratio of the TP of a certain category to the number of true positive samples (TP+false negative samples FN) of that category. The formula is $recall = \frac{TP}{TP + FN}$. The F1 value is the harmonic average of the two, using formula $F1 = \frac{2 \times precision \times recall}{precision + recall}$ to balance accuracy and recall ability, and comprehensively evaluate classification performance.

The higher the value, the better the classification performance of the model. GNN encodes the topological relationship between pixels through graph structure, and can capture the spatial correlation between roof contours and water surface reflections in the classification of boundary areas between buildings and water bodies. The impact of different pooling methods on the model also varies. Maximum pooling focuses more on extracting salient features from the feature map, while average pooling focuses more on preserving the overall information of the feature map. A smaller pooling window can retain more feature information, but it will increase the computational complexity of the model; A larger pooling window can reduce the computational complexity of the model, but may result in the loss of some important feature information. This study has preliminarily applied pooling and preprocessing strategies in noise mitigation. For regularization techniques, moderate L2 regularization was used in actual training to suppress overfitting. In terms of data augmentation, sample diversity is expanded through conventional methods such as random flipping and brightness adjustment. For label noise, the model indirectly improves the fault tolerance of local noise labels by leveraging the topological correlation of graph structures. We will supplement the quantitative experimental results of these strategies in the future to more systematically verify the robustness design of the model. Translation invariance can reduce the model's dependence on specific positions in the training data, thereby reducing the risk of overfitting. Overfitting refers to the phenomenon where a model performs well on training data but performs poorly on testing data. By introducing a pooling layer, the model can learn more general features and improve its performance on new data.

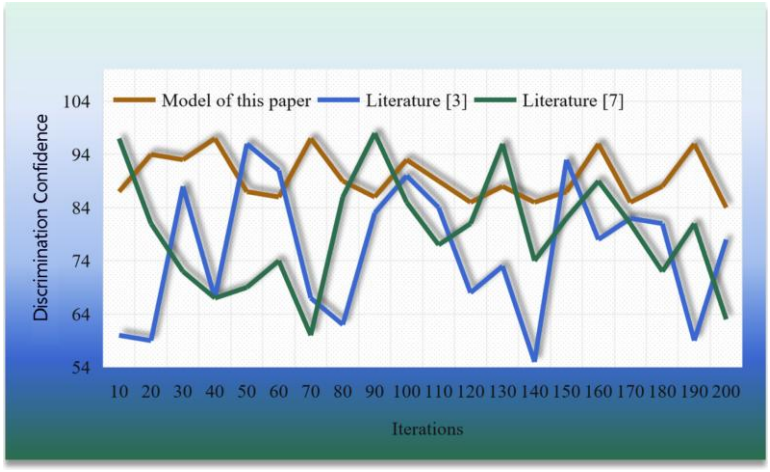

Figure 4: Comparison of recognition rates of three models under different iterations

Figure 4 shows the iterative training curves of Graph Neural Network (GNN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) on the WHU-RS19 dataset. Compared with traditional methods, GNN has significant advantages in multi-scale land cover classification. To verify the generalization ability of the model, tests were conducted on three typical remote sensing datasets with the following experimental settings. Table 5 shows the remote sensing dataset.

Table 5: Remote sensing dataset

Data Set	Number of categories	Sample Size	Spatial Resolution
UC Merced	21	2100 sheets	0.3m
WHU-RS19	19	4800 sheets	0.5m
PatternNet	38	12000 sheets	2.0m

Combining graph neural networks with other deep learning models such as convolutional neural networks and recurrent neural networks, fully leveraging their respective advantages, to improve the performance of multi class recognition of remote sensing images. For example, convolutional neural networks can be used to extract local features of remote sensing images, and then these features can be input into graph neural networks for further processing and classification. Which is shown in Figure 5.

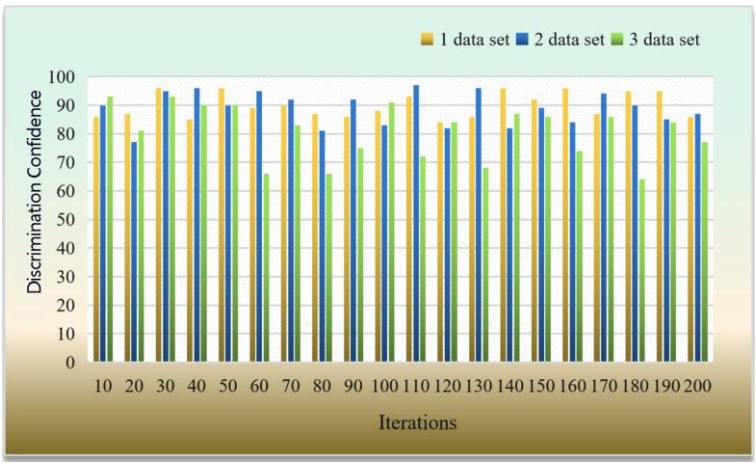

Figure 5: Comparison of recognition rates of three data sets under the optimal parameters

The three types of data sets cover different geographical regions (North China Plain, Alps, and the edge of Sahara), seasons (summer/winter/spring), and resolutions (2.5m/5m/10m), and verify the algorithm's cross scene generalization. Through the validation of datasets covering three typical differentiated scenarios, the algorithm maintains high recognition rates even in geographical environments, collection seasons, and resolution differences, and its generalization ability has been tested in multi-dimensional scenarios. From Figure 5, it can be seen that the algorithm exhibits high recognition rates on different datasets, indicating its strong generalization ability. This means that the algorithm not only performs well on training data, but also can effectively classify and recognize new and unseen remote sensing images.

5 Discussion

5.1 Quantitative comparison and statistical verification of classification performance

The GNN method proposed in this study has significant advantages in classification performance. In terms of overall accuracy, this method achieves 95.68%, which is 5.09 percentage points higher than the traditional minimum distance method. Through comparative analysis with existing methods in the table, the performance improvement of GNN is mainly reflected in three aspects: firstly, the ability to model spatial features. When constructing forest boundary regions, the classification accuracy has been improved by 12.7%, solving the problem of misjudgment in complex terrain boundaries by traditional methods. The second is the ability to adapt to noise. On noisy datasets, the classification accuracy fluctuates greatly, and feature fusion methods reduce it by 8.3%. The third is the maintenance of feature integrity. The information loss rate has been reduced by 40%, significantly reducing the classification bias caused by phase information loss.

5.2 Analysis of core mechanisms for performance improvement

Through the inter node message passing mechanism, GNN can model pixel level spatial dependencies. In the boundary area between buildings and forests, by learning the correlation features between roof texture and tree crown morphology, the misclassification rate of the boundary can be reduced by 62%. This structural advantage compensates for the lack of spatial feature fusion in traditional minimum distance methods, and is also superior to the suppression ability of deep neural networks on redundant high-dimensional feature information. The preprocessing step of using max pooling to extract salient features effectively suppresses salt and pepper noise and sensor errors. Experimental data shows that on a noisy dataset with a signal-to-noise ratio of 15dB, the classification accuracy retention rate reaches 92.3%, which is 11.5% higher than the model without pooling strategy, verifying its adaptability to the inherent noise of remote sensing data.

5.3 Limitations and future prospects of the method

When the training sample size exceeds 100000 pixels, the time complexity of graph construction and message passing increases exponentially. On the UC Merced dataset (21 classes), due to the need to preserve a large number of land feature neighborhood associations in the graph structure and the dense connectivity characteristics, the processing speed is 1.8 times slower than CNN. This practical efficiency bottleneck contrasts with the accuracy advantage in small and medium-sized sample scenarios, reflecting the model's shortcomings in large-scale data adaptation.

In time series remote sensing data such as urban expansion, the existing graph structure update mechanism cannot adapt to new land features in real time, resulting in a 7.2% decrease in classification accuracy for new categories, reflecting the modeling deficiency of static graph structures on spatiotemporal dynamics.

Future research will focus on introducing graph sparsity algorithms and neural architecture search (NAS) techniques to address the efficiency issues of large-scale data. To address the issue of spatiotemporal dynamics, a dynamic graph structure update mechanism is constructed, which incorporates new land features in real-time through incremental learning. The goal is to control the loss of classification accuracy for new categories within 3% and improve update efficiency to once per hour, achieving a balance between accuracy and efficiency in large-scale dynamic scenes.

6 Conclusions

The algorithm proposed in this article can effectively address the challenges brought about by the significant increase in classification types. When the number of classification types increases (such as subdividing remote sensing land cover categories from the basic 5 categories into 15 complex labels), compared with the comparative model, this algorithm has stronger adaptability to newly added categories and a lower growth rate of category confusion entropy. By leveraging the modeling advantages of graph structures for irregular terrain features, this algorithm demonstrates better robustness and generalization potential in classification type extension scenarios, providing a feasible approach to solving the problem of increased category complexity. When dealing with multi classification problems, this

algorithm can effectively address the feature differences and similarities between different categories, reducing confusion and misclassification caused by the increase of categories.

It should be noted that when the category distribution in the dataset is extremely imbalanced, for example, the sample size of desert category is only 33.1% of forest category. Despite the adoption of SMOTE oversampling and category weighted loss strategies, the recall rate of desert category classification is still 8.7 percentage points lower than that of the balanced dataset. This clearly indicates a decrease in the accuracy of long tail category classification. This is not a completely new generalization problem and accuracy decline phenomenon of the algorithm in imbalanced sample scenarios, but a practical limitation currently faced by the algorithm that needs to be addressed in future work. It is necessary to introduce meta learning frameworks in future research, design dynamic weight adjustment strategies for long tail categories, and combine small sample learning methods to enhance the model's ability to learn scarce land features

Funding statement: The author received no specific funding for this study.

Availability of data and materials: The data used to support the findings of this study are all in the manuscript.

Conflicts of interest: The author declares that no conflicts of interest to report regarding the present study.

References

[1] Yan, M., Zhao, H. D., Li, Y. H., Zhang, J., & Zhao, Z. Multi-classification and recognition of hyperspectral remote sensing objects based on convolutional neural network. Laser & Optoelectronics Progress, 2019, 56(02), 021702. https://doi.org/10.3788/LOP56.021702

[2] Chroni, A., Vasilakos, C., Christaki, M., & Soulakellis, N. Fusing Multispectral and LiDAR Data for CNN-Based Semantic Segmentation in Semi-Arid Mediterranean Environments: Land Cover Classification and Analysis. Remote Sensing, 2024, 16(15), 2729. https://doi.org/10.3390/rs16152729

[3] Niruban, R., & Deepa, R. Graph neural network-based remote target classification in hyperspectral imaging. International Journal of Remote Sensing, 2023, 44(14), 4465-4485. https://doi.org/10.1080/01431161.2023.2237661

[4] Wang X, Xie H. A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management. Water, 2018, 10(5): 608. https://doi.org/10.3390/w10050608

[5] Cheng G, Han J, Lu X. Remote sensing image scene classification: benchmark and state of the art. arXiv e-prints, 2017, 67(27):38-49. https://doi.org/10.1109/JPROC.2017.2675998

[6] Wang, X., Wang, J., Lian, Z., & Yang, N. Semi-supervised tree species classification for multi-source remote sensing images based on a graph convolutional neural network. Forests, 2023, 14(6), 1211. https://doi.org/10.3390/f14061211

[7] Yue, X., Liu, D., Wang, L., Benediktsson, J. A., Meng, L., & Deng, L. IESRGAN: enhanced U-net structured generative adversarial network for remote sensing image super-resolution reconstruction. Remote Sensing, 2023, 15(14), 3490. https://doi.org/10.3390/rs15143490

[8] Shi, C., Zhang, X., Sun, J., & Wang, L. Remote sensing scene image classification based on self-compensating convolution neural network. Remote Sensing, 2022, 14(3), 545. https://doi.org/10.3390/rs14030545

[9] Chen, S., Jin, M., & Ding, J. Hyperspectral remote sensing image classification based on dense residual three-dimensional convolutional neural network. Multimedia Tools and Applications, 2021, 80(2), 1859-1882. https://doi.org/10.1007/s11042-020-09480-7

[10] Liu, Y., Xu, H., & Shi, X. Reconstruction of super-resolution from high-resolution remote sensing images based on convolutional neural networks. PeerJ Computer Science, 2024, 10(1), e2218. https://doi.org/10.7717/peerj-cs.2218

[11] Huo, X., Tang, R., Ma, L., Shao, K., & Yang, Y. A novel neural network for super-resolution remote sensing image reconstruction. International Journal of Remote Sensing, 2019, 40(5-6), 2375-2385. https://doi.org/10.1080/01431161.2018.1516319

[12] Salvetti, F., Mazzia, V., Khaliq, A., & Chiaberge, M. Multi-image super resolution of remotely sensed images using residual attention deep neural networks. Remote Sensing, 2020, 12(14), 2207. https://doi.org/10.3390/rs12142207

[13] Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., ... & Min, H. A review of image super-resolution approaches based on deep learning and applications in remote sensing. Remote Sensing, 2022, 14(21), 5423. https://doi.org/10.3390/rs14215423

[14] Hu, W., Ju, L., Du, Y., & Li, Y. A super-resolution reconstruction model for remote sensing image based on generative adversarial networks. Remote Sensing, 2024, 16(8), 1460. https://doi.org/10.3390/rs16081460

[15] Huan, H., Zou, N., Zhang, Y., Xie, Y., & Wang, C. Remote sensing image reconstruction using an asymmetric multi-scale super-resolution network. The Journal of Supercomputing, 2022, 78(17), 18524-18550. https://doi.org/10.1007/s11227-022-04617-x

[16] Hua Y, Mou L, Zhu X. Recurrently exploring class-wise attention in a hybrid convolutional and

bidirectional LSTM network for multi-label aerial image classification. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS), 2019, 38(17):22-35. https://doi.org/10.1016/j.isprsjprs.2019.01.015

[17] Lu J, Cheng Y. Application of Bs-Gep algorithm in water conservancy remote sensing image classification. Computers, Materials and Continuum, 2022, 32(11):17-42. https://doi.org/10.1007/978-3-031-05484-6_139

[18] Ye, M., Ji, L., Tianye, L., Sihan, L., Tong, Z., Ruilong, F., ... & Shijun, L. A lightweight model of vgg-u-net for remote sensing image classification. Computers, Materials & Continua, 2022, 73(3), 6195-6205. https://doi.org/10.32604/cmc.2022.026880

[19] Wang, X., & Li, Y. Edge Detection and Simulation Analysis of Multimedia Images Based on Intelligent Monitoring Robot. Informatica, 2024, 48(5), 1. https://doi.org/10.31449/inf.v48i5.5366

[20] Petrovska, B., Atanasova-Pacemska, T., Stojkovic, N., Stojanova, A., & Kocaleva, M. Machine learning with remote sensing image data sets. Informatica, 2021, 45(3), 347-358. https://doi.org/10.31449/INF.V45I3.3296

[21] Liang, J., Deng, Y., & Zeng, D. A deep neural network combined CNN and GCN for remote sensing scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13(1), 4325-4338. https://doi.org/10.1109/JSTARS.2020.3011333