

# Auto Image Classification Based on Convolution Neural Network

Yong Wang <sup>a)</sup>, Dongdong Shen <sup>b)</sup> and Ying Wang

*School of computers, Guangdong University of Technology, Guangzhou 510003, China*

<sup>a)</sup> wangyong@gdut.edu.cn

<sup>b)</sup> Corresponding author: 393146198@qq.com

**Abstract.** Aiming at the low accuracy of vehicle image retrieval algorithm based on deep learning, an improved vehicle image classification retrieval model based on convolutional neural network is proposed. According to the complexity of the car image, using convolution neural network to extract the image features from Stanford Cars Dataset database, and use a local feature aggregation descriptor (vector of locally aggregated descriptors, VLAD) to represent a picture. Finally, SVM is used to classify the image of the car. The experimental results show that compared with the traditional visual feature classification algorithm, the accuracy of the model is higher and the retrieval effect is better.

**Key words:** Vlad; CNN; Vehicle image retrieval; SVM; Traditional feature extraction; SIFT.

## INTRODUCTION

With the improvement of people's economy and the increasing demand for vehicles, more and more Chinese families own cars. By the end of 2015, China's car ownership reached 172 million, and the trend of rapid growth remained. In Beijing, Chengdu, Shenzhen and other big cities, less than two families own a private car, China has become a veritable "country on wheels".

Due to the variety of vehicles and the complex background, the artificial design features are increasingly difficult to meet the actual demand with the rapid growth of image data. In recent years, traditional image processing technology has gradually been replaced by the convolutional neural network. Instead of using the features of manual extraction, the convolution neural network is used to study autonomously. For example, in image processing, is mostly based on CNN models, such as VGG-16, VGG-19, GoogleNet and so on. Although these models perform well on ImageNet, they are far from industrial applications in the classification of cars. The main reason is that the lack of automobile samples and the difference of the sample result in low accuracy.

This paper USES the convolution neural network to extract the image features, and then combines the traditional image processing technology, and adopts the model of CNN [1]+SVM for the problem of the shortage and difference of the automobile sample.

## RELATED TECHNOLOGIES

### Convolutional Neural Network.

Convolution neural network [2] is a special kind of neural network model of deep, its particularity embodied in two aspects, on the one hand, it's the connections between neurons are all connected, on the other hand some weights of connections between neurons in the same layer is Shared (i.e., the same). It's not all connections and weights of Shared network structure to make it more similar to the biological neural networks, reduces the complexity of the network model (for the deep structure of it is difficult to learn, this is very important), reduce the number of the weights.

### SIFT

The full name of SIFT [3] is the Scale Invariant Feature Transform, and the Scale Invariant Feature transformation is proposed by the Canadian professor David g.l. owe. The characteristics of SIFT are invariant to rotation, scale scaling and brightness change, which is a very stable local feature.

### SVM

The main ideas of SVM [8] can be summarized as two points:

1. It is in view of the linearly separable case analysis, in the case of linear inseparable, through the use of nonlinear mapping algorithm will undivided linear sample low-dimensional input space into a high-dimensional feature space linear separable, so as to make the high dimensional feature space by using linear algorithm for linear analysis on the nonlinear characteristics of samples is possible.

It is based on structural risk minimization theory to build the optimal hyperplane in the feature space, make learning editor to get the global optimization, and in the whole sample space expectations with a certain probability to meet a certain upper bound.

## CLASSIFICATION AND RETRIEVAL OF AUTO IMAGES BASED ON CONVOLUTION NEURAL NETWORK.

In this paper, a new classification retrieval model for automobile images is presented. The database used is Stanford Cars Dataset, with 16185 pictures and 196 categories. the convolution neural network is used to automatically learn the characteristics of the car, and the vlad is adopted to encode the features, and 196 SVM is used for classification. For the test image extraction feature and vlad encoding, the maximum value obtained by using SVM is the final category.

### Convolutional Neural Network.

In this paper, the CNN [9] training model is used to design the network framework as shown in FIG1, including the 4-layer convolutional layer, the three-layer full connection layer, and the input image size is 224 \* 224. On the basis of this framework and model, the whole connection layer structure is adjusted.

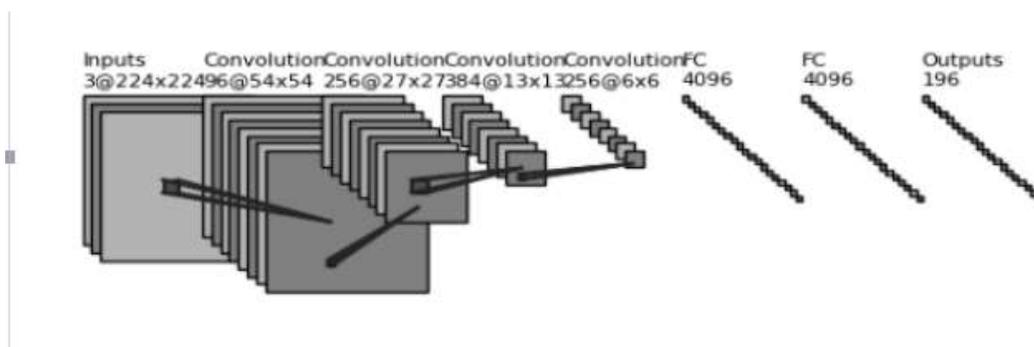


FIGURE 1. CNN the network structure

Due to the connection layer parameters fitting, can produce too much of a fitting to avoid network, combined with the car data sets, adjust the output for 196 of the last layer of the whole connections, which reduces the original parameters of the whole connection layer.

Lossfunction:

$$L = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K \left[ \max \left( 0, 1 - \delta \left( y^{(i)}, j \right) w^T x_i \right) \right]^p \quad (1)$$

As shown in figure 2-3, by comparing the features of CNN and SIFT, it is found that the characteristics of CNN learning are better. According to the extracted SIFT feature graph, SIFT can only extract the global information description of the car, and it is difficult to distinguish the characteristics of the class. Different from SIFT features, CNN can learn the characteristics of edge features through the connection between layers. The fourth layer of convolution can be used to extract all the global features, rather than the global characteristics of the shallow layer, according to a large number of data sample training. Therefore, the CNN model can effectively acquire deeper semantic features, which are very helpful to the classification of cars.

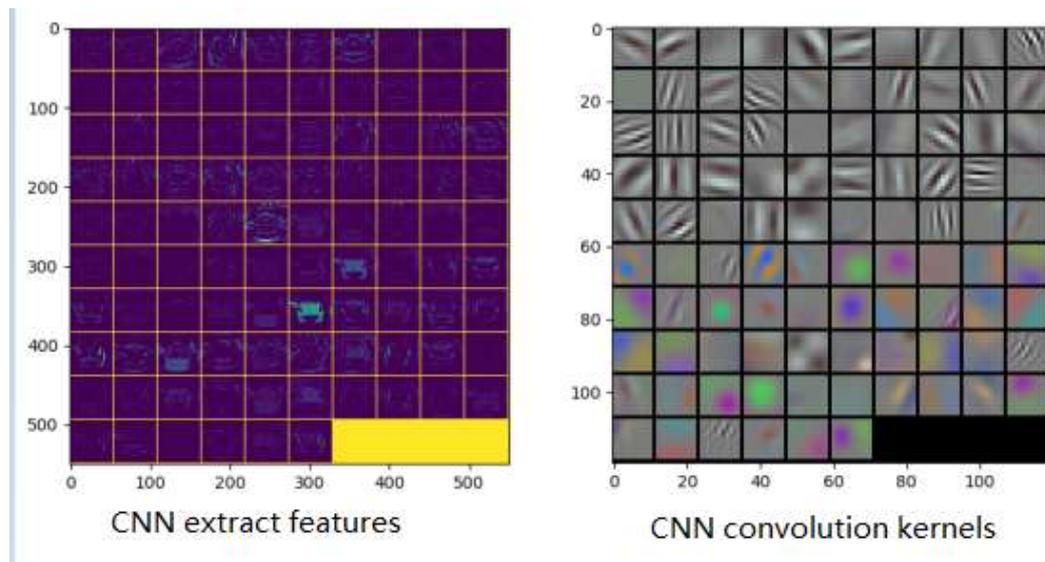


FIGURE 2. Cnn Feature visualization

### SVM classifier.

The database used is Stanford Cars Dataset, with 16185 pictures and 196 categories. Among them, 8144 is the training set and 8041 is the test set. By CNN to extract the image feature is 4096 d, using VLAD local feature descriptor (using 256. The codebook size ranging from 64-64 or even more, in theory, the greater the code book retrieval accuracy is higher), and then do a power normalization with the super long vector can slightly improve the retrieval accuracy, and then to the super-long vector normalization, do it again now the long column can be saved. 196 classifiers are made using SVM.

## RESULT AND ANALYSIS

### Experimental results

This experimental hardware is configured with i7 processor and NVIDIA GTX1060 GUP. Software configuration is Ubuntu16.0.4 system Caffe and CUDA 8.0, Cudnn6.5, Matlab2017a, Opencv3.0. The initial learning rate of CNN is 0.01. The experimental results are as follows:

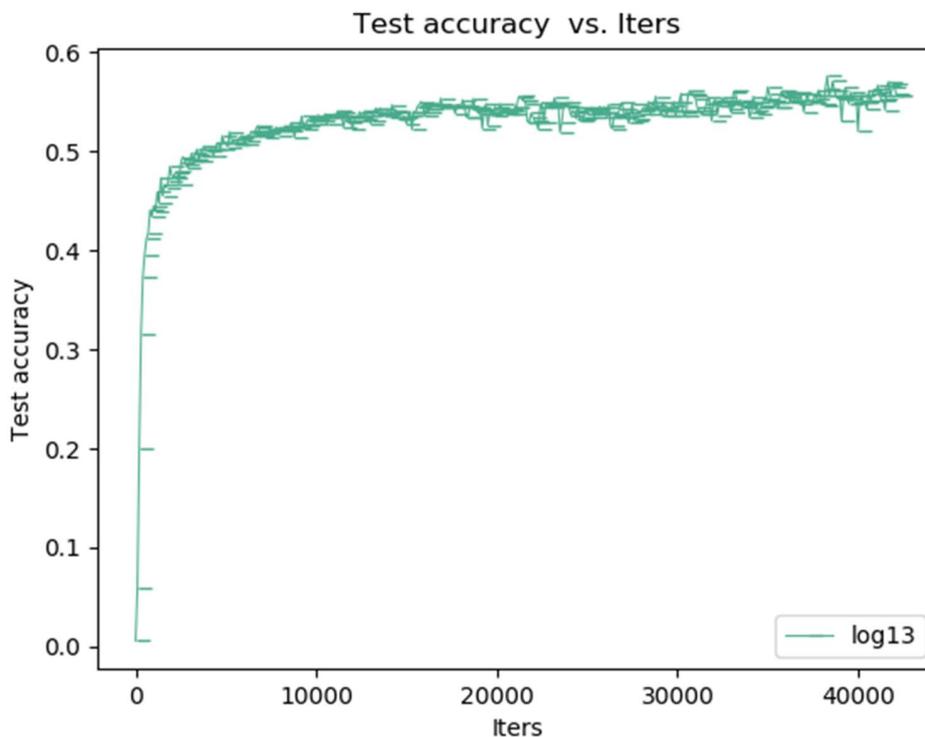


FIGURE 3. The experimental results are as follows

### Experimental analysis

The classification results of the test set in CNN+SVM reached 75% accuracy. The classification results of SIFT+SVM reached 60% accuracy. The results of the separate CNN classification reached 58 percent accuracy.

In terms of auto image classification retrieval, the algorithm of this paper is superior to traditional auto image classification retrieval, which is superior to single CNN algorithm. The image features extracted by convolutional neural network are superior to traditional visual feature extraction. However, in the case of insufficient data, a single convolutional neural network is similar to the traditional visual method in terms of accuracy. Therefore, the image classification retrieval algorithm proposed in this paper has a great advantage over other algorithms.

### CONCLUSION

In this paper, a classification retrieval algorithm based on convolutional neural network is proposed to classify cars accurately by means of deep learning and feature extraction using convolution feature. Experimental results show that the proposed algorithm is more accurate and better than traditional visual feature algorithms. But the amount of data is required for deep learning, as a result, the subsequent will increase the amount of training data, the depth of the research is more complex network structure to the training model, to further improve the accuracy of classification and retrieval

### REFERENCES

1. MH Nguyen, L Torresani, ITF De, C Rother. Weakly supervised discriminative localization and classification: a joint learning process. [C] IEEE International Conference on Computer Vision . 2009, 30(2):1925-1932

2. H Jegou, M Douze, C Schmid, P Perez .Aggregating local descriptors into a compact image representation.[J] Computer Vision & Pattern Recognition ,2010 , 238 (6)
3. AS Razavian, H Azizpour, J Sullivan, S Carlsson. CNN Features off-the-shelf: An Astounding Baseline for Recognition. [J] Computer Vision & Pattern Recognition Workshops. 2014 :512-519
4. J Donahue, Y Jia, O Vinyals, J Hoffman, N Zhang. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. [J]Computer Vision and Pattern Recognition (cs.CV).2013.6
5. Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]. Computer Vision and Pattern Recognition.2014.4
6. H Jegou, M Douze, C Schmid, P Perez. Aggregating local descriptors into a compact image representation. [C] Computer Vision & Pattern Recognition, 2010, 238 (6)
7. S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories [M]. CVPR 2006
8. MI Layton, MJF Gales .Augmented Statistical Models: Exploiting Generative Models in Discriminative Classifiers.[N] *Nips Workshops*, 2005
9. R Arandjelovic, P Gronat, A Torii, T Pajdla, J Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. [C] IEEE 2015, PP (99) :1-1