International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2024): 7.741

Image Retrieval Algorithm based on Multi-feature Fusion

Shijie He[1]

[1]North China Electric Power University, School of Control and Computer Engineering
No. 2 Beinong Road, Changping District, Beijing, China
2950234202@qq.com

Abstract: With the development of science and technology, more and more image information appears in all aspects of life. Image is a common unstructured data. In order to obtain more abundant information of image, this paper proposes an image retrieval algorithm based on multi-feature fusion. Through dimensionality reduction and weight calculation, different image features are fused to improve the accuracy and accuracy of image retrieval. Firstly, the image is preprocessed and the image features are extracted using three basic models, GLCM, SIFT and CNN. Then, the dimensionality of the image features is reduced using principal component analysis. Finally, the multi-scale attention mechanism is used to fuse the image features. Experiments show that the average retrieval accuracy of the proposed method is 93.74% and the accuracy is 87.84%. It is significantly higher than the single feature search results.

Keywords: Image retrieval; Feature fusion; Principal component analysis; Multi-scale attention mechanism

1. Introduction

With the rapid development of Internet technology, image information is becoming more and more abundant, and the demand for images in new media, transportation, medical and other fields is increasing rapidly. Images contain rich information, and how to efficiently and quickly retrieve the images you need from numerous images is the focus of current research in the image field.

Image retrieval is divided into two categories: text-based image retrieval (TBIR) and content-based image retrieval (CBIR). Text-based image retrieval is the retrieval of similar images through text information, which relies on manually marked text information, and has low retrieval efficiency. Traditional text-based retrieval methods have been unable to meet people's need for fast image information acquisition. Content-based image retrieval does not rely on the text information manually marked in advance, but directly analyzes the content features of the image, and reasonably expresses the image by extracting the color, texture, edge and other features of the image, so as to achieve image retrieval.

2. Related work

In recent years, most scholars in the image field have carried out research on image feature extraction and image retrieval. Gao Dihui et al proposed a joint feature method for image cross-modal retrieval. In the process of global feature optimization, the attention mechanism is used to filter redundant information, reduce the gap in fineness with local features, and improve the retrieval accuracy [1]. Zhang Qinglu et al. proposed an unsupervised hash retrieval algorithm based on feature co-occurrence, which reduces the interference of background factors in the process of feature fusion at the stage of deep feature extraction and improves the expression ability of image features. Finally, unsupervised hash generation is carried out to further improve the image information contained in the hash code through KL loss, which effectively improves the image retrieval accuracy [2]. Alefu et al. proposed an image retrieval method based on global attention by orthogonal fusion of image descriptors, which extracts the orthogonal values of local features on the global feature vector and then aggregates the global features to form the final descriptor. Compared with other descriptors, this method improves the image retrieval accuracy [3]. Existing methods only use global features to describe images or only use local features to describe images, and a single feature is not accurate enough to express image information, and there are limitations in the use of graphic information, and the retrieval performance needs to be further improved. This paper proposes a multi-feature fusion image retrieval algorithm to address the above problems. The main contributions of this paper are as follows:

(1) Multi-feature fusion is used to perform the final image retrieval task, avoiding the impact of insufficient information of single feature;

(2) Combined with the principal component analysis method to reduce the dimensionality of image features, reducing the complexity of features;

(3) Considering that different features have different effects on the accuracy of image retrieval, the multi-scale attention mechanism is used to calculate the weight of different features, and the importance of each feature is fully considered.

3. Proposed method

The model proposed in this paper uses gray level co-incidence matrix (GLCM), scale invariant Feature transform (SIFT) and convolutional neural network to extract image features respectively, and carries out dimensionality reduction of image features through principal component analysis and fusion of image features, and carries out image retrieval through the fused features. The image retrieval algorithm framework based on multi-feature fusion is shown in Figure 1 below:

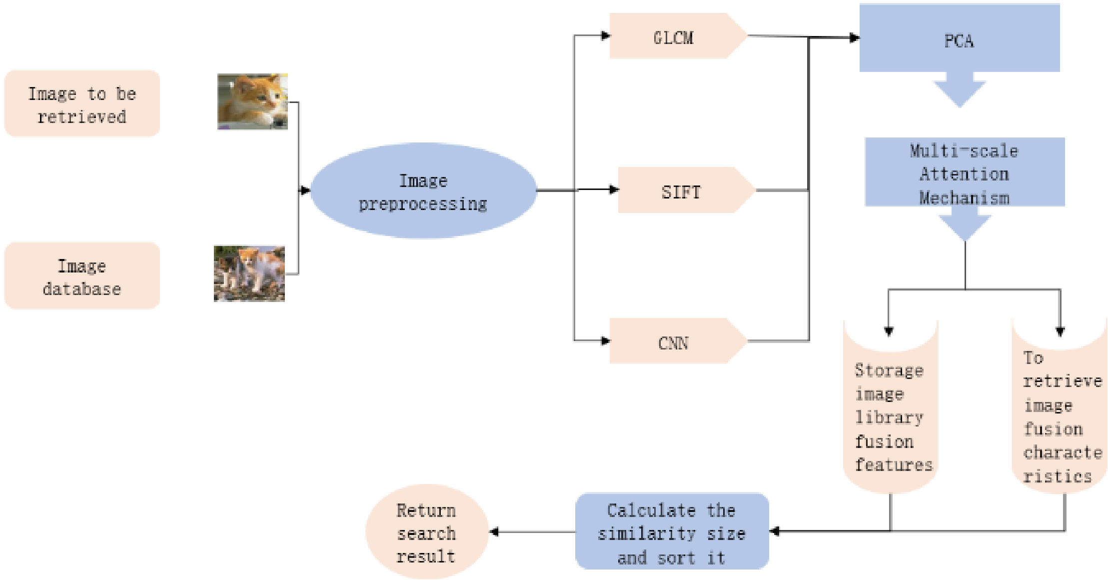

Figure 1 Multi-feature fusion image retrieval framework

The main steps of the model proposed in this paper are:

(1) Image preprocessing;

(2) Image features were extracted by different algorithms and dimensionality was reduced by principal component analysis;

(3) Calculate the weight of each feature and fuse the features;

(4) The similarity between the fused features and the image features in the image library is calculated;

(5) Feedback the search results.

Texture feature extraction adopts gray co-occurrence matrix (GLCM) method, which can be used to extract texture information in the image. Texture information mainly describes the texture level features of the image by analyzing the spatial relationship between pixels and gray distribution. Scale invariant Feature transform (SIFT) can extract key points in images and generate feature descriptors that are invariant to scale and rotation, which can accurately match some feature points in images at different scales and angles [4,5]. Convolutional neural networks can learn higher-level abstract features, shapes and edges of images, etc., with higher robustness and accuracy.

In feature extraction, principal component analysis (PCA) can focus on important feature parts and dimension reduction for features, which can effectively reduce the dimension of features.

Three basic models, texture feature extraction, scale-invariant Feature transform (SIFT) and convolutional neural network, extract image features respectively, and fuse multiple individual image features into a new image feature is an important problem in deep learning. The attention mechanism can dynamically learn to fuse the weights of different features [6]. It can make the model pay more attention to the important features of the current task, and adaptively integrate different feature expressions.

The combination of attention mechanism can reduce the feature dimension while retaining important feature information during the fusion of various features, thus enhancing the attention to the key features of the image, improving the model performance and generalization ability, and improving the accuracy of image retrieval through image features. The multi-scale attention mechanism can assist the model to pay attention to important information in the image from different scales [7]. By introducing multi-scale attention, the model can consider both global information and local details at the same time, so as to capture more comprehensive image features.

The extracted image features are expressed in the form of feature vectors. First of all, the vectors need to be standardized. The forward matrix of all indicator vectors is shown as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1b} \\ \vdots & \ddots & \vdots \\ x_{a1} & \cdots & x_{ab} \end{bmatrix} \tag{1}$$

The value of a ranges from 1 to n (n indicates the number of data items to be processed), and the value of b ranges from 1 to 3, with three feature vectors.

The normalized matrix is denoted as Y, where each element is calculated by:

$$y_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=0}^{a} x_{ij}^2}} \tag{2}$$

The numerator represents each element in the positive matrix, and the denominator is the sum of the squares of all the elements in each column. The normalized matrix is obtained as follows:

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1b} \\ \vdots & \ddots & \vdots \\ y_{a1} & \cdots & y_{ab} \end{bmatrix} \tag{3}$$

After standardized processing of image feature vectors, it is necessary to conduct feature dimensionality reduction through principal component analysis, and find covariance matrix for standardized data:

$$R = \frac{Y^T Y}{a} \tag{4}$$

Calculate the eigenequation of the covariance matrix to solve the eigenvalue:

$$R - \lambda E = 0 \tag{5}$$

To solve the$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_b > 0$

The corresponding feature vector is:

$$\beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \vdots \\ \beta_{b1} \end{bmatrix}, \beta_2 = \begin{bmatrix} \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \vdots \\ \beta_{b2} \end{bmatrix}, \cdots, \beta_1 = \begin{bmatrix} \beta_{1b} \\ \beta_{2b} \\ \beta_{3b} \\ \vdots \\ \beta_{bb} \end{bmatrix} \quad (6)$$

Set a threshold k, and the accumulative and higher eigenvalues of the principal components of the preceding p can be expressed as the original variable:

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq k \quad (7)$$

After the number of principal components is obtained, the eigenvector corresponding to the first p eigenvalues can be used to calculate the data after dimensionality reduction:

$$Z_i = \beta_i^T X = \beta_{1i} Y_1 + \beta_{2i} Y_2 + \cdots + \beta_{bi} Y_b,$$
$$i = 1,2,\cdots,p \quad (8)$$

After dimensionality reduction of image features, the weight of each feature is calculated for the subsequent fusion of image features. The probability matrix P of the proportion of the i sample value of the JTH index is as follows:

$$p_{ij} = \frac{\overline{x_{ij}}}{\sum_{i=1}^{a} \overline{x_{ij}}} \quad (9)$$

The weight of each vector is:

$$e_j = -\frac{1}{\ln(a)} \sum_{i=1}^{a} p_{ij} \ln(p_{ij}), j = 1,2,3,\cdots,b \quad (10)$$

Different features have different retrieval effects on images. After calculating the weight of features, the features are fused by summing, the fused feature vector is output, and the cosine similarity is used to calculate the similarity between the retrieved image and the image in the image library.

$$\sin(x,y) = cos\theta = \frac{\vec{x} * \vec{y}}{||x|| * ||y||} \quad (11)$$

4. Experiment

The image database used for experimental verification in this paper is CIFAR-10. The CIFAR-10 image dataset contains 10 categories of images, each category contains 6000 images, which can be used for model verification.

Firstly, feature extraction and image retrieval are carried out using three basic models, namely gray Scale Co-existing matrix (GLCM) method, scale invariant Feature Transform (SIFT) and convolutional neural network. The first four most similar images are returned, and the wrong images are marked with red boxes. The retrieval results are shown as a, b and c in Figure 2 below:

（a）

（b）

（c）

Figure 2. Single image feature retrieval results

According to the retrieval results, the retrieval accuracy of image features extracted by gray scale co-existence matrix is lower than that of scale invariant Feature transform (SIFT) and convolutional neural network, and the retrieval accuracy of neural network is relatively higher. In order to compare the retrieval accuracy of the model proposed in this paper, the weight of each feature is calculated, the features are fused, and the fused features are used for image retrieval. The retrieval results are shown in Figure 3 below:

Figure 3 Image retrieval results of multi-feature fusion

Through the retrieval results returned by the experiment, it can be found that the image retrieval algorithm based on multi-feature fusion proposed in this paper has high accuracy. The following is a more objective description of the performance of the proposed model by calculating mAP, accuracy rate and recall rate of different retrieval methods. As shown in Table 1:

Table 1 Performance comparison table of different retrieval methods

Method	mAP	Accuracy	Recall rate
GLCM+ Cosine similarity	66.29%	62.69%	61.27%
SIFT+ Cosine similarity	72.53%	68.32%	69.35%
CNN+ Cosine similarity	80.65%	75.86%	71.62%
The model proposed in this paper	93.74%	87.84%	85.98%

In the image retrieval experiments of single image features and fused features respectively, it can be found that the fused image features are significantly superior to the single feature retrieval performance in three aspects: accuracy, recall rate and mAP, indicating that feature fusion has excellent retrieval effect, and the combination of feature fusion and weight calculation has excellent performance. In the process of feature fusion, three different image features are fully combined, and different weights are assigned to different features through weight calculation, which effectively considers the influence of different features on image retrieval results.

5. Conclusion

In order to improve the accuracy of image retrieval, this paper proposes an image retrieval algorithm based on multi-feature fusion. Firstly, three basic models GLCM, SIFT and CNN are used to extract image features respectively, and dimensionality reduction operations are carried out on the features. After dimensionality reduction, the weight of each feature is calculated, and the impact of each feature on the accuracy of image retrieval is fully considered. A single image feature is fused to output a new feature for image retrieval. After verification on CIFAR-10 dataset, the image retrieval accuracy is as high as 87.84%, which is obviously better than the single image feature retrieval accuracy, which proves the effectiveness of the proposed method.

References

[1] GAO Dihui, Sheng Lijie, Xu Xiaodong et al. Joint feature Method for graphic cross-modal retrieval [J/OL]. Journal of Xidian University :1-11[2024-03-19].

[2] Zhang Qinglu, Chen Yihua. Unsupervised hash search algorithm based on feature co-occurrence [J/OL]. Microelectronics and Computers :1-10[2024-03-19].

[3] Ai Lefu, Tao Yong, Jiang Changyu. Orthogonal fusion image descriptor based on global attention [J/OL]. Journal of Graphics :1-11[2024-03-19].

[4] Xiong W ,Xiong Z ,Xu P , et al.Learning to disentangle and fuse for fine-grained multi-modality ship image retrieval[J].Engineering Applications of Artificial Intelligence,2024,133(PB):108150-.

[5] Pundhir A ,Sagar S ,Singh P , et al.Echoes of images: multi-loss network for image retrieval in vision transformers.[J].Medical biological engineering computing,2024,

[6] Hao W ,Zhiyi Z ,Zhilin Z .Cross-Checking-Based Trademark Image Retrieval for Hot Company Detection[J].Journal of Organizational and End User Computing (JOEUC),2023,36(1):1-12.

[7] Tian X ,Hou D ,Wang S , et al.An Adaptive Weighted Method for Remote Sensing Image Retrieval with Noisy Labels[J].Applied Sciences,2024,14(5):