Ecological Capacity Assessment of Forest Park Tourists by Integrating Attention Mechanisms and Remote Sensing Images

Qiaofeng Zhou
Cultural Creativity and Tourism Department, Yuncheng Vocational and Technical University, China
qiaofengzhou227@outlook.com

Abstract: *Traditional methods for assessing tourist ecological capacity in forest parks often use single indicators or static models, failing to capture dynamic ecological changes. This study introduces a novel approach leveraging fusion Attention Mechanisms (AM) and Remote Sensing (RS) images to evaluate the ecological capacity of visitors in forest parks. The method uses high-resolution RS images and deep learning to accurately measure the impact of tourist activities on the ecological environment. The proposed method includes an image reconstruction technique that integrates AM with RS data. Performance analysis and validation are conducted, showing an average accuracy of 98.76%, a recall rate of 98.17%, and an F1 score ranging from 96.23% to 98.25% in the Fujian Regional Remote Sensing Image Dataset for Scene classification (FJ-RSIDS). Applying this method to Qilian Mountain Park, the study predicts ecological capacities of 31.43 million for the environment, 104.83 million for space, 115.74 million for tourist psychology, 334,200 for facilities, and 64.84 million for tourism. These predictions closely align with observed values. This research provides a scientific basis and technical support for ecological protection and tourist management in forest parks, contributing significantly to the sustainable development of the tourism industry.*

Keywords: *Attention mechanisms, remote sensing imagery, tourist, ecological capacity.*

Received March 4, 2025; accepted June 23, 2025
https://doi.org/10.34028/iajit/22/5/9

1. Introduction

As the global tourism industry experiences rapid growth, forest parks, which are vital natural tourism assets, not only shoulder the crucial responsibility of ecological conservation but also encounter a surge in tourist demand [12]. Tourism environmental capacity, also referred to as tourism ecological capacity, denotes the maximum number of tourists that a tourist destination or area's environmental space can accommodate without inflicting irreversible harm on the ecosystem [29]. From past research, traditional methods for studying tourist ecological capacity mainly rely on field investigations and simple mathematical model construction. Early research focused on roughly estimating ecological capacity through field observations of tourist distribution and activity intensity, combined with basic data such as park area and number of attractions [18]. However, this method is time-consuming and labour-intensive, and is greatly affected by human factors, making it difficult to comprehensively and accurately reflect the true situation of the complex and changing park ecosystem in carrying tourists. As Remote Sensing (RS) technology booms, it has gradually been applied to the study of ecological capacity in forest parks due to its advantages such as large-scale synchronous observation and fast data acquisition [19]. By using RS images, researchers can obtain key information such as park land use types and vegetation coverage, and evaluate the health status and carrying capacity of park ecosystems at a macro level [22]. For example, Su *et al*. [23] provided strong data support for the dynamic assessment of ecological capacity by analyzing RS images from different periods to monitor changes in park vegetation. Liu *et al*. [16] innovatively proposed an efficient image analysis algorithm in the study of RS image change subtitles based on dual branch transformers, which can more sensitively capture the subtle changes in the park ecological environment at different time periods, providing a new technical path and method reference for in-depth analysis of the influencing factors of ecological capacity, greatly expanding the application depth and breadth of RS technology in forest park ecological capacity research. However, relying solely on RS image data for feature extraction and information mining in complex scenes still has certain limitations, making it difficult to effectively distinguish the impact of different ecological elements on tourist capacity.

Attention Mechanisms (AMs) is a resource allocation scheme whose core idea is to dynamically allocate "attention" based on the current context when processing input data, in order to highlight key information [3]. Introducing AM in ecological capacity research can enable the model to pay more attention to

key ecological features closely related to tourist ecological capacity, such as rare species habitats, ecologically fragile areas, etc. A comparative study by Zhang *et al*. [32] on wheat yield estimation using attention-based deep learning and transfer strategies showed that incorporating AMs into ecological environment assessment models can significantly improve the ability to identify and analyze important ecological factors. Ai *et al*. [1] successfully improved the ability to precisely identify and delineate aquaculture regions in complex coastal ecological environments through the use of self AMs in their research on assisted loss of coastal aquaculture area extraction. Although some progress has been made in evaluating the ecological capacity of forest park visitors, there is still a lack of a comprehensive research method that can fully integrate the rich information and AM advantages of RS images. Based on this, this study integrates AM and RS imaging technology to deeply explore the ecological capacity of forest park visitors, to furnish a more rigorous and precise scientific foundation and practical direction for fostering sustainable tourism growth in forest parks. The innovation of the research lies in utilizing the spectral similarity of the spectral grouping attention module to group the fused gradient features, calculating attention within the group to enhance the correlation between features, and then outputting reconstructed features. The cross layer non-local attention module inputs multiple layers of features, calculates non-local attention through cross layer interaction, effectively captures global dependencies, and integrates the results with spectral grouping results. The collaborative application of two AMs can efficiently process features, capture global and local information from all angles, and significantly improve the analysis and processing capabilities of complex ecological information compared to traditional models. This research only takes the content of the relevant literature review as the theoretical basis and has no overlap with the above-mentioned works.

2. Methods and Materials

2.1. Fusion of GGS and SGA for RS Image SR Reconstruction Method

Spatial Attention Mechanisms (SAM) is an AM that focuses on the spatial dimension of Feature Maps (FMs). It is mainly used to highlight the regions in images or FMs that contribute the most to the task, while suppressing irrelevant or redundant regions, thus enhancing the model's efficacy [9]. The illustrative representation of SAM is in Figure 1.

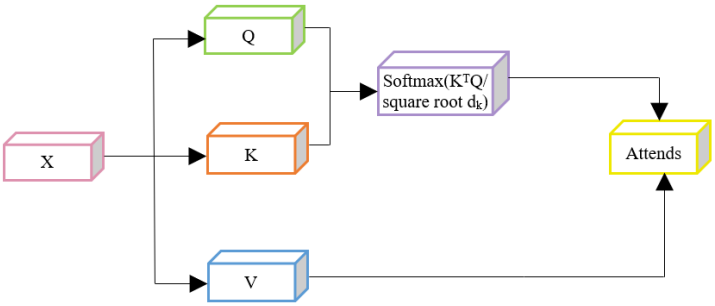


Figure 1. Illustrative representation of SAM.

In Figure 1, the input FM is represented by the feature extracted by the convolutional layer. The query vector is represented by symbol Q, the feature identification of the FM is represented by symbol K, and the actual features of the FM are represented by symbol V. The calculation of SAM is in Equation (1).

$$Attention = softmax\left(\frac{K^T Q}{d_k}\right)V \quad (1)$$

In Equation (1), the transpose of feature identification is represented by symbol K^T. The dimension of feature identification is represented by symbol d_k. To improve the performance of hyper-spectral image processing, the Spectral Group Attention (SGA) is studied and introduced. SGA is specifically designed for hyper-spectral image analysis. It enhances the model's ability to capture spectral features by implementing grouping and normalization strategies [8]. The schematic diagram of SGA structure is in Figure 2.

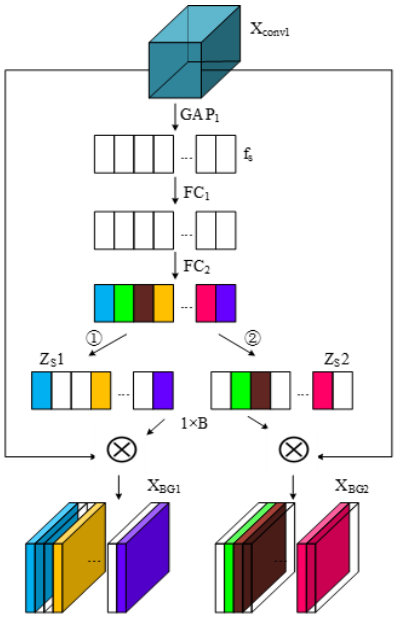


Figure 2. SGA architectural depiction.

In Figure 2, X_{BG1} and X_{BG2} refer to the spectral weighted FMs of two different spectral information contents. The mechanism groups input FMs by channel dimensions and processes each cluster via SAM [25]. The mechanism applies both global average pooling and global max pooling operations to the FMs within each cluster, and concatenates the results of global average pooling and global max pooling by channel to obtain the spectral vectors Z_{S1} and Z_{S2} [4]. The mechanism uses 1×1 convolution to compress the concatenated FMs, and normalizes the convolved FMs to the range of [0, 1] using the sigmoid activation function (AF) to obtain the spectral weight vector Z_S. The calculation formula for Z_S is in Equation (2).

$$Z_S = \sigma\left(\left(L_2 relu(L_1 f_S)\right)\right) \tag{2}$$

In Equation (2), σ represents the sigmoid AF. The weights of the first Fully Connected Layer (FCL) are represented by symbol L_1, and the weights of the second FCL are represented by symbol σ. The expression for the spectral vector Z_{S1} of the first group is in Equation (3).

$$Z_{S1} = relu(Z_S - \beta) \tag{3}$$

In Equation (3), β represents a constant vector. The expression for the spectral vector Z_{S2} of the second group is in Equation (4).

$$Z_{S2} = relu(\beta - Z_S) \tag{4}$$

Finally, the spectral attention weight matrix is multiplied element by element with the original grouped FM to obtain the weighted FM [15]. In the traditional single-branch Super Resolution (SR) network, there is usually only one main branch, which is responsible for restoring high-resolution images from low-resolution images. However, this method has limitations in detail restoration and edge sharpening. To overcome these limitations, the study introduces an additional gradient branch in the network architecture. This branch acts as a feature selector, capable of adaptively extracting structural features in the image and providing gradient prior information for the super-resolution branch [7]. The Gradient-Guided Strategy (GGS) is an innovative method. It utilizes gradient maps to guide the network to restore sharper edges and richer texture details in specific areas of the image, thereby significantly improving the quality of the recon-structed image [30]. The expression of the gradient graph $g(I_{LR})$ is in Figure 5.

$$g(I_{LR}) = \left\| I_{LR} * M_x, I_{LR} * M_y \right\|_2 \tag{5}$$

In Equation (5), the image to be reconstructed is represented by symbol I_{LR}, horizontal convolution is expressed as symbol M_x, and vertical convolution is represented by symbol M_y. The calculation formula for M_x and M_y is in Equation (6).

$$M_x = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}, M_y = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \tag{6}$$

This study uses convolution operation to extract shallow gradient feature F_0^{grand}, and the calculation of F_0^{grand} is in Equation (7).

$$F_0^{grand} = f_{conv}\left(g(I_{LR})\right) \tag{7}$$

The k-th gradient module of the gradient branch is a key component used to enhance image edges and texture details, as presented in Figure 3.

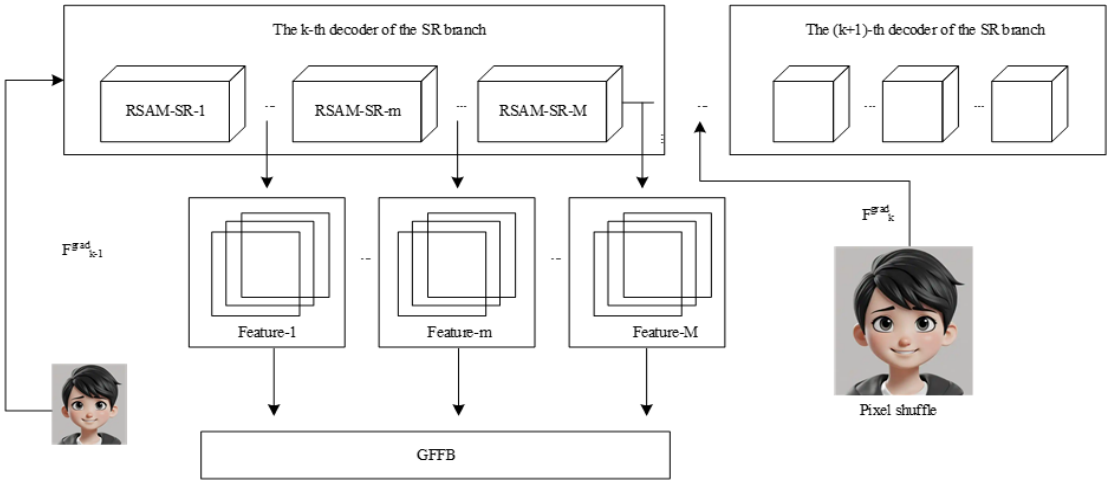


Figure 3. Illustrative representation of the kth gradient module of the gradient branch.

In Figure 3, F_{k-1}^{grand} and F_k^{grand} are input simultaneously, and then F_{k-1}^{grand} and F_k^{grand} are concatenated. The gradient information of the fused features is calculated using gradient calculation method, and further feature extraction is performed on the gradient FM through convolution operation [21]. Gradient guided loss is calculated to ensure effective utilization of gradient information. The FM processed by the gradient module F_{k-1}^{grand} will be used as the input for the k+1th gradient module. F_k^{grand} provides structural prior information for super-resolution reconstruction, helping to restore sharp edge and texture

features [11]. The calculation formula for F_k^{grand} is in Equation (8).

$$F_0^{grand} = f_{conv}\left(f_{subpix}\left(H_{GE}\left(F_{k-1}^{grand}, F_{de,k}^1, F_{de,k}^2, \ldots, F_{de,k}^M\right)\right)\right) \quad (8)$$

In Equation (8), $F_{de,k}^M$ represents the output function. The illustrative representation of the SR network integrating GGS and SGA is in Figure 4.

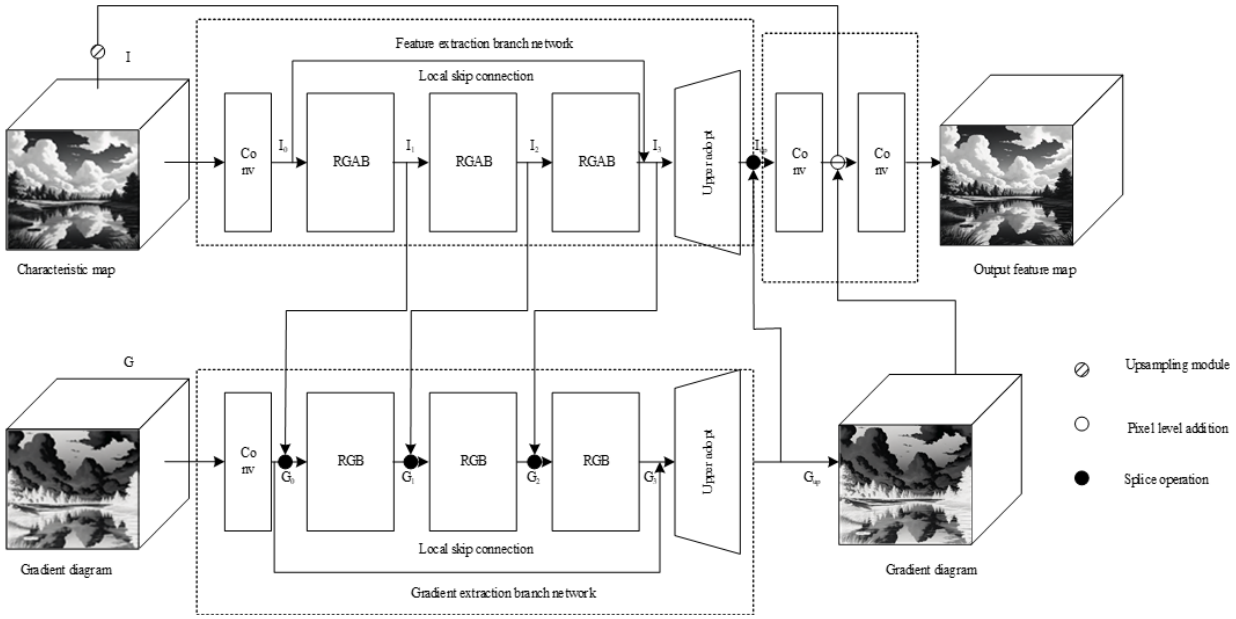


Figure 4. SR network diagram of fusion gradient guidance and grouping attention.

In Figure 4, the introduction of gradient information boosts the network's capacity for perceiving image edges and texture details [33]. The group AM divides feature channels into multiple groups and applies AMs independently within each group, allowing for more flexible processing of feature information from different channels. To better handle image features of various scales, the network designs a multi-scale fusion module. By integrating features of various scales, it can more comprehensively capture global and local information of the image, thereby improving the quality of reconstructed images.

2.2. Ecological Capacity Assessment of Forest Park Tourists Based on SGA and RS

The study enhances image edge and texture detail perception by introducing gradient information, flexibly processes feature channel information using group AM, and comprehensively captures global and local information of the image through multi-scale fusion module, thereby significantly improving the quality of reconstructed images. To achieve better image SR performance, a model construction of an SR network that integrates gradient guidance and group attention is studied, as shown in Figure 5.

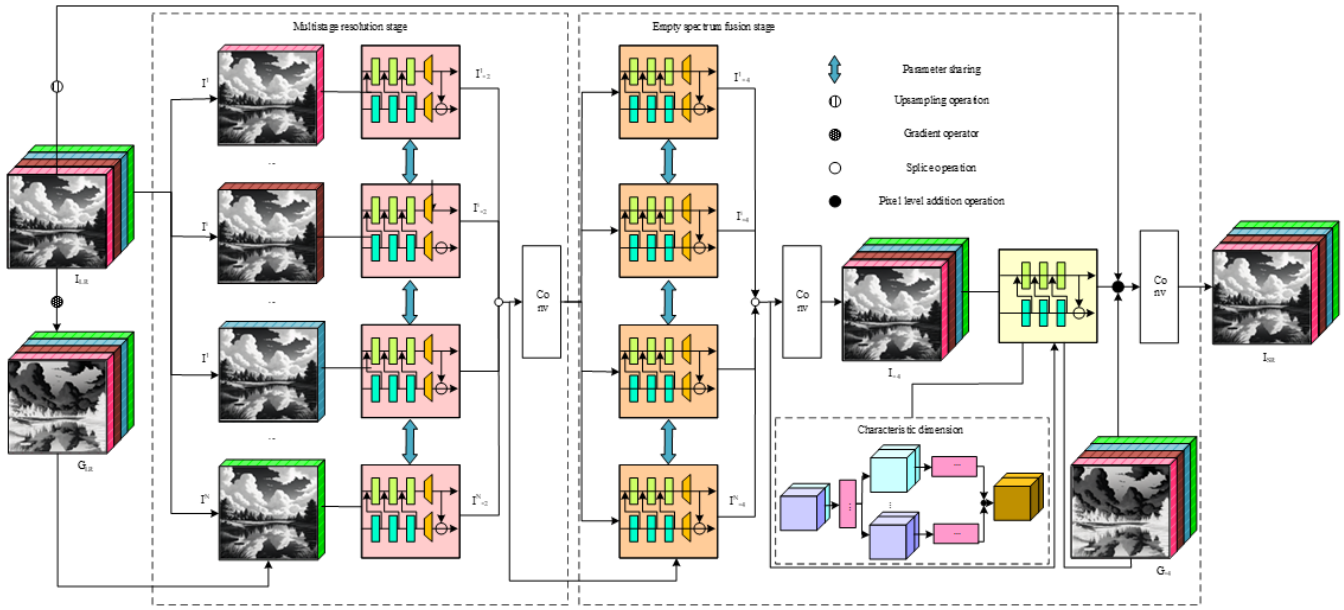


Figure 5. SR network model integrating gradient guidance and grouping attention.

In Figure 5, the study enhances the network's ability to perceive details by extracting multiple sub-pixel information from the image. Meanwhile, utilizing Transformer architecture to enhance feature representation further improves the network's global perception capability. Extracting structural details of the image through gradient branches and fusing them with the features of the main branch can enhance the preservation of edge information in the process of SR reconstruction [26]. In RS image SR tasks, utilizing self-similarity can significantly improve the quality of reconstructed images [27]. Self-similarity refers to the characteristic of similar patterns in an image repeatedly appearing at different scales or within the same scale. Gated units serve as a pivotal mechanism, extensively employed within Recurrent Neural Networks (RNNs) and their diverse variants to enable meticulous regulation of information flow [13]. By introducing gate control mechanisms, gate units can dynamically determine which information needs to be retained, updated, or forgotten, effectively solving the challenges of gradient disappearance and explosion encountered by conventional RNNs when dealing with extended sequence data [14]. Therefore, the study introduces the sub-unit and utilizes its repeated appearance characteristics at different scales to generate fused feature Z. Assuming that the dimension of N input FMs for different layers are N×H×W, the input FM is (x^i, x^j), i and j represent the indices of the input FMs, and the expression for the fused feature Z is in Equation (9).

$$Z = f_{gate}([x^1, x^2, \ldots, x^N]) \tag{9}$$

The formula for calculating the weighted sum y_i of features in the input FM (x^i, x^j) is in Equation (10).

$$y_i = \frac{1}{\theta(Z_i, X)} \sum_{x \in X} \sum_{\forall j} \exp\left(\mu(Z_i, x_j)\right) \phi(x_j) \tag{10}$$

In Equation (10), $\theta(Z_i, X)$ is the normalized normalization quantity, and its computation is in Equation (11).

$$\theta(Z_i, X) = \sum_{x \in X} \sum_{\forall j} \exp\left(\mu(Z_i, x_k)\right) \tag{11}$$

The single pixel of RS images usually covers a large actual area, leading to the erosion of fine details in the image, making SR reconstruction particularly difficult [10]. Therefore, the research integrates features from different depth levels through cross layer non local AMs, fully utilizing global and local information, and enhancing the capacity to express features in intricate scenes. The flowchart of the cross layer non local AM is in Figure 6.

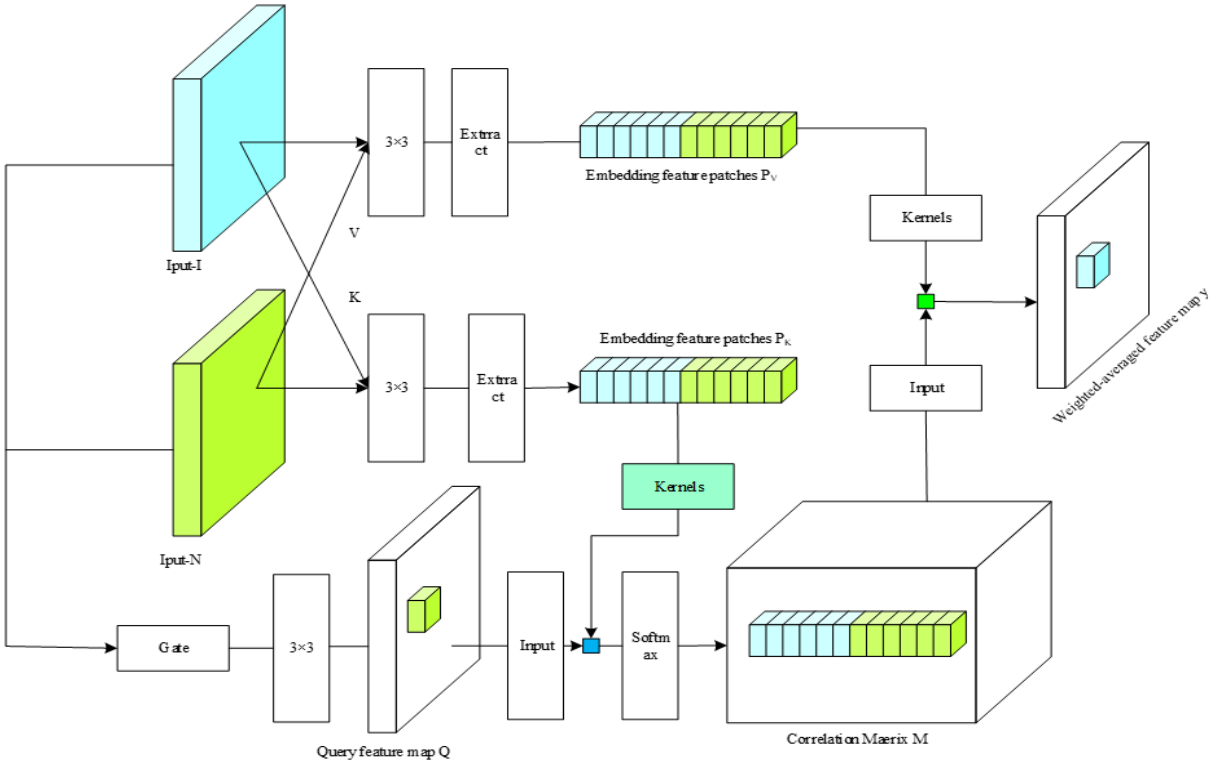


Figure 6. Flow chart of cross-layer non-local AM.

In Figure 6, in the cross layer non-local AM, the first step is to embed the input FM into a new space, which is achieved through 1×1 convolution to modify the channel count in the FM to a smaller dimension. Non-local AM is applied on the fused FM to capture global contextual information. Three FMs are generated for the fused FM through convolution or FCLs, Q, K, and V. The similarity matrix between Q and K is calculated, and residual connections between non-local attention features and the original fusion features are performed to obtain the final FM. As an important ecological tourism destination, the evaluation of tourist ecological capacity in forest parks is crucial for achieving environmental protection and sustainable development.

Tourist ecological capacity is a complex comprehensive indicator, and current calculations mainly focus on spatial capacity, neglecting other aspects of measurement. Therefore, the RS images are analyzed using a fusion of cross layer non-local AM and group AM. The process of the research model is in Figure 7.

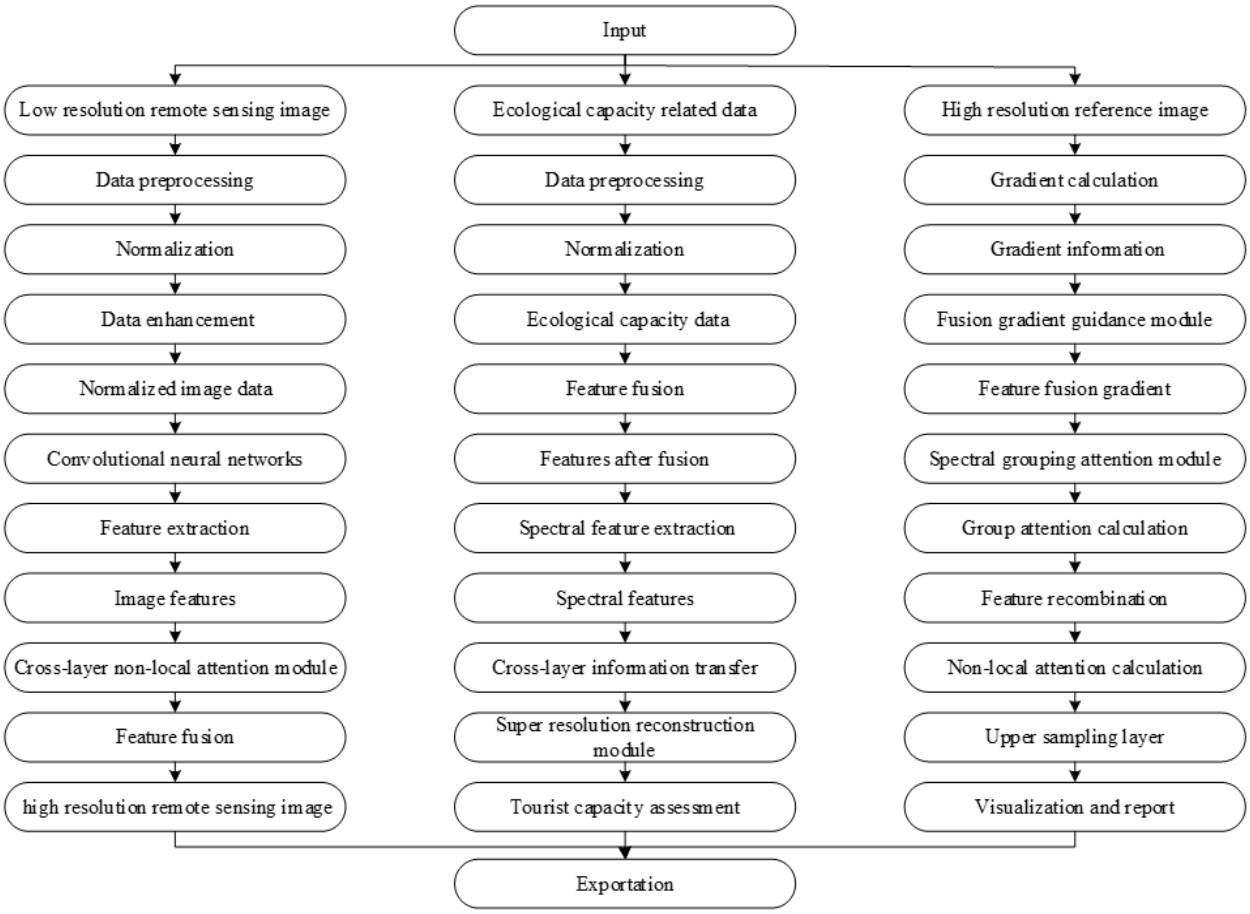


Figure 7. Model flow chart.

In Figure 7, the input module includes low resolution RS images, ecological capacity related data, and high-resolution reference images. The feature extraction module includes image feature extraction, ecological capacity feature fusion, and spectral feature extraction. The fusion gradient guidance module includes fusion gradient information and feature extraction output, enhanced detail perception, and output fusion gradient features. The spectral grouping attention module combines gradient features by grouping based on spectral similarity, calculates attention enhancement associations within the group, and outputs reconstructed features. The cross layer non-local attention module takes multiple layers of features as input, captures global dependencies through cross layer interaction and non-local attention, and integrates the fusion results with spectral grouping results. The SR reconstruction module includes input recombination features, upsampling amplification, and refinement through a reconstruction network to generate high-resolution images for ecological capacity assessment. Root Mean Square Error (RMSE) serves as a statistical measure to quantify the discrepancy between observed and actual values [24]. The study uses RMSE to measure the capability of the research model, and the computation is in Equation (12).

$$RMSE = \sqrt{\frac{1}{abc}\sum_{i,j,k}(I_{SR} - I_{HR})^2} \tag{12}$$

In Equation (12), I_{SR} represents the input hyper-spectral RS image, I_{HR} represents the output hyper-spectral RS image, a represents the length of the hyper-spectral RS image, b is the width of the hyper-spectral RS image, and c represents the number of bands in the hyper-spectral RS image. Peak Signal to Noise Ratio (PSNR) refers to the evaluation of image quality by comparing the disparities between the initial and processed images [28]. The PSNR calculation formula is in Equation (13).

$$PSNR = 20\log_{10}\frac{Max(I_{HR})}{\sqrt{MSE(I_{HR}, I_{SR})}} \tag{13}$$

In Equation (13), *Max* represents the maximum value of the real RS image. The Structural Similarity Index (SSIM) is an image quality assessment index based on the human visual system, used to measure the similarity between two images in terms of brightness, contrast, and structure. The calculation formula for *SSIM* is in Equation (14).

$$SSIM = 20\log_{10}\frac{2\rho_g\rho_o + D_1}{\rho_g{}^2 + \rho_o{}^2 + D_1} * \frac{2\eta_g\eta_o + D_2}{\eta_g{}^2 + \eta_o{}^2 + D_2} \quad (14)$$

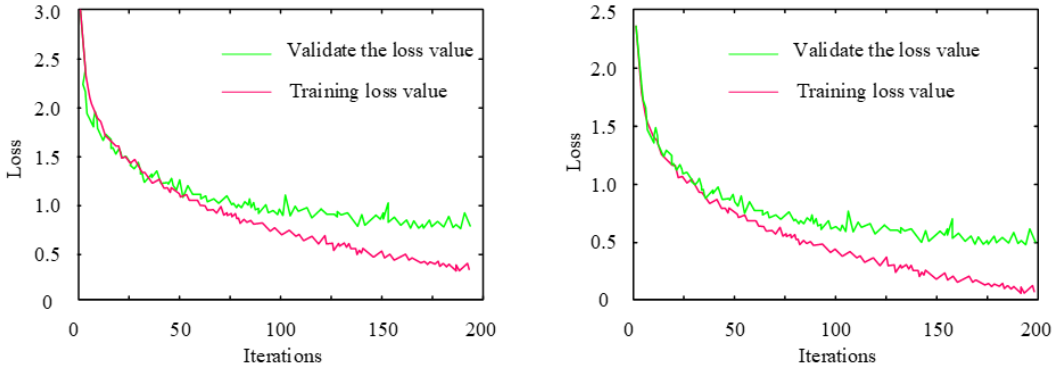
In Equation (14), the average brightness of images g and o is expressed as symbols ρ_g and ρ_o, the standard deviation of images g and o is expressed as symbols η_g and η_o, and the constants used to stabilize the denominator are expressed as symbols D_1 and D_2. SAM is a commonly used RS image classification method, which is based on the basic idea of maintaining consistent spectral information of the same type of land cover. The SAM expression is in Equation (15).

$$SAM = arccos\left(\frac{I_{SR}^T - I_{HR}}{\sqrt{I_{SR}^T I_{SR}} \cdot \sqrt{I_{HR}^T I_{HR}}}\right) \quad (15)$$

3. Results

3.1. Performance Testing of RS Image SR Network Integrating SGA and GGS

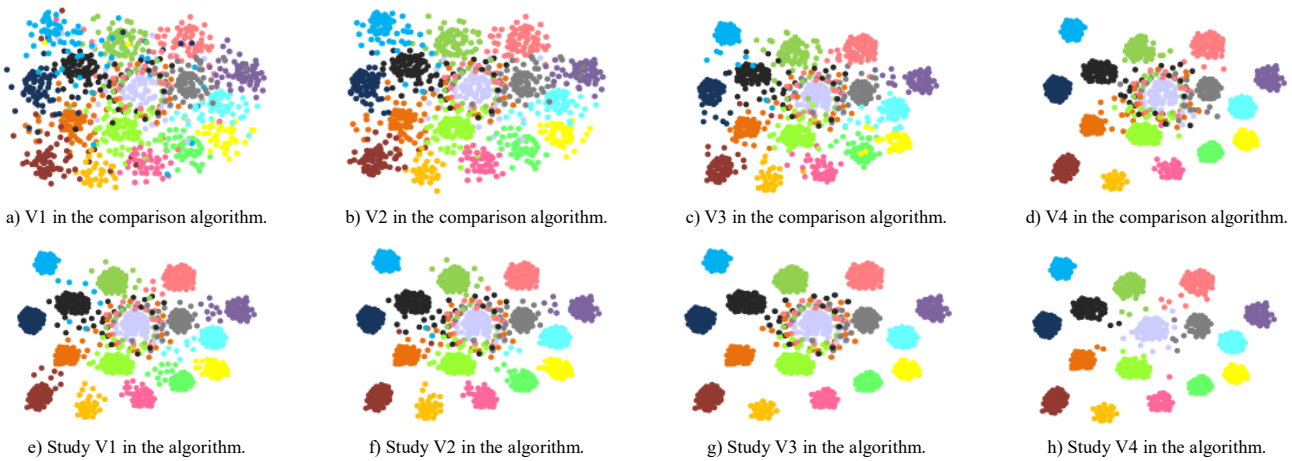
The Aerial Image Dataset (AID) dataset is a large benchmark dataset for RS image scene categorization, issued by Huazhong University of Science and Technology (HUST) and Wuhan University (WHU) in 2017 [5]. The dataset collected and annotated over 10000 high-resolution aerial scene images from Google Earth imagery [2]. To confirm the capability of the raised algorithm, the training ratios of the dataset were set to 20% and 50%, respectively, and the corresponding results were recorded, as shown in Figure 8.



a) 20 percent training ratio. b) 50 percent training ratio.

Figure 8. Loss curve generated by the research algorithm on AID dataset.

Figure 8-a) indicates the loss curve of the algorithm on the AID dataset under a 20% training ratio. From Figure 8-a), as the iteration count rose, both the training loss and validation loss of the algorithm showed a steady downward trend. When the iteration count was 200, the training loss and real loss reached a minimum of 0.4 and 0.9, respectively. Figure 8-b) shows the loss curve of the algorithm on the AID dataset under a 50% training ratio. From Figure 8-b), when the iteration count reached 200, the training loss reached the lowest value of 0.1, while the validation loss dropped to 0.5. To visually display the distribution of different categories of images in the feature space and explore the impact of AMs on different levels of features, ResNet18 was selected as the backbone network to extract features. On the AID dataset, different levels of feature vectors V1, V2, V3, and V4 were visualized using research algorithms and research algorithms without AMs. V1 represents the highest level feature vector, followed by V2, while V3 and V4 represent lower and lowest level feature vectors, respectively. The outcomes are in Figure 9.



a) V1 in the comparison algorithm. b) V2 in the comparison algorithm. c) V3 in the comparison algorithm. d) V4 in the comparison algorithm.

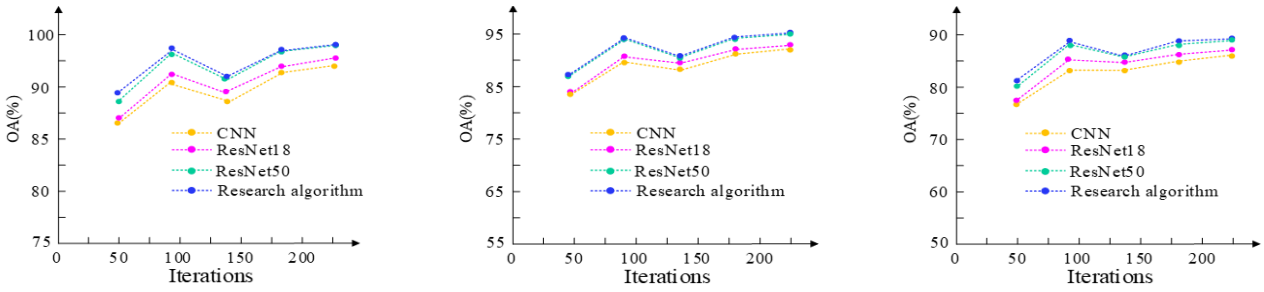
e) Study V1 in the algorithm. f) Study V2 in the algorithm. g) Study V3 in the algorithm. h) Study V4 in the algorithm.

Figure 9. Visualization of feature distribution at different scales of AID dataset.

Figure 9-a) is an illustrative representation of the V1 visualization of the comparative method, which showed

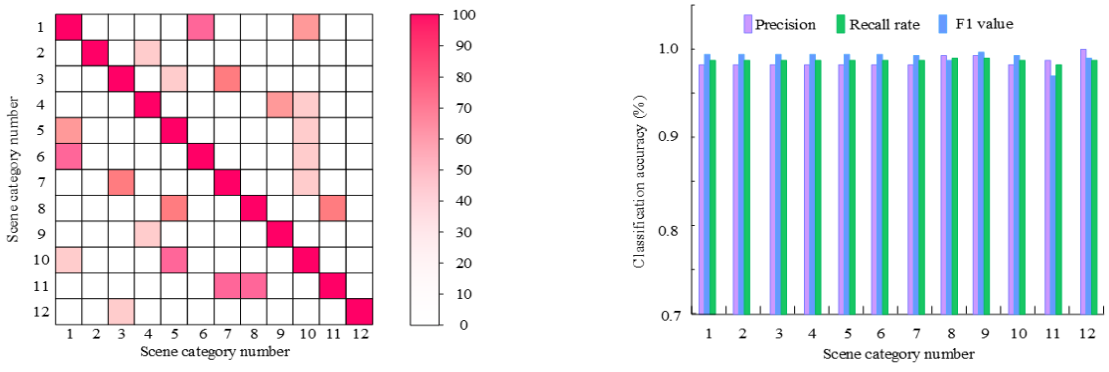
significant confusion and overlap in classification performance. Figure 9-b) is a visualization diagram of the V2 comparison method, which showed an improvement in classification performance compared to V1. Figure 9-c) is a schematic diagram of the V3 visualization of the comparative method, with its effect located before V2. Figure 9-d) indicates the V4 visualization diagram of the comparative method, which presented a better effect. Figure 9-e) indicates the visualization of V1 for the research method, which had a classification effect second only to V3 for the comparison method. Figure 9-f) indicates the visualization of V2 for the research method, which had a higher classification effect than V4 for the comparison method. Figure 9-g) indicates the visualization of V3 for the research method, which had a better classification effect; Figure 9-h) indicates the V4 visualization diagram of the research method, where the inter class distance of samples from different categories significantly increased and the intra class clustering was stronger. To examine the effects of varying degrees of FMs on overall classification accuracy, a comparison was made between the classification accuracy of different classification algorithms Convolutional Neural Network (CNN), ResNet18, ResNet50, and the classification accuracy of the research algorithm, as shown in Figure 10.

Figure 10-a) indicates the comparison results of the classification accuracy of different classification algorithms on low-level feature dimension maps. From Figure 10-a), the research algorithm had the highest classification accuracy, with an average of about 93.2%. Figure 10-b) indicates the comparison results of the classification accuracy of different classification algorithms on the mid-level feature dimension map. From Figure 10-b), the CNN algorithm had the lowest classification accuracy, with an average classification accuracy of about 86.5%, while the research algorithm had the highest average classification accuracy, at 92.8%. Figure 10-c) shows the comparison results of the classification accuracy of different classification algorithms on high-level feature dimension maps. From Figure 10-c), although the classification accuracy of the research algorithm in high-level feature dimension maps was lower than that in low to medium dimensions, it was much higher than the average classification accuracy of CNN, which was 89.3%. The Fujian Regional Remote Sensing Image Dataset for Scene classification (FJ-RSIDS) is a professional dataset focusing on the classification research of regional RS images. With Fujian region as the core research object, it provides rich high-resolution image resources for geospatial information analysis. This dataset contains approximately 5,000 RS images in total. The resolution of the original images' ranges from 0.5 to 2 meters per pixel, which can clearly capture the detailed features of ground objects. The data collection covers the entire seasonal cycle of spring, summer, autumn and winter. The vegetation coverage shows significant spatiotemporal variation characteristics in different phenological periods, providing diverse samples for the dynamic monitoring and classification research of vegetation. To verify the performance of the research algorithm in terms of classification performance, classification tests were conducted on the FJ-RIDS dataset, and the test results are in Figure 11.



a) Low level. b) Mesolevel. c) High level.

Figure 10. Accuracy difference of classifiers at different layers on the AID dataset.



a) Obfuscation matrix results. b) Accuracy, recall and F1 values.

Figure 11. The classification results of FJ-RSIDS dataset.

Figure 11-a) indicates the confusion matrix of the research algorithm on the FJ-RIDS dataset. From Figure 11-a), the confusion matrix values of playground 2, forest land 9, and salt field 12 were the highest, all at 99.7. Figure 11-b) shows the accuracy, recall rate and F1 value of the research algorithm in the FJ-RSIDS dataset. From Figure 11-b), the average accuracy of the research algorithm in the FJ-RSIDS dataset was 98.76, of which the accuracy of No. 11 coastal mudflat was the lowest, 95%. The average recall rate of the research algorithm in the FJ-RIDS dataset was 98.17%. The F1 value of the research algorithm was highest at 98.25% and lowest at 96.23. The research results indicated that the algorithm performed well on the FJ-RIDS dataset and could effectively handle SR classification tasks of hyper-spectral images.

3.2. Model Performance and Application Effect Analysis

To confirm the performance of the raised model, Bicubic and RCAN were used as comparative models for experiments, and the experimental out-comes are in Table 1.

According to Table 1, the RMSE index of the research model was lower than that of Bicubic and RCAN models in all datasets. In the Pavia Centre dataset, the RMSE value was only 0.0334, which was lower than Bicubic's 0.0432 and RCAN's 0.0352, indicating that the deviation between the forecasted outcomes of the research model and the true values was smaller and the accuracy was higher. For park management decisions, more accurate predictive data can help managers assess the current situation and changing trends of the ecological environment more accurately, thereby formulating more scientific and reasonable resource allocation and protection strategies and avoiding decision making mistakes caused by data deviations. In terms of PSNR index, the research model achieved 45.727 on the Houston dataset, higher than Bicubic's 40.672 and RCAN's 44.333, indicating that the image quality restored by the research model was higher and the noise impact was smaller. In the management of ecological parks, high-quality image data helps managers observe the vegetation coverage, topography and other conditions within the park more clearly, and then plan ecological tourism routes more effectively and assess the carrying capacity of the ecological environment, providing strong support for the balance between ecological protection and tourism development. In terms of SSIM indicators, the research model achieved 0.8495 on the Pavia Centre dataset, which was higher than Bicubic's 0.7353 and RCAN's 0.8329, indicating that the research model could better preserve the structural information of images. In the SAM index, the research model had a value of 2.7525 on the Chikusei dataset, which was lower than Bicubic's 3.6372 and RCAN's 4.1815, indicating that the research model was more accurate in spectral feature matching. For park managers, images that retain complete structural information can help them more accurately identify the ecosystem structure within the park, such as the distribution of wetlands, forests and other areas, thereby formulating more targeted ecological protection measures to safeguard the integrity and stability of the ecosystem. The computational complexity of different models is in Figure 12.

Table 1. Performance of various models on various datasets.

Data set	Pavia centre				Houston				Chikusei			
Model	Bicubic	RCAN	Research model	Increase by percentage	Bicubic	RCAN	Research model	Increase by percentage	Bicubic	RCAN	Research model	Increase by percentage
RMSE	0.0432	0.0352	0.0334	22.69%	0.0112	0.0075	0.0064	42.86%	0.0168	0.0170	0.0124	26.19%
PSNR	28.085	29.908	30.387	8.19%	40.672	44.333	45.727	12.43%	38.148	37.672	40.905	7.23%
SSIM	0.7353	0.8329	0.8495	15.53%	0.9568	0.9690	0.9779	2.20%	0.8942	0.8788	0.9348	4.54%
SAM	7.4967	6.7311	6.1359	18.15%	2.2782	2.1076	1.8356	19.43%	3.6372	4.1815	2.7525	24.33%

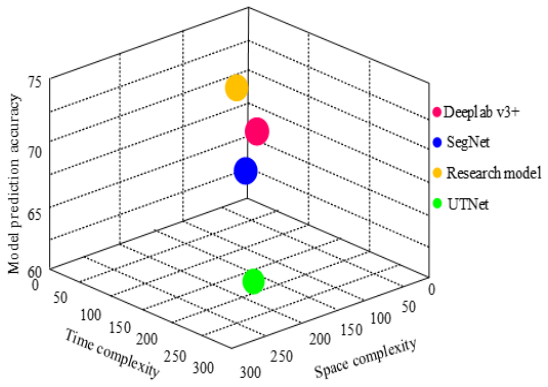
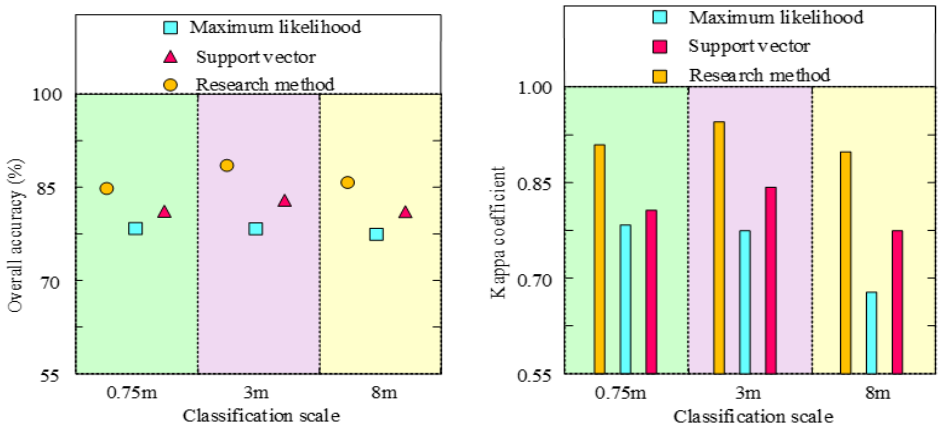


Figure 12. Computational complexity of various models.

As shown in Figure 12, the UTNet model had high time and space complexity, and its accuracy was substantially lower than the others. The SegNet model had a high prediction accuracy of 64, but had low time and space complexity, while the Deeplab v3+model had the fastest prediction speed. The research model not only had the lowest time and space complexity, but also had the highest prediction accuracy, at 69. In RS observation, the apearance of land phenomena in RS images presents different patterns with different scale units, and their spatial structure and information extraction accuracy also vary accordingly. Therefore, scale effect is a crucial factor that cannot be ignored when extracting land cover information from RS images. The study selected RS image data with resolutions of 0.75m, 3m, and 8m, and used maximum likelihood method and support vector method as comparison methods to explore the impact of scale effects on the extraction of ecological capacity

information for forest park visitors. The comparison of the extraction effects of various approaches on the ecological capacity information of forest park visitors is in Figure 13.



a) Overall Precision comparison diagram. b) Kappa coefficient comparison chart.

Figure 13. Comparison of the extraction effects of different methods on the ecological capacity information of tourists in forest parks.

Figure 13-a) shows a comparison of the over-all accuracy of extracting ecological capacity information for forest park visitors using different methods. From Figure 13-a), regardless of at 0.75m, 3m, or 8m, the research method had the highest overall accuracy in extracting ecological capacity information for forest park visitors compared to the comparison algorithm. When the resolution was 3m, the overall accuracy of the research algorithm in extracting ecological capacity information for forest park visitors was 91.2%. Figure 13-b) shows a comparison of Kappa coefficients for extracting ecological capacity information of forest park visitors using different methods. From Figure 13-b), at a 3-meter resolution, the Kappa coefficients of the three methods all reached the highest values, which were 0.94, 0.79, and 0.84, respectively. Among them, the method proposed in the study had the highest Kappa coefficient, further verifying the superiority of the research method in extracting ecological capacity information of forest park visitors. Ecological environment capacity pertains to the upper limit of tourists that a tourist area can carry without disrupting ecological balance. The spatial environmental capacity pertains to the upper limit of tourists that a tourist area can carry in space. The psychological environment capacity of tourists refers to the comfort and satisfaction they feel during the tourism process. Facility environmental capacity pertains to the upper limit of tourists that the facilities in a tourist area can serve. To confirm the effectiveness of the raised approach in forest parks, the Qilian Mountain Park was taken as an example to apply the research method to predict the ecological tourism environmental capacity of the park. The prediction outcomes are in Table 2.

According to Table 2, the research method demonstrated good predictive ability in predicting the ecological tourism environmental capacity of Qilian Mountain Forest national park. From the perspective of ecological environment capacity, the predicted value was 31.4268 million people. The predicted spatial environment capacity reached 104.8325 million people, which was highly consistent with the number of tourists that the park's vast geographic space could accommodate, reflecting the scientific and accurate consideration of spatial factors in the research method. The predicted psychological environment capacity of tourists was 115.7436 million, and the predicted facility environment capacity was 334200, which was closely related to the actual service capacity of the current facilities in the park, further verifying the reliability of the research method in predicting the facility dimension. The comprehensive calculation showed that the tourism environment capacity was 64.8402 million people, which was the result of the synergistic effect of accurate prediction of capacity in various dimensions. This strongly proved the validity and advantage of the research method in comprehensively predicting the ecological tourism environment capacity.

Table 2. Prediction outcomes of eco-tourism environmental capacity in Qilian Mountain Forest national park.

Ecotourism environmental capacity name	Real capacity (10,000 people)	Forecast capacity (10,000 people)
Ecological environmental capacity	3142.65	3142.68
Space environment capacity	10483.94	10483.25
Tourist psychological environment capacity	11574.75	11574.36
Facility environmental capacity	33.45	33.42
Tourism environmental capacity	6484.83	6484.02

4. Discussions

In the realm of studying the ecological capacity of visitors in forest parks, RS images offer a wealth of surface-related data. This invaluable information facilitates a precise assessment of various dimensions, including the ecological environment capacity, spatial environment capacity, tourists' psychological environment capacity, and facility environment capacity. However, how to extract effective features from complex and diverse RS images remains an

important challenge in current research. The study proposes a research method that integrates AMs, aiming to improve the accuracy of predicting the ecological capacity of forest park visitors by optimizing the feature extraction process.

From the perspective of algorithm performance, the research algorithm demonstrated good convergence in AID dataset testing. At different training ratios, as the number of iterations increased, both the training loss value and the validation loss value steadily decreased. When the training ratio was 20%, the training loss reached 0.4 after 200 iterations, and the validation loss was 0.9. When the training ratio was 50%, the training loss decreased to 0.1 after 200 iterations, and the validation loss was 0.5, indicating that the algorithm could effectively converge under different training conditions, providing a guarantee for its stability in practical applications. This is consistent with the results obtained by Peng *et al*. [20] in the study of the spatiotemporal feature extraction classification framework for arrhythmia grounded on the Seq2Seq model with AM. In terms of feature vector visualization, compared with the comparison method without AM, the research method significantly improved the inter class distribution and intra class distance of different levels of features by introducing residual attention module, making the discrimination between samples of different categories higher. In V4 feature vector visualization, the inter class distance was significantly increased, and the intra class aggregation was stronger, further enhancing the separability of features, fully demonstrating the effectiveness of AM in improving feature extraction and classification performance. Yu *et al*. [31] also obtained similar results in hyper-spectral feature extraction research based on deep spectral spatial feature fusion multi-scale adaptive attention network.

In tests on different datasets, the research model outperformed Bicubic and RCAN models in metrics such as RMSE, PSNR, SSIM, and SAM, fully demonstrating its outstanding performance in image reconstruction and feature extraction. This result is consistent with Chen *et al*.'s [6] research on digital elevation model SR in CNNs. The research method had significant advantages in extracting ecological capacity information for forest park visitors. In the prediction of ecological tourism environmental capacity in Qilian Mountain Forest national park, research methods showed good predictive ability. The predicted values of ecological environment capacity were 31.4268 million people, spatial environment capacity was 104.8325 million people, tourist psychological environment capacity was 115.7436 million people, and facility environment capacity was 334200 people, all of which closely mirrored the actual conditions of the park. The comprehensive calculation of the tourism environment capacity was 64.8402 million people, providing a key reference for park tourism planning. This aligns with the findings obtained by Lu *et al*. [17] in evaluating the sustainable development of the middle and lower reaches of the yellow river basin using multiple data sources.

In summary, the research method of integrating AM and RS images performed well in the study of ecological capacity of forest park visitors. By introducing residual attention module, the separability and classification accuracy of features were significantly improved, and its effectiveness and robustness were verified on multiple datasets. The research outcomes provided a scientific basis for the tourism planning and management of forest parks, demonstrating the enormous potential of this method in practical applications.

5. Conclusions

A method for measuring the ecological capacity of forest park tourists by integrating AM and RS imaging technology was proposed to address the balance between ecological protection and tourist carrying capacity in the current development of ecotourism. The method was tested, and the test results showed that in terms of classification accuracy for different levels of feature dimension maps, the average classification accuracy of the research algorithm for low-level feature dimension maps was about 93.2%. In the mid-level feature dimension map, the average classification accuracy of the research algorithm was still the highest, at 92.8%. On the high-level feature dimension map, the classification accuracy of the research algorithm also reached 89.3%. In the comparison of the extraction effect of ecological capacity information for forest park visitors, when the resolution was set to 3m, the overall accuracy of the research method for extracting relevant information was as high as 91.2%. At this resolution, the Kappa coefficients of the three methods reached their respective highest values, with the Kappa coefficient of the research method being 0.94. From a practical standpoint, research methods can offer valuable reference approaches and innovative ideas for investigating tourist ecological capacity within ecotourism regions. However, in some complex ecological scenarios, the ability of research methods to capture the dynamic changes of ecological factors still needs to be improved. The ecological environment of the ecotourism area changed significantly with the change of seasons. The research method failed to capture these dynamic characteristics that changed over time, resulting in the assessment results of ecological capacity being inconsistent with the actual situation. Tourist flow and behavioral patterns may fluctuate due to holidays, special events, or policy changes. Research methods cannot reflect these changes in real time, thereby affecting the accurate prediction of tourism environmental capacity. Future research can combine various data sources such as RS data, ground observation data, meteorological data and

socioeconomic data to obtain a more comprehensive and accurate assessment of ecological capacity.

References

[1] Ai B., Xiao H., Xu H., Yuan F., and Ling M., "Coastal Aquaculture Area Extraction Based on Self-Attention Mechanism and Auxiliary Loss," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, no. 1, pp. 2250-2261, 2023. DOI: 10.1109/JSTARS.2022.3230081

[2] Arpitha M., Ahmed S., and Harishnaika N., "Land Use and Land Cover Classification Using Machine Learning Algorithms in Google Earth Engine," *Earth Science Informatics*, vol. 16, no. 4, pp. 3057-3073, 2023. https://link.springer.com/article/10.1007/s12145-023-01073-w

[3] Brauwers G. and Frasincar F., "A General Survey on Attention Mechanisms in Deep Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3279-3298, 2023. DOI: 10.1109/TKDE.2021.3126456

[4] Chen B., Liu T., He C., Liu Z., and Zhang L., "Fault Diagnosis for Limited Annotation Signals and Strong Noise Based on Interpretable Attention Mechanism," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11865-11880, 2022. DOI: 10.1109/JSEN.2022.3169341

[5] Chen W., Ouyang S., Tong W., Li X., Zheng X., and Wang L., "GCSANet: A Global Context Spatial Attention Deep Learning Network for Remote Sensing Scene Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, no. 1, pp. 1150-1162, 2022. DOI: 10.1109/JSTARS.2022.3141826

[6] Chen Z., Han X., and Ma X., "Combining Contextual Information by Integrated Attention Mechanism in Convolutional Neural Networks for Digital Elevation Model Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 8, pp. 1-16, 2024. DOI: 10.1109/TGRS.2024.3423716

[7] Chi K., Yuan Y., and Wang Q., "Trinity-Net: Gradient-Guided Swin Transformer-based Remote Sensing Image Dehazing and Beyond," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, no. 12, pp. 1-14, 2023. DOI: 10.1109/TGRS.2023.3285228

[8] Guo R., Liu H., Xie G., Zhang Y., and Liu D., "A Self-Interpretable Soft Sensor Based on Deep Learning and Multiple Attention Mechanism: From Data Selection to Sensor Modeling," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6859-6871, 2023. DOI: 10.1109/TII.2022.3181692

[9] Hao Y., Wang S., Gao P., Gao X., and Xu T., "Attention in Attention: Modeling Context Correlation for Efficient Video Classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 10, pp. 7120-7132, 2022. DOI: 10.1109/TCSVT.2022.3169842

[10] He X., Zhou Y., Zhao J., Zhang D., Yao R., and Xue Y., "Swin Transformer Embedding UNet for Remote Sensing Image Semantic Segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 1, pp. 1-15, 2022. DOI: 10.1109/TGRS.2022.3144165

[11] Hussain M., Chen T., Titrenko S., Su P., and Mahmud M., "A Gradient Guided Architecture Coupled with Filter Fused Representations for Micro-Crack Detection in Photovoltaic Cell Surfaces," *IEEE Access*, vol. 10, no. 8, pp. 58950-58964, 2022. DOI: 10.1109/ACCESS.2022.3178588

[12] Ji Z., Lv T., and Yu Y., "A Dual-End Recommendation Algorithm Integrating User Intent and Knowledge-Aware Attention Networks," *The International Arab Journal of Information Technology*, vol. 22, no. 3, pp. 505-521, 2025, DOI: 10.34028/iajit/22/3/7

[13] Khataei A., Singh G., and Bazargan K., "Optimizing Hybrid Binary-Unary Hardware Accelerators Using Self-Similarity Measures," *in Proceedings of the IEEE 31st Annual International Symposium on Field-Programmable Custom Computing Machines*, California, pp. 105-113, 2023. DOI: 10.1109/FCCM57271.2023.00020

[14] Khataei A., Singh G., and Bazargan K., "SimBU: Self-Similarity-Based Hybrid Binary-Unary Computing for Nonlinear Functions," *IEEE Transactions on Computers*, vol. 73, no. 9, pp. 2192-2205, 2024. DOI: 10.1109/TC.2024.3398512

[15] Liang Z., Tao M., Xie J., Yang X., and Wang L., "A Radio Signal Recognition Approach Based on Complex-Valued CNN and Self-Attention Mechanism," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 3, pp. 1358-1373, 2022. DOI: 10.1109/TCCN.2022.3179450

[16] Liu C., Zhao R., Chen H., Zou Z., and Shi Z., "Remote Sensing Image Change Captioning with Dual-Branch Transformers: A New Method and a Large Scale Dataset," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 9, pp. 1-20, 2022. DOI: 10.1109/TGRS.2022.3218921

[17] Lu Y., Li J., Song Z., and Lu M., "Evaluating Sustainable Development in the Middle and Lower Reaches of the Yellow River Basin Using Multiple Data Sources," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, no. 2, pp. 1-18, 2025. DOI:

10.1109/TGRS.2024.3509934

[18] Luo Z., Xu A., and Gao S., "Research on the Coupling Coordination Relationship between Tourism Carrying Capacity and the High-Speed Railway Network: A Case Study in China," *Access*, vol. 11, no. 6, pp. 20426-20444, 2023.

[19] Lv Z., Huang H., Li X., Zhao M., Benediktsson J., and Sun W., "Land Cover Change Detection with Heterogeneous Remote Sensing Images: Review, Progress, and Perspective," *Proceedings of the IEEE*, vol. 110, no. 12, pp. 1976-1991, 2022. DOI: 10.1109/JPROC.2022.3219376

[20] Peng X., Shu W., Pan C., Ke Z., Zhu H., and Zhou X., "DSCSSA: A Classification Framework for Spatiotemporal Features Extraction of Arrhythmia Based on the Seq2Seq Model with Attention Mechanism," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, no. 5, pp. 1-12, 2022. DOI: 10.1109/TIM.2022.3194906

[21] Preethi P. and Mamatha H., "Region-based Convolutional Neural Network for Segmenting Text in Epigraphical Images," *Artificial Intelligence and Applications*, vol. 1, no. 2, pp. 119-127, 2023. https://doi.org/10.47852/bonviewAIA2202293

[22] Roy S., Deria A., Hong D., Rasti B., Plaza A., and Chanussot J., "Multimodal Fusion Transformer for Remote Sensing Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, no. 12, pp. 1-20, 2023. DOI: 10.1109/TGRS.2023.3286826

[23] Su X., Li J., and Hua Z., "Transformer-based Regression Network for Pansharpening Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, no. 6, pp. 1-23, 2022. DOI: 10.1109/TGRS.2022.3152425

[24] Tang D., Cao X., Hou X., Jiang Z., Liu J., and Meng D., "CRS-Diff: Controllable Remote Sensing Image Generation with Diffusion Model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 3, pp. 1-14, 2024. DOI: 10.1109/TGRS.2024.3453414

[25] Tutek M. and Snajder J., "Toward Practical Usage of the Attention Mechanism as a Tool for Interpretability," *IEEE Access*, vol. 10, no. 2, pp. 47011-47030, 2022. DOI: 10.1109/ACCESS.2022.3169772

[26] Wang W., Zhang L., Fu S., Ren P., Ren G., and Peng Q., "Gradient Guided Multiscale Feature Collaboration Networks for Few-Shot Class-Incremental Remote Sensing Scene Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 5, pp. 1-12, 2024. DOI: 10.1109/TGRS.2024.3369083

[27] Xiao Y., Yuan Q., Jiang K., He J., Jin X., and Zhang L., "EDiffSR: An Efficient Diffusion Probabilistic Model for Remote Sensing Image Super-Resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, no. 5, pp. 1-14, 2024. DOI: 10.1109/TGRS.2023.3341437

[28] Xiao Y., Yuan Q., Jiang K., He J., Lin C., and Zhang L., "TTST: A Top-k Token Selective Transformer for Remote Sensing Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 33, no. 7, pp. 738-752, 2024. DOI: 10.1109/TIP.2023.3349004

[29] Yan R., Han Y., Li F., and Li P., "Evaluation of Sustainable Potential Bearing Capacity of Tourism Environment Under Uncertainty: A Multiphase Intuitionistic Fuzzy EDAS Technique Based on Hamming Distance and Logarithmic Distance Measures," *Access*, vol. 12, no. 9, pp. 8081-8095, 2024.

[30] Yang S., Tian Y., He C., Zhang X., Tan K., and Jin Y., "A Gradient-Guided Evolutionary Approach to Training Deep Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 9, pp. 4861-4875, 2022. DOI: 10.1109/TNNLS.2021.3061630

[31] Yu W., Huang H., and Shen G., "Deep Spectral-Spatial Feature Fusion-based Multiscale Adaptable Attention Network for Hyperspectral Feature Extraction," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, no. 1, pp. 1-13, 2023. DOI: 10.1109/TIM.2022.3222480

[32] Zhang S., Qi X., Duan J., Yuan X., Zhang H., and Feng W., "Comparison of Attention Mechanism-based Deep Learning and Transfer Strategies for Wheat Yield Estimation Using Multisource Temporal Drone Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 4, pp. 1-23, 2024. DOI: 10.1109/TGRS.2024.3401474

[33] Zhao J., Yu C., Shi Z., Liu Y., and Zhang Y., "Gradient-Guided Learning Network for Infrared Small Target Detection," *IEEE Geoscience and Remote Sensing Letters.*, vol. 20, no. 6, pp. 1-5, 2023. DOI: 10.1109/LGRS.2023.3308783

Qiaofeng Zhou has been working as a teacher in the field of Tourism Management. Over an eight-year teaching career in Tourism Management, she has maintained a keen interest in the Development Planning of Tourist Attractions and the study of Eco-Logical Tourist carrying capacity within these areas. In my published papers, she has discussed to some extent the green development and sustainable development of enterprises, as well as researched the synergistic development of tourism.