Tourism Promotion Mechanism Based on Virtual Reality Technology for Real-life Interactive Experience

Xinxin Nie[1], Jing Pu[2*]
[1]College of Literature and Media, Chengdu Jincheng College, Chengdu 611731, China
[2]School of Arts and Media, Sichuan Agricultural University, Ya'an 625014, China
E-mail: niexinxin@cdjcc.edu.cn, pujing325@163.com
[*]Corresponding author

Keywords: virtual reality technology, tourism recommendations, image retrieval, residual network, panoramic splicing

Received: May 28, 2024

The development of the national economy has promoted the development of the tourism industry, and outdoor tourism has gradually become the main leisure and entertainment method for people during holidays. However, the current tourism promotion mechanism only considers tourist preferences, resulting in tourists receiving recommendations. The unique features of the tourist attractions in the destination area are extremely similar to those of previous tourist areas. Therefore, the research is based on the Visual Geometry 19 model, using skip connections and soft threshold functions to improve the Visual Geometry 19 network. A residual network is designed for feature extraction, and panoramic stitching technology is used to construct virtual scenes of tourist attractions. The findings indicated that the residual network designed in the study demonstrated convergence after 50 iterations of training and testing on the tourism map dataset, with an accuracy of scene reconstruction reaching 85%. The designed tourism location image reconstruction could be completed in about 90 seconds, effectively improving the speed of the virtual reconstruction of tourism scenes. The designed tourism attraction promotion mechanism based on virtual reality technology effectively solves the problem of single tourism promotion and improves the experience of national outbound tourism.

Povzetek: Razvit je sistem za promocijo turizma, ki uporablja virtualno resničnost in slikovno analizo za interaktivno izkušnjo destinacij, kar izboljšuje priporočila turističnih krajev in preprečuje estetsko utrujenost obiskovalcev.

1 Introduction

The continuous growth of the global economy has led to an increase in people's disposable income, and the frequency of people traveling has gradually increased [1]. With the growth of the middle class, especially in the Asian region, the demand for tourism consumption has significantly increased [2]. Many countries and regional governments recognize the importance of tourism in stimulating economic growth. Consequently, they have implemented a range of policies designed to facilitate the expansion of the tourism sector. These include measures to streamline visa procedures, invest in tourism infrastructure, and promote cultural tourism [3-4] In addition, technological advancements and networking in transportation such as aviation, railways, and highways have made tourism more convenient and efficient [5]. The rise of low-cost airlines has also greatly reduced travel costs, making long-distance travel affordable for everyone [6]. The development of the Internet, mobile payment, social media, and online travel services has greatly facilitated the travel planning and booking process. Passengers can easily compare different tourism products and services and make online reservations [7-8]. However, due to the wide variety of tourist destinations, it is difficult for tourists to choose their favorite attractions. The current common tourist destination recommendation algorithms are mostly based on the introduction of scenic spots or tourist preferences. This recommendation method can easily cause aesthetic fatigue among tourists, leading to a poor travel experience. Therefore, to improve the tourism recommendation service experience of tourists, this study proposes to use a residual network (ResNet) structure to retrieve and extract features from relevant images of tourist attractions. The combination of virtual reality (VR) technology with a simulated virtual scene of a scenic location enables tourists to gain a firsthand understanding of the characteristics of the location, thus assisting them in making an informed choice regarding their preferred destination.

This study innovatively combines VR and image processing technology to build simulated scenes of scenic spots, allowing tourists to directly experience some of the features of the scenic spots, thereby helping them choose their desired attractions. The primary contribution is to address the aesthetic fatigue that tourists often experience as a result of being consistently recommended similar attractions during their travels. This is achieved by enhancing their travel experience, promoting the advancement of the tourism industry, and stimulating the growth of related VR devices.

2 Related works

Tourism is one of people's daily entertainment activities. Nitu et al. proposed a travel recommendation system to provide customized travel destinations that meet the specific needs and preferences of users. The system identified travel tweets by analyzing Twitter data and user social networks, combined with machine learning classifiers, and considered time-sensitive recent weights. This model was superior to existing models, with an overall accuracy of 75.23% [9]. Gao et al. proposed the DeepTrip model to improve the understanding of human mobility in travel recommendation systems. This model used a travel encoder and a trip decoder, combined with an adversarial network, to better model the transition distribution of POI in human motion patterns. DeepTrip was effective both theoretically and empirically, outperforming the advanced baselines [10]. Liu et al. proposed a multi-task deep learning-based Hydra recommendation system to improve the user experience of transportation recommendations, implementing multi-modal transportation planning and considering contextual environments. By deploying on Baidu Maps, the system has effectively improved user click through rates [11]. Huang et al. proposed a multi-task deep travel route planning framework to improve tourism route planning and meet the diverse needs of tourists, integrating rich auxiliary information such as interest point attributes and user preferences. This framework demonstrated flexibility and superiority in route recommendation [12]. Wang et al. proposed an Internet of Things (IoT) system based on 5G and AI to address the challenges faced by the IoT in smart tourism. The system employed 5G technology to facilitate efficient data transmission and artificial intelligence (AI) to enable intelligent data processing, thereby enabling the development of smart tourism applications. The case study showed that the proposed method performed well in POI recommendation, verifying its effectiveness and excellent performance [13].

VR technology is widely used in various fields and has a promising future market. Lv et al. proposed an intrusion detection model for industrial control networks based on Class and Sample Weighted C-Support Vector Machine (CSWC-SVM) to protect industrial security and simulate VR environments. Simulation experiments showed that the CSWC-SVM algorithm exhibited high recognition accuracy and low false alarm rate under different kernel functions, and its accuracy remained above 90% under different sample sizes [14]. To explore the potential of 360° VR videos and real VR settings in teaching, Pirker and Dengel conducted a systematic evaluation of VR and proposed research on its application in education. 360° VR videos were beneficial for multiple disciplines and could enhance the learning experience [15]. Image retrieval technology (IRT) helped to build VR scenes. Dubey et al. conducted a comprehensive survey of relevant IRTs to explore content-based IRTs based on deep learning. A classification method was proposed that includes supervision, network, descriptor types, and retrieval types. Deep learning has shown outstanding performance in automatically learning image features, which contributed to the further development of IRT [16]. Fernandez-Beltran et al. proposed a probabilistic latent semantic hashing model to address the limitations of existing unsupervised hashing methods in processing complex semantic content in remote sensing images. It effectively learned hash codes through three steps: data grouping, topic calculation, and hash code generation. This method was significantly superior to the best unsupervised hashing methods [17]. Liu et al. proposed a similarity Siamese convolutional neural networks (CNNs) model based on unsupervised transfer learning to overcome the problems of sparse labeled samples and cumbersome CNNs in remote sensing image retrieval. This model was superior to existing CNN-based methods [18].

In summary, in the current global economic environment and situation, local tourism has become the main leisure and entertainment method for people during their holidays. Common attraction recommendation algorithms often use specially captured attraction images to recommend tourist areas to visitors. The utilization of this recommendation method has the potential to result in aesthetic fatigue among tourists. Conventional tourist attraction recommendations do not facilitate a comprehensive understanding of the distinctive attributes of the scenic spots, limiting tourists to a general perception of the types of scenery present in these locations. VR technology can build simulation scenes for scenic spots, allowing tourists to experience the unique features of the attractions firsthand. Therefore, this study proposes to use IRT to assist in building VR scenes for tourist attractions (VRSFTA) and recommend them to tourists.

Table 1: Summary of relevant work survey results

Author	Contribute
Nitu et al.	Improved the accuracy of customized recommendations
Gao et al.	Analyzed the flow characteristics of tourist tourism

Liu et al.	Improved the user experience of the recommendation system
Huang et al.	Improved the flexibility of recommending travel routes
Wang et al.	Proposed a new type of tourism recommendation method
Lv et al.	Strengthened the protection technology of virtual networks
Pirker and Dengel	Proposed evaluation methods for the role of VR systems in other fields
Dubey et al.	A deep learning-based image feature retrieval technique has been proposed
Fernandez-Beltran et al.	A novel image semantic content processing technology has been proposed
Liu et al.	Proposed an image feature retrieval technique

3 Methods and materials

3.1 Image feature extraction based on ResNet

The construction of VRSFTA requires a large amount of scenic feature data. The Visual Geometry Group (VGG) 19 model is one of the architectures of CNN models, which has high localization ability [19-20]. When constructing VRSFTA, it is necessary to locate the high correlation features of scenic spots in the VR scene based on photos from different angles. However, the VGG19 model is prone to network degradation or gradient disappearance when extracting features from scenic spot images [21]. Therefore, this study improves the VGG19 model and designs a ResNet. The ResNet structure is shown in Figure 1. ResNet adopts batch normalization (BN) and Dropout mechanisms. BN is a widely used technique in deep neural networks that can improve the stability and speed of model training [22-23]. When performing BN processing on data in a network model, it is necessary to subtract the batch mean and divide it by the batch standard deviation to standardize each small batch of data, making the input distribution of each layer of the network more stable. At the same time, the study introduces a soft threshold function in the network to process the interference noise in image feature extraction. The soft threshold function is shown in equation (1).

$$\begin{cases} soft(x,T)=\begin{cases} x+T, & x\le T \\ 0, & |x|\le T \\ x-T, & x\ge T \end{cases} \\ \dfrac{\partial y}{\partial x}=\begin{cases} 1, & x\le -T \\ 0, & |x|\le -T \\ 1, & x\ge T \end{cases} \end{cases} \tag{1}$$

In equation (1), x represents the input variable. T represents the noise processing threshold. y represents the output result of the soft threshold function. BN helps alleviate the problem of internal covariate shift in deep network training, which is the problem of the input distribution of the network layer constantly changing with the training process. During the one-way transmission of each batch of data, the BN operation is usually placed after the fully connected or convolutional layer and before the activation function. This approach ensures that the hidden activation inputs of the network are normalized, thereby reducing differences between different training batches [24-25]. In BN, the training batch mean needs to be calculated first, as shown in equation (2).

$$\mu_B \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \tag{2}$$

In equation (2), μ_B represents the mean of a small batch of input sets. m represents the size of the small batch input set. x_i represents the input set elements. After calculating the mean of each small batch, the variance of each small batch can be calculated according to equation (3).

$$\sigma_B^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}\left(x_i-\mu_B\right)^2 \tag{3}$$

In equation (3), σ_B^2 represents the variance of a small batch of inputs. According to equation (4), the small batch data is normalized.

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{4}$$

In equation (4), $\hat{x}_i$ represents the normalized input data. ε is to prevent positive numbers with zero denominators in normalization calculations. After normalizing the input data, the final adjustment can be made to the input data using equation (5).

$$y_i \leftarrow \gamma \hat{x}_i + \beta \tag{5}$$

In equation (5), y_i represents the input data after completing BN. γ and β represent transformation reconstruction parameters, where γ is the scale transformation reconstruction factor, and β is the translation reconstruction factor. Dropout is a commonly used regularization technique in deep learning, mainly used to prevent overfitting in neural networks during operation [26]. During the training process, during each forward propagation, Dropout randomly selects a batch of neurons and sets their output to zero. In each iteration, the structure of the network is different. Activation functions in neural networks can increase the expressive power of the model and affect the learning speed of the network. Different activation functions have different effects on the network [27-28]. Common activation functions include rectified linear unit (ReLU), parametric rectified linear unit (PReLU), and exponential linear unit (ELU). The definition of the PReLU function is shown in equation (6).

$$f(x_i) = \begin{cases} x_i, & x_i > 0 \\ a_i x_i, & x_i \le 0 \end{cases} \tag{6}$$

In equation (6), x_i represents the input of the activation function. a_i represents the negative half axis slope parameter. When parameter a_i in equation (6) is 0, the PReLU function is the ReLU function. If parameter a_i is a normal number $\leqslant 0.1$, the PReLU function is the Leaky ReLU function. The definition of the ELU function is given by equation (7).

$$f(x_i) = \begin{cases} x_i, & x_i > 0 \\ a\left(e^{x_i} - 1\right), & x_i \le 0 \end{cases} \tag{7}$$

In equation (7), a represents the parameter. The above functions can effectively avoid the problem of gradient vanishing in the learning process of neural networks. ResNet is composed of stacked residual blocks, and the common residual block structure is shown in Figure 2 [29-30].

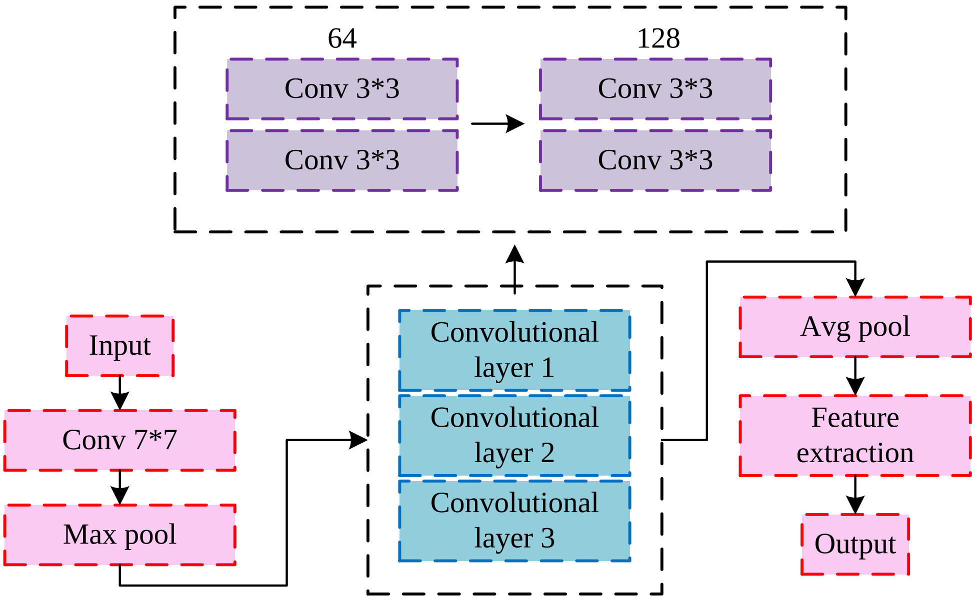


Figure 1: ResNet structure based on improved VGG19

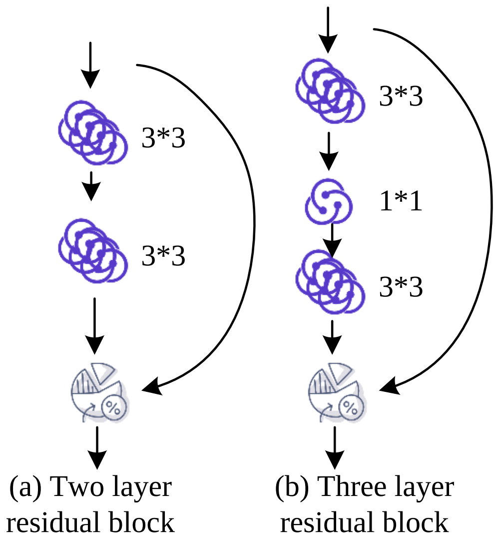


Figure 2: Residual block structure

The two-layer residual block structure contains two 3*3 convolutional structures. The three-layer residual structure consists of two 1*1 and one 3*3 convolutional structures. The expression for the residual structure is given by equation (8).

$$x_{l+1} = x_1 + F\left(x_l, W_l\right) \quad (8)$$

In equation (8), x_l represents shallow units. $F\left(x_l, W_l\right)$ represents the residual function. The feature expression of deep units in ResNet is given by equation (9).

$$x_L = x_l + \sum_{i=1}^{L-1} F\left(x_i, W_i\right) \quad (9)$$

In equation (9), x_L represents a deep unit. In the design of residual modules, a division into channel-specific threshold modules and inter-channel shared threshold modules can be implemented, depending on the complexity of the network and the requirements of the task at hand. In different threshold modules for each channel, each channel has an independent threshold, which means that each channel can adjust its threshold according to its own characteristics to adapt to different feature responses. In the inter-channel shared threshold module, all channels share the same threshold, which means that regardless of the characteristics of the input data, all channels will use the same threshold for processing. The two structures are shown in Figure 3.

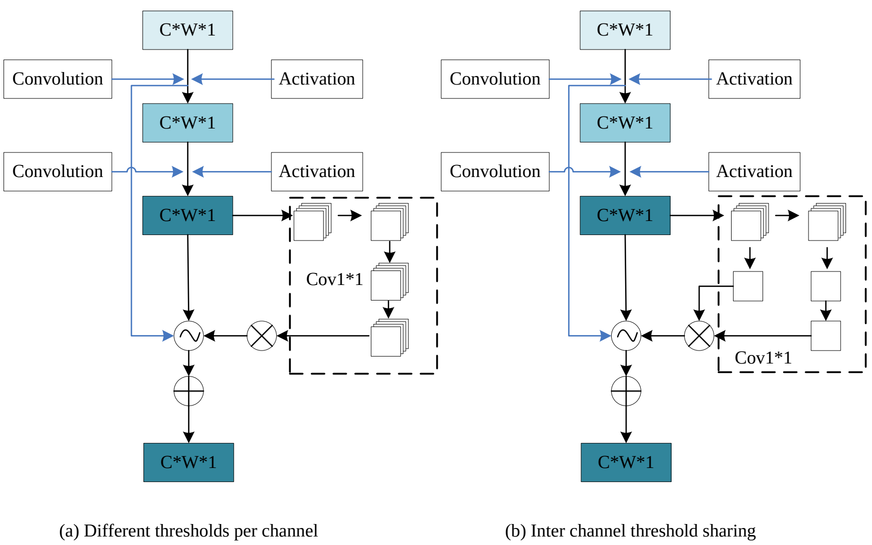


Figure 3: Different threshold modules by channel share threshold modules between channels

3.2 Tourism scene reconstruction based on PIS technology

ResNet can effectively reduce the impact of unimportant features on tourism scene reconstruction in scenic spot captured images. Panoramic image stitching (PIS) technology is the technique of stitching images from different angles in the same region, discarding overlapping parts, and synthesizing panoramic images of the region [31-32]. Therefore, this study first uses ResNet to extract features from tourist attraction images, removing non important features from the captured images, and then uses PIS technology to synthesize panoramic images of tourist attractions. In PIS technology, most images used for stitching are imaged using fish eye images, as shown in Figure 4.

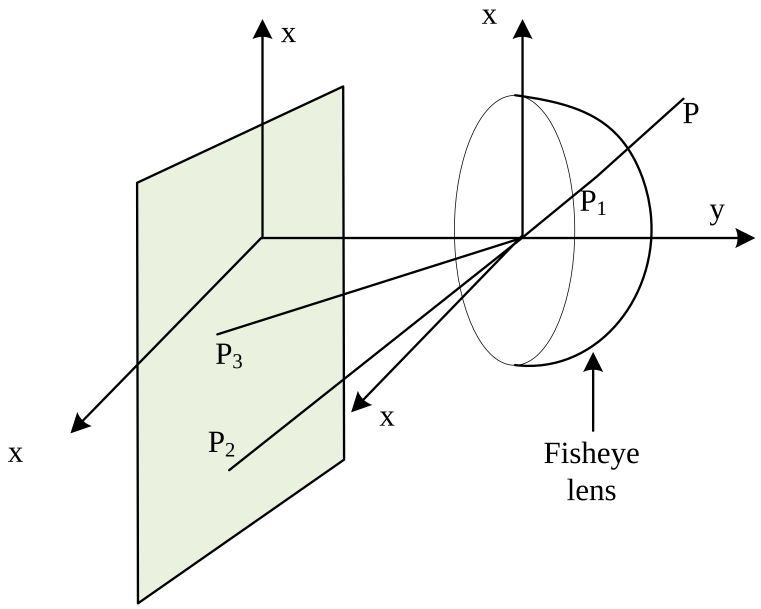


Figure 4: Imaging principle of fisheye images

Any point situated outside the lens is associated with the camera's position. Subsequent to traversing the fish eye lens, the camera's light path undergoes a shift, and the position of the imaging point on the imaging plane also undergoes a corresponding change. The imaging models for fish eye images include four types: isometric projection models (IPM), stereo angle projection models (SAPM), orthogonal projection models (OPM), and stereo projection models (SPM). The expression for the IPM is given by equation (10).

$$l = f\theta \tag{10}$$

In equation (10), l represents the imaging distance. θ represents the angle between the incident light and the longitudinal axis of the optical axis. f represents the camera focal length. The SAPM expression is given by equation (11).

$$l = 2f\sin\left(\frac{\theta}{2}\right) \tag{11}$$

The viewing range of IPM and SAPM is both 360°. The imaging distance in IPM is positively correlated with the angle between the light rays. Compared to IPM, when the imaging distance is the same, the light angle of SAPM will be larger. The viewing range of OPM is only 180°, and the distortion of its imaging pattern is more obvious, as expressed in equation (12).

$$l = f\sin\theta \tag{12}$$

The perspective range of SPM is greater than that of OPM, but lower than that of IPM and SAPM, as expressed in equation (13).

$$l = 2f\tan\left(\frac{\theta}{2}\right) \tag{13}$$

The four imaging models of fish eye images can be uniformly expressed, as shown in equations (14) [33-34].

$$l=\theta\left(1+k_1\theta^2+k_2\theta^4+k_3\theta^6+k_4\theta^8\right) \quad (14)$$

In equation (14), k_i represents the imaging parameters of the image. Fish eye images may experience radial distortion, tangential distortion, or thin prism distortion during shooting due to their large viewing angle range and planar imaging. Radial distortion refers to the basic absence of distortion at the center point of an image, typically exhibiting barrel shaped distortion that is stretched outwards or pillow shaped distortion that is compressed towards the center [35]. The correction of radial distortion is given by equation (15).

$$\begin{cases} x_c = x\left(1+m_1r^2+m_2r^4+\cdots\right) \\ y_c = y\left(1+m_1r^2+m_2r^4+\cdots\right) \end{cases} \quad (15)$$

In equation (15), (x_c, y_c) represents the corrected radial distortion point. (x, y) represents the radial distortion point before correction. m_n represents the distortion coefficient of the fish eye image. The tangential distortion correction model is equation (16) due to the mechanical error caused by the equipment.

$$\begin{cases} x_c = 2p_1xy+P_2\left(r^2+2x^2\right) \\ y_c = 2p_2xy+P_1\left(r^2+2y^2\right) \end{cases} \quad (16)$$

In equation (16), p_n represents the tangential distortion coefficient in the camera hardware facilities. Thin prism distortion is similar to tangential distortion and is also caused by mechanical errors in hardware facilities, but can be basically ignored in imaging. The panoramic projection models of tourist attractions are divided into planar panoramic, cylindrical panoramic, and spherical panoramic. A planar panoramic image uses feature point matching and image registration techniques to project and stitch photos onto the same plane, suitable for scenes with limited viewing angles. The cylindrical panoramic view uses a fixed fulcrum and is captured by horizontal rotation of the camera, projecting the image onto the cylindrical surface to achieve a 360° horizontal viewing angle. However, the vertical viewing angle is less than 180°, making it suitable for scenes with wider viewing angles. A spherical panoramic image provides a panoramic view by projecting the image onto a spherical surface. Both horizontal and vertical viewing angles can reach 360° and 180°, making it suitable for virtual scene roaming that requires all-round observation. This splicing technique is more complex, but the effect is more outstanding. These three methods each adapt to different application requirements, and through different projection and stitching techniques, can achieve a limited to comprehensive perspective expansion. This study uses fish eye images for VRSFTA reconstruction, and uses ResNet to extract and filter features from fish eye images. The specific PIS steps are shown in Figure 5. When constructing VRSFTA, it is necessary to first bring a fish eye imaging device to the scenic spot for photo collection. After completing the collection, the fish eye image is corrected to conform to human vision, and the image is transformed into linear storage. After preprocessing the image, ResNets can be used for feature extraction, which requires constructing the image scale space. When constructing the scale space, a Gaussian pyramid structure is used, with the original image placed at the bottom of the pyramid. After Gaussian blur processing, the original image is placed on the upper layer of the original image, and so on. The image scale space is shown in equation (16).

$$L(x,y,\sigma)=\mathrm{G}(x,y,\sigma)*I(x,y) \quad (17)$$

In equation (16), $L(x,y,\sigma)$ represents the two-dimensional image scale space. σ represents the scale space parameter. $\mathrm{G}(x,y,\sigma)$ represents a two-dimensional Gaussian function. $I(x,y)$ represents the converted grayscale image.

After preprocessing the image, ResNet can be used for feature extraction. After completing feature extraction, it is necessary to observe whether there is overlap between the images. If there is no overlapping area, the images with other overlapping areas are replaced. After there are overlapping areas in the image, image registration can be performed. The erroneous matching points in the image are removed, and the image with the proposed erroneous matching points is projected onto a sphere, which is then reflected as a planar image. Finally, all images can be fused. After completing the reconstruction of the virtual scene, collision detection is also required to ensure the authenticity of the scene. This study uses the bounding box method (BBM) for collision detection, and the process is shown in Figure 6.

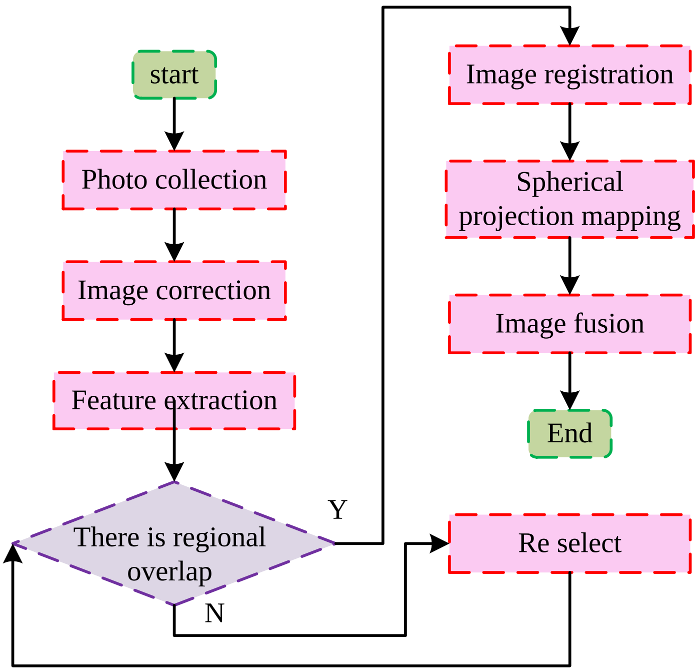


Figure 5: VR scene reconstruction process of tourist attractions

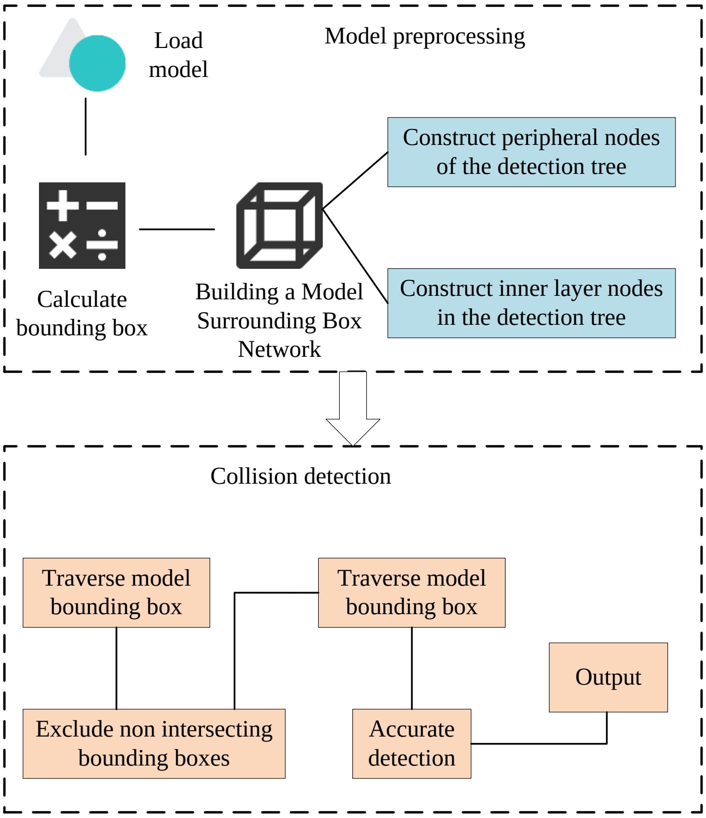


Figure 6: Process of model collision detection method based on BBM

When using BBM for collision detection, it is necessary to first initialize the detection scene, determine the set of objects that need collision detection, and then calculate the bounding boxes of all objects to be detected. The bounding boxes of all objects are compared to be detected, and if the bounding boxes do not intersect, it is considered that no collision has occurred. If the bounding boxes intersect, it is considered a collision and the virtual scene needs to be rearranged based on the detection results.

4 Results

4.1 Experimental parameters and environmental settings

The experiments are all conducted in the Python framework, with the device operating system being Windows 10 Pro 64bit and the device processor being Intel (R) Core (TM) i5-10300H CPU@ 2.50GHz. The operating memory of the device is 32.0GB, and the GPU of the device is NVIDIA GeForce GTX 1660Ti. The algorithm training and testing process adopts SGD optimizer with cross entropy loss function, and the initial learning rate is 0.01. After every 20 iterations, the learning rate decreases by 0.1, the batch size is 128, and the maximum number of iterations is 200. Table 2 shows the specific configuration of the environment.

Table 2: Experimental environment setting

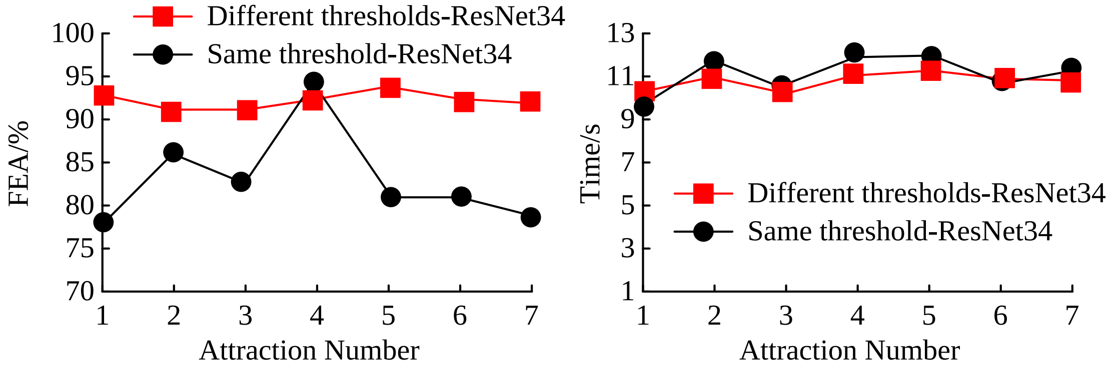
Hardware configuration		Software configuration	
CPU	Intel(R) Core (TM) i5-10300H	Operating system	Windows 10 Pro 64-bit
GPU	NVIDIA GeForce GTX 1660Ti	Deep learning framework	PyTorch 1.7.1
RAM	32 GB DDR4	Python version	Python 3.8
Storage	1 TB SSD	CUDA Version	CUDA 11.2
/	/	CuDNN version	CuDNN 8.1

The dataset used for simulation training and testing of ResNets is the CIFAR-10 dataset. The CIFAR-10 dataset contains 60,000 32×32-pixel color images, divided into 10 categories with 6,000 images per category. These categories include airplanes, cars, birds, cats, deer, dogs, frogs, horses, boats, and trucks. These data come form a publicly available dataset for image feature extraction and processing, which can be directly used for feature processing of image data. The dataset is divided into 50,000 training images and 10,000 test images. This dataset is commonly used for image recognition and classification tasks in computer vision research. The study divides the training dataset into 5 batches, each containing 10,000 images. The hardware facilities for building VR scenes are the same as those for image feature extraction experiments. The VR system development engine is the Unity engine, and the system's new energy optimization adopts LOD technology. In the virtual scene, the architectural structure of the scenic area is modeled using 3D models, and the landscape and climate are integrated using 2D images. In the modeling of building structures, 3D models mainly focus on the structural contours, and structural details are implemented using textures.

4.2 Verification of image feature extraction effect based on ResNet

The constructed ResNet34 is based on the VGG16 network. To compare the extraction effects of different threshold-ResNet34 (DT-RN) modules and same threshold-ResNet34 (ST-RN) on important features of tourist attraction images, the feature extraction effects using DT-RN and ST-RN structures are compared. The results are shown in Figure 7.

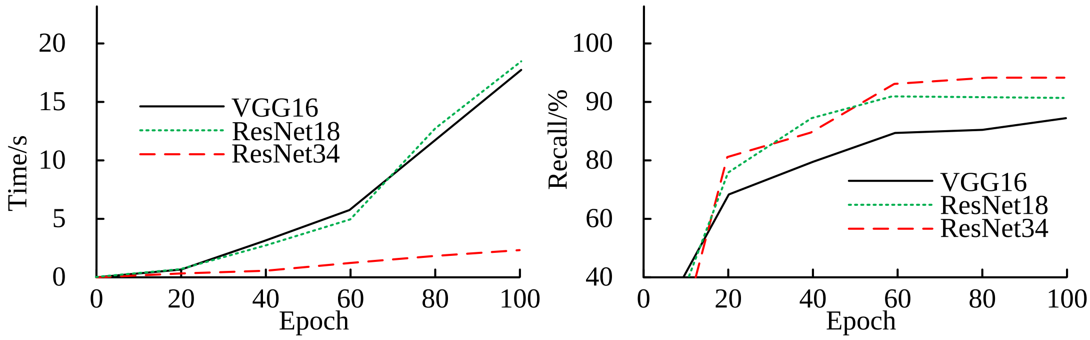
Figure 7 (a) shows the comparison of feature extraction accuracy (FEA) between two residual structures. When using the DT-RN structure, the FEA of the network for scenic spot images is stable between 90%-95%. When using the ST-RN module, ResNet has a wide range of FEA fluctuations for scenic spots. After analysis, when using ST-RN, the feature extraction performance of complex images is poor, making it difficult to distinguish between important and non important features of the image. Figure 7 (b) compares the feature extraction efficiency (FEE) of two residual structures. The FEE of the two feature extraction methods is basically the same. This study compares the training efficiency and recall of ResNet34, VGG16, and ResNet18 networks on the TMD, as shown in Figure 8. Figure 8 (a) shows the comparison of training efficiency among three models. As the number of training iterations increases, the training time of all three networks begins to increase. The training time variation of ResNet34 is the smallest, followed by VGNN16. As for the time required to complete 100 iterations, ResNet34 is 2.6 seconds, VGG16 is 17.2s, ResNet18 is 16.8s. The training efficiency of ResNet34 network is the highest, with only 2.6s required for 100 iterations of training. Figure 8 (b) shows a comparison of the training recall rates of the model. As iterations increase, the recall rates of all three models are increasing. The recall rates of ResNet18 and ResNet34 can both increase to around 90%. For the highest recall rate, ResNet18 is 90.8%, ResNet34 is 95.2%, wile the VGG16 is only 88.7%. The recall rates of all three models can reach a high level, while the ResNet34 network has the best recall level. The testing time and recall results of ResNet18, ResNet34, and VGG16 on the TMD are displayed in Figure 9.



(a) Comparison of FEA for different residual structures

(b) Comparison of FEE for different residual structures

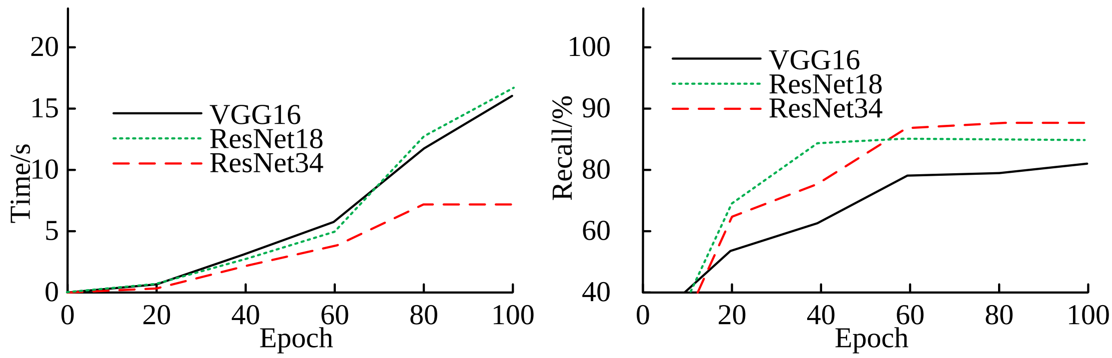
Figure 7: Comparison of feature extraction effects of different structures



(a) Training efficiency comparison

(b) Comparison of recall rates

Figure 8: Analysis of neural network training time and recall rate



(a) Training efficiency comparison

(b) Comparison of recall rates

Figure 9: Analysis of neural network testing time and recall rate

Figure 9 (a) shows a comparison of the testing time for each neural network. As the number of network detection iterations increases, the time required for each network to complete testing also increases significantly. ResNet34 has the least increase in time consumption, and completing 100 iterations of testing for this network takes about 7s. The time consumption of VGG16 is much higher than ResNet34 but slightly lower than ResNet18, and it takes about 15.5s for the network to complete 100 iterations. ResNet18 takes the most time to complete testing, but it is very close to VGG16, taking about 16s to complete 100 tests. Figure 9 (b) shows the changes in recall rates of three networks during testing. The recall rates of ResNet18 and ResNet34 tend to stabilize at the 40th and 60th iterations, with stable recall rates approaching 85% and 89%, respectively. The recall rate of VGG16 dies not show a stable trend and continues to rise in the 100th iteration. At this time, the recall rate of VGG16 testing is only about 80%. The fusion of residual modules can effectively accelerate the training and testing efficiency of the network, and improve the recall rate of the network to the output results. To further verify the effectiveness of ResNet34 network in extracting features of tourist attractions, this study compares the loss values and accuracy of VGG16 and ResNet34. The training and testing loss values of VGG16 and ResNet34 on the TMD are exhibited in Figure 10. In Figure 10 (a), the change in loss values of the VGG16 network during training on this dataset is similar to the training results on the CIFAR-10 dataset. On the TMD, the change in loss values is more pronounced, with greater fluctuations. The test results of VGG16 on two datasets are completely different. When tested on the TMD, the model does not show a convergence trend and shows a certain upward trend as the iterations increase. In Figure 10 (b), the ResNet34 model converges well during training, completing convergence directly in the 50th iteration, and the network loss value after convergence tends to 0. The testing performance of this network on the TMD is poor, but overall, it shows a significant downward trend. After the 30th iteration, the loss value during network testing begins to fluctuate around 1.5. Compared to the VGG16 network, ResNet34 has better convergence performance on complex TMDs. Figure 11 shows the accuracy changes of VGG16 and ResNet34 on the TMD.

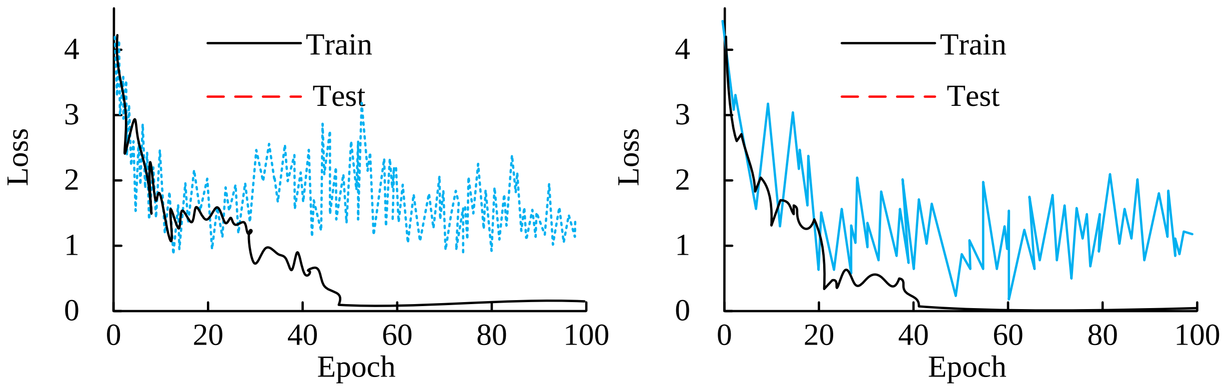
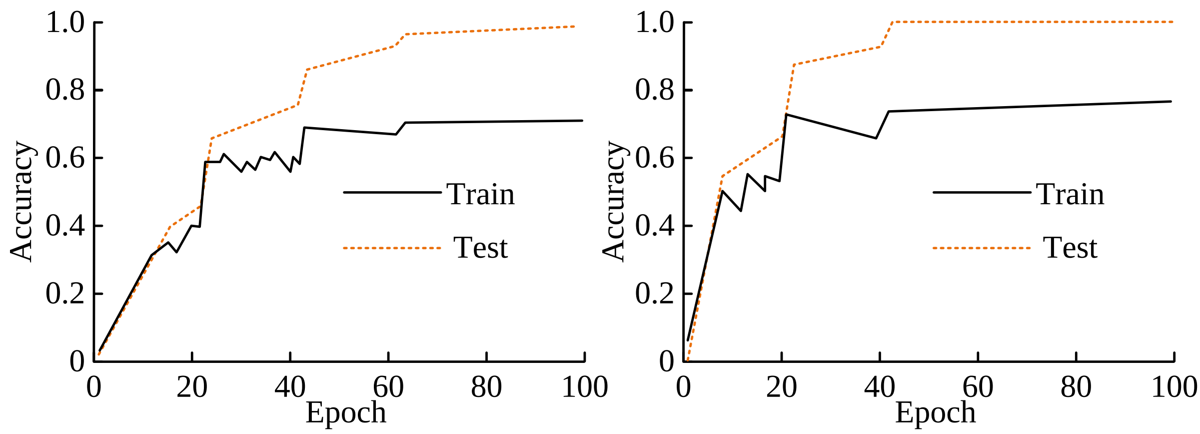


Figure 10: VGG 16 and ResNet34 changes in loss values in the TMD



(a) Travel photos-VGG16 (b) Travel photos-ResNet34

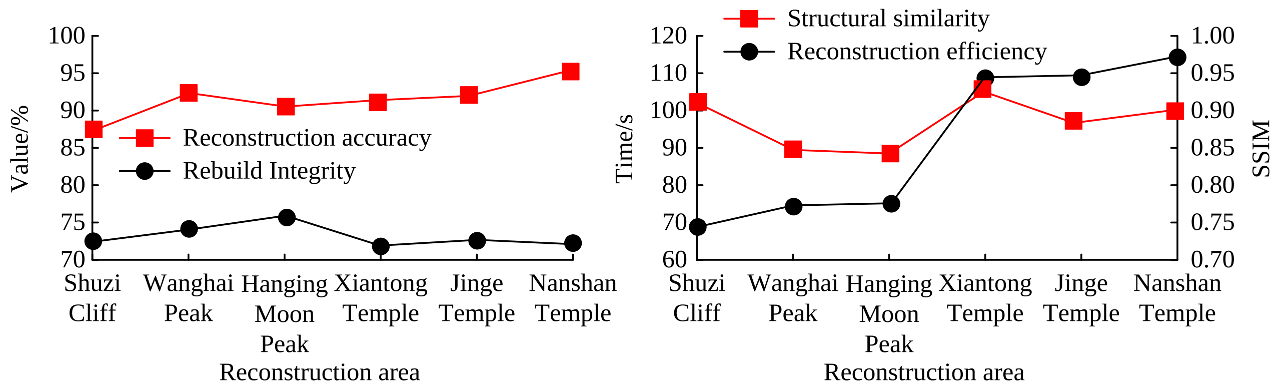
Figure 11: Changes in the accuracy of VGG16 and ResNet34 on the TMD

In Figure 11 (a), when VGG16 is trained on the TMD, the accuracy of the network increases with the increase of iterations. At the 70th iteration, the training accuracy of the network tends to 100%. Before reaching 70 iterations, the training accuracy of the network shows a phased increase. In the first 20 iterations, the accuracy of the network increases the fastest, from 0 to about 45%, and then increases by about 10% in each stage. The testing accuracy of this network on the TMD is at a relatively low level. After 70 iterations, its testing accuracy is only about 70%, far lower than the training accuracy. In Figure 11 (b), during training, compared to VGG16, at the 50th iteration, ResNet34's accuracy approaches 100%. When the iteration is less than 50, the accuracy of the ResNet34 network has also shown a phased increase. ResNet34 reaches its highest accuracy during the 30th iteration of training, and then its accuracy decreases to around 60% before the 50th iteration after the 30th iteration, and then increases to around 75% after the 50th iteration.

4.3 Verification of tourism promotion effect based on VR

When analyzing the promotion effect of tourist attractions based on VR tourism promotion mechanism, this study first determines the reconstruction effect of VRSFTA. This study takes the Mount Wutai tourist attractions as an example to analyze the reconstruction accuracy, reconstruction efficiency, reconstruction integrity and structural similarity of different areas of the scenic spots. The results are shown in Figure 12.

Figure 12 (a) shows the comparison of reconstruction accuracy and reconstruction integrity of virtual scenes in different regions of Mount Wutai. The reconstruction accuracy of the constructed virtual scenes of scenic spots in various regions is maintained at over 85%, and the reconstruction completeness of all positions is also maintained at around 70%. Figure 12 (b) shows the comparison between the reconstruction efficiency of virtual scenes and the structural similarity of the reconstructed scenes. The reconstruction of the internal structure of the Mount Wutai building takes the highest time, about 110s. The reconstruction of other locations takes about 90s. The structural similarity of all scenes in the scenic spot can reach 0.7 or above. The interior reconstruction effect of the scenic building is shown in Figure 13. Figures 13 (a) and (b) show the virtual reconstruction effect of the left and front structures inside the building. In Figure 13, this scenic spot is a temple type attraction. The reconstruction effect of the left side chapel structure is good, and the reconstruction scene can better display the main structure of the temple's side chapel, such as the two door pillars and structures similar to door curtains. The main goal of the reconstruction of the front structure is to worship Buddha statues and their protectors in temples. The details of the Buddha statues in the reconstruction scene are relatively comprehensive, but the reconstruction effect of the protectors is slightly poor. After analysis, the reason is that during the extraction of image features, the Buddha statue protector statue was mistakenly identified as non-important features such as tourist flow in the scenic area, resulting in feature loss in the reconstruction of the front structure. The visualization results of VR reconstruction of tourist attractions are shown in Figure 14.



(a) Analysis of Reconstruction Accuracy and Reconstruction Integrity

(b) Reconstruction efficiency and structural similarity analysis

Figure 12: Analysis of the reconstruction effect of the VRSFTAs

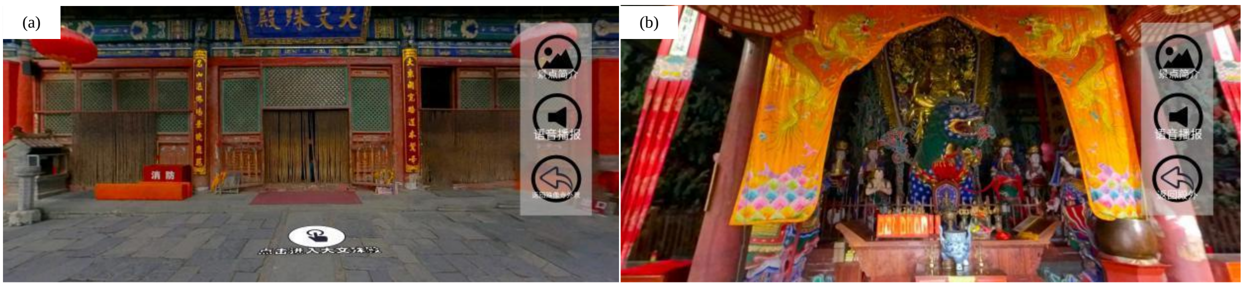


Figure 13: Visualization results of VR reconstruction scenes within scenic spots

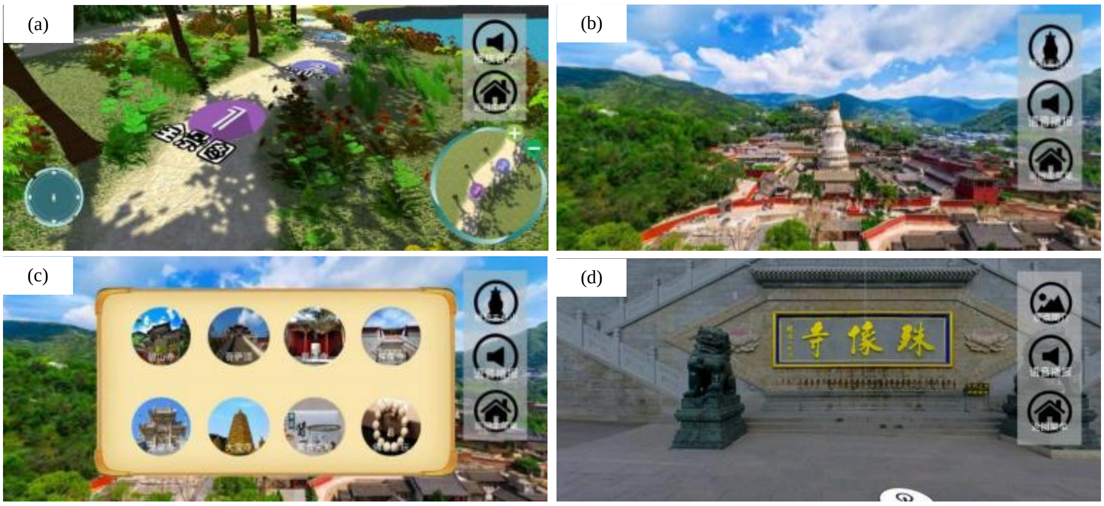


Figure 14: Visualization results of VR reconstruction of tourist attractions

This study investigates the comprehensive evaluation of tourists on VR attraction promotion, and the results are shown in Table 3.

Table 3: Scenic spot promotion effect evaluation

Tourist ID	Recommended methods	Count the number of people	Satisfaction rating	Surprise rating	Experience effect
1	This paper	50	89***	92***	Good
	Traditional method	50	72	72	Generality
2	This paper	50	87***	96***	Excellent
	Traditional method	50	70	71	Generality
3	This paper	50	88***	95***	Excellent
	Traditional method	50	71	69	Generality
4	This paper	50	87***	94***	Excellent
	Traditional method	50	70	68	Generality
5	This paper	50	89***	92***	Good
	Traditional method	50	69	72	Generality
6	This paper	50	92***	91***	Good
	Traditional method	50	71	72	Generality

Note: *: P<0.05%; **: P<0.01%; ***: P<0.001%.

In Table 2, the volunteers are relatively satisfied with the comprehensive rating of the promotion mechanism for the tourist attraction, all of which remain above 85 points. Among them, participant 6 rates the promotion mechanism above 90 points. All participants in the test show a high level of surprise towards the recommended attractions through the promotion mechanism, with scores consistently above 92. Participants 2, 3 and 4 rate the experience of VR scenes in scenic spots as excellent, while participants 1, 5 and 6 rate the spots as good.

5 Discussion

A new method for tourism promotion has been proposed, which utilizes VR technology and ResNet for image feature extraction and scenic spot recommendation. Compared with the advanced methods outlined in related works, several key differences and advantages have emerged. Nitu et al. [9] enhanced the precision of personalized recommendations through the examination of Twitter data and the application of machine learning classifiers. However, the proposed method exceeds this by offering an immersive virtual reality experience. This not only helps with decision-making but also enhances the overall appeal of the destination through more attractive and realistic previews. Compared with the method proposed by Gao et al. [10], the proposed method focuses on visual and experiential aspects, providing virtual tours of potential destinations for tourists and helping them make more informed and personalized travel choices.

The experimental results show that the proposed ResNet34 model has a fast convergence speed and high accuracy in scene reconstruction, which is superior to other models such as VGG16 and ResNet18. This outstanding performance can be attributed to the skip connections in ResNet, which helps alleviate the problem of gradient vanishing and enable deeper network architectures. VR reconstruction technology also demonstrates a high degree of detail and structural similarity, providing almost realistic experiences for tourist attractions.

Integrating VR technology into tourism promotion mechanisms has a profound impact on the industry. Firstly, it has the potential to change the way tourists interact with and perceive tourist destinations, providing more immersive and informative decision-making tools. Secondly, achieving product differentiation through virtual experiences can enhance the competitiveness of tourism enterprises. Finally, it can reduce the impact on the environment and achieve more sustainable tourism practices by reducing the need for on-site visits during the decision-making stage.

Although the proposed method has performed well in personalized travel recommendations for tourists, there are still several areas that need improvement in the future. The representation of long-range structures in VR scenes still needs improvement and currently lacks details. Advanced image processing techniques and deeper neural networks can be explored to solve this problem. In addition, the integration of real-time data, such as weather conditions and crowd levels, can provide more value to tourists. Finally, developing more interactive and user-customizable VR experiences can further enhance user engagement and satisfaction.

6 Conclusion

To optimize the current tourism recommendation mechanism and change the current situation where recommendations are only based on tourists' past preferences or recent browsing content, this study used VR to design a scenic spot recommendation mechanism based on tourists' own experience. This technology used ResNet to extract features from scenic spot images, and then utilized panoramic stitching technology to construct VR scenes of relevant scenic spots. This allowed tourists to experience certain attractions in advance, thereby attracting them to the actual location of the attraction. The results verified that the designed ResNet34 could converge after 50 iterations of training on both the TMD and the CIFAR-10 dataset, and the loss values after convergence tended to 0. When tested on CIFAR-10, the converged loss value of ResNet34 was about 0.4, and the lowest loss value on the TMD was about 0.1. All tourists gave a comprehensive rating of 85 or above for the promotion mechanism, with a surprise rating of 90 or above for the promotion. The VR reconstruction technology designed in this project could clearly display the texture, structure, and other details of the target building structure when reconstructing virtual scenes of scenic spots. However, this technology has a poor reconstruction effect on distant structures when reconstructing panoramic views of scenic spots, and cannot show more details to tourists. In the future, it is possible to add a distance dimension to virtual scenes, so that tourists can intuitively experience all the characteristics of the scenic spot when experiencing them in the virtual scene.

References

[1] X. Zheng, B. Han, and Z. Ni, “Tourism route recommendation based on a multi-objective evolutionary algorithm using two-stage decomposition and pareto layering,” IEEE/CAA Journal of Automatica Sinica, vol. 10, no. 2, pp. 486-500, 2023. https://doi.org/10.1109/JAS.2023.123219

[2] W. Tu, K. Mai, Y. Zhang, Y. Xu, J. Huang, M. Deng, L. Chen, and Q. Li, "Real-time route recommendations for e-taxies leveraging GPS trajectories," IEEE Transactions on Industrial Informatics, vol. 17, no. 5, pp. 3133-3142, 2021. https://doi.org/10.1109/TII.2020.2990206

[3] Y. Wu, K. Li, G. Zhao, and X. Qian, "Personalized long- and short-term preference learning for next POI recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 4, pp. 1944-1957, 2022. https://doi.org/10.1109/TKDE.2020.3002531.

[4] Y. Liu, F. Wu, C. Lyu, X. Liu, and Z. Liu, "Behavior2vector: Embedding users’ personalized travel behavior to vector," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 8346-8355, 2022. https://doi.org/10.1109/TITS.2021.3078229

[5] Y. Liu, H. Wu, K. Rezaee, M. R. Khosravi, O. I. Khalaf, A. A. Khan, D. Ramesh, and L. Qi, "Interaction-enhanced and time-aware graph convolutional network for successive point-of-interest recommendation in traveling enterprises," IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 635-643, 2023. https://doi.org/10.1109/TII.2022.3200067

[6] Y. Li, F. Xiong, Z. Wang, Z. Chen, C. Xu, Y. Yin, and L. Zhou, "Spatial-temporal deep intention destination networks for online travel planning," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 2688-2700, 2022. https://doi.org/10.1109/TITS.2021.3091534

[7] H. Mezni, D. Benslimane, and L. Bellatreche, "Context-aware service recommendation based on knowledge graph embedding," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 11, pp. 5225-5238, 2022. https://doi.org/10.1109/TKDE.2021.3059506

[8] Y. Yin, N. Zhang, Z. Chen, M. Li, Y. Li, H. Gao, and L. He, "Exploiting user preferences for multiscenarios in query-less search," IEEE Transactions on Computational Social Systems, vol. 9, no. 6, pp. 1794-1806, 2022. https://doi.org/10.1109/TCSS.2022.3181271

[9] P. Nitu, J. Coelho, and P. Madiraju, "Improvising personalized travel recommendation system with recency effects," Big Data Mining and Analytics, vol. 4, no. 3, pp. 139-154, 2021. https://doi.org/10.26599/BDMA.2020.9020026

[10] Q. Gao, F. Zhou, K. Zhang, F. Zhang, and G. Trajcevski, "Adversarial human trajectory learning for trip recommendation," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 4, pp. 1764-1776, 2023. https://doi.org/10.1109/TNNLS.2021.3058102

[11] H. Liu, Y. Tong, J. Han, P. Zhang, X. Lu and, and H. Xiong, "Incorporating multi-source urban data for personalized and context-aware multi-modal transportation recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 2, pp. 723-735, 2022. https://doi.org/10.1109/TKDE.2020.2985954

[12] F. Huang, J. Xu, and J. Weng, "Multi-task travel route planning with a flexible deep learning framework," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 7, pp. 3907-3918, 2021. https://doi.org/10.1109/TITS.2020.2987645

[13] W. Wang, N. Kumar, J. Chen, Z. Gong, X. Kong, and W. Wei., "Realizing the potential of the internet of things for smart tourism with 5G and AI," IEEE Network, vol. 34, no. 6, pp. 295-301, 2020. https://doi.org/10.1109/MNET.011.2000250

[14] Z. Lv, D. Chen, R. Lou, and H. Song, "Industrial security solution for virtual reality," IEEE Internet of Things Journal, vol. 8, no. 8, pp. 6273-6281, 2021. https://doi.org/10.1109/JIOT.2020.3004469

[15] J. Pirker, and A. Dengel, "The potential of 360° virtual reality videos and real VR for education - a literature review," IEEE Computer Graphics and Applications, vol. 41, no. 4, pp. 76-89, 2021. https://doi.org/10.1109/MCG.2021.3067999

[16] S. R. Dubey, "A decade survey of content-based image retrieval using deep learning," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 5, pp. 2687-2704, 2022. https://doi.org/10.1109/TCSVT.2021.3080920

[17] R. Fernandez-Beltran, B. Demir, F. Pla, and A. Plaza, "Unsupervised remote sensing image retrieval using probabilistic latent semantic hashing," IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 2, pp. 256-260, 2021. https://doi.org/10.1109/LGRS.2020.2969491

[18] Y. Liu, L. Ding, C. Chen, and Y. Liu, "Similarity-based unsupervised deep transfer learning for remote sensing image retrieval," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 11, pp. 7872-7889, 2020. https://doi.org/10.1109/TGRS.2020.2984703

[19] H. Zhai, S. Lai, H. Jin, X. Qian, and T. Mei, "Deep transfer hashing for image retrieval," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 2, pp. 742-753, 2021. https://doi.org/10.1109/TCSVT.2020.2991171

[20] M. Jankowski, D. Gündüz, and K. Mikolajczyk, "Wireless image retrieval at the edge," IEEE Journal on Selected Areas in Communications, vol. 39, no. 1, pp. 89-100, 2021. https://doi.org/10.1109/JSAC.2020.3036955

[21] H. Lu, M. Zhang, X. Xu, Y. Li, and H. T. Shen, "Deep fuzzy hashing network for efficient image retrieval," IEEE Transactions on Fuzzy Systems, vol. 29, no. 1, pp. 166-176, 2021. https://doi.org/10.1109/TFUZZ.2020.2984991

[22] C. Liu, J. Ma, X. Tang, F. Liu, X. Zhang, and L. Jiao, "Deep hash learning for remote sensing image retrieval," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 4, pp. 3420-3443, 2021. https://doi.org/10.1109/TGRS.2020.3007533

[23] C. Yan, B. Gong, Y. Wei, and Y. Gao, "Deep multi-view enhancement hashing for image retrieval," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 4, pp. 1445-1451, 2021. https://doi.org/10.1109/TPAMI.2020.2975798

[24] Z. Xia, L. Wang, J. Tang, N. N. Xiong, and J. Weng, "A privacy-preserving image retrieval scheme using secure local binary pattern in cloud computing," IEEE Transactions on Network Science and Engineering, vol. 8, no. 1, pp. 318-330, 2021. https://doi.org/10.1109/TNSE.2020.3038218

[25] P. Li, L. Han, X. Tao, X. Zhang, C. Grecos, A. Plaza, and P. Ren, "Hashing nets for hashing: a quantized deep learning to hash framework for remote sensing image retrieval," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 10, pp. 7331-7345, 2020. https://doi.org/10.1109/TGRS.2020.2981997

[26] L. Zhang, Z. He, Y. Yang, L. Wang, and X. Gao, "Tasks integrated networks: joint detection and retrieval for image search," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44,

no. 1, pp. 456-473, 2022. https://doi.org/10.1109/TPAMI.2020.3009758

[27] N. Pellas, A. Dengel, and A. Christopoulos, "A scoping review of immersive virtual reality in STEM education," IEEE Transactions on Learning Technologies, vol. 13, no. 4, pp. 748-761, 2020. https://doi.org/10.1109/TLT.2020.3019405

[28] F. Hu, Y. Deng, W. Saad, M. Bennis, and A. H. Aghvami, "Cellular-Connected Wireless Virtual Reality: Requirements, Challenges, and Solutions," IEEE Communications Magazine, vol. 58, no. 5, pp. 105-111, 2020. https://doi.org/10.1109/MCOM.001.1900511

[29] M. Melo, G. Gonçalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa, "Do multisensory stimuli benefit the virtual reality experience? a systematic review," IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 2, pp. 1428-1442, 2022. https://doi.org/10.1109/TVCG.2020.3010088

[30] L. Zhang, and J. Chakareski, "UAV-assisted edge computing and streaming for wireless virtual reality: analysis, algorithm design, and performance guarantees," IEEE Transactions on Vehicular Technology, vol. 71, no. 3, pp. 3267-3275, 2022. https://doi.org/10.1109/TVT.2022.3142169

[31] X. Liu, Y. Deng, C. Han, and M. D. Renzo, "Learning-based prediction, rendering and transmission for interactive virtual reality in ris-assisted terahertz networks," IEEE Journal on Selected Areas in Communications, vol. 40, no. 2, pp. 710-724, 2022. https://doi.org/10.1109/JSAC.2021.3118405

[32] O. Postolache, D. J. Hemanth, R. Alexandre, D. Gupta, O. Geman, and A. Khanna, "Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality," IEEE Journal on Selected Areas in Communications, vol. 39, no. 2, pp. 562-573, 2021. https://doi.org/10.1109/JSAC.2020.3020600

[33] H. Oagaz, B. Schoun, and M. -H. Choi, "Performance improvement and skill transfer in table tennis through training in virtual reality," IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 12, pp. 4332-4343, 2022. https://doi.org/10.1109/TVCG.2021.3086403

[34] K. Bang, Y. Jo, M. Chae, and B. Lee, "Lenslet VR: thin, flat and wide-fov virtual reality display using fresnel lens and lenslet array," IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 5, pp. 2545-2554, 2021. https://doi.org/10.1109/TVCG.2021.3067758

[35] F. Guo, F. R. Yu, H. Zhang, H. Ji, V. C. M. Leung, and X. Li, "An adaptive wireless virtual reality framework in future wireless networks: a distributed learning approach," IEEE Transactions on Vehicular Technology, vol. 69, no. 8, pp. 8514-8528, 2020. https://doi.org/10.1109/TVT.2020.2995877