International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 5895-5907

https://doi.org/10.70517/ijhsa464427

Intelligent Recognition and Recommendation of Tourist Attractions by Integrating Artificial Intelligence Systems and Image Recognition Technology

Naiyuan Jiang[1,2], Zhaojie Wang[3,4,*] and Mengya Li[1]
[1] School of Business Administration, Dongbei University of Finance and Economics, Dalian, Liaoning, 116025, China
[2] School of Tourism and Geography, Baicheng Normal University, Baicheng, Jilin, 137000, China
[3] College of Tourism and Service, Nankai University, Tianjin, 300071, China
[4] College of Tourism Management, Guilin Tourism University, Guilin, Guangxi, 541006, China
Corresponding authors: (e-mail: 602319499@qq.com).

Abstract Aiming at the issues of low accuracy and poor robustness in the recommendation system for cultural tourism attractions, this article adopted a combination model of multimodal visual geometry group 16 (VGG16) and neural collaborative filtering (NCF) to study the intelligent identification and recommendation of cultural tourism attractions. Firstly, the convolutional neural network (CNN) VGG16 model was adopted for feature extraction of scenic spot images, and multimodal data was combined to help recommendation systems better understand the characteristics of scenic spots and improve the accuracy of recognition and classification of scenic spot images. Then, a neural collaborative filtering model was introduced to fully consider the relevant information of tourists for personalized recommendation of tourist attractions, improving tourist satisfaction and recommendation accuracy. By comparing the recommendation performance of four models, NCF, content filtering, collaborative filtering, and matrix factorization, on a self built dataset, the test outcomes indicate that the recommendation accuracy of NCF model reaches 97.21%, which is 8.07% higher than collaborative filtering, and the recommendation coverage reaches 99.41%, with a response speed of only 64.1ms, which improves the accuracy and adaptability of the recommendation of the system of cultural and tourist attractions in different situations.

Index Terms Cultural Tourist Attractions, Artificial Intelligence Systems, Image Recognition, Personalized Recommendations, Multimodal Information

I. Introduction

With the rapid growth of artificial intelligence technology and the urgent need for the development of the tourism industry, a variety of tourist attraction recommendations have begun to appear in public places like crazy, and different groups recommend different personalized solutions. Currently, the tourist attraction recommendation system fails to fully consider the interests and preferences of different tourists, and the extraction of tourist attraction features is not comprehensive, resulting in poor recommendation accuracy, especially for new tourists. There is a cold-start problem, which makes it difficult for the model to fully learn the relevant information of tourists. In addition, it is less robust and cannot adapt to the identification and recommendation of tourist attractions in different contexts. Accurate identification and recommendation of tourist attractions, as well as real-time updating of tourist attraction information, can provide tourists with a satisfactory travel experience and further boost the growth of the tourism industry.

The application fields of intelligent recognition and recommendation have gradually become widespread, and researchers have currently achieved a large number of research results in this field. Zhang K and other scholars used artificial intelligence (AI) methods to identify tourist photos in order to explore effective ways for tourists to recognize scenic spots. Natural language processing (NLP) methods were used to statistically analyze the views of different tourists on scenic spots [1], [2]. Scholars such as Han S used an improved particle swarm optimization algorithm to design a smart tourism path recommendation method, with an average absolute error value of 8.13, overcoming the problems of increased fluctuations in smart tourism data and fuzzy optimal solutions [3]. Scholars such as Liao Y aimed to solve the problem of inadequate personalized tourism information push and proposed a hotspot analysis method based on trajectory stop point spatial clustering to achieve accurate push of tourism information [4]. Wei C provided a reference basis for tourists to choose scenic spots by using big data technology to analyze passenger data [5]. In order to accurately explore customer needs, Novianti S and other scholars proposed an extended model of the theory of planned behavior (TPB), which improved the satisfaction of system recommendations [6]. Scholars such as Kim J used deep learning models and image feature vector clustering to automatically classify photos of tourist attractions, improving the recognition and classification performance of scenic buildings [7]. Giglio S and other scholars used clustering analysis to allow people to automatically identify clusters around points of interest (POI), providing a reference for others to choose interest points in tourist attractions and promoting tourist decision-making [8], [9]. From the above literature, it can be seen that scholars have improved their accuracy in recommending and recognizing tourist attractions, but the accuracy of their recommendation systems is still not ideal.

To adapt to the present demand of the society and to respond to the personalized recommendation needs of tourists, many studies have been carried out by researchers to improve the accuracy of tourist attraction recommendations. Zhang Q and other scholars adopted image processing based tourist attraction location recognition and personalized recommendation methods to achieve good retrieval results on tourist attraction images with text labels [10].To improve the accuracy of system recommendations, An H adopted a scenic area sentiment analysis recommendation system on the basis of the CNN-long short-term memory (CNN-LSTM) model, which combined weather information, comments, etc., for intelligent recommendations [11]. In order to better recognize buildings in photos and reflect the characteristics of the scenic area, scholars such as Kang Y used a deep learning model based on geographical markers for image recognition and classification, with an accuracy rate of 85.77% [12]. Fudholi D H and other scholars adopted a mobile travel recommendation system based on deep learning, using the state-of-the-art mobile deep learning architecture EfficientNet-Lite for training, with an average model accuracy of over 85% [13]. In order to address data sparsity, scholars such as Li G implemented a mixed recommendation system for tourist attractions on the basis of stratified sampling statistics and multimodal visual Bayesian personalized ranking, which improved detection performance [14]. Scholars such as Javed U adopted a recommendation system based on a combination of context aware and content filtering strategies, fully considering user interests and improving recommendation matching accuracy [15]. Al-Ghobari M and other scholars adopted the K Nearest Neighbor (KNN) location-aware personalized traveler assistance (LAPTA) system, and the experimental findings indicated that LAPTA improved the reliability and accuracy of scenic spot recommendation [16], [17]. To enhance the efficiency of system decision-making, scholars such as Forouzandeh S proposed a new method for tourism industry recommendation system that combines the Artificial Bee Colony (ABC) algorithm and Techniques for Order of Preference by Similarity to Ideal Solution (TOPSIS). The experimental results achieved good recommendation satisfaction [18]. The above scholars have achieved good accuracy in recommending tourist attractions, but their robustness is poor. In summary, it can be seen that the combination model based on multimodal VGG-16 and collaborative filtering is feasible for intelligent recognition and recommendation of tourist attractions.

To address the issues of low accuracy and poor robustness of the cultural tourism attraction recommendation system, this paper adopted a multimodal VGG16 and NCF combination model to intelligently identify and recommend cultural tourism attractions. Firstly, the self built dataset was subjected to preprocessing operations such as image annotation and encoding, grayscale processing, random rotation, and cropping. Then, the CNN VGG16 model was utilized to extract features from scenic spot images, improving the accuracy of recognition and classification of scenic spot images. On the basis of the collaborative filtering model, neural networks were introduced to fully consider the relevant information of tourists for personalized recommendation of tourist attractions, improving tourist satisfaction and recommendation accuracy. Finally, the recommendation performance of four models, NCF, content filtering, collaborative filtering, and matrix factorization, as well as the recognition performance of scenic spot images by VGG16, ResNet, Inception, and AlexNet, were compared. The findings indicated that the recognition accuracy of the VGG16 model reached 98.62%, which was 3.28% higher than ResNet. The recommendation accuracy of the NCF model reached 97.21%, which was 8.07% higher than collaborative filtering. The recommended coverage rate reached 99.41%, with a response speed of only 64.1ms. The satisfaction score of tourists was as high as 9 points, which improved the accuracy and adaptability of cultural tourism attraction system recommendations to different situations.

II. Experimental Data

II. A. Experimental Dataset

The dataset of this article comes from collecting relevant scenic spot images online and freely shooting relevant scenic spot images, as well as conducting on-site surveys to collect tourists' personal wishes and interests. The relevant image information of tourist attractions includes images, historical and cultural information, specialty cuisine, geographical location, etc. The tourist data information includes the intended tourist attractions style, historical travel records, etc. There are a total of 6000 image information and survey information for 500 tourists.

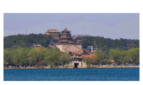
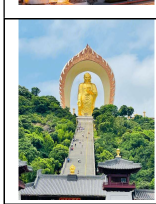
The partitioning method uses the ten fold cross validation method [19] to partition the training set and dataset. The original dataset is shown in Figure 1, and some tourist data is displayed in Figure 2.

Figure 1: Partial original dataset

Name	Intended attraction style	Intended area	Historical travel records
Li*	Cultural tourism	Jiangxi	Shanghai disney
Wang*	Nature adventure	Chongqing	Wuhan happy valley
Liu*	Nature adventure	Tibet	Zhangjiajie
Sun*	Cultural tourism	Hangzhou	Forbidden city
Gong*	Beach vacation	Fujian	Sanya
Zhang*	Cultural tourism	Beijing	Xi'an datang evernight city
Wan*	City tour	Chongqing	Hongyadong
Shi*	Beach vacation	Hainan	Gulangyu island
Chen*	Food tourism	Wuhan	Longevity palace

Figure 2: Partial tourist data

Table 1: Labeling of scenic area information

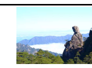
Image	History and culture	Food nearby	Location	Image	History and culture	Food nearby	Location
	One of the first batch of national first-level museums, The province's patriotism education base	Like the third child	Nanchang, Jiangxi province		Imitation of song dynasty architectural style, Famous building	Lushan stone fish	Jiujiang city, Jiangxi province
	Royal garden, World cultural heritage	Beijing old hot pot	Beijing		Romantic, residential characteristics	Lijiang baba	Li jiang,Yun nam
	Shaohua mountain, Mesoproterozoic realm	Braised sunfish	Shangrao city, Jiangxi province		Yuzhang ancient civilization, Preface to tengwang pavilion	Nanchang mixed powder	Jiangxi province
	Buddhist holy land, Jixue wotang	Yak meat	Tibet autonomous region		Guanyin of the south china Sea	Wenchang chicken	Sanya, Hainan
	Donglin buddha, a holy place of pilgrimage	Jiujiang glutinous rice cake	Jiujiang city, Jiangxi province		Military fortifications, Beacon fire plays on princes	Peking duck	Beijing

II. B.Dataset Preprocessing

(1) Image annotation and encoding

For the convenience of experimental processing, attraction information is annotated and encoded. The original scenic spot information annotation information is shown in Table 1. Binary encoding is adopted. The corresponding rules are shown in Table 2.

Table 2: Encoding table for scenic area information

Annotation encoding					
Category classification	Natural attractions	Historical and cultural attractions	Theme park	Cultural and artistic attractions	Scenic area
Picture					
Tags	Nanhai	Summer palace	Disney	Museum	Sanqingshan
Coding	0001	0010	0011	0100	0101
Style classification	Cultural tourism	Nature adventure	Beach vacation	Food tourism	City tour
Coding	0110	0111	1000	1001	1010

(2) Grayscale processing

In this study, in order to reduce feature dimensions and computational complexity, and avoid color interference, the image is subjected to grayscale preprocessing. Firstly, the color image is loaded and the grayscale averaging method is utilized. Based on the pixel values of three channels in the Red Green Blue (RGB) color space, the corresponding grayscale values are calculated. Finally, the color values of each pixel in the image are converted into corresponding grayscale values, as shown in Figure 3.

Figure 3: Grayscale processing image

(3) Image enhancement

To further balance the dataset, after unifying the size, the image is enhanced [20], specifically by performing random rotation and cropping operations on the segmented image. The images of tourist attractions are uniformly rotated counterclockwise by 30 degrees, and the action images are randomly cropped while maintaining image size. The random cropping action image is shown in Figure 4.

Figure 4: Random rotation and cropping of images

III. Neural Collaborative Filtering Model

In the traditional collaborative filtering [21], [22] studied by scholars such as Chen Si, matrix factorization (MF) uses dot product operations to model the connections between tourists and tourist attractions, making it difficult to learn nonlinear relationships. This experiment is conducted using a neural collaborative filtering model [23], [24]. Among them, the deep enhanced matrix factorization (DEMF) model adopts an element-wise product to model the relationships between the same embedding spatial dimensions, and adopts a linear multi-layer structure to implicitly learn the relationships between different dimensions. On this basis, a nonlinear multi-layer perceptron (MLP) structure [25] is introduced to train and learn high-order relationships between embeddings, outputting interaction functions.

In the linear representation stage of the interaction function, the left linear representation layer uses p_u^l to represent the user's input implicit vector and q_i^l to represent the input implicit vector of the item. The specific calculation is shown in Formula (1). Among them, $\odot$ represents the multiplication of vectors element by element, and $\emptyset_1^l$ represents the mapping function during the interaction function process.

$$z_0^l = \emptyset_1^l(p_u^l, q_i^l) = p_u^l \odot q_i^l \tag{1}$$

The linear representation layer on the right is mainly used to extract linear cross relationships between different embedded dimensions. Therefore, a MLP neural network is introduced, as shown in Formula (2).

$$c_0 = \begin{bmatrix} p_u^r \\ q_i^r \end{bmatrix} \tag{2}$$

Among them, p_u^r represents the implicit vector of the user, and q_i^r represents the implicit vector of the item.

$$z_0^r = \emptyset_X^r(C_{X-1}) = a_X(W_X^T C_{X-1} + b_X) \tag{3}$$

W_X^T represents the weight of the X-th layer in the linear feature extraction of the interaction function. In addition, b_X represents the corresponding bias vector, and a_X represents the activation function.

In the interactive nonlinear learning stage, the complete linear representation of the interaction function is fed into the deep neural network, which is represented in the hidden layer as shown in Formulas (4) and (5).

$$z_0 = \begin{bmatrix} z_0^r \\ z_0^l \end{bmatrix} \tag{4}$$

$$\emptyset_Y(Z_{Y-1}) = a_Y(W_Y^T Z_{Y-1} + b_Y) \tag{5}$$

Among them, z_0 represents the input vector in this stage, and Z_Y represents the representation vector of the nonlinear interaction function of the hidden layer Y.

$\emptyset_Y()$ represents the corresponding mapping function, and W_Y^T represents the corresponding weight.

In the prediction layer, a recommendation list is generated after predicting and scoring. Firstly, the optimization model objective function is solved, and then the adaptive learning algorithm Adam is used for optimization. Among them, the loss function calculation in training is shown in Formula (6), and the final prediction score calculation is shown in Formula (7).

$$L = -\sum_{(u,i \in Y^+ \cup Y^-_{sampled})} y_{ui} \log \widetilde{y_{ui}} + (1 - y_{ui}) \log(1 - \widetilde{y_{ui}}) \tag{6}$$

$$\widetilde{y_{ui}} = \sigma(h^T \emptyset_Y(Z_{Y-1})) \tag{7}$$

Among them, in Formula (6), Y^+ represents the positive sample set for training; $Y^-_{sampled}$ represents the partial adoption of results from the overall negative sample set; L represents the loss function optimized by the model during the training process. In Formula (7), σ represents the probability function and h^T represents the weight matrix of the output layer. Among them, σ uses the sigmoid function, and the calculation is shown in Formula (8).

$$\sigma(x) = \frac{1}{1+\exp(-x)} \tag{8}$$

To further improve the performance of the model, based on DEMF and combined with the nonlinear ability of MLP modeling interaction functions, a neural network enhanced matrix factorization (NNEMF) model is constructed to enhance the learning and prediction ability of the model. The prediction factor vector output by the MLP model is represented by Formula (9).

$$\varphi^{MLP} = a_L(\ldots a_2(W_2^T(\begin{bmatrix} p_u^M \\ q_i^M \end{bmatrix} + b_2))\ldots) \tag{9}$$

Among them, φ^{MLP} represents the prediction factor vector output by the DEMF model.

$$h = \begin{bmatrix} \alpha h^{DEMF} \\ (1-\alpha)h^{MLP} \end{bmatrix} \tag{10}$$

$$\widetilde{y_{ui}} = \sigma(h^{T} \begin{bmatrix} \varphi^{DEMF} \\ (\varphi^{MLP} \end{bmatrix}) \tag{11}$$

Among them, h^{DEMF} represents the DEMF output layer h vector, and h^{MLP} represents the MLP output layer h vector. α is the hyperparameter between the two models, which is used to balance the performance effect between DEMF and MLP.

IV. CNN VGG16

CNNs [26] perform well in image recognition and classification. This article adopts VGG16 as the model architecture. Among them, convolutional layers have more layers and deeper network layers compared to other architectures. VGG16 has a total of 16 layers, including 13 convolutional layers and 3 fully connected layers [27], [28]. The specific process is as follows. Firstly, a standard sized grayscale image is input, In addition, the initial model adopts convolutional processing with two convolutional kernels, and after activating the rectified linear unit (RELU), maximum pooling is utilized. The second model adopts convolutional processing with 128 convolutional kernels twice. The same as the first activation pooling operation, 128 and 256 convolutional kernels are used for processing in the third and fourth times, respectively. The final model balances the data into a pile of vectors through the Flatten function. After passing through two layers of 1*1*1*4096 and one layer of 1*1*2 fully connected layers, the model also uses RELU activation to output prediction results through the softmax function.

V. Intelligent Recognition and Recommendation Experiment for Scenic Spot Images

Experimental Process：Firstly, the photos collected online and taken by oneself are synthesized into a self built dataset, which is used as input for the model. After inputting the dataset, the scenic spot images are preprocessed with image annotation and encoding, grayscale processing, random rotation, and cropping. The experiment is divided into two parts. ImageNet is utilized for pre training. The model was initially trained on a self-built dataset for 30 epochs, with the image size set to 50 and a batch normalization layer added. The optimizer used was Adam (lr=0.0001, betas=(0.9, 0.999)), and the loss function was cross entropy. The classification used softmax with a threshold of 0.5. The number of training rounds was dynamically increased when adjusting the parameters (in multiples of 2). Finally, real-time recommendations were made to tourists under the NCF framework, and the accuracy and satisfaction of the recommendations were evaluated.

The experimental flowchart is shown in Figure 5.

This study evaluates the results using accuracy, precision, recall, and comprehensive evaluation indicators f1-score [29], [30]. The following is an evaluation of action recognition and recommendation in cultural tourism attractions based on VGG and NCF models.

VI. Experimental Results and Discussions on Intelligent Recognition and Recommendation of Scenic Spot Images

VI. A. Experimental Results of Scenic Spot Image Recognition and Classification

Table 3: Recognition and classification results of scenic spot images

	Image	Forecast style	Prediction category		Image	Forecast style	Prediction category		Image	Forecast style	Prediction category
1		Cultural tourism	Cultural and artistic attractions	5		City tour	Historical and cultural attractions	9		Beach vacation	Natural attractions
2		Cultural tourism	Historical and cultural attractions	6		City tour	Historical and cultural attractions	10		Cultural tourism	Historical and cultural attractions
3		Nature adventure	Scenic area	7		Cultural tourism	Cultural and artistic attractions	11		Cultural tourism	Scenic area
4		Nature adventure	Scenic area	8		Nature adventure	Theme park	12		Nature adventure	Natural attractions

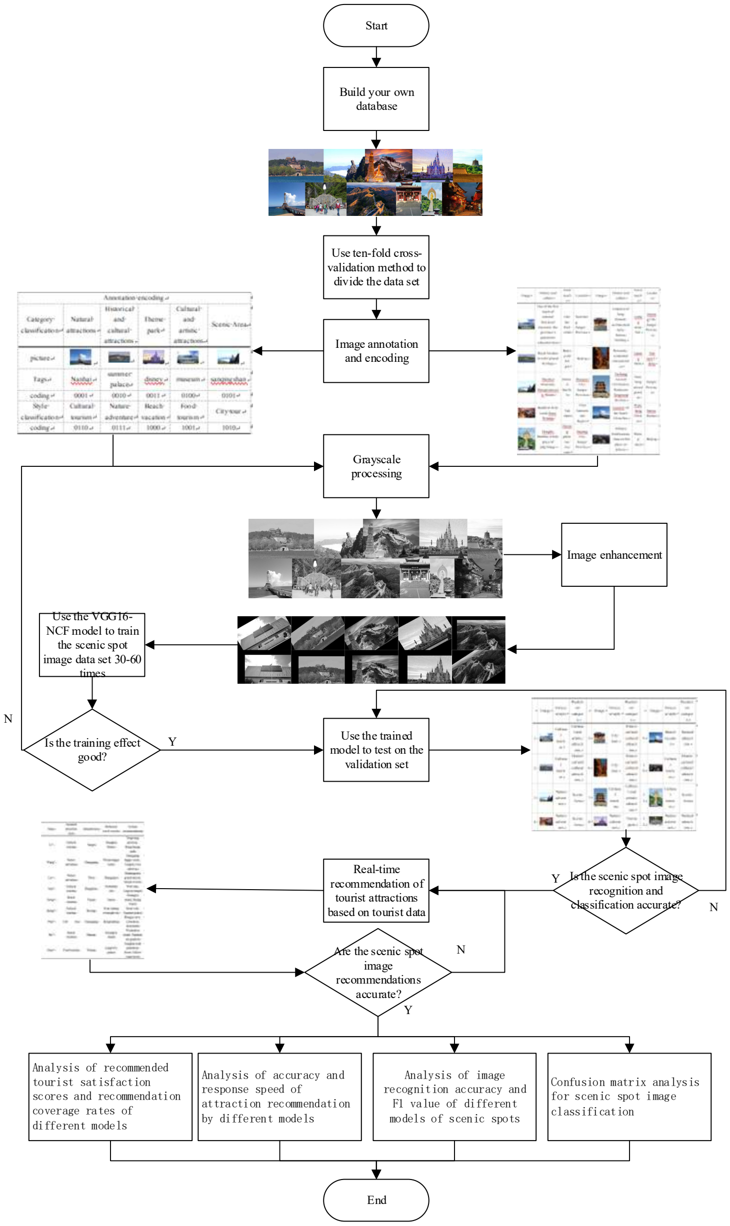


Figure 5: Experimental flowchart

VI. B. Experimental Results of Scenic Spots Recommendation

Table 4: Partial experimental results

Name	Intended attraction style	Intended area	Historical travel records	System recommendation
Li*	Cultural tourism	Jiangxi	Shanghai Disney	Tengwang pavilion, Rongchuang park
Wang*	Nature adventure	Chongqing	Wuhan happy valley	Chongqing happy valley, Yangtze river cableway
Liu*	Nature adventure	Tibet	Zhangjiajie	Brahmaputra grand canyon, Mount everest
Sun*	Cultural tourism	Hangzhou	Forbidden city	West lake, Lingyin temple
Gong*	Beach vacation	Fujian	Sanya	Gulangyu island, Dajing beach
Zhang*	Cultural tourism	Beijing	Xi’an datang evernight city	Great wall, Summer palace
Wan*	City tour	Chongqing	Hongyadong	Hongya cave, Liberation monument
Shi*	Beach vacation	Hainan	Gulangyu island	Wuzhizhou island, Nanshan sea guanyin
Chen*	Food tourism	Wuhan	Longevity palace	Jianghan road pedestrian street, Yellow crane tower

VI. C. Experimental Discussion

(1) Recognition and recommendation errors in scenic spot images

The results of scenic spot recognition and classification obtained through experiments are shown in Table 3. Overall, the model can effectively recognize and classify scenic spot images. In sequence number 2, the image shows the Summer Palace in Beijing, and the style is predicted to be cultural tourism. In sequence number 9, the image shows Nanhai Guanyin, and the style is predicted to be a beach vacation. However, it should be a cultural tourism and mistakenly predicted to be a beach vacation. Due to the need to extract more ocean features during recognition and the lack of extraction of main features, improvements would be made in conjunction with attention mechanisms in the future. In sequence number 4, the image shows a night view of Lhasa, and the system predicts it as a natural adventure. Due to the systematic learning that Lhasa yearns for more freedom and exciting sports, urban tourism is predicted as a natural adventure. In sequence number 5, the image shows Xunyang Tower in Jiujiang, with a predicted style of urban tourism and actual cultural tourism. However, in sequence number 10, the image shows the Great Wall, but the actual cultural tourism is predicted to be cultural tourism, indicating that the system can still recognize the image features of the scenic area well. However, due to the intersection of styles, some errors may occur.

For prediction types, there are some results that result in recognition errors due to type crossing. For example, in sequence number 7 of Table 3, the image is Pavilion of Prince Teng in Nanchang, and the category is predicted to be culture and art, but it is actually a historical and cultural scenic area. However, in other images, such as the Jiangxi Provincial Library in sequence number 1, it can be well recognized as a cultural and artistic type by the system. Other types are as follows: in sequence number 12, the image is Zhangjiajie, and the prediction type is natural scenic spot; in sequence number 3, the image is Mount Sanqing, which is predicted to be a scenic spot and can be well recognized and classified by the system.

For the phenomenon of biased recommendations, as shown in some experimental results in Table 4, Li *'s intended tourist attraction style is cultural tourism, and his intended area is Jiangxi. He has visited Shanghai Disneyland and has systematically recommended Pavilion of Prince Teng and Rongchuang Park. After statistical analysis of tourists' evaluation of recommendation results, there are some deviations. After analysis, both the system and tourists believe that it is more suitable. Tourists believe that Sunac Land is not what they thought. This is because the system has analyzed Li *'s previous travel records and related adventure style experiences, and thus recommends Sunac Land. For Chen *, the recommended results are Jianghan Road Pedestrian Street and Huanghe Tower, with some deviation in the recommended Huanghe Tower. Chen *'s intended tourist attraction style is actually food tourism, which is also recommended as coverage and has certain rationality.

(2) Attraction image classification confusion matrix

The confusion matrix for the different attraction styles is shown in Figure 6. As can be seen from the analysis in Fig. 6, the highest percentage of samples actually belonging to the matching actual styles matched the predicted styles, mainly focusing on beach vacation. The percentage of correctly forecasted samples reaches 99%, which is a very significant categorization effect. The lowest concentration is in Gourmet Tourism and City Tourism. In the Gourmet Tourism style, the percentage of samples correctly forecasted amounted to 95%, which is higher than the number of misclassifications. Among them, 2% of the samples were wrongly forecasted as Nature Adventure; 2% of the samples were wrongly forecasted as Beach Vacation; and 1% of the samples were forecasted as City Tourism. In addition, for city tourism, only 95% of the samples were confirmed as correctly forecasted, of which 2% were wrongly forecasted as cultural tourism and 3% were wrongly forecasted as gastronomic tourism. For the nature adventure style, only 97% of the samples were correctly forecasted, with 1% being forecasted as cultural tourism and 2% as urban tourism. Other styles are better categorized and can meet the demand of actual application.

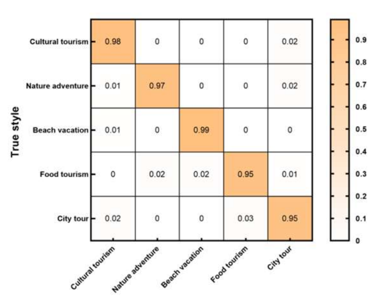


Figure 6: Classification confusion matrix for different scenic spot styles

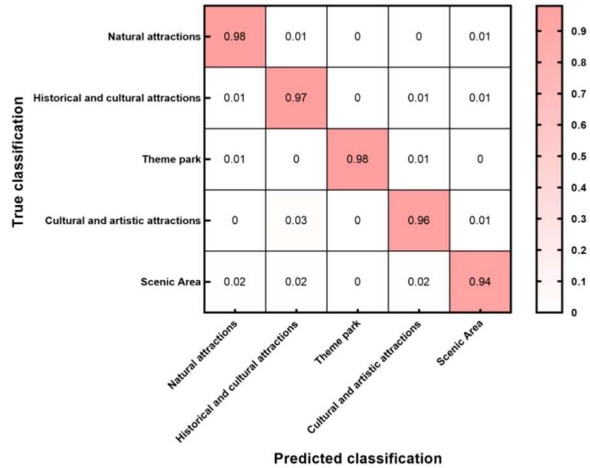


Figure 7: Confusion matrix for different types of scenic spots

The confusion matrix for different types of attractions is illustrated in Figure 7. As can be seen from the analysis in Fig. 7, the samples actually belonging to the corresponding actual categories have the highest percentage of being forecasted as the corresponding predicted categories, which are mainly focused on the types of natural attractions and theme parks. The percentage of correctly forecasted samples reaches 98%, and the categorization effect is very significant. The lowest concentration is in scenic attractions, where the percentage of correctly forecasted samples reaches 94%, which is high than the number of classification errors. Among them, 2% were wrongly forecast as natural attractions; 2% were wrongly forecast as historical and cultural attractions; and 2% were forecast as cultural and artistic attractions. Due to the difficulty of categorizing cultural and artistic attractions, the forecast rate was only 96%. Among them, 1% were forecasted as scenic spots and 3% were forecasted as historical and cultural spots. The classification results of other categories are better and can meet the experimental requirements.

(3) Accuracy and F1 value of scenic spot image recognition in different models

To better compete the validity of various models, the scenic image recognition accuracy and F1 value of different models were analyzed. The line graph is shown in Figure 8. In terms of scenic image recognition accuracy, overall, the VGG16 model has the best accuracy, while the AlexNet model has the least accuracy, with a difference of 18.25%. It can be seen that the range is quite obvious and the model has a clear advantage. Specifically, the ResNet model has an accuracy of 95.34%, which is 3.28% lower than VGG16 and worse. the VGG16 model achieves 98.62%, which is 6.43% better than the Inception model and better. The worst case is the AlexNet model with only 80.37%, which is not favorable for this experiment.

The F1 value of the scene recognition is shown in Figure 8. Among them, the F1 value of ResNet model reaches 94.12%, which is 0.54% better than Inception; the F1 value of VGG16 model is the highest, which reaches 97.91%, which is 3.79% better than ResNet model; and the AlexNet model, which is the worst performer, is only 72.64%, which is 25.27% lower than VGG16. In conclusion, the VGG16 model performs well in the recognition and classification of scenic spots.

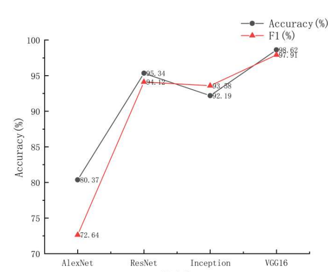


Figure 8: Analysis of the accuracy and F1 value of scenic spot image recognition in different models

(4) Recall and precision of scenic spot images in different models

Fig. 9 shows a comparison of the recall and precision of the different models for landscape point images. In terms of precision, overall, the VGG16 model has the longest histogram and the AlexNet model has the shortest histogram. Among them, the precision of VGG16 model reaches 98.12%, which is 15.53% higher than AlexNet model; ResNet model reaches 94.20%; and Inception model reaches 93.38%. In terms of recall, the higher the recall, the better the model can perform, and the VGG16 model has the best performance with 97.24%, which is 2.12% higher than the ResNet model and 9.81% higher than the Inception model, while the AlexNet model has the worst result with only 69.19%. It can be observed that the VGG16 model achieves high precision and recall in scenic spot recognition.

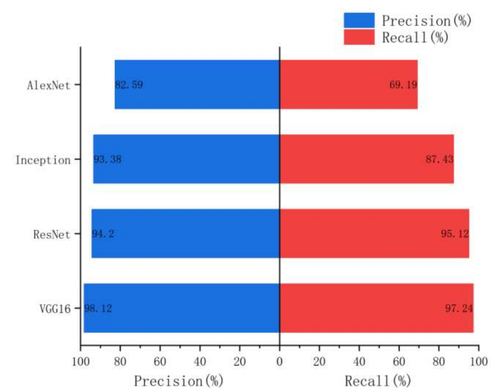


Figure 9: Analysis of recall and precision of scenic spot images in different models

(5) Accuracy and response speed of scenic spot recommendations in different models

To verify the real-time performance of the model and the accuracy of the recommendation, a comparison was made as shown in Figure 10, with a bar chart representing accuracy and a line chart representing response speed. From Figure 10, it can be seen that the corresponding recommendation accuracy showed a decreasing trend overall. As an experimental recommendation model, the NCF of this model achieved good results, with a recommendation accuracy of 97.21%, which was 4.53% higher than the content filtering model, and the effect was more ideal. In addition, collaborative filtering achieved 89.14%, a decrease of 8.07% compared to NCF. It can be seen that the model with the introduction of neural networks performed better in recommendation accuracy than without, with the worst being the matrix factorization model, which was only 73.27%.

For the response speed, it can be analyzed from Figure 10 that collaborative filtering had the fastest response speed, only requiring 46.7ms, which was 17.4ms less than NCF. It can be seen that the introduction of neural networks has improved the complexity of this model and extended its response speed. However, the NCF model has improved compared to the content filtering model, reducing 28.1ms compared to the content filtering model and 66.3ms compared to the matrix factorization model. Overall, NCF has extended its response speed, but has achieved good results in recommendation accuracy, and can be optimized through lightweight measures in the future.

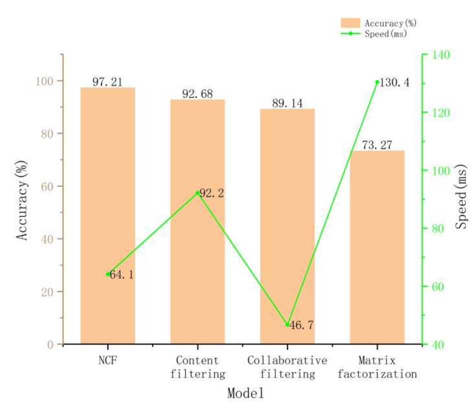


Figure 10: Analysis of accuracy and response speed of scenic spot recommendations for different models

(6) System recommendation tourist satisfaction scores and recommendation coverage for different models

To verify the actual effectiveness of the model, a real-time survey of tourist feedback was conducted and the satisfaction scores were summarized as shown in Table 5. Among them, the NCF coverage rate reached 99.41%, which can fully learn various images and text data. Compared to content filtering, it has increased by 1.58%, and compared to collaborative filtering, it has increased by 7.81%. The matrix factorization model had the worst coverage rate. The satisfaction of the developed system for the model was discussed in three aspects: system page aesthetics, system recommendation satisfaction, and convenience of system usage. In terms of system page aesthetics, according to statistics, NCF satisfaction reached 9 points; both the content filtering model and collaborative filtering model achieved a satisfaction score of 8 points; the matrix factorization model only scored 5 points. In summary, it can be seen that tourists are more inclined towards the aesthetics of the system pages developed by the NCF model. In terms of system recommendation satisfaction, the NCF and content filtering models both achieved a recommendation satisfaction score of 8 points, while the recommendation system of the worst matrix factorization model only scored 6 points. In addition, in terms of the convenience of system usage, the NCF model had the highest tourist rating of 9 points, followed by the system generated by the content filtering model. Overall, the satisfaction score of the system generated by the NCF model was the highest, reaching 26 points, while the satisfaction score of the system generated by the matrix factorization model was the lowest, only 14 points. In summary, tourists are more inclined towards the recommendation system generated by the NCF model.

Table 5: Analysis of system recommendation tourist satisfaction scores and recommendation coverage for different models

	Coverage	System page aesthetics (10points)	System recommendation satisfaction (10points)	Convenience of system usage (10points)
NCF	99.41%	9	8	9
Content filtering	97.83%	8	8	7
Collaborative filtering	91.60%	8	7	6
Matrix factorization	65.62%	5	6	3

(7) Robustness (different seasons, weather conditions, and populations)

To verify the robustness of the model, a comparative analysis was conducted on the recommendation accuracy and AUC under different conditions in different seasons, weather conditions, and populations, as shown in Figure 11. For different seasons, the summer model had the highest recommendation accuracy, reaching 98.25%, an increase of 0.33% compared to spring, and 97.85% in autumn. The worst-case scenario was winter, with a recommendation accuracy of only 85.20% due to frequent snow and adverse weather affecting scenic spot image recognition. On AUC, the highest accuracy was 0.97 in summer and 0.83 in winter. For different weather conditions, the recommended accuracy for snowy days was 73.49%, and for sunny days it reached 98.05%, which was 12.41% higher than for rainy days. On AUC, the highest recommended accuracy for sunny days was 0.98, while for rainy days it was only 0.88. For different groups, the recommendation accuracy of the student group was more accurate, reaching 95.38%, while the group of office workers was broader, with a recommendation accuracy of 87.64%, a decrease of 7.74% compared to the student group. On AUC, the student group reached 0.94, far exceeding the office group. In summary, it can be seen that this model exhibits good robustness in different seasons, weather conditions, and populations.

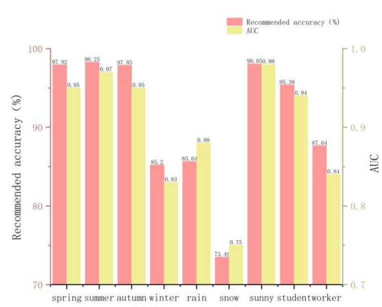


Figure 11: Robustness analysis

VII. Conclusions

In this paper, a multimodal VGG16 model combined with Neural Collaborative Filtering (NCF) is used to study the intelligent recognition and recommendation of cultural tourist attractions. CNN VGG16 model is used to extract features from scenic spot images. Combined with multimodal data, the neural collaborative filtering model is introduced to realize the personalized recommendation of tourist attractions by adequately considering the tourists' relevant information. The test outcomes indicate that the VGG16 model has sufficient feature extraction capability, which enhances the accuracy of the model in recognizing and classifying the scenic spots. The NCF model obtains good behavior in terms of recommendation accuracy and coverage, and it can adapt well to different situations. However, there are some shortcomings in the article, such as the dataset is not comprehensive enough, and the response speed cannot reach a good performance. In the future, the model will be lightened by enriching the experimental dataset.

Funding

This work was supported by: Research Project of Jilin Provincial Department of Education Humanities and Social Science Research Project of Jilin Provincial Department of Education "Research on the Coordinated Development of Rural Tourism and Ecological Economy in Western Jilin from the Perspective of Rural Revitalization" (JJKH20230025SK);Research Project on Higher Education Teaching Reform in Jilin Province: Research and Practice on the Cultivation Model of Collaborative Driven Innovative Talents in Geographic Information Science in Local Universities under the Background of the Construction of "New Engineering" (20224BR01C500HE); Ministry of Education Industry-University Cooperative Education Project: Application of BOPPPS+ Curriculum Ideological and Political Teaching Mode in Tourism Economics Course (230814175707198); 2025 Macro Decision-making Topics for Culture and Tourism, No. 2025HGJCG12. Convergence or Differences? Research on the multi-dimensional mechanism and governance path of cultural representation of tourism and leisure blocks from the perspective of digital inclusion.

References

[1] Zhang K, Chen Y, Li C. Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Bei g[J]. Tourism Management, 2019, 75(3): 595-608.https://doi.org/10.1016/j.tourman.2019.07.002

[2] Guerrero-Rodriguez R, Álvarez-Carmona M Á, Aranda R, Lopez-Monroy A P. Studying online travel reviews related to tourist attractions using nlp methods: the case of guanajuato, mexico[J]. Current issues in tourism, 2023, 26(2): 289-304.https://doi.org/10.1080/13683500.2021.2007227

[3] Han S. A new intelligent method for travel path recommendation based on improved particle swarm optimisation[J]. International Journal of Computing Science and Mathematics, 2020, 12(1): 36-50.https://doi.org/10.1504/IJCSM.2020.108802

[4] Liao Y. Hot spot analysis of tourist attractions based on stay point spatial clustering[J]. Journal of Information Processing Systems, 2020, 16(4): 750-759.https://doi.org/10.3745/JIPS.04.0177

[5] Wei C, Wang Q, Liu C. Research on construction of a cloud platform for tourism information intelligent service based on blockchain technology[J]. Wireless Communications and Mobile Computing, 2020, 2020: 1-9.https://doi.org/10.1155/2020/8877625

[6] Novianti S, Susanto E, Rafdinal W. Predicting Tourists' Behaviour Towards Smart Tourism: The Case in Emerging Smart Destinations[J]. Journal of Tourism Sustainability, 2022, 2(1): 19-30.http://dx.doi.org/10.35313/jtospolban.v2i1.30

[7] Kim J, Kang Y. Automatic classification of photos by tourist attractions using deep learning model and image feature vector clustering[J]. ISPRS International Journal of Geo-Information, 2022, 11(4): 245-264.https://www.mdpi.com/2220-9964/11/4/245#

[8] Giglio S, Bertacchini F, Bilotta E, Pantano P. Machine learning and points of interest: typical tourist Italian cities[J]. Current Issues in Tourism, 2020, 23(13): 1646-1658.https://doi.org/10.1080/13683500.2019.1637827

[9] Pantano E, Priporas C V, Stylos N, Dennis C. Facilitating tourists' decision making through open data analyses: A novel recommender system[J]. Tourism Management Perspectives, 2019, 31: 323-331.https://doi.org/10.1016/j.tmp.2019.06.003

[10] Zhang Q, Liu Y, Liu L, Lu S,Feng Y,Yu X. Location Identification and Personalized Recommendation of Tourist Attractions Based on Image Processing[J]. Traitement du Signal, 2021, 38(1).205-214,,doi:10.18280/TS.380121.

[11] An H, Moon N. Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM[J]. Journal of Ambient Intelligence and Humanized Computing, 2022,13: 1-11.https://doi.org/10.1007/s12652-019-01521-w

[12] Kang Y, Cho N, Yoon J, Park S,Kim J. Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos[J]. ISPRS International Journal of Geo-Information, 2021, 10(3): 137-157.https://doi.org/10.3390/ijgi10030137

[13] Fudholi D H, Rani S, Arifin D M, Stayatama M R. Deep learning-based mobile tourism recommender system[J]. Scientific Journal of Informatics, 2021, 8(1): 111-118.DOI:10.15294/sji.v8i1.29262

[14] Li G, Zhu T, Hua J, Yuan T,Niu Z,Li T,Zhang H. Asking images: Hybrid recommendation system for tourist spots by hierarchical sampling statistics and multimodal visual Bayesian personalized ranking[J]. IEEE Access, 2019, 7(1): 126539-126560.https://doi.org/10.1109/ACCESS.2019.2937375

[15] Javed U, Shaukat K, Hameed I A, Iqbal F,Alam T M,Luo S. A review of content-based and context-based recommendation systems[J]. International Journal of Emerging Technologies in Learning (iJET), 2021, 16(3): 274-306.https://www.learntechlib.org/p/219036/

[16] Al-Ghobari M, Muneer A, Fati S M. Location-Aware Personalized Traveler Recommender System (LAPTA) Using Collaborative Filtering KNN[J]. Computers, Materials & Continua, 2021, 69(2).1553-1570.http://dx.doi.org/10.32604/cmc.2021.016348

[17] Wang W, Chen J, Wang J, Chen J,Liu J,Gong Z. Trust-enhanced collaborative filtering for personalized point of interests recommendation[J]. IEEE Transactions on Industrial Informatics, 2019, 16(9): 6124-6132.https://doi.org/10.1109/TII.2019.2958696

[18] Forouzandeh S, Rostami M, Berahmand K. A hybrid method for recommendation systems based on tourism with an evolutionary algorithm and topsis model[J]. Fuzzy Information and Engineering, 2022, 14(1): 26-50.https://doi.org/10.1080/16168658.2021.2019430

[19] Derdouri A, Osaragi T. A machine learning-based approach for classifying tourists and locals using geotagged photos: the case of Tokyo[J]. Information Technology & Tourism, 2021, 23(4): 575-609.https://doi.org/10.1007/s40558-021-00208-3

[20] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of big data, 2019, 6(1): 1-48.https://doi.org/10.1186/s40537-019-0197-0

[21] Chen Si, Tian Jingyang. Research on Tourist Attraction Recommendation Model Based on Collaborative Filtering Algorithm [J]. Modern Electronic Technology, 2020,43 (11): 132-135.DOI:10.16652/j.issn.1004-373x.2020.11.031.

[22] Li Guangli, Zhu Tao, Yuan Tian, Hua Jin, Zhang Hongbin. A Study on the Tourist Attraction Recommendation Model Based on Mixed Layered Sampling and Collaborative Filtering [J]. Data Collection and Processing, 2019, 34 (03): 566-576.DOI:10.16337/j.1004-9037.2019.03.020.

[23] Chen W, Cai F, Chen H, Rijke M D. Joint neural collaborative filtering for recommender systems[J]. ACM Transactions on Information Systems (TOIS), 2019, 37(4): 1-30.https://doi.org/10.1145/3343117

[24] Sivaramakrishnan N, Subramaniyaswamy V, Viloria A, Vijayakumar,N.Senthilselvan. A deep learning-based hybrid model for recommendation generation and ranking[J]. Neural Computing and Applications, 2021, 33(1): 10719-10736.https://doi.org/10.1007/s00521-020-04844-4

[25] emali Hounmenou C G, Gneyou K E, Kakaı R G. Empirical determination of optimal configuration for characteristics of a multilayer perceptron neural network in nonlinear regression[J]. Afrika Statistika, 2020, 15(3): 2413-2429.http://dx.doi.org/10.16929/as/2020.2413.166

[26] Wang W, Yang Y, Wang X, Wang W,Li J. Development of convolutional neural network and its application in image classification: a survey[J]. Optical Engineering, 2019, 58(4): 040901-040901.https://doi.org/1117.1/58.OE.4.040901.

[27] Tammina S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images[J]. International Journal of Scientific and Research Publications (IJSRP), 2019, 9(10): 143-150.http://dx.doi.org/10.29322/IJSRP.9.10.2019.p9420

[28] Ramadhan A A, Baykara M. A Novel Approach to Detect COVID-19: Enhanced Deep Learning Models with Convolutional Neural Networks[J]. Applied Sciences, 2022, 12(18): 9325-9340.https://doi.org/10.3390/app12189325

[29] Sari A W, Hermanto T I, Defriani M. Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine[J]. Sinkron: jurnal dan penelitian teknik informatika, 2023, 8(3): 1366-1378.https://doi.org/10.33395/sinkron.v8i3.12447

[30] Srivastava S, Divekar A V, Anilkumar C, Naik I,Kulkarni V,Pattabiraman V. Comparative analysis of deep learning image detection algorithms[J]. Journal of Big data, 2021, 8(1): 1-27.https://doi.org/10.1186/s40537-021-00434-w