

Research Article

Large-Scale Image Retrieval of Tourist Attractions Based on Multiple Linear Regression Equations

Yinping Song 

Xi'an International University, Xi'an, 710077 Shaanxi Province, China

Correspondence should be addressed to Yinping Song; xaiu00010@xaiu.edu.cn

Received 26 August 2021; Revised 17 September 2021; Accepted 22 September 2021; Published 12 October 2021

Academic Editor: Miaochoao Chen

Copyright © 2021 Yinping Song. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents an in-depth study and analysis of large-scale tourist attraction image retrieval using multiple linear regression equation approaches. This feature extraction method often relies on the partitioning of the grid and is only effective when the overall similarity of different images is high. The BOF model is borrowed from the method for text retrieval, which generally extracts the local features of an image by the scale-invariant feature transform algorithm and clusters them using k -means to obtain a low-dimensional visual dictionary and characterizes the image features with a histogram vector based on the visual dictionary. However, when there are many kinds of images, the dimensionality of the visual dictionary will be large and it is not convenient to construct the BOF model. The last fully connected layer is taken as the image feature, and it is dimensionalized by the principal component analysis method, and then, the low-dimensional feature index structure is constructed using the locality-sensitive hashing- (LSH-) based approximate nearest neighbor algorithm. The accuracy of our graph retrieval has increased by 8%. The advantages of feature extraction by a convolutional neural network and the high efficiency of a hash index structure in retrieval are used to solve the shortcomings of traditional methods in terms of accuracy and other aspects in image retrieval. The results show that compared with the above two algorithms, for most of the attractions, the method has a relatively obvious advantage in the accuracy of retrieval, and when there are few similar images of a particular attraction in the attraction image library, the accuracy of the query results is not much different from the first two methods.

1. Introduction

With the rapid development of computer technology, various computer-related disciplines have emerged. Computer vision, as an important discipline, involves a wide range and extensive research. Image feature extraction is an important concept in computer vision, and during years of research, many scholars have also identified several of the most basic image features for images. These features are widely used because they are easy to extract, are widely applied, and can accurately describe images, among other properties. Image feature extraction is the key technology for image processing, and the extracted image features are used to perform digital image matching and image retrieval based on the matching results [1]. The existing image retrieval technology provides technical support for the research and implementation of the tour guide system. With

the advent of the era of big data, how to collect, store, and process big data has gradually become an urgent problem in all walks of life, so in the face of such a large-scale image of tourist attractions, how to quickly and accurately retrieve the relevant information of the corresponding attractions is of great significance, which will help to improve the operation of China's tourism industry, optimize the industrial structure, and improve the competitiveness of China's tourism industry in the international arena [2]. In conclusion, with the rapid development of the tourism industry and the arrival of the era of big data, it is of great practical value to realize the fast and accurate retrieval of large-scale images of tourist attractions, which will certainly promote the further development of China's tourism industry.

This paper applies the recommendation system to travel websites, which not only can improve the user's satisfaction with the recommendation but also has important

application value for the improvement of related travel websites. From the theoretical point of view, this paper combines the latent semantic space analysis model with visual Bayesian personalized ranking (VBPR) for the first time and proposes a new hybrid recommendation model multimodal visual Bayesian personalized ranking, which is a contribution to the theoretical aspect of the recommendation system [3] (sparsity problem) and improves user interaction experience, thus contributing to better development of tourism industry. The analysis of an ideal intelligent scenic spot construction system needs to be done from a tourist perspective. Tourists occupy the main position in the tourism industry, and solving the contradiction between tourism supply and demand is the communication bridge between the two [4]. In the scenic spot, intelligent tourism should be built with tourists as the center, and every time tourists visit a scenic spot, it is like putting themselves on the Internet of Things, using cloud computing and other information technology to make big data statistics and analysis of tourists' needs, and then starting from precise data, providing tourists with intelligent scenic spots' starting points, destinations, plans, and itinerary. Before visitors leave, they can also have a virtual experience of the attraction, which means that they can view the attraction and have a comprehensive understanding of it on the Internet, producing ways to attract visitors, such as tourist cartoons and caricatures; visitors to scenic spots can scan QR code e-tickets by using mobile devices, which will automatically save the visitor's information and ensure that the tickets will not be lost.

If we can use the cell phone, which is so intelligent now, to design and realize a powerful and convenient tour guide system software, it will be able to provide a great convenience for the travelers as follows: reduce the cost of travel, make the most suitable for their travel plans, and significantly improve the quality of tourism and the service level of cell phone information technology. The research work of this paper consists of two main parts: the mobile application is the mobile program of the intelligent tourism guide system developed by using Android native development technology, which is mainly responsible for realizing the functions of image acquisition and preprocessing, including compact feature extraction of CDVS of scenic spot images and display of retrieval results and push information from the server-side application; the server-side application is developed based on C++ technology, and the server-side application is mainly responsible for receiving requests from the mobile terminal for attraction retrieval, completing accurate attraction retrieval based on the image features and GPS information sent from the mobile terminal, and pushing the retrieval results and related attraction information to the mobile terminal for display. By introducing the visual retrieval technology based on CDVS features into the intelligent tourist guide system, this paper achieves precise location and accurate identification of attractions greatly improves the accuracy of attraction location and identification, provides technical guarantee for the subsequent accurate pushing of tourist attraction information and surrounding commercial information, and thus improves the user experience of the tourist guide system. Thus, the

work of this paper has greater theoretical research value and practical application value.

2. Current Status of Research

For the currently used content-based large-scale image retrieval, most of the images are characterized by extracting the underlying features of the images, which are used as the basis for image retrieval, and can be divided into the global feature and local feature retrieval methods according to the underlying features [5]. The global features of the image are to treat the image as a whole and extract its global feature information directly by some global feature description method, which makes feature extraction convenient and facilitates retrieval, but the features extracted by this method only focus on the image as a whole and ignore the image detail information, which often has high requirements on the image quality [6]. This feature extraction method has achieved good research, generally by using some feature extraction method to get multiple local features of the image, which were then integrated to describe the whole image; using this method can get the details of the image features but ignore the overall information of the image, and often, the extracted local features are more complex, in large-scale image retrieval [7]. Multithreading plays a very important role in this part. The use of multithreaded programming allows the program to perform multiple tasks at the same time. The interface can be switched when extracting features, and the results of extracting features can be displayed at the same time when sending a request. This multistep simultaneous method can ensure the improvement of the overall operating efficiency of the system. It has a certain impact on the retrieval efficiency in large-scale image retrieval. It is very simple to use: you only need to input the destination information or its orientation information, its surrounding food and beverage information, gas stations, and tourist areas, and a series of information can be very clearly displayed, so this electronic tour guide can be accepted by many tourists and can be widely put into use [8]. However, for a picture, which presents the scenic spots that tourists are interested in and want to reach, it is difficult to get the corresponding tourist information and is even unable to determine the destination and make a travel plan [9]. There exist many tour guide tools as well as tour guide systems that are applied in the travel process. There is a lack of a system that can be used to determine travel locations and facilitate travel planning for tourists.

In large-scale image retrieval, usually, a low-dimensional indexing method needs to be constructed for the feature vector after extracting the global features of the image, and most of the indexing methods constructed based on the description of global features are using the nearest neighbor or approximate nearest neighbor retrieval methods [10]. One is the classical K-D (k -dimensional) tree method, which is a classical nearest neighbor retrieval method, mainly based on the original feature data to build a hierarchical index structure, and we then search from the root node of the tree to find the child node like the query image; this method is better in the retrieval efficiency when the dimensionality is

relatively low than when the data dimensionality is high [11]. The retrieval complexity is relatively high. The other one is to perform cluster quantization analysis on feature vectors. Fujita used k -means to perform cluster analysis on the established tree structure, which results in lower dimensionality [12]. Based on this, Keane et al. proposed a hierarchical k -means tree index structure to compare the similarity between leaf nodes, which is suitable for large-scale image retrieval [13]. Another method is the locality-sensitive hashing- (LSH-) based method proposed in recent years to use hash learning in image retrieval, which is an approximate nearest neighbor search method [14] and improves the user's interactive experience, thereby helping to better develop the tourism industry. The analysis of the ideal intelligent scenic spot construction system needs to start from the perspective of tourists. The main method is to map the original space to the Hamming space and construct multiple hash functions under the corresponding conditions, i.e., the probability that two vectors that are adjacent or similar in the original space are similar enough after they are mapped to the Hamming space by hashing. The hash function can encode the high-dimensional vectors into binary form, which has obvious advantages over the K-D tree method in large-scale image retrieval, and the LSH method occupies less storage space and has more obvious advantages when the data dimension is larger [15].

Firstly, we introduce the background and significance of the current research on the mobile tour guide, analyze the current situation of the system at home and abroad, and introduce the research content of this paper given its problems; we design and implement an intelligent tour guide system. The overall design framework is outlined; then, the design ideas of each submodule in the framework are expanded and introduced, specifically including image acquisition and visual feature extraction on the mobile side, global feature generation on the mobile side, distributed inverted index construction on the server-side, attraction location and information push based on real-time image retrieval, etc.; finally, the release and operation of the designed system are introduced. Based on the interest preferences of the target user's most similar neighbors, the target user's preference level towards the recommended object is predicted, and the system makes recommendations to the target user according to the preference level. The biggest advantage of this recommendation system is that there is no special requirement for the recommended objects, and it can recommend objects that are difficult to represent in text structure. However, since the user rating data for the recommended objects are small compared to the total number of items, this type of system has the problem of data sparsity.

3. Multiple Linear Regression Equation for Large-Scale Tourist Attraction Image Retrieval Design

3.1. Multiple Linear Regression Equation Algorithm Analysis. Nonnegative matrix factorization (NMF) is the decomposi-

tion of a given original matrix into the product of two matrices, both of which are nonnegative, and the result of multiplying the decomposed two matrices together is approximately equal to the original matrix.

$$R \approx PQ^2. \quad (1)$$

Suppose R is an $m \times n$ matrix, and P and Q are $m \times k$ -dimensional and $k \times n$ -dimensional nonnegative matrices, respectively. k generally satisfies $(n \times m)k \leq mn$. Because the decomposed matrix adds the nonnegative constraint, the product of the two decomposed matrices is hardly equal to the original matrix, and the decomposed matrix can only be made equal to the predecomposition matrix as much as possible. Therefore, the nonnegative matrix decomposition is transformed into the following optimization problem [16]:

$$\begin{cases} \phi_{\min}(R, P, Q), \\ P \leq 0, \\ Q \leq 0, \end{cases} \quad (2)$$

$$Q = \begin{cases} y_i = b_0 - \sum_{j=1}^9 b_j x_{ij}, \\ \xi_i^{iid} = N(0, \sigma^2). \end{cases} \quad (3)$$

The Laplacian of Gaussian- (LoG-) based interest point detection algorithm used by CDVS uses a polynomial to approximate the LoG filtering effect, which is called the low-order polynomial (ALP).

$$p(x, y, \sigma) = \sum_{k=0}^{k-1} a_k L_k(x, y) \sigma^3 - \sum_{k=0}^{k-1} a_k L_k(x, y) \sigma^2. \quad (4)$$

The four-octave image is filtered by Laplacian of Gaussian filtering normalized to the scale of the original image. The ALP algorithm uses the point with the first-order derivative of the polynomial at scale 0 as the point of interest by computing the polynomial and later compares the polar values with the eight neighboring pixel points in the plane around the point of interest, as shown in Figure 1.

The stratified sampling statistical model classifies the samples in strata according to the characteristic distribution of the overall units to reduce the differences within each stratum while increasing the differences between strata. On this basis, a certain number of samples are drawn from each stratum separately to portray the distribution of the stratum and constitute the overall sample. Since the samples are reasonably stratified, the stratified sampling statistical model can better capture the travel preferences of users and lay the foundation for a more accurate generation of recommendation lists [17]. The server-side application is developed based on C++ technology. The server-side program is mainly responsible for receiving the scenic spot retrieval request from the mobile terminal and completes the accurate retrieval of the scenic spot according to the image characteristics and GPS information sent by the mobile terminal.

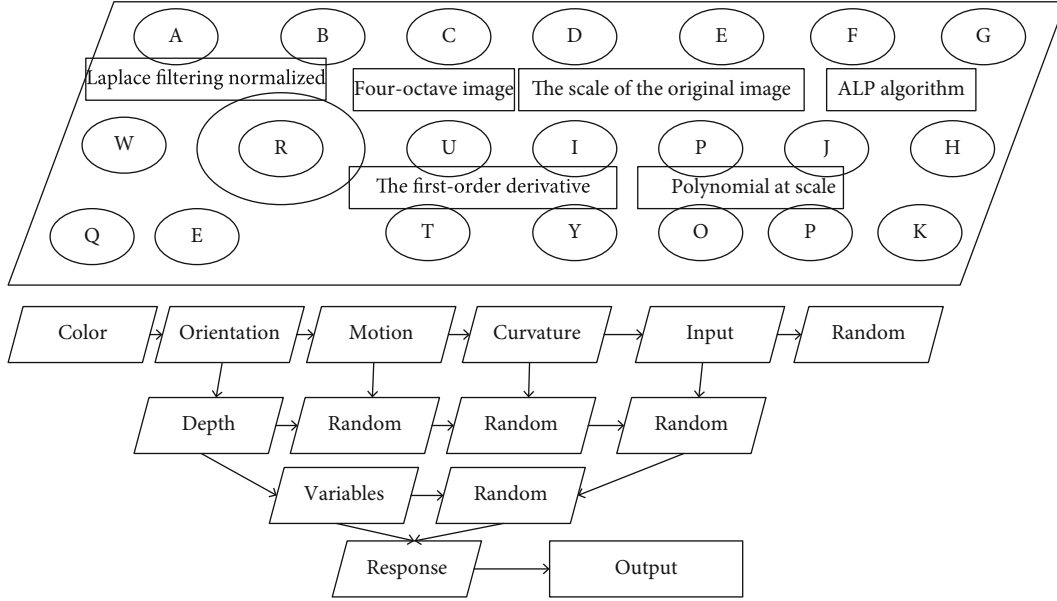


FIGURE 1: Algorithmic framework of multiple linear regression equations.

Setting the weights of sampling statistics based on the subjective assignment evaluation method, hierarchical analysis is applied to adjust the weights of different user attributes; i.e., the relative importance of the same level is compared to establish a new discriminant matrix, and the weights of each user attribute are determined according to the discriminant matrix.

$$G = \begin{bmatrix} G_1 & G_2 & G_3 \\ 1 & \frac{1}{2} & 2 \\ 2 & 1 & \frac{1}{2} \end{bmatrix}. \quad (5)$$

An image $g(x, y)$ of size is filtered with a two-dimensional Gabor filter set of scale p and direction q . This step is essentially a convolution operation with the image using this filter set, respectively, where the mathematical expression of the two-dimensional filter can be expressed as follows:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \cos (2\pi f_0 - \varphi). \quad (6)$$

Texture features are another important global descriptive feature. Texture features describe the local spatial distribution in an image, including information about local light intensity, and are often used to distinguish between images that are rich in information, have similar colors, or are not easy to segment. This feature can describe the overall texture of the image, based on the relationship of the grayscale image response to the full image region values, will not be affected by local extremes and cause mismatching, and has rotational invariance and is not significantly affected by noise, and its research results have been applied to many

important fields. Due to the above properties of texture features, texture features can play a good role in distinguishing landscapes and buildings in different countries and regions. Texture features of images are usually reflected in the form of a cogeneration matrix, and the grayscale cogeneration matrix reflects the grayscale of graphics in the direction, the correlation between adjacent pixels, and the transformation amplitude value, which is a more common method to analyze the distribution of local texture of images.

$$P(i, j) = f(x_1, y_1). \quad (7)$$

The BOF model is a feature that is widely used in the field of image processing. Drawing on a document representation method previously used in the retrieval of information such as text, the BOF model replaces text in the text retrieval model with images and uses the same idea to classify or retrieve images. In the bag-of-words model for text information retrieval, for a document, regardless of the word order and sentence syntax in the content of that document, it is treated as a combination of many individual words only, and the occurrence of each word in the document is random; i.e., the occurrence of any word in the document is independent of other words, and the content in the document is disordered.

$$F_{S_i} = \frac{s^2}{mn} \sum_{(x,y) \in S_i} F_{S_i}(x, y). \quad (8)$$

s represents the area of the graph. If we compare an image to a document, such that the set of images is equivalent to the set of documents and the features of the images are equivalent to the words in the documents, i.e., the image can be understood here as an ensemble of many “visual words,” and there is no order among these visual words.

Then, we can apply the methods for text retrieval to image retrieval, for example, in the field of large-scale image retrieval, and use the efficiency in text processing to improve the speed in large-scale image retrieval.

For images, the “words” in images are not like the words that usually appear in the text, but images are usually multi-dimensional datasets. Therefore, the first thing we need to do is to extract independent “visual words” from the images, which constitute a visual vocabulary [18]. The commonly used method for extracting features is SIFT-based feature extraction, which is suitable for feature extraction to form a vocabulary vector because of the good uniqueness of the features extracted by the SIFT algorithm and the richness of the image information it contains. Using the SIFT algorithm, visual vocabulary is extracted from each type of image set, and all the visual vocabulary is formed into a visual vocabulary vector, as shown in Figure 2.

Due to the natural properties of the image, the statistical properties in one part of the image are the same as in other regions of the image. So, we can use the features learned in one part of the image in other regions, which enables parameter sharing. For example, when we select an arbitrary area from an image and get a small sample, we learn a set of features from this small sample, and then, this set of features is applied to any other part of the image, and the features are used as a detector to convolve with the original image, and we can get different feature values at different locations of the image. In a convolutional neural network, the convolution kernel of each convolutional layer does the convolution operation on the image to get the features of a certain aspect of the whole image. Each convolution kernel will share the same parameters so that when feature extraction is performed on the image, it is not necessary to know the location of the local features of the image, and it also reduces the parameters of the network convolutional neural network.

$$f(z) = \frac{1}{1 - e^{-z}}, \quad (9)$$

$$z_j^{l+1} = \lim_{S_l \rightarrow \infty} \sum_{i=1}^{S_l} a_i^l W_{ij}^{l+1}, \quad (10)$$

$$a_j^{l-1} = f(z_j^{l-1}). \quad (11)$$

After the convolutional features are obtained by the above calculation, they can theoretically be used directly as the feature basis for training the classifier, but they are very computationally intensive and prone to overfitting. Considering the inherent properties of natural images, the features of one region are also applicable to another region; therefore, the features at different locations can be aggregated statistically, and the average or maximum value of features in a local region can be calculated to characterize the specialization of the whole region. The edge extraction algorithm is generally adopted in the implementation process, and the edges are connected to extract the shape of the object. Natural landscape images generally do not have regular shapes, and the shape features extracted from landscape images that

exist in cities are not representative and do not distinguish well between different regions. This feature extraction method does not have application value compared with other features in the study of this paper.

$$r_{u,i} = \alpha - \beta_u - \beta_i, \quad (12)$$

$$D_S = \{(u, i, j) | u \in U\}. \quad (13)$$

The server-side module is the key to the operation of the whole system, which first accepts the images sent from the mobile client and extracts features from the images, extracting RGB color features, texture features, and GIST features of the images, respectively, and using the perceptual hash algorithm to rank the features of the images and perform image matching according to the feature values. By combining various basic image features and filtering them at different levels, the content-based image retrieval is completed, and a more desirable retrieval result can be obtained. In the second stage of “fine sorting,” the RANSAC algorithm described in the previous section is generally used, but this algorithm is computationally intensive and takes a long time to compute.

$$S_{\text{LDR}} = \left\{ \log \left(\frac{\text{dict}(l_{q,i}, l_{q,m})}{\text{dict}(l_{q,i}, l_{q,m})} \right) | (i,j) \in M \right\}. \quad (14)$$

These scores are then quantified into a histogram, and later, the largest value in the histogram is taken as a measure of the score between the two images, with higher scores indicating greater similarity, and then, the candidate images are reordered to obtain the result.

$$C_{\text{LDR}}(\alpha) = \sum_{z \in S_{\text{LDR}}} I \left(\frac{\alpha}{c} \leq z \leq \frac{\alpha - 1}{c - 1} \right). \quad (15)$$

To ensure the security of information, first, the user logs in, and the server-side makes a judgment based on the username and password entered by the user, whether the login is successful or not. After successful login, users can select images for query and retrieval: images can come from local albums, or users can take pictures of images they are interested in through cell phone cameras, and after selecting the image to be retrieved, send the image to the server, and check the results returned from the server check.

3.2. Experimental Design for Large-Scale Tourist Attraction Image Retrieval. In the process of guiding, the user is the main one, and the design of the tourist guide system must meet the needs of the user and provide maximum satisfaction to the user. The designed tourist guide system should meet the following design principles. The user is mobile, and he can stay at any location in any attraction. When using a cell phone, wireless communication is needed between the user side and the central processing site, which allows the user to better experience the convenience [19]. The aggregation process of calculating the feature average or maximum value of a certain local area to characterize

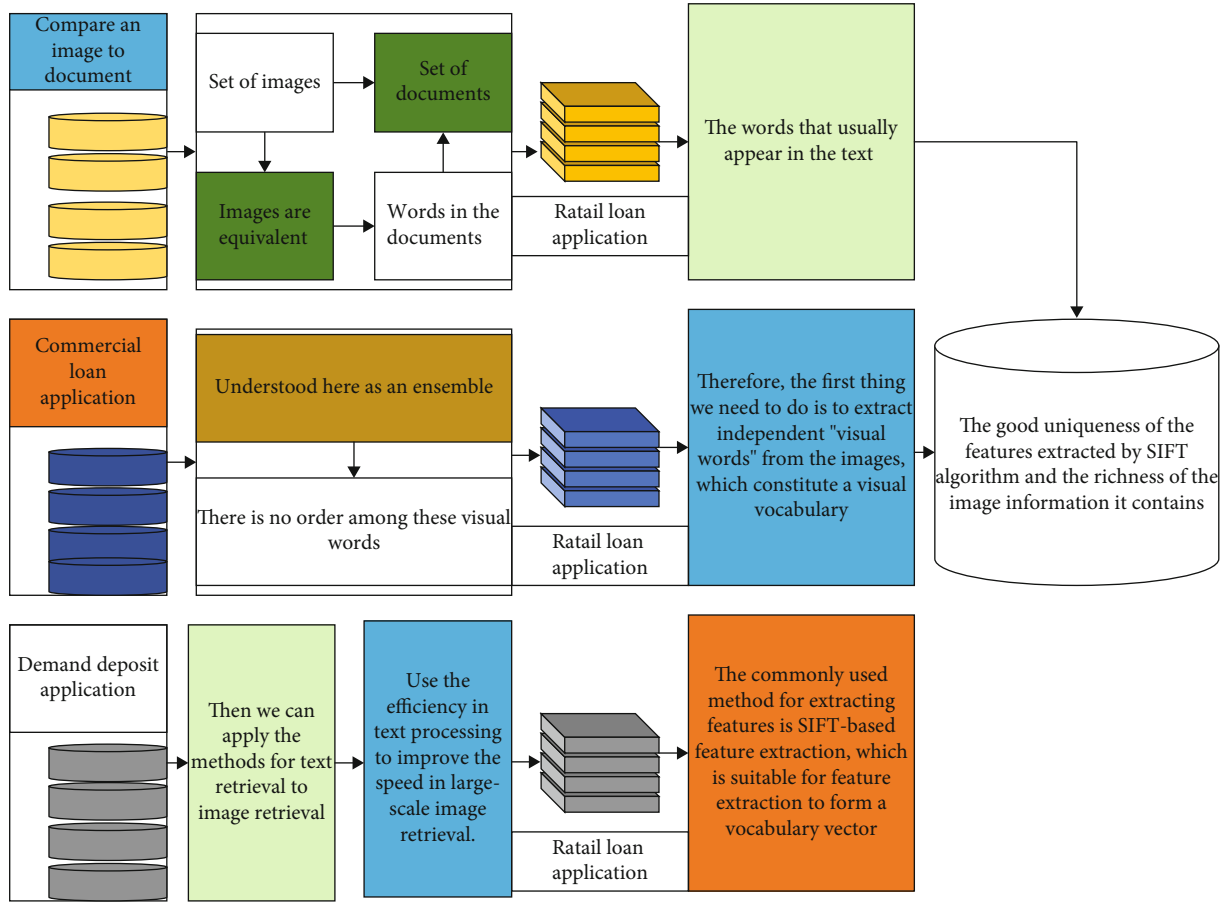


FIGURE 2: Schematic diagram of local connection.

the specialization of the entire area is called pooling or downsampling. Real time is reflected in two aspects. One is the rapid identification of the user's location because the user is mobile; if the user's location is not quickly identified, it may appear that the user has moved to the next attraction; the system only identified the last attraction, which is not feasible. Another is to respond promptly to changes in attraction infrastructure maintenance, environmental changes, and changes in surrounding stores. The user community includes people of all ages and knowledge levels, which means that the user-side interface should be user-friendly and able to be operated effectively by the elderly and even older children.

Once the tour guide system based on mobile visual search technology is put into operation, the number of users will proliferate, and many unavoidable MIS operations will occur during the use process according to the different cognitive levels of different users. A reliable and stable tour guide system based on mobile visual search technology must solve the MIS operations of users in time to ensure the stable and normal use of the system. It is worth noting that real time and accuracy are trade-offs. This is because to improve accuracy, a large amount of computational processing is required, which increases processing time and decreases real-time performance. The reverse is also true; if real time is to be improved, fast recognition is required, which will reduce accuracy, as shown in Figure 3.

As an innovative concept, smart tourism has been accepted by increased tourism enterprises. In addition to the ticketing system, the scenic guide system is also a breakthrough in smart tourism, which gives tourists a better touring experience while also realizing the transparent display of scenic information. The scenic guide system needs to combine artificial intelligence, mobile Internet, Internet of Things, and other technologies so that tourists can get one-to-one in-depth guide service through cell phones, to meet the demand of tourists to find information about scenic spots; to help scenic spots to achieve panoramic display, voice explanation, route planning, information transfer, and other integrated guide services; and to improve the quality of service in scenic spots and improve the tour experience of tourists [20]. Scenic areas cannot implement accurate marketing, high conversion costs, and low tourist independent experience. Scenic areas have a high cost of the manual guide, low efficiency, and no intelligent guide service for foreign tourists. Using audio, video, pictures, text, etc., as the main presentation methods, the scenic spot information is displayed to tourists, and the problems of passenger flow guidance, information lag, and play guides are solved. The intelligent guide system helps scenic spots to provide intelligent self-service for tourists. Through the network control system formed by the electronic guide hardware equipment and the central database in the background, the information

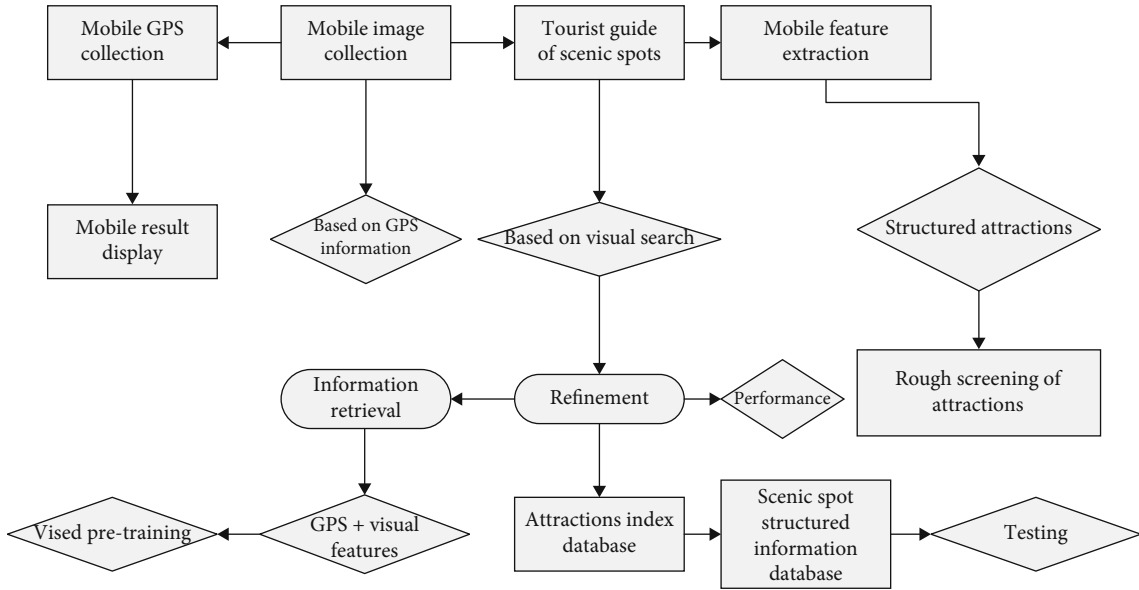


FIGURE 3: Overall framework of the tour guide system.

of scenic spots is displayed to tourists with audio, video, pictures, and text as the main presentation methods, solving the problems of passenger flow guidance, information lag, and tour guides. From the whole situation, it helps the scenic spot to guide the flow intelligently; prevent the tourist attractions, line congestion, and other problems; and improve the tourists' touring experience.

The information processing on the server-side consists of two main steps: the first one is to retrieve attractions, and the second one is to add attraction-related information. Based on these two steps, the database is divided into an attraction index database and an attraction structured information database. Using the attraction index database, the attraction retrieval step is performed by the GPS information-based attraction cluster filtering module and the visual search-based refinement module. Based on the retrieved attraction information and the attraction structured information database, the attraction structured information retrieval module adds attraction-related information and downlinks this information to the mobile terminal. The information display on the mobile terminal is done by the mobile terminal result display module and the attraction tour guide module. The mobile terminal result display module gives basic information about the attraction, and the attraction tour guide module introduces further information about the attraction, such as the humanities and the surrounding business information related to the attraction. In this paper, a grayscale co-occurrence matrix is selected and applied to the system. This method counts the relationship between the grayscale in each direction of the image, and between different pixels, and reflects the characteristics of the spatial distribution of the image color and light intensity, as shown in Figure 4.

In this paper, a rapid and accurate method is designed using the acceleration based on a hash algorithm for various descriptors, different types of classifiers based on a human visual perspective based on the characteristics of different

image features, and the calculation of image similarity using scientific methods, which improves the accuracy of image retrieval and makes this system more application-worthy. Image retrieval is performed using annotated images downloaded on Flickr, and the data given in the experiments are tested on 2000 randomly selected images [21]. All the images are divided into 16 image collections according to their features to speed up their retrieval by classifying the images. Based on human vision, this paper classifies the images using the grayscale correlation of the images and the entropy value of the images, i.e., the first classification of the images using the texture distribution of the images and the complexity of the images.

4. Analysis of Results

4.1. Analysis of the Multiple Linear Regression Algorithm. Based on human vision, this paper uses the grayscale correlation of images and the entropy value of images to classify images, i.e., the first classification of images using the texture distribution of images and the complexity of images. When using correlation for classification, this paper divides the original dataset into 4 subsets, and after many data tests, the boundary values that can divide the subsets are obtained. After evaluating all the image correlation values accordingly, a range of approximate values is obtained, and 3 values should be finally determined to classify the subsets. To facilitate the statistics of the data, a value greater than a certain value is used for comparison, and the number of images such as 500, 1000, and 1500 is obtained, and the equal cases are divided into the next subset for statistics. Based on this, a second division is performed, the same as the previous method, and the final range is determined by continuous experiments with different values. The results of the test on the image set were also extremely accurate, and according to the correlation shown in Figure 5, 1001 images met the

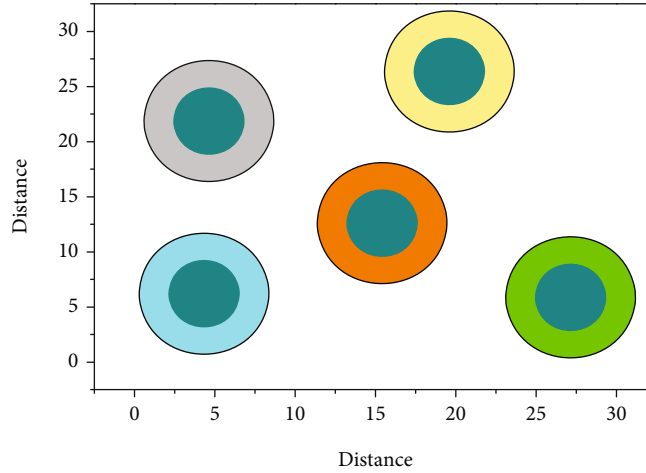


FIGURE 4: Feature map.

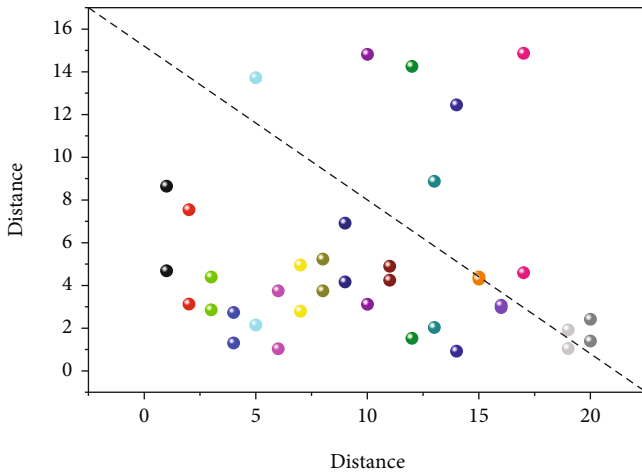


FIGURE 5: Multiple linear regression performance.

correlation with a value greater than 0.156 so that the division of the first two groups was completed.

For values greater than 0.076, 1501 images are obtained. Finally, four subsets containing 500 images each can be obtained, and experiments on larger datasets can also yield better results. Based on the above data, the values determined in this paper can satisfy the average distribution of the image sets. In the next step, the entropy value is used for further classification, and the subset obtained in the previous step is again divided into four subsets. Both the average precision rate and the average recall rate are decreasing, mainly since the amount of data in the image database of different scenic spots is different and the number of similar pictures in the image database is different from that of the query image. The accuracy rate is slightly reduced. The entropy value can effectively describe the complexity of the images, so it can play a good role in classifying the natural environment and urban buildings, people, animals, etc. However, if a lot of human factors are added when the images are taken, it will affect the classification effect of

this layer, and this effect will be reduced in the case of very large image sets. Based on the normalization of the color features, the similarity calculation is performed simultaneously in groups of 8 numbers. In the process of calculation, if the specified degree of similarity is not satisfied, the image calculated now is discarded and the color information of the next image is calculated so that the efficiency and quality of the returned information can be guaranteed at the same time. A similarity level of 45 does result in fewer results, but more than half of the images are not like the original. For example, if the main scene depicted in the original image is a mountain peak, then more than half of the images have a mountain peak. In this step of the retrieval process, the similarity is also used to determine the number of images, which is passed into the third feature of the retrieval.

In large-scale tourist attraction image retrieval, we need to query the accuracy and completeness of different attractions to calculate the average accuracy and average completeness of multiple attractions, which is used to characterize the overall performance of query results. For a total of 400,000 images of 1740 attractions online, we return 30 similar images for each attraction in order of similarity; query 100, 200, 500, 700, and 1000 attractions, respectively; and calculate the average accuracy and average completeness of the returned results belonging to the queried attractions, as shown in Figure 6.

From Figure 6, we can find that as the number of query sites increases, the average search accuracy and average search completeness decrease, mainly due to the different data volume of different sites in the image database and the different number of images like the query images in the image database, which leads to a slight decrease in accuracy. However, we can also find that the overall decrease of the average accuracy and average completeness is not significant, so we can consider that the method used in this paper is stable for large-scale tourist attraction image retrieval.

4.2. Graphical Retrieval Results of Large-Scale Tourist Attractions. In Figure 7, the recall rate of the improved

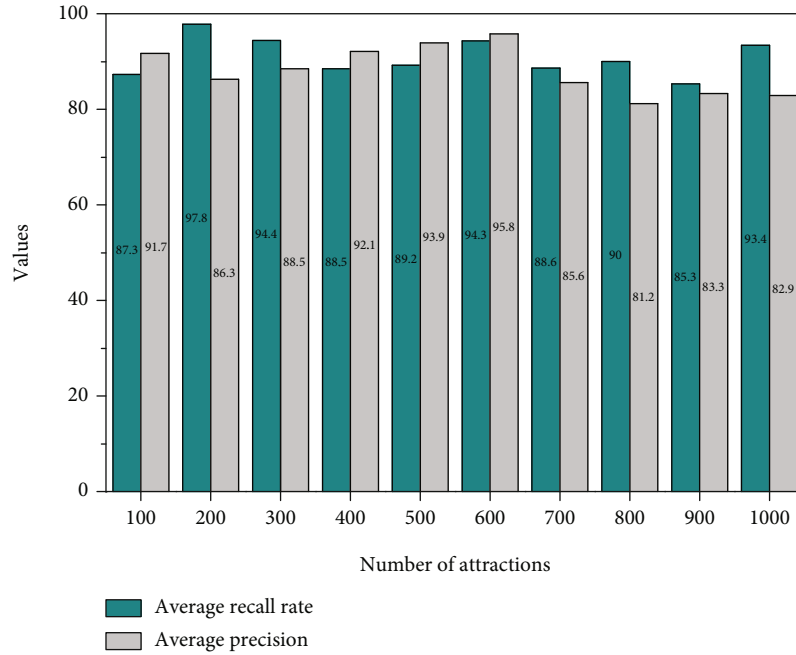


FIGURE 6: Average accuracy rate and average completeness rate.

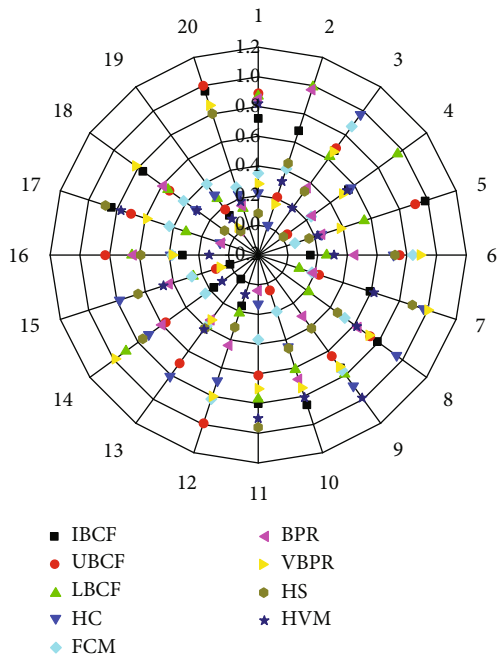


FIGURE 7: Comparison of recall values for recommended systems.

VBPR model is significantly better than the comparison baseline for the same reason as above. As the number of recommended attractions increases, its recall rate steadily increases and is less volatile compared to NMF and KNN. This indicates that the improved VBPR model is better at recommending multiple attractions. The hybrid recommendation (HVM) makes full use of the user travel preference information and the recommendation results of the improved VBPR model, and its recommendation perfor-

mance is optimal; i.e., the user travel preferences better smoothen the final recommendation results. The VBPR model plays the main role, while the user travel preference information based on the stratified sampling statistical model plays a secondary role.

The recommendation system of tourist attractions based on stratified sampling statistics and an improved VBPR model can effectively improve the recommendation performance and meet the user requirements to a greater extent, thus alleviating the data sparsity problem. And the stability of recommendations is better. Although the improved VBPR model can complete the recommendation well, the image features used in this chapter are all independent features, and the multimodal semantic correlation between different image features is not fully explored, so the recommendation performance still needs to be improved.

When the feature extraction is finished, APP will generate a string with the extracted features and send it to the server. Since the retrieval result returned by the server is the URL of the image, when the APP receives the returned result, it sends a request to the server again to get the image and then displays it. Each network request here requires a multithreaded operation to ensure operational efficiency and user experience at the same time. Multithreading plays a very important role in this part. Using multithreaded programming allows the program to perform multiple tasks at the same time, converting the interface when extracting features and displaying the results of extracted features when sending a request at the same time. Compared with other methods, our method improves the efficiency and accuracy by about 8%. This multistep simultaneous approach ensures that the overall system runs more efficiently and that the program does not get stuck due to certain operations that take a long time to compute.

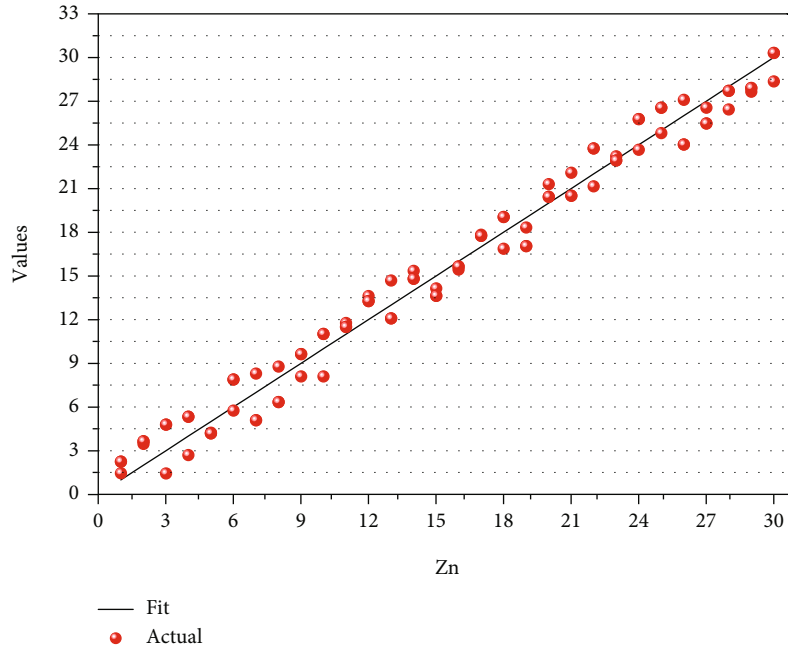


FIGURE 8: Search accuracy results.

This multistep simultaneous method can ensure the improvement of the overall operating efficiency of the system and objectively describe the rich semantic information contained in the scenic spot images. Therefore, it is necessary to deeply explore the multimodal correlation between different image features to better portray the visual content of attraction images from the perspective of multimodal feature fusion. The DCA-VBPR model plays this role well, as it utilizes label information and suppresses correlations between classes in the feature mapping process, so different image features are fully fused, feature discriminability is continuously improved, content analysis of objects is very successful, and recommendation performance is improved. The DCA model can generate high-quality recommendation results compared with scheme 1 and scheme 2. In the image upload interface, we select an image to upload for retrieval; at this time, we select the familiar Badaling Great Wall attraction image; we click the upload button; the image is uploaded to the server; the server will, according to the received image, carry out the corresponding feature extraction and use the extracted features to carry out the corresponding matching; and the initial matching result will be filtered to get the retrieval result, and the detection result will be returned to the cell phone. The detection results are viewed in the View Results option, and we click the View Results button, and there will be the results of the images we have queried, as shown in Figure 8.

Visitors take pictures of the attractions by using the mobile client, use CDVS descriptors to extract the corresponding texture features from the pictures, and upload the feature information and GPS information to the server. Based on the received information, the server performs image matching using CDVS feature-based visual retrieval and rearrangement techniques, performs image retrieval

based on the matched results, filters the matched results, and returns them to the mobile client. A front-to-back separation method is used in the server module, which improves retrieval efficiency. This paper's travel guide system, based on mobile visual search technology, facilitates the travel needs of tourists, eliminates unnecessary expenses, makes travel easier and faster, and improves the quality of tourism in general.

5. Conclusion

This paper applies the retrieval of large-scale tourist attraction images. The WAN method is a more classical algorithm; of course, these two algorithms have certain shortcomings; for example, the fully connected layer in the convolutional network structure, due to too many parameters in the fully connected layer, often leads to overfitting; in the locality-sensitive hashing, we hope that the adjacent data can be integrated into the same bucket after hashing, while the nonadjacent data fall into different buckets. However, in practice, the adjacent data will fall into different buckets and the nonadjacent data will fall into adjacent buckets, which we do not want to see. Therefore, we can improve the accuracy of retrieval and reduce the error rate by improving the structure of optimized convolutional networks (e.g., using more network layers and reducing the number of fully connected layers) and the structure of locality-sensitive hashes (e.g., using more hash pylons within a hash table or building more hash tables). In summary, on the one hand, the accuracy of image retrieval can be improved by building a larger and better image database of tourist attractions, and on the other hand, the error rate of image retrieval can be reduced by continuously optimizing

the feature extraction algorithm and index structure, to optimize the image retrieval of large-scale tourist attractions.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] M. Ren, H. Q. Vu, G. Li, and R. Law, "Large-scale comparative analyses of hotel photo content posted by managers and customers to review platforms based on deep learning: implications for hospitality marketers," *Journal of Hospitality Marketing & Management*, vol. 30, no. 1, pp. 96–119, 2021.
- [2] D. Li, L. Deng, and Z. Cai, "Statistical analysis of tourist flow in tourist spots based on big data platform and DA-HKRVM algorithms," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 87–101, 2020.
- [3] Y. Liang, S. Gao, Y. Cai, N. Z. Foutz, and L. Wu, "Calibrating the dynamic Huff model for business analysis using location big data," *Transactions in GIS*, vol. 24, no. 3, pp. 681–703, 2020.
- [4] C. Tosun, B. B. Dedeoğlu, C. Çalışkan, and Y. Karakuş, "Role of place image in support for tourism development: the mediating role of multi-dimensional impacts," *International Journal of Tourism Research*, vol. 23, no. 3, pp. 268–286, 2021.
- [5] S. du, S. du, B. Liu, X. Zhang, and Z. Zheng, "Large-scale urban functional zone mapping by integrating remote sensing images and open social data," *GIScience & Remote Sensing*, vol. 57, no. 3, pp. 411–430, 2020.
- [6] Q. Wang, J. Gao, W. Lin, and X. Li, "NWPU-Crowd: a large-scale benchmark for crowd counting and localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2141–2149, 2021.
- [7] S. Giglio, F. Bertacchini, E. Bilotta, and P. Pantano, "Machine learning and points of interest: typical tourist Italian cities," *Current Issues in Tourism*, vol. 23, no. 13, pp. 1646–1658, 2020.
- [8] J. Kang, R. Fernandez-Beltran, Z. Ye, X. Tong, P. Ghamisi, and A. Plaza, "Deep metric learning based on scalable neighborhood components for remote sensing scene characterization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8905–8918, 2020.
- [9] K. L. van Leeuwen, R. A. Hill, and A. H. Korstjens, "Classifying chimpanzee (*Pan troglodytes*) landscapes across large-scale environmental gradients in Africa," *International Journal of Primatology*, vol. 41, no. 6, pp. 800–821, 2020.
- [10] N. Tadi Bani and S. Fekri-Ershad, "Content-based image retrieval based on combination of texture and colour information extracted in spatial and frequency domains," *The Electronic Library*, vol. 37, no. 4, pp. 650–666, 2019.
- [11] X. Li, "Space-time distribution model of visitor flow in tourism culture construction via back propagation neural network model," *Personal and Ubiquitous Computing*, vol. 24, no. 2, pp. 223–235, 2020.
- [12] H. Fujita, "AI-based computer-aided diagnosis (AI-CAD): the latest review to read first," *Radiological Physics and Technology*, vol. 13, no. 1, pp. 6–19, 2020.
- [13] A. Keane, J. F. Lund, J. Bluwstein, N. D. Burgess, M. R. Nielsen, and K. Homewood, "Impact of Tanzania's Wildlife Management Areas on household wealth," *Nature Sustainability*, vol. 3, no. 3, pp. 226–233, 2020.
- [14] F. Z. Benkhattab, M. Hakkou, I. Bagdanavičiūtė et al., "Spatial-temporal analysis of the shoreline change rate using automatic computation and geospatial tools along the Tetouan coast in Morocco," *Natural Hazards*, vol. 104, no. 1, pp. 519–536, 2020.
- [15] M. Ligerio, O. Jordi-Ollero, K. Bernatowicz et al., "Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis," *European Radiology*, vol. 31, no. 3, pp. 1460–1470, 2021.
- [16] S. E. Rigby, T. J. Lodge, S. Alotaibi et al., "Preliminary yield estimation of the 2020 Beirut explosion using video footage from social media," *Shock Waves*, vol. 30, no. 6, pp. 671–675, 2020.
- [17] N. Hor and S. Fekri-Ershad, "Image retrieval approach based on local texture information derived from predefined patterns and spatial domain information.," vol. 8, no. 6, pp. 246–254, 2019, <https://arxiv.org/abs/1912.12978>.
- [18] K. A. Shinde, "Religious theme parks as tourist attraction systems," *Journal of Heritage Tourism*, vol. 16, no. 3, pp. 281–299, 2021.
- [19] S. O. Petrovan, C. G. Vale, and N. Sillero, "Using citizen science in road surveys for large-scale amphibian monitoring: are biased data representative for species distribution?," *Biodiversity and Conservation*, vol. 29, no. 6, pp. 1767–1781, 2020.
- [20] M. T. Falk and E. Hagsten, "Visitor flows to World Heritage Sites in the era of Instagram," *Journal of Sustainable Tourism*, vol. 29, no. 10, pp. 1547–1564, 2021.
- [21] K. E. Callahan, C. J. Clark, A. F. Edwards et al., "Automated frailty screening at-scale for pre-operative risk stratification using the electronic frailty index," *Journal of the American Geriatrics Society*, vol. 69, no. 5, pp. 1357–1362, 2021.