Research Article

Image Retrieval of Tourism Landscape in Rural Revitalization Based on Wireless Communication Network

Ou Xie

Academy of Fine Arts, Hubei Normal University, Huangshi 435002, Hubei, China

Correspondence should be addressed to Ou Xie; xieou0561@hbnu.edu.cn

Received 5 December 2021; Revised 5 January 2022; Accepted 17 January 2022; Published 11 February 2022

Academic Editor: Haibin Lv

Copyright © 2022 Ou Xie. 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.

With the rapid development of information technology, people's acquisition of tourism and other information is increasingly dependent on images and other information. Aiming at the low efficiency of traditional image retrieval methods in processing massive image data, an image retrieval method based on wireless communication network is proposed. Based on the salient area and wireless communication network combined with the hash method to extract local CNN features, simulate INRIA Holidays Data set and Oxford Buildings Data set, and calculate the accuracy and recall rate of search results based on the returned results of tourist attractions pictures. This article designs an experiment to verify the accuracy and recall rate of search results. By comparing the feature hash function to generate the hash code and the Hamming distance between each hash code in the image library, the image is queried, and the final search result is obtained: The more searches, the lower the accuracy and recall rate. This also proves to a certain extent that the CNN feature extraction technology can be used for travel image search, improving the search accuracy by 20%. The wireless communication network is still of great significance to the future social development. It is necessary to conduct in-depth research, not only the image retrieval of tourist attractions proposed in this article but also the potential value of wireless communication networks from multiple angles and more comprehensively.

1. Introduction

Since the beginning of the 21st century, electronic information technology has developed rapidly, and almost every household has more or less intelligent mobile terminals such as computers and mobile phones. The capacity of storage devices has also been refreshed again and again, and these have gradually changed the way people live. The content on the Internet is varied and brings a lot of bad information while bringing convenience to people. These include a large number of advertisements, unsightly pornography, and information that contradicts social values. The amount of image content is also growing at an explosive rate, which also poses a problem for us to quickly locate the desired image [1]. At the same time, thanks to the rapid development of hardware devices in recent years, wireless communication network technology, especially in the field of image processing, has received increasing attention from researchers. Especially as tourism has become an increasingly popular trend to meet people's spiritual entertainment needs, more

and more people are willing to travel. However, finding suitable tourist locations has become a major obstacle to the development of tourism. With massive amounts of information, how to accurately retrieve images has become a major research hotspot. Therefore, the research on image retrieval system based on wireless communication network still has important theoretical and practical significance [2].

In recent years, domestic scholars and experts have also conducted useful research on content-based image retrieval technology. Compared with foreign research results, its theoretical research and application lags behind, but with the emphasis on Chinese research, the gap is getting smaller and smaller. As far as the current application status is concerned, domestic content-based image retrieval technology is mainly used to retrieve images from professional image libraries, and is being applied in the fields of industry, agriculture, medicine, national defense and entertainment. In general, with CBIR as an example, CBIR technology is far from intelligent and fully automated, and the technology is still immature. There are still many key technologies that need

improvement. For example, feature extraction, semantic gap description, and performance optimization still have many issues that require more in-depth research [3].

An image retrieval system usually includes an image retrieval part, an image retrieval library creation part, a model creation part and a system maintenance part. Among them, image retrieval is the core of the function, the image retrieval database and retrieval model are the foundation of the entire system, and system maintenance is the guarantee [4]. Image retrieval means that the computer automatically retrieves the image most similar to the image to be retrieved in the image library. Therefore, we need to first create an image library that we want to retrieve, extract all the images in the library through the neural network model we build, and then save. Then input the image to be retrieved, extract high-dimensional features through the neural network model, and then perform feature matching with the image features in the library, and output similar images from high to low, thereby completing image-based retrieval on the wireless communication network. In order to achieve image retrieval of tourist attractions based on wireless communication networks [5], this paper proposes a local CNN feature extraction algorithm based on salient regions from the perspective of image understanding.

This article is based on the wireless communication network image retrieval system research, the advantages are: 1. Construct a convolutional neural network model to extract image features. 2. Use the hash function to generate a hash code with the help of image features, and compare the Hamming distance between the experimental image and the standard image hash. 3. Calculate the accuracy and recall rate of the image, and measure the image search results.

2. Method

2.1. Wireless Communication Network. In the field of information communication in recent years, wireless communication technology has been the fastest growing and most widely used [6]. Common wireless communication application equipment is shown in Figure 1.

The wireless communication realized on the move is also commonly referred to as mobile communication [7, 8]. People call the two together as wireless mobile communication. The application field of wireless communication technology is shown in Figure 2.

2.1.1. Cooperative Communication Technology. Cooperative communication technology originated from the research work on the information theory characteristics of relay channels in the 1970s [9]. The relay channel model they analyzed is shown in Figure 3, which includes three nodes: source node S , relay node R and destination node D , and pointed out that this model can be decomposed into broadcast channel and multiple access channel [10, 11].

2.1.2. Cognitive Wireless Network. Cognitive wireless network is the networking of CR [12]. The overall study of cognitive characteristics and wireless communication

network is carried out. The working mechanism is shown in Figure 4. Cognitive wireless network has cognitive functions [13], can distinguish the current network status, and then adaptively learn, make decisions, and respond according to these network statuses. Its ultimate goal is to achieve end-to-end performance.

Dynamic spectrum management in cognitive wireless networks, also known as dynamic spectrum allocation, mainly uses information obtained from spectrum detection to perform spectrum analysis and spectrum decision-making. Its goal is to combine power control to provide an effective and adaptive use of wireless spectrum Resource methods and strategies [14, 15]. The wireless cognitive network is composed of a group of primary users (PU) and a group of secondary users (SU). PU is an authorized user in the network and has a higher priority than SU, and SU cannot interfere with PU's communication. At present, dynamic spectrum management research mainly involves unlicensed models, secondary utilization models of licensed spectrum, and market models, as shown in Figure 5. The unlicensed model requires SU to use idle frequency bands in accordance with a certain spectrum etiquette; the secondary utilization model of licensed spectrum allows SU and PU to share the licensed spectrum when the SU does not cause harmful interference to the PU [16]; the market model allows the trading of licensed spectrum, The right to part of the authorized spectrum can be transferred according to market rules formulated in advance.

QoS (quality of service) refers to the ability of a network to use various basic technologies to provide better service capabilities for specified network communications. It is a security mechanism for the network and a technology used to solve problems such as network delay and congestion. In the process of spectrum decision-making, first extract the characteristics of each frequency band according to the statistical information of the authorized network and the local observations of the SU [17], and then select the best frequency band according to the user's QoS [18] requirements and frequency band characteristics. Figure 6 shows the rule-based dynamic spectrum decision-making.

The cooperative relationship takes place between SU and PU, as shown in Figure 7. As a cooperative relay of PU, SU actively participates in the communication process of PU. While ensuring the communication quality of PU, it may obtain the opportunity to transmit its own information [19].

Cooperative diversity technology is applied in cognitive wireless network [20], so that two or more SUs cooperate with each other, as shown in Figure 8. In the Overlay-type [21] spectrum sharing system, SU can only opportunistically access idle frequency bands.

2.1.3. Wireless Communication Technology. Bluetooth [22, 23], ZigBee, and Wi-Fi are three popular short-range wireless communication protocol standards. The network standards adopted by the three wireless communication technologies of Bluetooth, ZigBee, and Wi-Fi are IEEE802.15.1, IEEE802.15.4, and IEEE802, respectively, and their main features are shown in Table 1.

FIGURE 1: Common wireless communication equipment.

FIGURE 2: Application areas of wireless communication technology.

FIGURE 3: Three-node relay channel model.

FIGURE 4: The working mechanism of cognitive wireless network.

FIGURE 5: Classification of spectrum management.

FIGURE 6: Rule-based dynamic spectrum management.

FIGURE 7: Schematic diagram of the cooperation between SU and PU.

FIGURE 8: Schematic diagram of writing between SU.

TABLE 1: The main characteristics of the three wireless communication technologies.

Communication technology	Bluetooth	ZigBee	Wi-Fi
Frequency band	2.4 GHz	2.4 GHz	2.4 GHz
Transfer speed	24 Mb/s	250 kb/s	54 Mb/s
Transmission distance	10 m	10–100 m	100 m
Maximum transmission rate	0.72	0.25	54
Bit propagation time	1.39 μ s	4 μ s	0.0185 μ s
Coding efficiency	94.41	76.52	97.18

As a layered reference model, the TCP/IP protocol is different from OSI's layering. Figure 9 shows the comparison between the OSI reference model and the TCP/IP protocol model.

2.2. Wireless Communication Network Convolutional Neural Network AlexNet Model. Wireless communication network mainly includes convolutional neural network (CNN), recurrent neural network (RNN) [24, 25] sparse self-coding network, deep belief network, and other models. Convolutional neural network (CNN) is a type of feedforward neural network that includes convolution calculations and has a deep structure. Recurrent neural network (RNN) is a type of recurrent neural network that takes sequence data as input, recursively in the evolution direction of the sequence, and all nodes are connected in a chain. It has certain advantages when learning the non-linear characteristics of the sequence. The sparse autoencoder is an unsupervised machine learning algorithm that continuously adjusts the parameters of the autoencoder by calculating the error between the output of the autoencoder and the original input and finally trains the model. Deep belief network is a network of graphical knowledge representation and reasoning model based on probability composed of multiple layers of neurons. Among the different types of wireless communication network models, the convolutional neural network (CNN) is the most in-depth research and application. A convolutional network refers to a neural network that uses convolution operations instead of general matrix multiplication operations in the network structure [26, 27].

Convolutional neural networks make full use of local features such as the data itself by combining local perception areas, shared weights, and spatial or temporal pool down-sampling. Sharing weight means that for an input picture, use a convolution kernel to scan the picture. The number in the convolution kernel is called the weight. Each position in this picture is swept by the same convolution kernel, so the weight is the same, that is, sharing. Sampling is divided into downsampling and upsampling. Downsampling is to reduce the image, it can make the image fit the size of the display area and generate a thumbnail of the corresponding image. Upsampling is to enlarge the image, which enlarges the original image so that it can be displayed on a higher-resolution display device. Due to this structural feature, it is particularly suitable for machine learning of large image data, which can reduce the number of image recognition problems.

The main structure of a convolution neural network includes a convolution part and a fully connected part. The convolution part includes a Convolutional Neural Layer, an

FIGURE 9: Comparison of OSI reference model and TCP/IP protocol model.

activation layer and a down-sampling layer. Features are extracted through the superposition of the convolution part; the fully connected part connects the feature extraction and output calculation loss to complete the recognition and classification functions.

Convolutional neural network is the first robust wireless communication network method that successfully adopts multilayer hierarchical network. Convolutional neural networks are highly adaptable and good at mining local features of data, making convolutional neural networks one of the research hotspots in many scientific fields.

This paper uses the 8-layer convolutional neural network AlexNet model of Alex et al. The AlexNet model is a model designed by Hinton, the winner of the 2012 ImageNet competition, and his student Alex Krizhevsky. Its innovation lies in the successful application of the ReLU activation function, the Dropout mechanism, GPU accelerated training and data enhancement strategies, and the proposed LRN (local response normalization). It also uses overlapping maximum pooling instead of average pooling, successfully avoiding the blurring effect for image retrieval. AlexNet network structure parameters are shown in Table 2. Assuming that the weight and offset are represented by w and b , respectively, the linear prediction of the i th category can be expressed as

TABLE 2: AlexNet model structure parameters.

Layer name	Input data size (Pixel)	Convolution template size (Pixel)	Feature size (Pixel)
Convolutional neural layer 1	227*227	11	27*27*96
Convolutional neural layer 2	27*27	5	13*13*256
Convolutional neural layer 3	13*13	3	13*13*384
Convolutional neural layer 4	13*13	3	13*13*256*384
Convolutional neural layer 5	13*13	3	6*6*256
Fully connected layer 6	6*6*256	—	4096
Fully connected layer 7	4096	—	4096
Fully connected layer 8	4096	—	1000

$$\begin{aligned}
z_i &= w_i^T + b, \\
\sigma(z) &= e(\sigma_1(z), \sigma_2(z), \dots, \sigma_m(z)), \\
\sigma_i(z) &= \frac{\exp(z_i)}{\sum_{j=1}^m \exp(z_j)}, \quad j = 1, 2, \dots, m.
\end{aligned} \tag{1}$$

2.3. Local CNN Feature Extraction Based on Saliency Regions. The sift feature is a very common local descriptor [28]. It condenses image information into 128-dimensional feature vectors by constructing scale space, extracting key points, and generating descriptors. Using image understanding theories and models related to wireless communication networks, the salient areas of the image are extracted and characterized, and local features similar to sift are generated. That is, local CNN features extraction based on salient regions (LCF-SR) [29]. The sift feature extraction operation can be divided into five steps: one is the generation of the scale space; the second is the detection of extreme points in the scale space; the third is to accurately locate the extreme points; the fourth is to specify the direction parameters for each key point, and the fifth is the key point description. The generation of the child.

2.3.1. Image-Based Understanding Training CNN + RPN + LSTM Model. This model is used to extract areas of interest in the image. The network structure consists of CNN network, RPN positioning layer, simple identification network, and LSTM language model.

(1) *CNN Network.* CNN uses the VGG-16 structure, which consists of 13 layers of 3×3 convolutions and 5 layers of 2×2 maximum pooling [30]. In this paper, the final fully connected and classified layers are all deleted. Thus, an $3 \times W \times H$ input image becomes a collection of $C \times W' \times H'$ series of feature maps after passing through the CNN network, among them:

$$\begin{aligned}
C &= 512, \\
W' &= \left\lfloor \frac{W}{16} \right\rfloor, \\
H' &= \left\lfloor \frac{H}{16} \right\rfloor.
\end{aligned} \tag{2}$$

(2) *Positioning Layer.* After receiving the input feature map, the positioning layer locates the region of interest according to the feature map and extracts a representation of the appropriate length from each region. The input of the positioning layer is $C \times W' \times H'$ generated by the CNN network, and then outputs three kinds of information: (1) A $B \times 4$ matrix contains candidate regions of the region boundary information; (2) region score, the higher the score, the more likely it is selected as the saliency region and; (3) the region feature: the output of this layer is $B \times C \times X \times Y$ characteristic stream, each region is represented as a feature of $C \times X \times Y$.

(3) *Identifying the Network.* The identification network is actually a fully connected neural network that is used to process regional features from the localization layer. Each $C \times X \times Y$ region feature is first expanded into a vector and then passed through two layers of fully connected layers. The fully connected layer contains ReLU (rectified linear units) activation functions and dropout regularization [31]. ReLU, also known as modified linear unit, is an activation function commonly used in artificial neural networks and usually refers to a non-linear function represented by a ramp function and its variants. Dropout reduces the size of the neural network by randomly dropping some neurons during the training process to prevent overfitting. When dropout propagates forward, the activation value of a certain neuron will stop working with a certain probability p , which can make the model more generalized because it will not rely too much on some local features. Finally, each region feature is encoded into a feature vector of $D = 4096$ dimensions. The feature vectors of the selected salient regions form a matrix of $B \times D$ dimensions. This feature matrix is an input to the LSTM language model and is a feature matrix for encoding in image retrieval tasks.

(4) *The LSTM Language Model.* Training models and the selection of salient regions are based on image understanding [32]. The label of an image is a lot of sentences described in natural language, so end-to-end training based on image understanding can be achieved. The LSTM language model follows the recognition network and enters the feature matrix of the $B \times D$ dimension. Specifically, for the marked training sequence $s_1, s_2, \dots, s_T$, we use $T+2$ word vectors $x_{-1}, x_0, x_1, \dots, x_T$ as input to the LSTM. Among them, $x_{-1} = \text{CNN}(I)$, the area coding of the linear layer is connected to the non-linear RELU layer, x_0 is a special start mark, and x_t encodes each training feature s_t , $t = 1, \dots, T$. The LSTM calculates the hidden state sequence h_t and the

output vector $y_t, h_t, y_t = f(h_{t-1}, x_t)$, by a recursive formula. The length of y_t is $|V| + 1$, V is the dimension of the marker sequence, and the extra length 1 is the end symbol.

2.3.2. Significant Region Selection and Coding

(1) *Bounding Box Regression*. Each point in the dimension $W' \times H'$ convolution feature map is mapped to the original image of dimension $W \times H$. Assuming that there are k target regions in the original image, k frames of different aspect ratios are selected around each projection point of $W' \times H'$. The positioning layer gives a confidence score and 4-dimensional position information to the k regions, center point coordinates (x, y) , width and height (w, h) . This produces an output of the $5k \times W' \times H'$ dimension, including confidence scores and position information for each region.

The general target window G 's ground truth is represented by a four-dimensional vector (x, y, w, h) that represents the coordinates of the center point and the width and height of the target window. Assuming that the original window obtained by the prediction is P , then we need a relationship during the training so that the original input window P is mapped to obtain a regression window G' which is closer to the real window G . That is, given (P_x, P_y, P_w, P_h) , a mapping f is sought such that

$$f(P_x, P_y, P_w, P_h) = (G'_x, G'_y, G'_w, G'_h). \quad (3)$$

and there is

$$(G'_x, G'_y, G'_w, G'_h) \approx (G_x, G_y, G_w, G_h). \quad (4)$$

The idea of getting the transformation is

(1) First translate the input window

$$\begin{aligned} &(\Delta x, \Delta y), \\ \Delta x &= P_w d_x(P), \\ \Delta y &= P_h d_y(P). \end{aligned} \quad (5)$$

to get:

$$\begin{aligned} G'_x &= P_w d_x(P) + P_x, \\ G'_y &= P_h d_y(P) + P_y. \end{aligned} \quad (6)$$

(2) Scale the input window

$$\begin{aligned} &(S_w, S_h), \\ S_w &= P_w d_w(P), \\ S_h &= P_h d_h(P). \end{aligned} \quad (7)$$

The corresponding transformation is

$$\begin{aligned} G'_x &= P_w \exp(d_w(P)), \\ G'_h &= P_h \exp(d_h(P)). \end{aligned} \quad (8)$$

The input in the border regression algorithm is the four-dimensional coordinate $P = (P_x, P_y, P_w, P_h)$ of the original window, the window coordinate of the Ground Truth, and

the prediction window G' is obtained by the four transformations of $d_x(P), d_y(P), d_w(P), d_h(P)$.

$$G = (G_x, G_y, G_w, G_h). \quad (9)$$

(2) *Screening Area*. Next, we need to select B significant areas from the k areas. Nonmaximum suppression, referred to as NMS algorithm for short. It is an effective method to obtain the local maximum. During training, if the same target area appears to be surrounded by multiple candidate frames, a nonmaximum suppression algorithm is needed to remove the lower score candidate frames to reduce the overlap box. The candidate area is divided into a positive area and a negative area. IoU (intersection over union) is defined as the overlap ratio of two candidate boxes, $\text{IoU} = (A \cap B) / (A \cup B)$. If an area and a target area have an $\text{IoU} > 0.7$, it is a positive area; if it and all target areas have IoU values less than 0.3, it is a negative area. Based on the credibility score for each region, a nonmaximum suppression algorithm is used to screen out B significant regions. Nonmaximum suppression is widely used in object detection. The main purpose is to eliminate redundant candidate frames with low confidence to ensure that the best detection position is found.

(3) *Bilinear Interpolation*. In training, gradients can propagate from output features to input features, but not in saliency regions. Therefore, bilinear interpolation is used instead of the RoI pooling layer. Bilinear means to perform linear interpolation in two directions (first linear interpolation is used in one direction, and then linear interpolation is used in the other direction to perform bilinear interpolation. Although each step is in sample value and position All are linear, but the interpolation is not linear in general, but quadratic at the sampling position. Feature representations of the same size are extracted for each significant region, which can be linked to subsequent recognition networks and LSTM language models.

Set

$$G_{i,j} = (x_{i,j}, y_{i,j}). \quad (10)$$

to convolve with core k :

$$V_{c,i,j} = \sum_{i'=1}^W \sum_{j'=1}^H U_{c,i',j'} k(i' - x_{i,j}) k(j' - y_{i,j}). \quad (11)$$

In bilinear interpolation, the kernel function k has the following representation:

$$k(d) = \max(0, 1 - |d|). \quad (12)$$

Thus, the sampling grid becomes a linear function of the saliency region, and the gradient can predict and train the candidate salient regions by backpropagation. Bilinear interpolation can extract a fixed size feature of $B \times C \times X \times Y$ for all candidate regions for input to the recognition network.

2.3.3. *Area Coding, Extracting Feature Vectors*. To get the features of the BCXY dimension, you need to encode it. The identification network is a fully connected network that can

FIGURE 10: Hash image retrieval process.

perform preliminary coding of features to obtain features of $B \times 4096$ dimensions. Encoding is performed using sum pooling.

The principle of the sampling algorithm is for the feature C of the $B \times 4096$ dimension, B represents the number of significant regions, and 4096 represents the feature dimension of each significant region. First, for each feature of the feature, calculate the sum of its eigenvalues over all significant regions:

$$F'_k = \sum_{i=1}^B C_i \sum. \quad (13)$$

The feature is then encoded as the ratio of the sum of each feature dimension, i.e.,

$$F_k = \frac{F'_k}{\sum_{k=1}^{4096} F'_k}. \quad (14)$$

The dimension of the eigenvector F is 4096, and then we reduce it by PCA and adjust it to get the final eigenvector. Finally, according to the above method, feature vectors are extracted from all sample images to construct a search library.

2.4. Signal Reconstruction Algorithm. The signal reconstruction of the compressed sensing theory refers to the reconstruction of the original signal a or its equivalent sparse representation d from the observation vector b according to the known observation matrix ϕ and the transformation basis matrix ψ , which is to find the satisfaction

$$b = \phi a = \phi \psi d. \quad (15)$$

The reconstruction algorithms that have appeared so far can be divided into three categories: algorithms based on convex optimization, greedy pursuit algorithms, and combined algorithms.

As long as the reconstruction algorithm based on convex optimization obtains the sparsest solution by adding constraints, the norm constraint is the most commonly used. First give the definition of p -norm:

The p -norm of the vector $a = \{a_1, a_2, \dots, a_k\}$ is

$$a_p = \left(\sum_{i=1}^K |a_i|^p \right)^{1/p}. \quad (16)$$

In the traditional processing method, the method of solving the minimum l_2 -norm is generally used to solve this recovery problem, namely:

$$\begin{aligned} \hat{d} &= \min d_{l_2}, \\ b &= \phi \psi d. \end{aligned} \quad (17)$$

This optimization can get a fixed solution

$$\hat{d} = \delta^T (\delta \delta^T)^{-1} b. \quad (18)$$

However, the solution that the minimum l_2 -norm optimization method cannot obtain does not have N -sparseness, but a vector containing many nonzero elements.

When $p=0$, the l_2 -norm is obtained, which means the number of nonzero elements is in a . On the premise that the signal a is sparse or compressible, the problem of finding the definite solution of the underdetermined equations $b = \delta d$ can be transformed into the problem of finding the minimum l_0 -norm:

$$\begin{aligned} \hat{d} &= \min d_{l_0}, \\ b &= \phi \psi d. \end{aligned} \quad (19)$$

However, the solution process needs to list all possible linear combinations of Z_K^N for all nonzero term positions in the signal a in order to obtain the optimal solution. Therefore, the numerical calculation of (20) and (21) is extremely unstable and it is NP difficult.

Solving a simpler l_1 optimization problem can produce equivalent solutions:

$$\hat{d} = \min d_{l_1}, \quad (20)$$

$$b = \phi \psi d. \quad (21)$$

2.4.1. Wireless Communication Network Combined with Hash Method. In order to solve the problem of high-dimensional retrieval in large image libraries, an image-based hashing method has emerged by measuring retrieval speed and storage space. The hashing technique can use a hash function to extract a sequence of binary bits representing image features, also called hash encoding, from the image data. Since the hash code is identical to the binary representation used by the computer, the retrieval efficiency and storage space are greatly improved. In the image processing process, various transformations are usually performed on the image, and the traditional hash function is very sensitive to this transformation, which is not conducive to image processing.

FIGURE 11: INRIA Holidays query map.

FIGURE 12: Examples of related photos.

FIGURE 13: INRIA Holidays partial query results.

In order to apply hash coding in the content-based image retrieval process, we require the hash function to satisfy as much as possible: (1) the hash coding of similar images is as similar as possible; (2) the hash coding of different content images is as different. In the retrieval process, the search results are obtained by calculating the Hamming distance between the retrieved image and the training image in the image library and sorting by size. It can be seen that the hashing technique has great convenience.

The structure of the image retrieval using the hash method is as shown in Figure 10. The image feature extraction technique extracts features from the image in the query image and the image library, generates hash codes by using these feature hash functions, and then compares the query images with The Hamming distance between each hash code of the image in the image library, and the final search result is obtained.

3. Data Sources

In order to evaluate the performance of the proposed method for the image of tourist attractions, we conducted experiments on two image data sets. The data sets are INRIA Holidays Dataset and Oxford Buildings Dataset. INRIA Holidays Dataset is a dataset used by Herve Jegou et al., which is a photograph taken by the Institute during regular holidays (landscape dominated). A total of 1491 images, 500 queries (one map and one group) and corresponding to 991 related images. Oxford Buildings Dataset is a VGG group in Oxford that collected images of 5062 building attractions from Flickr. These images can be applied to the research of image retrieval technology of tourist attractions based on wireless communication network.

4. Results and Discussion

4.1. Result 1: INRIA Holidays Data Set. Select a church main entrance tourist attraction image in the data set as the query image, as shown in Figure 11.

Among the pictures belonging to this attraction, the similar pictures have the following 4 pictures (excluding Figure 11).

The image query is carried out with Figure 12 as input, and the image retrieval result of the tourist attraction is obtained based on the local CNN feature of the saliency region. Due to the large number of search results, the results of the search here are taken as examples, and some of the query results are shown in Figure 13 (not including Figure 11 as the query image).

4.2. Result 2: The Oxford Buildings Data set. The data selected this time comes from The Oxford Buildings. In the same way, select a picture of a tourist attraction as a query picture, as shown in Figure 14.

Feature extraction of image in query image and image library by image feature extraction technology and hashing of these features by hash function. The Hamming distance between the search image and the hash code of the image in the image library is then compared to obtain the final search result. Part of the results of the image retrieval of the tourist attraction is shown in Figure 15.

4.3. Result 3: Precision and Recall Results

- (1) After many experiments, with the different number of scenic spot recognition pictures, this article calculates

FIGURE 14: Oxford Buildings query map.

FIGURE 15: Oxford Buildings partial query results.

TABLE 3: The return rate and accuracy of experimental results for different numbers of received search information.

Return the number of search results (frequency)	Precision rate (%)	Recovery rate (%)
3	0.30	0.20
5	0.40	0.40
8	0.375	0.60
10	0.40	0.80

From the table, we can get that with more and more search information received, the accuracy of recognizing images will become smaller and smaller, and the return rate will also increase.

the odd accuracy and return rate, the numbers are 3, 5, 8, 10, and the results are shown in Table 3.

- (2) In the image retrieval of tourist attractions, it is necessary to search the recall and precision rates obtained from different scenic spots to calculate the

average recall rate and precision rate of multiple scenic spots, which are used to indicate the overall performance of the search results. Under the premise of returning the same number of pictures after each attraction search, multiple tourist attractions are retrieved, respectively, and the number of searches is 50, 100, 200, 400, 800. The average recall rate and average precision of tourist attractions are calculated separately. The results obtained are shown in Table 4 and Figure 16.

It can be seen from the results in Table 3 that as the number of retrieved tourist attractions increases, the average accuracy rate and average recall rate are both decreasing. It shows that the tourism image retrieval technology based on wireless communication network is stable and feasible. But this also suggests some shortcomings of this technique: the accuracy of the results is

TABLE 4: Average and precision recall rates for the number of different attractions.

Search for the number of attractions (frequency)	Average precision (%)	Average recall (%)
50	63	85
100	61.2	81.9
200	59.7	79.4
400	57.9	77.6
800	55.3	77.0

FIGURE 16: Average and precision recall rates for the number of different attractions.

not high. While reducing the number of image retrievals, we should also optimize the image retrieval technology to further improve the accuracy rate and meet people's needs for accurate image retrieval.

5. Conclusions

Based on the above background, this paper studies the application of wireless communication network in image retrieval of tourist attractions. Based on the related theories of image retrieval and wireless communication network, this paper first proposes a local CNN feature extraction algorithm based on saliency regions from the perspective of image understanding. It mainly studies the extraction of local CNN features, how to use the local detail information contained in the image to improve the accuracy of image retrieval, and realizes the localization and description of the key areas in the image. It was verified and analyzed in the same category of image retrieval tasks. Second, wireless communication network combined with hashing. The experiment first calculates the classification effect of the hash code, and then calculates the average precision by calculating the Hamming distance between the hash code of the image feature and the hash code of the image feature in the image library. By comparing the results, the precision and recall of the algorithm are analyzed. Through experiments, the above method has achieved better retrieval results in the image retrieval task, which can further improve the accuracy of image retrieval and has strong

practicability. Wireless communication network will still be of great significance to the future social development. But there are also certain problems: the samples selected in the experiment are large and may not be representative. The experiment is established on the basis of previous experiments, and its scientificity needs to be evaluated. The experimental data have not been verified, and there may be a problem of unreliable data. It is necessary to conduct in-depth research, not only in the image retrieval of tourist attractions proposed in this paper but also to develop the potential value of wireless communication network from multiple angles and more comprehensively.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the major project of Philosophy and Social Sciences Research of Colleges and Universities in Hubei Province: "Research on environmental construction strategies and methods of villages and towns in Hubei from the perspective of antiurbanization" (20zd061).

References

- [1] Q. Zhao and J. F. Cao, "Joint optimization of feature selection and parameters for multi-class SVM in skin symptomatic recognition," in *Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence (AICI'09)*, vol. 1, pp. 407–411, Shanghai, China, November 2009.
- [2] J. Deng, W. Dong, R. Socher, L. J. Li, and K. Li, "ImageNet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, Florida, FL, USA, June 2009.
- [3] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [4] P. K. Sankar and C. V. Jawahar, "Probabilistic reverse annotation for large scale image retrieval," in *Proceedings of the CVPR*, pp. 1–6, Minnesota, MN, USA, June 2007.
- [5] Y. Sun, X. G. Wang, and X. O. Tang, "Deep learning face representation by joint identification-verification," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1891–1898, Ohio, OH, USA, June 2014.
- [6] W. J. Li, S. Wang, and W. C. Kang, "Feature learning based deep supervised hashing with pairwise labels," in *Proceedings of the CVPR*, Boston, MA, USA, June 2015.
- [7] J. R. Bach, C. Fuller, and A. Gupta, "The virage image search engine: an open framework for image management," in *Proceedings of the SPIE, Storage and Retrieval for Still Image and Video Database IV*, pp. 76–87, San Jose, CA, USA, February 1996.
- [8] A. P. Pentland and R. W. Picard, "Tools for content-based manipulation of image databases," in *Proceedings of the Storage and Retrieval for Image and Video Database II*, pp. 34–47, Washington, DC, USA, February 1996.
- [9] W. Li and H. Song, "ART: an attack-resistant trust management scheme for securing vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 960–969, 2016.
- [10] W. Niblack, B. Ron, and E. Will, "QBIC project: querying images by content using color, texture and shape[C]," in *Proceedings of the SPIE Storage and Retrieval for Image and Video Databases*, W. Niblack, Ed., vol. 1908, pp. 173–187, SPIE Press, San Jose, CA, USA, February 1993.
- [11] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: content-based manipulation of image databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233–254, 1996.
- [12] H. Song and M. Brandt-Pearce, "A 2-D discrete-time model of physical impairments in wavelength-division multiplexing systems," *Journal of Lightwave Technology*, vol. 30, no. 5, pp. 713–726, 2012.
- [13] A. Rao, R. K. Srihari, and Z. Zhang, "Spatial color histogram for content-based image retrieval," in *Proceedings of the 11th IEEE International Conference on Tools with AI*, pp. 183–186, Illinois, IL, USA, November 1999.
- [14] G. Lin, C. Shen, and Q. Shi, "Fast supervised hashing with decision trees for high-dimensional data[C]," in *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, pp. 1971–1978, Ohio, OH, USA, June 2014.
- [15] W. Y. Ma and B. S. Manjunath, "NeTra: A toolbox for navigating large image databases," in *Proceedings of the International Conference on Image Processing*, vol. 1, pp. 568–571, Santa Barbara, CA, USA, October 1997.
- [16] B. Han, J. Li, J. Su, M. Guo, and B. Zhao, "Secrecy capacity optimization via cooperative relaying and jamming for WANETs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 1117–1128, 2014.
- [17] H. Lai, Y. Pan, and Y. Liu, "Simultaneous feature learning and hash coding with deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3270–3278, Boston, MA, USA, June 2015.
- [18] M. Douze, H. Jegou, and H. Sandhawalia, "Evaluation of gist descriptors for web-scale image search," in *Proceedings of the International Conference on Image and Video Retrieval*, Xi'an, Shaanxi, 2012.
- [19] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp. 209–226, 1977.
- [20] P. Shan, "Image segmentation method based on K-mean algorithm," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 81, 2018.
- [21] K. Fukunage and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Transactions on Computers*, vol. C-24, no. 7, pp. 750–753, 1975.
- [22] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, June 2006.
- [23] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *Proceedings of the 25th International Conference on Very Large Data Bases*, Morgan Kaufmann Publishers Inc., New York, NY, USA, June 1999.
- [24] A. Alexandr and P. Indyk, "Near-Optimal hashing algorithms for approximate nearest-neighbor in high dimensions," *Foundations of Computer Science Annual Symposium on*, vol. 51, no. 1, pp. 459–468, 2006.
- [25] D. Shakhnarovich, T. Darrell, and P. Indyk, "Nearest-neighbor methods in learning and vision theory and practice," *Pattern Analysis & Application*, 2006.
- [26] K. Grauman and R. Fergus, "Learning binary hash codes for large-scale image search," in *Machine Learning for Computer Vision*, pp. 49–87, Springer Berlin Heidelberg, 2013.
- [27] R. Datta and J. Dhiraj, "Image retrieval: ideas, influences, and trends of the new age," *ACM Computing*, vol. 40, no. 2, pp. 1–60, 2008.
- [28] G. Hinton, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [29] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [30] D. Silver, A. Huang, and C. Maddison, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2006.
- [31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105, MIT Press, Lake Tahoe, NV, USA, December 2012.