Research on Vehicle Classification Method Based on Improved AlexNet

Bocheng zhu¹, Mengfan zhang¹, yusu¹ and xinping hu¹,²*

¹School of Information Science and Technology, Nantong University, NanTong, JiangSu, 226000, China
²Library of Nantong University, NanTong, JiangSu, 226000, China
* Corresponding author’s e-mail: xphu@ntu.edu.cn

Abstract: In the actual traffic intersection environment, vehicles adopted for traffic monitoring miss detection due to various reasons, and because the difference between different types of vehicles is very small, the traditional method cannot effectively distinguish the types of vehicles, and the deep learning image processing method can be used for automatic recognition of vehicle types. We propose an improved AlexNet¹ (ProAlexNet) intersection vehicle classification method by improving and reconstructing the hierarchy and parameters of the AlexNet convolutional neural network structure. In addition, we used a self-made high-quality vehicle category data set, which included 5000 pictures of cars, trucks, buses, motorcycles, and vans. In the experiment, we carried out comparative experiments on three indicators between ProAlexNet network and traditional AlexNet method, and carried out comparative experiments on three traditional recognition algorithms of ProAlexNet. Experimental results show that our improved algorithm has strong competitiveness.

1. Introduction
In recent years, the number of vehicles on the road has gradually increased. Road vehicle supervision has become an important task. Automatic vehicle detection and classification has become an important part of intelligent transportation system. Vehicle detection and classification include image acquisition, target detection, vehicle classification and other technologies. With the development of Internet and artificial intelligence, there are many image recognition algorithms based on machine learning. In this paper, the typical convolutional neural network model AlexNet is improved, and a vehicle detection and classification algorithm based on improved AlexNet is proposed².

With the development of science and technology, the image processing ability of GPU (Graphics Processing Unit) has been improved, and it is easier to train deep and complex neural network. Initially, researchers proposed deep learning models such as AlexNet, VGGNet, GoogleNet and ResNet, and used regularization, batch standardization and residual learning to avoid problems such as overfitting or gradient dispersion³. These network models laid the foundation for the development of deep learning later. However, considering that the deep network on the equipment with limited computing power cannot realize real-time vehicle detection, and the data tasks involved are relatively simple, do not need VGG, GoogleNet and other deep network models, so the AlexNet network model is chosen as the basic structure of the vehicle detection algorithm in this paper.
2. Method model

2.1. The traditional AlexNet network:
AlexNet was first proposed for image classification, and its network structure is shown in Figure 1. AlexNet has 8 layers, in which the first 5 layers C1~C5 are convolutional layers, the last 3 layers F1~F3 are fully connected layers, and the output of the last fully connected layer is the feature class with 1000 outputs[4]. The final optimization objective is to maximize the average Multinomial Logistic Regression. The convolution layer is used for feature extraction, the pooling function is dimensionality reduction, and the full connection layer is used for image classification[5].

![Fig. 1. AlexNet model](image)

2.2. ProAlexNet (The improved AlexNet):
The vehicle classification method based on improved AlexNet consists of three parts: The first part preprocesses the data. The second part, input training image set training model; In the third part, the image of the vehicle to be classified is input into  the trained model for classification. The method flow chart is shown in Figure 2:

![Fig. 2. The flow chart of PRoAlexNet](image)

AlexNet network model structure is relatively complex, the parameters are relatively many, so the computer hardware configuration requirements are higher, and will affect the time cost of model training. Therefore, we can reduce the time of model training and improve the accuracy of classification by designing network parameters. Specific improvements are as follows:

(1)Originally, the size of the picture input in the AlexNet network model was 227×227×3[6]. Considering the specifications of the car pictures collected, the size of the picture input was unified as...
256×256×3.

(2) As we detect and classify the types of cars, we modify the original 1000 neurons[7] in the output layer of AlexNet into 5 neurons, namely, they are divided into five categories: cars, trucks, buses, motorcycles and vans. Because the output categories are greatly reduced, the number of required output feature graphs can be reduced correspondingly, so as to improve the computing speed of the computer.

(3) Convolution is the most important operation in CNN[8]. The convolution calculation of two-dimensional image is mapped to the continuous sliding convolution window, and the corresponding convolution value is obtained. In CNN, each feature graph is convolved by multiple input feature graphs[9]. For the input x of the ith convolution layer, it is calculated as:

\[ h_{ic} = f(W_i * x) \] (1)

Where, * represents the convolution operation, \( W_i \) represents the convolution kernel of this layer, and \( f \) represents the activation function. \( W_i = [W_i^1, W_i^2, ..., W_i^K] \), K is the number of convolution kernels at this layer. Each convolution kernel \( W_i^K \) is a weight matrix of \( M \times M \times N \), where M is the window size and N is the number of input channels. The first five layers are the convolution layers of the feature extractor[10]; the change we made is to reduce the number of convolution kernels of these layers by half, that is, the first layer uses 48 11×11 convolution kernels. The second layer uses 128 5×5 convolution kernels. The third and fourth layers are provided with 192 3×3 convolution kernels; The fifth layer uses 128 3×3 convolution kernels. The nodes of the sixth and seventh layers, which are fully connected layers, are also halved correspondingly, from the original 4,096 to 2,048 [11].

(4) The ReLU[12] activation function can adaptively learn the parameters of the rectifier and improve the accuracy without increasing the extra cost. The graph of the function is shown in the figure:

\[ \text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \] (2)

The loss function measures the difference between the predicted result and the input label, and is defined as:

\[ E(W) = -1/n \sum_{i=1}^{n} \sum_{k=1}^{K} [y_{ik}\log P(x_i = k) + (1 - y_{ik})\log (1 - P(x_i = k))] \] (3)

Where \( W \) is the weight matrix of the convolution and fully connected layer, \( n \) is the number of training samples, \( i \) is the index of training samples, and \( K \) is the index of the class. If the \( i \)th sample belongs to class \( K \), then \( y_{ik} = 1 \); Otherwise \( y_{ik} = 0 \). \( P(x_i = K) \) is the probability of input \( x_i \) of class \( K \) predicted by the model, and is a function of parameter \( W \), so the loss function takes \( W \) as the parameter.

This paper uses a stochastic gradient descent (SGD) algorithm, where \( W \) iteration is updated as:

\[ W_k = W_{k-1} - \alpha(\partial E(W)/ \partial W) \] (4)

Where \( \alpha \) is the learning rate, which is a very important parameter, and it determines the step size of learning, \( K \) is the index of class.
3. The experiment is introduced

Experimental environment and parameters are as follows: Inter(R) Core(TM) I7-9750H 4210 CPU, GeForce GTX3090 24GB GPU, 16GB operating memory, and Windows10 operating system.

In our experiment, we used a self-made data set consisting of 5000 images of cars, trucks, buses, motorcycles, and vans. In the experiment, the data set is divided into three parts, namely, training set, test set and verification set. The ratio of the three was 8:1:1.

Since the number of vehicles in the dataset is not equal among the five types, different performance indicators need to be listed in order to provide a clear assessment of system performance. In this case, this paper uses the Positive Rate, the False Rate and the Accurate Rate of each class as the evaluation index.

4. The experimental results

4.1. Original AlexNet classifier

In the first part of the experiment, we used the original AlexNet classifier experiment. We randomly selected 150 car pictures and classified them into five categories: 0, 1, 2, 3 and 4, corresponding to cars, trucks, buses, vans and motorcycles respectively. And 30 of them were inverted to test the pretreatment effect. Table 1 shows the classification results using the original AlexNet. It can be seen that the experimental results are not very good, with the correct classification reaching 85% (total accuracy).

| Auto classification | Positive Rate | False Rate | Accurate rate |
|---------------------|---------------|------------|---------------|
| The car             | 82.3          | 1.43       | 79.1          |
| The truck           | 89.4          | 0.55       | 90.6          |
| The coach           | 87.6          | 0.64       | 88.3          |
| The van             | 81.6          | 3.00       | 78.4          |
| The motorcycle      | 92.3          | 0.36       | 93.7          |

4.2. ProAlexNet classifier

The same experiment was carried out using the PROALEXNET classifier. As can be seen from Table 2, the classification accuracy of the PROALEXNET classifier reached 96%, which was significantly better than the traditional ALALNET network structure.

| Auto classification | Positive Rate | False Rate | Accurate rate |
|---------------------|---------------|------------|---------------|
| The car             | 92.4          | 0.30       | 93.1          |
| The truck           | 97.4          | 0.10       | 98.6          |
| The coach           | 96.8          | 0.26       | 97.3          |
| The van             | 93.1          | 0.33       | 94.4          |
| The motorcycle      | 98.3          | 0.16       | 98.7          |

By comparing the two tables, it can be seen more intuitively that the result of ProAlexNet classification in the case of small samples is more excellent than that of traditional AlexNet classification.

On the basis of the above experiments, we conducted a comparative experiment on the original test set between ProAlexNet and other traditional recognition algorithms. The experimental results are shown in Table 3. Positive Rate is used as the experimental evaluation index.

| Recognition algorithm | The car | The truck | The coach | The van | The motorcycle | The average |
|-----------------------|---------|-----------|-----------|---------|----------------|-------------|
| ResNet                | 79.5    | 86.4      | 88.3      | 86.3    | 93.2           | 86.7        |
| SIFT                  | 73.2    | 77.5      | 75.2      | 69.4    | 78.9           | 74.8        |
| AlexNet               | 82.3    | 89.4      | 87.6      | 81.6    | 92.3           | 86.5        |
| ProAlexNet            | 92.4    | 97.4      | 96.8      | 93.1    | 98.3           | 95.6        |

Accurate recognition rate of SIFT algorithm of minimum, resulting in low recognition rate of the
main reason is: SIFT algorithm by comparing the image also is the key point in the local characteristics to detect the similarity of two images, for a change in bright and has good noise robustness, but the different kinds of vehicles with similar local characteristics, is difficult to accurately identify lead to SIFT algorithm. The accuracy of our proposed algorithm reaches 95.6%, which is 10.1% higher than that of traditional AlexNet. And because the number of convolution cores in ProAlexNet convolution layer is only half that of AlexNet, it takes less time in training and recognition than AlexNet.

5. Conclusion
In this study, we use image classification to deal with the problem of vehicle classification at the intersection. Based on the traditional ALEXNET, we propose an improved intersection vehicle classification and recognition algorithm, PROALEXNET. In order to verify the effectiveness of PROALEXNET, we select the data set of 5000 car pictures to verify it. In three performance indexes compared with the traditional AlexNet, the experiment shows that ProAlexNet is more competitive. In addition, compared with other recognition algorithms on the index of Positive Rate, it also shows that Positive Rate is more competitive. In the future work, we will further improve the performance of Positive Rate.

Reference
[1] Jie Yang; Fan Liu. Modulation Recognition using Wavelet Transform based on AlexNet. 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)
[2] Shuguang Sun; Tingting Zhang; Qin Li; Jingqin Wang; Wei Zhang; Zhitao Wen; Yao Tang. Fault Diagnosis of Conventional Circuit Breaker Contact System Based on Time–Frequency Analysis and Improved AlexNet. IEEE Transactions on Instrumentation and Measurement
[3] Maoshuang Feng; Dongfang Chen; Xiaofeng Wang. Classification of Fabric Patterns Image Based on Improved Log-AlexNet. 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)
[4] Ebrahim Farahbakhsh; Ehsan Kozegar; Mohsen Soryani. Improving persian digit recognition by combining data augmentation and AlexNet. 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP)
[5] Xu Zhang; Wei Pan; Perry Xiao. In-Vivo Skin Capacitive Image Classification Using AlexNet Convolution Neural Network. 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)
[6] Yang Lu. Image Classification Algorithm Based on Improved AlexNet in Cloud Computing Environment. 2020 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI)
[7] Qing Li; Libiao Jin; Siqi Liu. Multi-task Digital Verification Code Recognition Based on Improved ALEXNet. 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)
[8] Holger R. Roth; Christopher T. Lee; Hoo-Chang Shin; Ari Seff; Lauren Kim; Jianhua Yao; Le Lu; Ronald M. Summers. Anatomy-specific classification of medical images using deep convolutional nets. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)
[9] Ankit Titoriya; Shelly Sachdeva. Breast Cancer Histopathology Image Classification using AlexNet. 2019 4th International Conference on Information Systems and Computer Networks (ISCON)
[10] Yanpeng Guo; Zhenjiang Pang; Jun Du; Fan Jiang; Qilong Hu. An Improved AlexNet for Power Edge Transmission Line Anomaly Detection. IEEE Access
[11] Fangliang Huang; Lei Yu; Tongping Shen; Li Jin. Chinese Herbal Medicine Leaves Classification Based on Improved AlexNet Convolutional Neural Network. 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)
[12] Weiyong Gong; Xinman Zhang; Bohua Deng; Xuebin Xu. Palmprint Recognition Based on Convolutional Neural Network-Alexnet. 2019 Federated Conference on Computer Science and Information Systems (FedCSIS)