Research on Vehicle Recognition Algorithm based on Convolution Neural Network

Pai Zhang¹*, Hanqing Chen², Qinrui Li³

¹School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, 150000
²School of Mathematics&Statistic, Anhui Normal University, Wuhu, Anhui, 241000
³School of Computer Engineering and Science, Shanghai University, Shanghai, 200444

*Corresponding author: pai_zhang@hrbust.edu.cn

Abstract. In order to effectively identify vehicle types in intelligent transportation system, based on the analysis of Inception V3 model, a deep learning model of vehicle classification based on transfer learning theory is proposed in this paper. In this model, the last full connection layer is removed on the basis of Inception V3 model, and the parameter optimization layer is added, and then Dropout and global average pooling layer are adopted. Theoretical analysis and experimental results show that the performance of this model is better than that of VGG-16-based vehicle classification model, Xception-based vehicle classification model and Resnet-50-based vehicle classification model. The experimental results show that the training accuracy of the method proposed in this paper is more than 96.48% and the test accuracy is more than 83.86%.

Keywords: Intelligent transportation, Vehicle recognition, Deep learning model, Transfer learning

1. Introduction

Vehicle type information has become the basis of data processing and analysis in public transport services, traffic operation supervision, safety protection and other industries, therefore, the research of image-based vehicle type recognition method has become an urgent problem to be solved in the vehicle management and maintenance of intelligent transportation system. Professor Geoffrey Hinton proposed an improved method of model training to break the bottleneck that the number of neural network layers can not be too many. He puts forward two viewpoints: one is that the multi-layer neural network has stronger feature learning ability and can get the deep features which are more favorable for classification; the other is that the training problem of the deep neural network can be solved through layer-by-layer training. After the development of single-layer perceptron, multi-layer perceptron, BP neural network and deep neural network, deep learning is born. Deep learning has become a hot spot of artificial intelligence, and has made remarkable achievements in the fields of image detection, image classification and natural language processing. In 2015, Rong et al proposed to use automatic sparse encoders to generate convolution kernels [1], use the convolution kernels to...
generate convolution features, and then carry out pooling operations, and repeat this step to get a deep network framework to achieve the purpose of vehicle classification. Dong et al. proposed a semi-supervised convolution neural network model based on the front image of the vehicle [2]. The sparse Laplace filter is introduced to learn the unlabeled data, and only a small part of the data is used to train the Soft-max function of the classification layer. In 2017, Wang et al used deep transfer learning to classify vehicle images [3] and established a convolution neural network model that can be used for both network images and monitoring images. In 2018, Chen et al proposed a convolution neural network row classification model based on the tail image [4], which normalized the vehicle tail image to $32 \times 32$ and sent it to the neural network for vehicle classification.

In order to effectively improve the accuracy of vehicle type recognition, a deep learning network model for vehicle type classification is constructed based on transfer learning theory.

2. Convolution neural network (CNN)

Convolution neural network (CNN) is a neural network specially used to deal with grid structure data. It has excellent performance in processing typical network structure data such as image data, time series data and so on. Typical convolution neural networks include convolution layer, pooling layer and full connection layer. As the name implies, the convolution layer is the convolution operation of the network of this layer. In the term definition of convolution network, the first parameter of convolution is usually called input, the second parameter is usually called kernel function, and the output is also called feature mapping. From the definition of convolution layer, we can see that the operation process of convolution layer is a process of automatic feature extraction, so convolution neural network is not only a feature extractor but also a classifier compared with the traditional fully connected neural network. Can learn features independently without the need for manual pre-extraction. Convolution neural network uses two important ideas to improve the traditional neural network algorithm: sparse interaction and parameter sharing. The structure of the common convolution neural network is shown in figure 1.

![Structure Diagram of convolution Neural Network](image)

Figure 1. Structure Diagram of convolution Neural Network

3. Deep Learning Network Model of vehicle Classification based on transfer Learning Theory

3.1. Transfer learning theory

Transfer learning is to apply the models and parameters learned in one scene to another scene, which should belong to similar areas and have similar tasks. Because of the high cost of learning the new scene directly, transfer learning is used to find the similarity between the new scene and the old scene, which is also the core of transfer learning. The existing knowledge is called the source domain, and the new knowledge to be learned is called the target domain. Transfer learning is to transfer the knowledge from the source domain to the target domain, and there is usually a certain association between the source domain and the target domain. Under the condition of the change of data
distribution, feature dimension and model output, the target domain is better modeled based on the knowledge in the source domain, and it is not necessary to model from scratch. In the case of lack of calibration data, this learning method can make good use of the calibrated data in related fields to complete the calibration of data. When there are few data samples, it can be retrained on the basis of relevant domain models.

The classical convolution neural network usually uses convolution operation in the lower layer, and then uses the full connection layer for classification, which can be simply understood that the convolution layer is extracting image features, and the full connection layer is the traditional neural network classifier. However, the large number of parameters in the full connection layer leads to two bottlenecks: first, the configuration pressure of the training machine is large, and it takes a long time, which reduces the training efficiency; second, it is easy to lead to over-fitting, and the generalization ability of the network is not satisfactory. Therefore, this paper adopts the parameter optimization strategy while migrating and learning the existing mature network, in order to reduce the number of parameters and prevent over-fitting. This paper adopts two strategies: Dropout and global average pooling. Dropout means that in the deep neural network training, the hidden nodes in the network are discarded temporarily according to a certain probability, and the discarded nodes can be considered not to belong to the network structure temporarily. Theoretically, Dropout can be regarded as a kind of model averaging.

3.2. Inception V3 vehicle Classification Model based on transfer Learning Theory

The biggest difference between the Inception network and the traditional convolution neural network is that the traditional convolution neural network, such as VGG, stack the convolution network layer upon layer, which is a typical vertical structure, while the Inception network not only vertically stacks the convolution layer, but also proposes the Inception layer, while the Inception unit is formed by the superposition of convolution kernels of different sizes. This superposition process not only increases the "width" of the network, but also increases the adaptability of the network to the scale [6]. The typical Inception unit is shown in figure 2. Taking Inception An as an example, the structure of this unit is based on 1×1 convolution kernel, 3×3 convolution kernel, 5×5 convolution kernel and pool kernel, superimposed vertically and horizontally, and the current scale characteristics are obtained directly by using 1×1 convolution kernel. After using 1×1 convolution kernel, the shallow convolution feature can be obtained by 5×5 convolution calculation. The deep convolution detail characteristics can be obtained by using 1×1 convolution kernel and 2 times 3×3 convolution calculation, and the characteristics after downsampling can be obtained by 1×1 convolution calculation after pooling. In a word, after the calculation of Inception unit, the characteristic maps of different scales can be output. The low-layer structure of Inception V3 network is still a classical convolution layer stack, which adopts the structure of convolution layer-pool layer-convolution layer-convolution layer-pool layer. After pooling, three Inception A units are added, one Inception B unit is connected, and four Inception C units are connected. Connect 1 Inception D unit and 2 Inception E units, then adopt average pooling treatment and Dropout processing, and then use full connection layer for Softmax classification.
The vehicle classification model based on Inception V3 proposed in this paper is to remove the last full connection layer on the basis of Inception V3, and add the parameter optimization layer, using Dropout and global average pool layer. The specific structure is shown in Table 1. The input of the whole network model is 299×299RGB three-channel image, each block represents a group of calculations, the first two groups are convolution calculations, the first group of convolution calculations contains several convolution layers and a pool layer, and the convolution core is 3×3. After completing the two Block convolution calculations, it is connected to the Inception unit. Three Inception A units are used in Block 3, one Inception B unit is used in Block 4, four Inception C units are used in Block 5, one Inception D unit is used in Block 6, and two Inception E units are used in Block 7. After that, it is connected to the global average pooling layer, and then uses the Dropout layer with a discard rate of 0.5. Finally, it uses the Softmax function to connect with the output node.

4. Test and analysis
This experiment is based on the vehicle type image database of Southeast University, including 9850 vehicle images. The vehicle images are divided into six types: passenger cars, minibuses, minivans, cars, SUV and trucks, as shown in figure 3. The algorithm models used in this experiment are all based on Python3.6 and Keras deep learning framework (TensorFlow as the back end) in the Ju-upyter Notebook (Ipython) environment, and the computer CPU is configured as Intel Core i7-6700HQ 2.6 GHz and 8GB. ROM is 1TB. The video card is configured with NVIDIA GeForce GTX960M, video memory as 4G. In the experiment, 60% of the vehicle type image database of Southeast University is used as the training image, 20% as the verification image, and the remaining 20% as the test image.

**Figure 2.** Inception unit structure
Figure 3. Image examples of vehicle types in Southeast University

Figure 4. Resnet50-VCM Optimizer=Adam training process

Figure 5. Resnet50-VCM Optimizer=RMSprop training process
Figure 4 shows the training process of Resnet50-VCM using Adam optimizer, figure 5 shows the training process of Resnet50-VCM using RMSprop optimizer, train-loss graph in the upper left corner represents training loss with the number of iterations, and validation-loss graph in the lower left corner represents the change of verification loss with the number of iterations. The train-accuracy diagram in the upper right corner represents the change of the training accuracy with the number of iterations, and the validation-accuracy diagram in the lower right corner represents the change of the verification accuracy with the number of iterations. It can be seen that whether it is the Adam optimizer or the RMSprop optimizer, the model tends to converge after 8 iterations of epoch, and the training accuracy and verification accuracy of the model convergence are basically the same, but the accuracy of the RMSprop optimizer changes slowly with the increase of the number of iterations, and the verification loss decreases obviously. On the contrary, the training loss and verification loss of Adam optimizer decreased slowly after two times of epoch, and both verification accuracy and training accuracy showed an upward trend.

As can be seen from figure 6, the accuracy of VGG16-VCM is the lowest among the four models, and the test accuracy is less than 0.7, which shows that VGG16-VCM can not fit the vehicle classification problem well, which is consistent with the conclusion of previous loss analysis. In addition to VGG16-VCM, Resnet50-VCM, Inception-VCM and Xception-VCM all get better training accuracy, and the Inception-VCM, test accuracy is 83.85%, which is the best among the four models. After training, five vehicle classification models are obtained, and the test set is tested. In order to test the stability and robustness of the model, the test set and verification set are used as the test sample database, and 100 cross-validation tests are carried out. In each experiment, 1 800 images are randomly selected from the test sample database, and the accuracy of the model is calculated. The distribution of the precision box diagram of the model is shown in figure 9 (b). As can be seen from the figure, the accuracy of InceptionV3-VCM model is high. Inception V3-VCM and Table 2 accuracy comparison of different optimizer models optimizer=RMSprop optimizer=Adam train-acc val-acc VGG16-VCM 0.703 1 0.7396 0.6302 0.724 Resnet50-VCM 0.983 3 0.826 6 0.957 7 0.788 4 InceptionV3-VCM 0.977 9 0.822 9 0.964 8 0.838 6 Xception-VCM 0.961 8 0.777 8 0.916 7 0.791 7 Fig. 9 Accuracy of the model is almost the same as that of Resnet50-VCM. The accuracy of Inception V3-VCM and Xception-VCM models are distributed in a relatively small range, indicating that the model has good stability. VGG16-VCM and Resnet50-VCM have several outliers, indicating that the stability of the model is not good. Comprehensive comparison shows that Inception V3-VCM is better than the other three models.

5. Conclusion

In order to effectively identify vehicle types for intelligent transportation system, based on the analysis of Inception V3 model, a deep learning model of vehicle classification based on transfer learning theory is proposed and compared with the vehicle classification model based on VGG-16, Xception...
and Resnet50. Experimental research is carried out based on the vehicle type image database of Southeast University. Theoretical analysis and experimental results show that the performance of the vehicle classification deep learning model based on transfer learning theory is better than that based on VGG-16, Xception and Resnet50.

References
[1] RONG H L, XIA YX. A Vehicle Type Recognition Method Based on Sparse Auto Encoder [J]. Advances in Computer Science Research, 2015, 18 (6): 323-326
[2] DONG Z, WU Y, PEI M, Et al. Vehicle Type Classification Using a Semi-supervised Convolutional Neural Network [J]. IEEE Transactions on Intellig. Ent Transportation Systems, 2015, 16 (4): 2247-2256
[3] WANG J T, ZHENG H, HUANG Y, Et al. Vehicle Type Recognition in Surveillance Images from Labeled Web-Nature Data Using Deep Transfer Learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19 (9): 2913-2922
[4] CHEN Y J, ZHU W X, YAO D H, Et al. Vehicle Classification Based on Convolutional Neural Network [C] 2017 Chinese Automation Congress. Jinan: IEEE, 2017: 1898-1901
[5] IN M, CHEN Q, YAN S C. Network In Network [C] 100% International Conference on Learning Representations. Banff: ICLR, 2013: 1-10
[6] SZEGEDY C, VANHOUCKE V, IOFFE S, Et al. Rethinking the Inception Architecture for Computer Vision [C] IEEE Conference on Computer Science and Pattern Recognition. Tion. Las Vegas: IEEE, 2016: 2818-2826