PAPER • OPEN ACCESS

Image retrieval based on convolutional neural network and linear discriminate analysis

To cite this article: Wei Xia *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **569** 052052

View the article online for updates and enhancements.

Image retrieval based on convolutional neural network and linear discriminate analysis

Wei Xia[1,2], Cong Hu[1,2], Xianghui Cheng[1,2], Peng Chen[1,2], Xing Jin[1,2*]

[1]School of Automation, China University of Geosciences, Wuhan 430074, China

[2]Hubei key laboratory of Advanced Control and Intelligent Automation for Complex System, Wuhan, 430074, China

jinxing@cug.edu.cn

Abstract. Image retrieval is a hot research topic in the field of computer vision image processing, and the user queries the image database for similar images and produces a list of recommendations. The paper firstly sets forth the research status of image retrieval, then the convolution neural network is briefly introduced. Due to the traditional image retrieval and recommendation system use manual extraction of image features is relatively cumbersome, and the retrieval accuracy is not high research status, the paper proposes an image retrieval method based on the improved convolutional neural network and linear discriminate analysis. Caltech256 and CIFAR-10 datasets were trained using the model in this paper, experimental, results show that the proposed method can effectively improve the performance of retrieval.

1. Introduction

With the popularity of digital cameras, mobile phones and other electronic devices, as well as the booming development of mobile internet and social network, the massive growth of image information on the internet[1]. As an important information carrier, image has many advantages over text information. For example, the information presented by image is more intuitive and abundant, which can not only accurately reflect the shape of objects, but also reflect the color, texture and other information. The rapid growth of image data on the internet also brings with it the rapidly growing demand for image information retrieval[2]. It follows that research on the lightweight prefabricated composite floor needs to be further studied. Hence, the lightweight composite slab was put forward, which consists of H-type thin-walled steel beams and lightweight concrete including the precast panels and the post-pouring layer. The composite slabs also have advantages of no formwork support, being lighter weight and factory prefabrication, et al. For application of the slabs in engineering, the mechanical properties of it is studied on basis of experiments, and a set of valuable suggestions are obtained.

The research on image retrieval started in the 1970s[3]. Image retrieval are mainly based on text. Relevant images are matched by input text information or relevant keywords, which relies on manually adding text descriptions to images in the retrieved database. Although this method can achieve a certain effect, due to the rich information contained in the image itself and the strong subjectivity of manual annotation, it is unable to accurately and comprehensively express the information contained in the image with keywords.

Content-based image retrieval indexes the image according to the underlying features (such as color, shape, texture, etc.). Without the semantic information of the image, the retrieval effect needs to

be further improved. The retrieval process is mainly divided into three steps[4]:(1)to extract the image underlying features;(2)design feature fusion method;(3)similarity matching, return characteristics similar results. Image retrieval based on image content obtained the widespread application, the intellectual property rights, medical image processing, remote sensing image processing, as well as public security and other fields have image retrieval has been widely used, but the traditional content-based image retrieval technologies have some shortcomings, for example the accuracy is not high.

With the continuous upgrading of computer hardware and the development of the internet, and to improve the running speed of software applications and huge data. Deep learning technology[5] have drawn the attention of the researchers in the field of image processing, image retrieval based on deep learning is a good way to solve the image of the semantic gap between low-level features and high-level semantic problem. It can effectively improve the retrieval performance, image retrieval is applicable quantity demand.In 2006, professor Hinton and his student first put forward the theory of deep learning[6], and then deep learning entered a period of rapid development. Microsoft, Google, Baidu and other well-known large IT companies vigorously promote the research and application of deep learning, and have achieved good results in speech recognition, image classification retrieval and natural language processing. Deep learning can learn image features autonomously through models. By combining the bottom features of the image, a relatively abstract high-level feature is formed, and then the feature distribution of the image data is found and expressed, which makes the image classification and recognition more efficient and accurate.

2. Image retrieval algorithm based on convolutional neural network
In digital image data processing applications, convolutional neural networks are mainly used. The traditional image recognition algorithm extracts features and reconstruction data are relatively complex. The convolutional neural network can automatically extract features such as image color, texture and shape and topology, and has good robustness. The main components of CNN are input layer, convolution layer, pooling layer, fully connected layer and output layer[7].

In the CNN model, the image is first entered directly at the input layer or simply pre-processed. Then, a convolution operation is performed on the convolution layer to extract features, and local region sampling is performed in the pooling layer. After multiple convolution transformations and pooling, the structure information of the extracted image data is comprehensive. After the image features are further connected to the layer, the final result is output at the classification layer[8].

2.1. AlexNet network model
Alex et al. used the 8-layer convolutional neural network AlexNet model to successfully reduce the error rate of more than 1 million image classification tops in 1000 categories in the computer vision competition ILSVRC in 2012, winning the competition champion. AlexNet consists of five convolutional layers and three fully connected layers, multiple convolution kernels can extract interesting features from an image. In general, the size of the kernels in the same convolution layer is same.

2.2. Improve the AlexNet network model
In the AlexNet model, a region of 227*227 size is truncated in the original. RGB image, and a convolution layer is sent for convolution to extract image features. The first convolution layer extracts 96 features. The Relu activation function is used for nonlinear activation, the maximum pool is performed at the pooling layer, then the local response normalization processing is performed, and finally the classification result is output at the fully connected layer. The number of neurons output by the five convolutional layers is 96, 256, 384, 384, and 256, and the number of neurons output by the three fully connected layers is 4096, 4096, and 1000, respectively. In order to effectively avoid the over-fitting problem, the first and second fully connected layers use the regularization technique dropout algorithm to randomly set half of the neurons in the network to 0, which greatly improves the adaptability of the parameters. The third fully connected layer is the classification layer, and the

classification function softmax is used to classify the image. Assuming that the weights and offsets are represented by W and b respectively, the linear prediction of the i-th category can be expressed as formula (1).

$$z_i = W_i^T x + b_i \tag{1}$$

The softmax function can be expressed as formula (2).

$$\sigma i(z) = \frac{\exp(z_i)}{\sum_{j=1}^{m} \exp(z_j)} \tag{2}$$

The difficulty of neural network training becomes larger with the deepening of depth. In order to improve the training efficiency, the AlexNet network model is slightly improved. The local response normalization layer after the pooling layer is eliminated, and the more popular adaptive parameterization algorithm BatchNorm[9] is used. So that the input of each layer is consistent in training. BatchNorm is a fixed issue of activation input distribution proposed by the Google team in 2015 to solve deep neural network hidden layer nodes. Each time a random gradient of the hidden layer neurons of the neural network is performed, the activation function is normalized by mini-batch, so that the input values of the neurons are normally distributed, the gradient disappears, and the learning convergence speed and training speed are accelerated. Mini-batch can be expressed as formula (3).

$$B = \{x_1, x_2, ..., x_m\} \tag{3}$$

assume formula (4)

$$y^k = \alpha^k x^k + \beta^k \tag{4}$$

the mean of mini-batch can be expressed formula (5).

$$\mu_\beta = \frac{1}{m} \sum_{i=1}^{m} x_i \tag{5}$$

the variance of mini-batch can be expressed formula (6)

$$\sigma_\beta = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (x_i - u_\beta)^2} \tag{6}$$

2.3 LDA dimension reduction
The image characteristics of the first two fully connected layers are all 4096 dimensions, and it takes a long time to perform similarity calculation on large image data sets. In order to reduce the amount of calculation and improve the retrieval speed, this paper uses linear discriminant analysis(LDA) algorithm to linearly reduce the obtained CNN features. LDA is a dimensionality reduction technique for supervised learning. Each sample of the data set has a category output. After projection, the variance within the class is the smallest, and the variance between classes is the largest. We want to project the data in a low dimension. After projection, we want the projection points of each category data to be as close as possible, and the distance between the category centers of different categories of data is as large as possible. The dimensionality reduction process of the LDA algorithm is expressed as follows: input: The dataset $D = \{(x_1, y_1), (x_2, y_2)...,(x_m, y_m)\}$, where any sample x_i is a n dimension vector, $y_i = \{C_1, C_2, ...C_k\}$, and dimension d to dimension reduction, output: sample set D' after dimensionality reduction, LDA dimension reduction process is shown in Table 1.

Table 1. LDA dimension reduction

LDA Dimension Reduction Process
1. Calculate the intra-class divergence matrix S_w.
2. Calculating the inter-class divergence matrix S_b.
3. Calculation matrix $S_w^{-1}S_b$.
4. Calculate the maximum d eigenvalues of $S_w^{-1}S_b$ and the corresponding d eigenvectors to obtain the projection matrix W.
5. Convert each sample characteristic x_i in the sample set to a new sample $Z_i = W^T x_i$
6. Get the output sample set $D' = \{(z_1, y_1),(z_2, y_2),...,(z_m, y_m)\}$

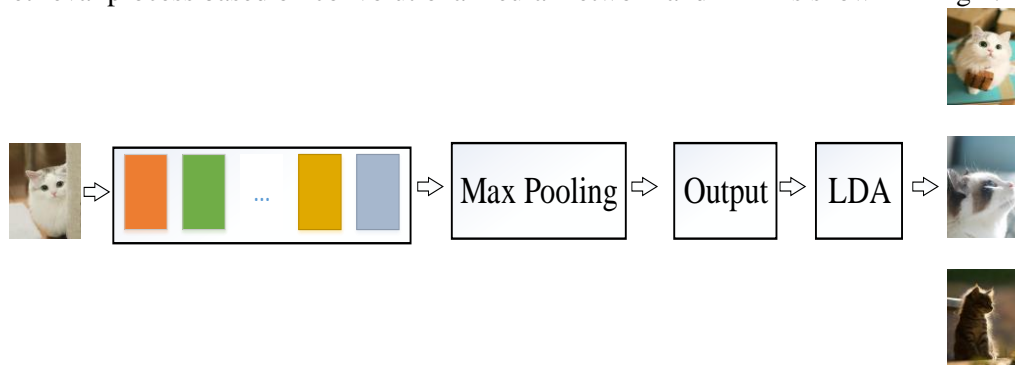The retrieval process based on convolutional neural network and LDA is shown in Fig 1.

Fig 1. Image retrieval process base on CNN and LDA

The sample data is simply processed and input into the CNN model for repeated iterative training. The image features outputted in the full connection layer of the model are composed into an image feature library, and the corresponding dimensionality reduction processing is performed according to the retrieval needs. Enter the image to be retrieved and extract the image feature, perform a similarity measure to calculate the distance between the image to be queried and the feature library. The metric algorithm uses Euclidean distance, and the results are sorted according to the similarity size.

3.   Experimental results and analysis

The experimental platform system configuration is: 64-bit Windows 7, Intel(R) Xeon E5-1620v4 @3. 50GHz, 64GB RAM, NVIDIA GeForee GTX 1080GPU. The improved model of AlexNet is used to improve the training speed and correct rate. The output of the fully connected layer of the improved model is extracted as the CNN feature, and the LDA algorithm is used to reduce the dimension of the extracted features. TensorFlow[10] is a deep learning framework for Google open source that runs on CPUs and GPUs and supports C++ and Python programming languages, so you can run code freely on different computers. TensorFlow not only supports deep learning, but also supports tools for reinforcement learning and other algorithms. The calculation graph of TensorFlow is pure Python, so the calculation slower.

3.1. Experimental dataset and performance evaluation index

The experimental datasets use the moderately sized CIFAR-10 and Caltech-256 datasets. The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. The Calteth-256 dataset consists of over 30,000 images in 256 object categories. Three evaluation indexes of MAP (average accuracy), P@K and R@K were used as the evaluation criteria for image retrieval experiments. Among them, P@K indicates the precision of the system search, and R@K indicates the recall rate of the system search. The similarity measure uses Euclidean distance.

3.2. Experiment and result analysis

In the iterative training process of the AlexNet model, the weights are dynamically updated through the learning rate. The training results show that when the learning rate is 0.0001, the accuracy is higher and the convergence speed of the network is faster. In the two datasets of CIFAR-10 and Caltech-256, five categories were randomly selected as experimental tests to compare the retrieval performance based on CNN features extracted from this paper and the retrieval performance based on GIST features, the experimental retrieval performance without using LDA for feature dimensionality reduction is shown in table 2. It can be seen that the method to improve the model is obviously better than the GIST characteristics.

Table 2. Retrieval performance comparison

Dataset	Methods	MAP	P@K=5	R@K=5
CIFAR-10	GIST	0.21	0.34	0.06
	Proposed method	0.45	0.75	0.18
Caltech-256	GIST	0.18	0.35	0.09
	Proposed method	0.82	0.78	0.16

In order to compare the retrieval performance in different dimensions, a total of 500 images in 10 categories were selected as test images in the Caltech256 dataset. LDA algorithm was used to conduct dimensionality reduction processing on the CNN features extracted from the model. The retrieval performance of the system in different dimensions is shown in table 3. The experimental results show that the retrieval performance after dimension reduction is better than that without dimension reduction. Retrieval performance is best when reduced to 128 dimensions.

Table 3. Retrieval performance table for different dimensions

The dimension	MAP	P@K=5	P@K=10	R@K=5	R@K=50
4096	0.38	0.76	0.63	0.16	0.45
1024	0.42	0.77	0.65	0.17	0.52
512	0.46	0.78	0.66	0.18	0.54
256	0.48	0.78	0.68	0.19	0.58
128	0.51	0.80	0.70	0.23	0.61
64	0.49	0.78	0.69	0.22	0.59

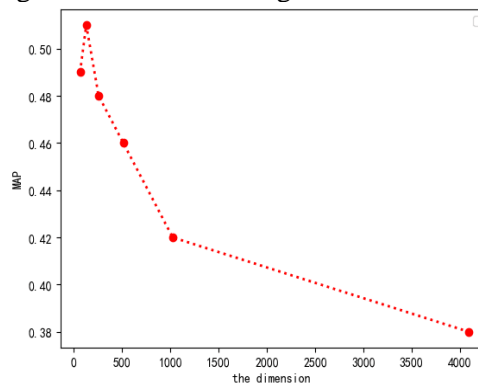MAP as the dimension changes as shown in the Fig 2.

Fig 2. The change curve of MAP in different dimensions

4.   Conclusion

The image retrieval algorithm based on CNN and LDA can effectively utilize image semantic information and improve image retrieval performance. It is very suitable for solving large-scale image data query requirements and is valued by researchers. In this paper, the convolutional neural network model AlexNet provided by the deep learning framework TensorFlow is improved and optimized. At the same time, LDA dimensionality reduction method is used to conduct linear dimensionality reduction for the extracted CNN features. The experimental results show that, the retrieval effect by

CNN is better than the traditional GIST feature retrieval effect, and the retrieval efficiency after dimension reduction has also been effectively improved, it has certain practicability.

References
[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems, 2012.
[2] Lai H, Yan P, Shu X, et al. Instance-Aware Hashing for Multi-Label Image Retrieval. IEEE Transactions on Image Processing. , 2016, 25(6):2469-2479.
[3] Gordo A, Almazán J, Revaud J, et al. Deep Image Retrieval: Learning Global Representations for Image Search European Conference on Computer Vision. 2016.
[4] Gordo A, Almazán J, Revaud J, et al. Deep Image Retrieval: Learning Global Representations for Image Search Computer Vision – ECCV 2016. 2016.
[5] Paulin M, Mairal J, Douze M, et al. Convolutional Patch Representations for Image Retrieval: An Unsupervised Approach. International Journal of Computer Vision, 2016, 121(1):1-20.
[6] Hoang T, Do T T, Tan D K L, et al. Selective Deep Convolutional Features for Image Retrieval. 2017.
[7] Liu X, Tizhoosh H R, Kofman J. Generating Binary Tags for Fast Medical Image Retrieval Based on Convolutional Nets and Radon Transform. 2016.
[8] Ke Y, Wang Y, Liang D, et al. CNN vs. SIFT for Image Retrieval: Alternative or Complementary? Acm on Multimedia Conference. 2016.
[9] Wei Y, Yang K, Yao H, et al. Exploiting the complementary strengths of multi-layer CNN features for image retrieval. Neurocomputing, 2016, 237:235-241.
[10] Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning[J]. 2016.