Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 107 (2017) 715 – 720

Procedia Computer Science

International Congress of Information and Communication Technology (ICICT 2017)

A Deep Convolution Neural Network Model for Vehicle Recognition and Face Recognition

Xingcheng Luo[a*], Ruihan Shen[b], Jian Hu[c], Jianhua Deng[d], Linji Hu[e] and Qing Guan[f]

[a]*Xingcheng Luo,* [b]*Ruihan Shen, School of Information and Software Engineering*

University of Electronic Science and Technology of ChinaCountry Chengdu, 610054, China P.R.C

** Corresponding author: 2014220601014@std.uestc.edu.cn*

Abstract

In recent years, vehicle recognition has become an important application in intelligent traffic monitoring and management. In this paper, we proposed a deep convolution neural network which is no less than nine layers. A vehicle data set is employed which is collected from multiple perspectives and the deep learning framework *Caffe* is used to verify the proposed algorithm. Comparing with traditional vehicle recognition based on machine learning which needs vehicle location and has low accuracy of shortcomings, the proposed model uses deep convolution neural network has a better performance.

Keywords: vehicle recognition, face recognition, nine-layer network, deep learning

1. Introduction

With the development of information technology, artificial intelligence has gradually become more and more important in daily life. Therefore, it is important to develop artificial intelligence. However, traditional machine learning algorithms have some flaws, such as low recognition efficiency, low accuracy of recognition efficiency. After deep learning proposed, both the recognition accuracy and efficiency have a greatly improve. More and more researchers take up deep learning. In fact, neuron network has been proposed since several years ago. Due to deep neuron network needs massive data, so it has not been able to become popular. Luckily, current *GPUs*, paired with a highly-optimized implementation of 2D convolution which are powerful enough to facilitate the training of deep neuron networks.

Some CNN models have been proposed in last few years, they have their own differences and innovations. Such

Peer-review under responsibility of the scientific committee of the 7th International Congress of Information and Communication Technology

doi:10.1016/j.procs.2017.03.153

as *Alexnet* [1], it has five convolution layers and three fully-connected layers. In addition, *Resnet* [2] is a hot spot recently. Many researchers also employ them in face recognition [3], detection [4] and video tracking [5]. To the best of our knowledge, deep learning algorithms are widely used for face recognition.

The network models all need tools to implement, so some deep learning frameworks are proposed. They support different programming languages interface, greatly accelerated the models operation speed. And the frameworks such as *Caffe*, *Torch*, *Theano* and so on, also can improve the performance in terms of accuracy.

In this paper, we employ *Caffe* framework to determine the different performance based on eight-layer network and nine-layer network. The Top-1 accuracy about nine-layer network reached 92.25% compared to the accuracy, below 80%, of the network where the layers are less than 8 layers Besides, the same nine-layer network with the same parameters is employed to determine the face *dataset*, where the performance in terms of accuracy only reached 80.5%.

The remainder of this paper is organized as follows. Section 2 introduces some related work of current research and section 3 describes the nine learned layers network and vehicle recognition process. Some experiment results will be reported by employing the deep learning framework *Caffe* in section 4, and finally, make a summary of the paper in section 5.

2. Related Works

The reason of deep learning popular is that it is able to independently learn a useful feature from data. Besides, a large-scale *dataset* is employed to study more complicated concepts, and massively parallel computing is used to do the optimization. So *Dr Jia Yangqing* who graduated from UC Berkeley invented a clear and efficient deep learning framework called *Caffe* which is convenient for researchers to implement their deep learning algorithms.

Caffe has a lot of advantages. Firstly, model and optimization are in the form of text rather than code form, and *Caffe* also gives the definition of the model, optimization settings and training weights. Secondly, *Caffe* used conjunction with *cuDNN*. Therefore, *Caffe* can run the best models and massive amounts of data [6]. And *Caffe* is easy to expand new tasks and settings. Researchers can use a simple language (e.g. *google protobuffer*) to define the new network structure by using CPU or GPU to implement it. Most importantly, researchers can publish their models, and share their own research results.

Furthermore, different CNN networks have different applications. *Alexnet* is an eight learned networks whose structure is shown in Fig. 1. The first two convolution layers with kernels size of 11×11 and 5×5, where has pooling layer. The convolution layers (i.e. 3^{th} 4^{th}) with kernels size of 3×3 each and without pooling layer. But the fifth convolution layer uses kernels size of 3×3 with a pooling layer. Besides, the first two fully-connected layers also use dropout to prevent over-fitting.

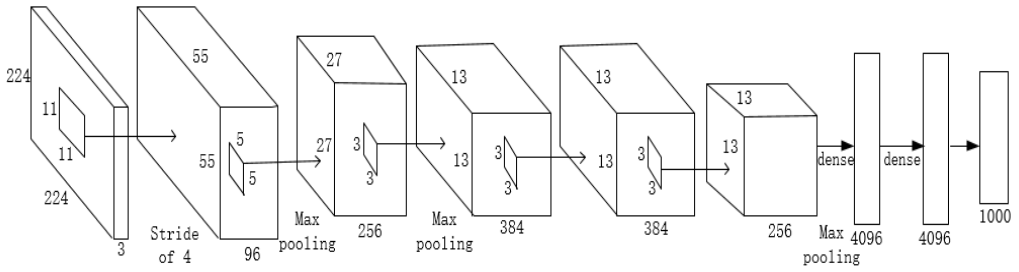

Fig. 1. The structure of *Alexnet* on *Caffe* framework.

3. Network Introduction

In this section, the architecture of nine-layer network is summarized in Fig. 2. It contains six convolution layers and three fully-connected layers. Then vehicle recognition model and process are described as follows.

The proposed nine learned layers network model consists of two parts. One is shown in Fig. 2, there are convolution layers, pooling layers and full-connected layers. Convolution layers are responsible for the feature extraction and noise remove. Pooling layers play a role in reducing parameters and improving the operation speed. Besides, fully-connected layers are responsible for classification and regression.

The other one also has some layers. As is known to all, activation function is very important for deep learning which is no longer a linear combination of the input, but can approximate arbitrary function. It makes hidden layers meaningful. However, how to choose an activation function is a difficult problem. Here, the proposed model use *ReLU* function, there are some advantages about *ReLU*. First, using *ReLU* activation function can save a lot of computation. Second, it can avoid the disappearance of the gradient [7]. Second, *ReLU* can cause the sparse nature of network and reduce the parameters, alleviate the over-fitting problem. LRN layer imitates biological neural system mechanism which create competition of local neuronal activity, and make response to a larger value is relatively larger, improve the model generalization ability [8]. In addition, over-fitting is a major problem in deep learning, so the dropout is necessary, which set to zero the output of each hidden neuron with probability 0.5 [9]. We use dropout in the first three fully-connected layers of Fig. 2.

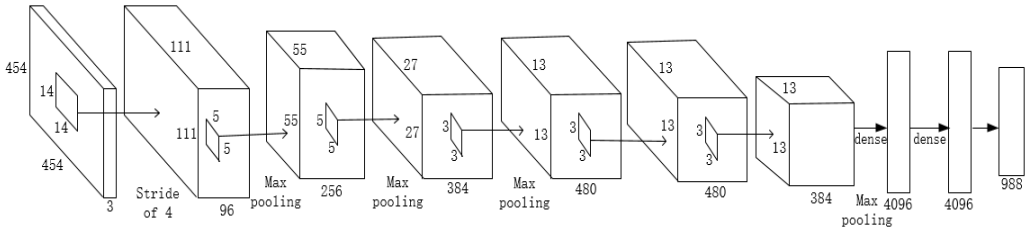

Fig. 2. The structure of nine learned layers *Caffe* Framework

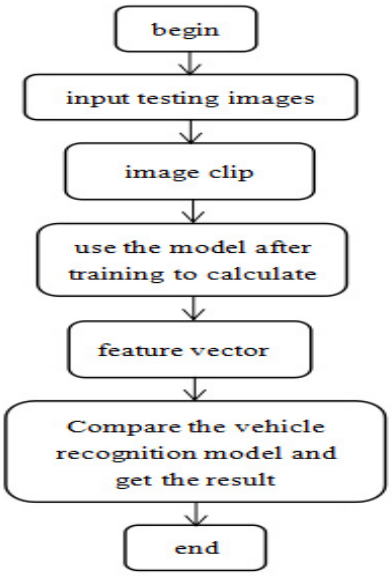

Fig. 3. Vehicle recognition process

As shown in fig. 2, the first convolution layer has the 454×454×3 input image, 96 kernels of size 14×14 and a pooling layer which uses Maximum pooling method [10]. The second convolution layer has 256 kernels of size 5×5 and a pooling layer by using maximum value method. The third convolution layer has 384 kernels of size 3×3 and also has a pooling layer by using maximum value method. Besides, the next three convolution layers (i.e. 4th 5th)

without any pooling layer and normalization layer [11]. But the fourth convolution layer has 480 kernels of size 3×3, the fifth convolution layer has 380 kernels of size 3×3, the sixth convolution layer has 384 kernels of size 3×3 and a pooling layer by using maximum value method. In addition, the fully-connected layers have 4096 neurons each and we use dropout in first two fully-connected layers. Finally, the 988 classes will be gotten.

Besides, to implement vehicle recognition, it still needs several steps which are shown in Fig. 3. Firstly, some testing images are employed and the images are clipped into 454×454. Secondly, the model after training is used to calculate. Then, we can get some feature vectors and compare the vehicle recognition model. Finally, the result will be gotten.

4. Experiment Results

In this section, we evaluate two scenarios. Then, we collect a large number of vehicle *dataset* from different sides, the vehicle *dataset* includes 998 classes, and nearly 90000 images, and are divided into 72000 training images and 18000 validation images. The face *dataset* includes 500 classes with over 5000 images, and are divided into 4000 training images and 1000 validation images. We evaluate both top-1 and top-5 accuracy rates.

All the experiments we use the same learning parameters setting which is shown in Table 1. First, the training images are randomly sampled into crops of size 454×454. And using SGD [12] with a batch size of 128, The learning rate initiated at 0.02 and is divided by 10 each 55000 iterations, and the model are trained for up to $45\times \mathbf{1O}^{4}$ iterations. We use a weight decay of 0.0005 and a momentum of 0.9.

Table 1. *Caffe* learning parameters setting

Parameter	Value	Meaning
batch size	128	number of pictures per training
base_lr	0.02	initial learning rate
lr_policy	step	trend of change in learning rate
stepsize	55000	learning rate for each change require lteration number
max_iteration	450000	maximum number of iterations
momentum	0.9	learning parameter
weight_decay	0.0005	learning parameter

4.1. Eight and Nine Learned Layers Network Based on Vehicle Dataset

We first evaluate eight learned layers network (*Alexnet*) and nine learned layers network, and make a comparison. The results in Table 2 show that the nine-layer net has a higher accuracy than the that of eight-layer net based on vehicle *dataset*. In other words, the performance in terms of accuracy, nine-layer network is better. To reveal the reasons, in Fig. 4 we compare their accuracy during the training procedure.

Table 2. Top-1 and Top-5 accuracy in eight-layer net and nine-layer net

	Top-1 accuracy	Top-5 accuracy
eight layers	79.86%	90.57%
nine layers	92.25%	97.51%

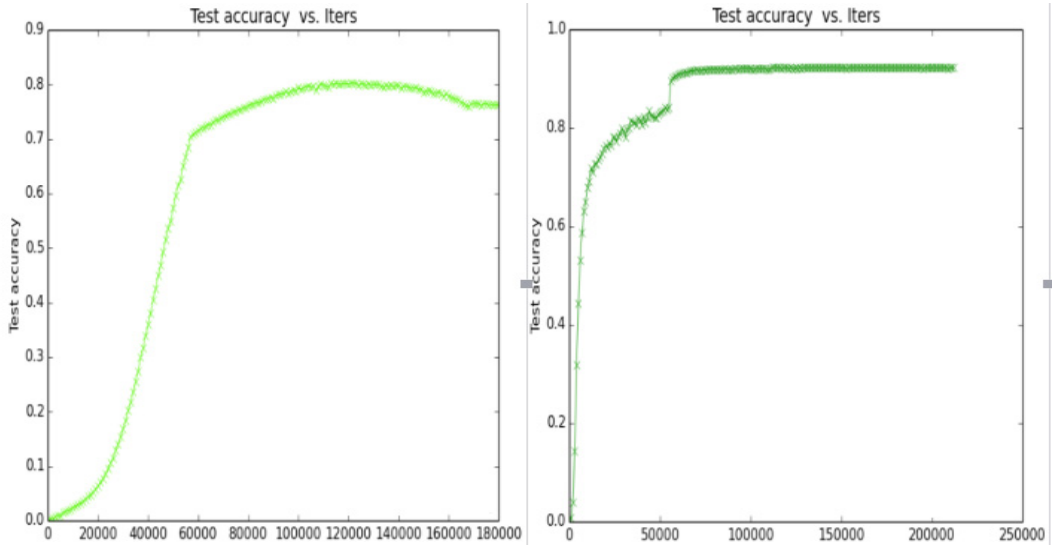

Fig. 4. The accuracy in training procedure about two nets. Left: eight-layer net. Right: nine-layer net.

As shown in Fig. 4, the performance in terms of accuracy, the nine-layer net is better, which over than 12%. This shows that add one layer of network, the accuracy has a significantly improved.

4.2. Nine Learned Layers Network Based on Vehicle and Face Dataset

Furthermore, we use the nine-layer net to test face *dataset*. Cutting out the image into 454×454. Besides, the actual face only occupies a small part of the whole image. But the vehicle occupies a most part of the whole image. The results on face *dataset* are summarized in Table 3.

Table 3. The accuracy on face *dataset*

	Top-1 accuracy	Top-5 accuracy
Face dataset (nine-layer net)	80.55%	91.22%

It is obvious that the accuracy based on face *dataset* is much lower than that on vehicle *dataset*. We estimate the reason is that the proportion of the effective parts in the whole image is different. The problem will be studied in the future.

5. Summary

To the best of our knowledge, more and more researches focus on computer vision for recognition, such as face recognition and vehicle recognition. In this paper, we proposed a nine-layer network, and the vehicle *dataset* is employed which is from multiple perspectives. The results are shown that a nine-layer net which achieve record-breaking results on vehicle recognition, and the accuracy of vehicle recognition reached over 92.2% by using the deep learning framework *Caffe*. In addition, the performance in terms of recognition accuracy, the eight-layer net is much lower than the nine-layer net.

Furthermore, there are also many potential researches which deserve our ceaseless exploration and excavation. We constantly optimize the algorithm model to improve recognition accuracy and speed. Finally, we will use the new model on other computer visions for recognition in the future.

References

1. K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun. What is the best multi-stage architecture forobject recognition? In International Conference on Computer Vision, pages 2146–2153. IEEE, 2009.
2. J. Sánchez and F. Perronnin. High-dimensional signature compression for large-scale image classification. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1665–1672. IEEE,2011.
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition. In Computer Vision and Pattern Recognition (CVPR), 2016
4. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks (NIPS), 2012
5. Steve Lawrence, C. Lee Giles, Ah Chung Tso, Andrew D. Back. Face Recognition: A Convolutional Neural Network Approach. IEEE Transactions on Neural Networks, 1997, 8(1):98-113
6. Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pages 253–256. IEEE, 2010.
7. Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, et al. Handwritten digit recognition with a back-propagation network. In Advances in neural information processing systems, 1990.
8. S. Gidaris and N. Komodakis. Object detection via a multi-region & semantic segmentation-aware cnn model. In ICCV, 2015.
9. K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In ECCV, 2014.
10. Haoxiang Liy, Zhe Linz, Xiaohui Shenz, Jonathan Brandtz, Gang HuayA Convolutional Neural Network Cascade for Face Detection. In CVPR, 2015.
11. AJ Spink，RAJ Tegelenbosch，MOS Buma，LPJJ Noldus. The EthoVision video tracking system—A tool for behavioral phenotyping of transgenic mice. Physiology &Behavior, 2001, 73(5):731-44
12. S Salti，A Cavallaro，SL Di. Adaptive appearance modeling for video tracking: survey and evaluation.《IEEE Transactions on Image Processing》, 2012, 21(10):4334-48
13. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, vZ. Huang, A. Karpathy, A.Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. arXiv:1409.0575, 2014.
14. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Le-Cun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In ICLR, 2014.
15. J Tsitsiklis，D Bertsekas，M Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms《IEEE Transactions on Automatic Control》, 1984, 31(9):803-812