Vehicle Attribute Recognition Algorithm Based on Multi-task Learning

Jingying Sun
School of Computer and Communication Engineering
University of Science and Technology Beijing
Beijing, China
913919621@qq.com

Chengzhe Jia
College of Mechanical Engineering and Applied Electronics Technology
Beijing University of Technology
Beijing, China
15601325420@163.com

Zhiguo Shi
School of Computer and Communication Engineering
University of Science and Technology Beijing
Beijing, China
szg@ustb.edu.cn

***Abstract*—Vehicle attribute recognition in urban traffic monitoring is the core task in urban intelligent transportation system, Vehicle attribute recognition mainly includes sub-tasks such as vehicle type identification, vehicle color recognition, and vehicle brand recognition. Most of the current solutions are single-task learning based on a single attribute, and there are few studies on complex learning tasks with multiple attributes. This paper constructs a vehicle multi-attribute data set for the multi-attribute characteristics of vehicles, and based on the training mode of multitasking learning, Separate vehicle brand recognition network and vehicle color recognition network that are more suitable for their respective characteristics, and integrate vehicle multi-attribute identification network into the same model structure for training. By comparing the current popular neural network with several attributes, the final experiment shows that the vehicle multi-attribute recognition model trained by the algorithm can obtain better recognition results and higher accuracy.**

***Keywords*—multi-task learning; convolutional neural network ; vehicle multi-attribute identification**

I. Introduction

The informatization and intelligence of traffic manageme nt are the key project established by the National Public Security Traffic Management Department and incorporated into the long-term strategic planning. During the past ten years, the average annual growth rate of motor vehicles and drivers has exceeded 10%, and the problem between transportation supply and demand has become increasingly serious. With the increasing number of cars, various types of vehicles travel through the streets of the city, followed by car-related secur ity incidents and traffic events. Facing a huge amount of vehi cle characteristics, it could be pretty hard for the traditional tr affic system to process. For various vehicle features, just rely ing on traditional manual identification and processing, not o nly wastes a lot of manpower and material resources, but also prone to errors. Therefore, vehicle feature attribute recogniti on is crucial for Intelligent Transportation Systems(ITS)[1], a mong many studies, vehicle multi-attribute recognition based on computer vision is an important part of vehicle feature rec ognition system research.

Many researchers have proposed many related recognition algorithms for vehicle attribute recognition. The traditional vehicle attribute recognition method mainly extracts manual image features (such as SURF[2], LBP, HOG[3], etc.), using texture, edge, contour and other features, through the classification algorithm (such as SVM, Adaboost, random forest, etc.) training model to achieve vehicle attribute recognition, Zhang et al. [4] used the manual extraction of vehicle image Haar features to construct a global appearance model, and then trained by SVM classifier. Krause et al. [5] used vehicle frame points for vehicle type identification. However, the traditional manual extraction features are shallow features. In the problem of vehicle attribute recognition, there are subtle differences between different models, and the artificially designed feature extraction method has limited ability to deal with such strong identification information problems, and it is generally impossible to accurately distinguish such nuances, and the traditional feature extraction process has many problems such as large amount of calculation, low efficiency, and poor adaptability. Therefore, it cannot be widely applied in the identification of fine-grained feature attributes.

In recent years, due to the improvement of hardware computing ability and algorithm, deep learning [6] has achieved great success in both speech processing and computer vision [7]. Among them, a large number of methods based on convolutional neural networks are used for target detection and image classification tasks, significantly improving the performance of many intelligent processing tasks. Image classification based on deep convolutional neural network can perform complex nonlinear modeling of the original input signal to the desired output information through hierarchical connection, thereby obtaining a more abstract and more essential representation of the data, realizing an end-to-end learning, with greater flexibility and universality. A more essential representation, an end-to-end learning, with greater flexibility and universality. Classic deep convolutional neural network models include LeNet, AlexNet [8], GoogLeNet [9], VGGNet, and depth residuals. These models have achieved good results in various data sets such as MNIST, CoCo [10], and ImageNet [11]. Hu C et al. [12] used the method of deep convolutional neural network to extract the vehicle color features, and combined the feature context to identify the vehicle color, and achieved a higher recognition accuracy, Huang et al. [13] used CNN to extract landmark features for identification, These methods avoid the complex feature extraction process of the traditional method and achieve better recognition accuracy.

There are a lot of researches in the field of vehicle attribute recognition, most learning tasks are single-task learning based on a single attribute, There are few studies on complex learning tasks with multiple attributes. For vehicles, it also has multiple attribute descriptions. For example, according to the vehicle brand, it can be described as: Mercedes-Benz, BMW, Volkswagen, etc; according to the vehicle color, it can be described as: red, white, black, etc. this paper aims at the multi-attribute characteristics of vehicles, we constructs vehicle multi-attribute data set by searching data through network, and constructs a vehicle brand recognition network and

978-1-7281-3488-8/19/$31.00 ©2019 IEEE
DOI 10.1109/SmartIoT.2019.00029

vehicle color recognition network to adapt to their respective characteristics, and based on the soft parameter sharing mode of multi-task learning [14,15], the vehicle multiple attribute recognition tasks are put together for training, multi-attribute recognition of the vehicle is carried out, and the generalization ability of the model is improved.

II. Related Work

A. Deep Learning Model Based on CNN

The Convolutional Neural Network (CNN) [16] is an end-to-end system that has achieved outstanding results in the fields of image recognition and speech recognition. The convolutional neural network uses the original image as input, alternates through the convolutional layer and the pooling layer, and finally passes through the fully connected layer, Train the network model by learning the loss of the task, using the back propagation algorithm, updating the network weights and biases. The convolution layer is used to convolve the feature map of the previous layer with multiple convolution kernels, extract features and generate output feature maps. The pooling layer generally follows the convolutional layer and performs sampling operations on the feature map output of the previous layer. The main purpose of the pooling operation is to reduce the resolution of the feature map and reduce the feature dimension. After multiple sets of convolution-pooling structures, several fully connected layers are connected as linear classifiers at the end of the neural network. Generally, the neural network selects softmax loss as the loss function of the network.

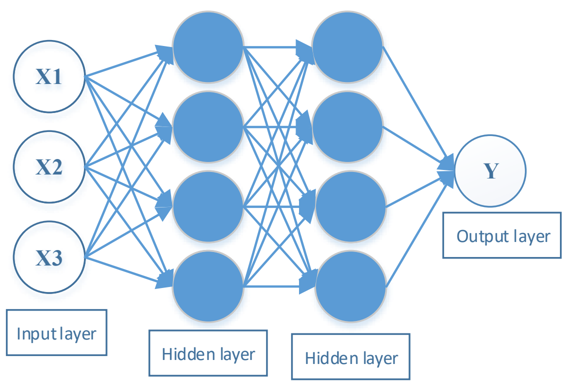

Fig. 1. CNN infrastructure

In recent years, with the rapid development of deep learning, convolutional neural networks have a place in the field of computer vision with extremely high performance. CNN's success on ImageNet stems from three main factors: One is large-scale training data. The second is a more complex model, the network structure is more complex, deeper, and more parameters; the third is the GPU's acceleration of the calculation, so that the training process that used to take several weeks can be completed in one day or even several hours. Because CNN has many parameters, it must rely on large-scale training data to prevent overfitting. In the case of a small amount of data, there are two solutions: one is called Data Augmentation, which relies on existing images to generate more images through changes in rotation, translation, and deformation. The second is to use Transfer Learning [17]. The idea is to train the parameters of the CNN as an initial value by training on another large-scale data set, and then training the parameters to be fine-tuned on the target data set.

The principle of transfer learning is that certain features are versatile on different training data sets. For CNN, the first layer is to extract local features, and the subsequent layers expand the sensing area by down-sampling, and then the layer sensing area is larger, and the obtained features are more abstract. Features in the first few layers are usually not directly related to a specific classification task, but are similar to Gabor Filter, edges, and direction-related features. These features are relatively generic, so they can be trained on a data set and applied to a similar data set.

In the field of computer vision, the "level" of features becomes higher as the depth of the network increases. Therefore, the depth of the network greatly affects the effect of image recognition and classification. However, the deep network of training often has problems of gradient dispersion and explosion, which leads to the inability to converge [18]. By using the methods of normalized initialization [19] and input normalization of each layer can increase the depth of the convergence network, but there is a phenomenon of network degradation, that is, increasing the number of network layers leads to greater errors. In response to the above problems, He et al. proposed the Deep Residual Network (DRN) [20] structure, the main idea is to add a jump around some layers of the connection on the standard feedforward convolution network. The residual network reconstructs the learning process and re-directs the information flow in the deep neural network, which well resolves the contradiction between the level and accuracy of the deep neural network.

(1)Traditional convolution: Use the output of the first layer as the input of the next layer.

$$\mathrm{x}_l = H_l(\mathrm{x}_{l-1}) \tag{1}$$

(2)ResNet: Here l represents the layer, and x_l represents the output of the l layer, indicating a nonlinear transformation. So for ResNet, the output of layer l is the output of $l-1$ layer plus the nonlinear transformation of the output of $l-1$ layer.

$$\mathrm{x}_l = H_l(\mathrm{x}_{l-1}) + \mathrm{x}_{l-1} \tag{2}$$

(3)DenseNet: $[\mathrm{x}_0,\ \mathrm{x}_1, \cdots, \mathrm{x}_{l\text{-}1}]$ means that the output feature of layer 0 to layer $l-1$ is concatenation the merger of the channels. ResNet is the sum of the values, and the number of channels is constant. H_l including convolution of BN, ReLU, and 3*3 convolution.

$$\mathrm{x}_l = H_l([\mathrm{x}_0, \mathrm{x}_1, \ldots, \mathrm{x}_{l-1}]) \tag{3}$$

B. Neural Network Model Based on Multi-task Learning

The traditional deep learning algorithm theory is mostly based on the single task learning mode [21]. For multiple learning task problems, the multitasking problem is generally disassembled into multiple single task problems, and then each task is studied separately. Multi-task learning [22,23] can learn the shared representation of multiple tasks. The learned shared representation has strong abstract representation ability, can adapt to multiple different but related goals, and make the main task get better generalization ability. Especially in the case of fewer samples, multitasking is significantly better than single task learning. Multitasking can make better use of limited data, and training time is relatively short. Deep neural networks can learn multi-layered abstract representations of data, and have a stronger ability to interpret the intrinsic feature relationships and task relationships between multitasking. Deep neural networks can benefit from related tasks by sharing parameters with other networks.

Two methods commonly used in multi-task learning based on deep neural networks: hard sharing and soft sharing of hidden layer parameters. (1) Hard share mechanism of parameters: The hard share mechanism of parameters is the most common way of multi-task learning in neural networks, In general, it can be applied to all hidden layers of all tasks while retaining the task-related output layer. The advantage of hard parameter sharing is that the structure design is simple, only the structure of the hidden layer is designed, and the network parameters are shared among different tasks without accurately modeling the task relationships in these layers. The disadvantage is that most tasks require high correlation, but the actual application scenarios are limited, and the feature relationships of many tasks are subject to mandatory constraints and are discarded, if this hard structure does not work in the association of deep networks, then it is likely that a negative migration will occur.

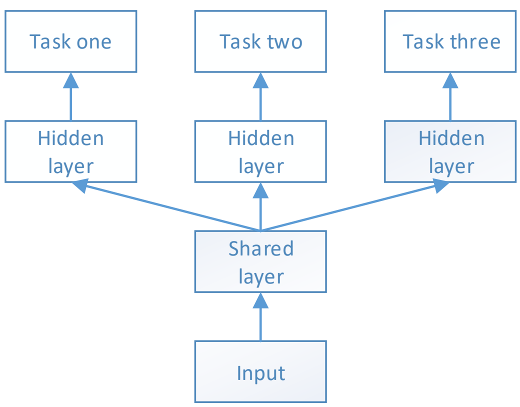

Fig. 2. Hard parameter sharing of deep neural network multitasking learning

(2) Soft sharing mechanism for parameters: In many scenarios, hard parameter structures often rely on artificially predefined shared structures, However, the internal links between many tasks are not very close. The increase in the number of tasks increases the diversity between tasks. It is difficult to include all the unique models on the task. Therefore, many scholars have proposed a soft parameter sharing neural network that does not specify a shared structure. Each task has its own model and its own parameters, The deep neural network structure contains a large number of hidden layers. It is difficult to divide the corresponding shared layers between different tasks. Unlike the hard parameter sharing structure, the soft shared structure retains the structure of the respective task networks, and all layers are bound to the same. The parameters are adjusted to assume that all the feature layers exist in a shared space, and the distance of the model parameters is constrained by the regularization method to ensure that the tasks are similar rather than identical. Similar to the unified feature structure under regularization constraints in multi-task discriminant models. The soft shared structure removes the restriction that information is not exchanged between certain layers in the hard parameter sharing, and the sharing strategy is completely implemented in a data-driven manner, which is a network with parameter constraints. The constraint of the soft sharing mechanism used in deep neural networks is largely influenced by the regularization technique in traditional multi-task learning.

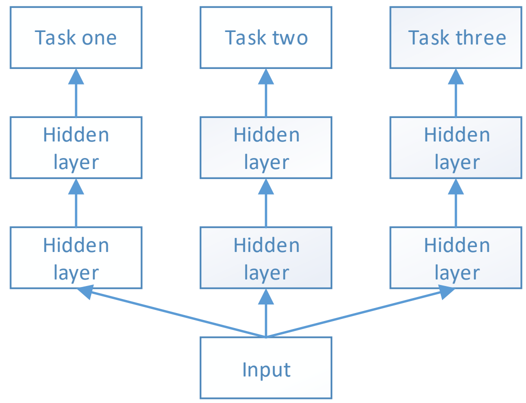

Fig. 3. Soft parameter sharing of deep neural network multitasking learning

III. VEHICLE ATTRIBUTE RECOGNITION ALGORITHM BASED ON MULTI-TASK LEARNING

A. Overall Framework

As the important attribute of the vehicle, the vehicle brand and color can describe the basic characteristics of the car. Firstly, we analyzes the difficulties and problems in the process of solving each task. The vehicle color recognition task mainly has the following difficulties:

1)The surface of the vehicle is easily affected by light, which affects the color recognition task.

2)The color of the vehicle's surface is unevenly distributed, and multiple colors may exist in the same vehicle.

The vehicle brand recognition task mainly has the following difficulties:

1) There are small differences between sub-categories of vehicle brands, and there are large gaps within the class and small gaps between classes.

2) There are a large number of vehicle brand categories, and relevant data is difficult to collect .

Based on the classification and identification of the various attributes of the vehicle, The main idea of this paper is to use the multi-task learning mode to train the convolutional neural network and use the network to complete the vehicle brand and color recognition task. First, By analyzing the problems of vehicle brand and vehicle color classification and recognition tasks, we design CNN models that adapt to their respective characteristics, and integrate the two structures into the same network. The network is trained through a multitasking learning framework, and supports both the vehicle brand and the vehicle color training at the same time. Update the parameters by simultaneously propagating the network through the supervision information of the vehicle brand and the vehicle color, and the final learning features include both brand and color information, making it possible to use one model for both purposes. The overall framework of the multitasking learning model is shown in Figure 4.

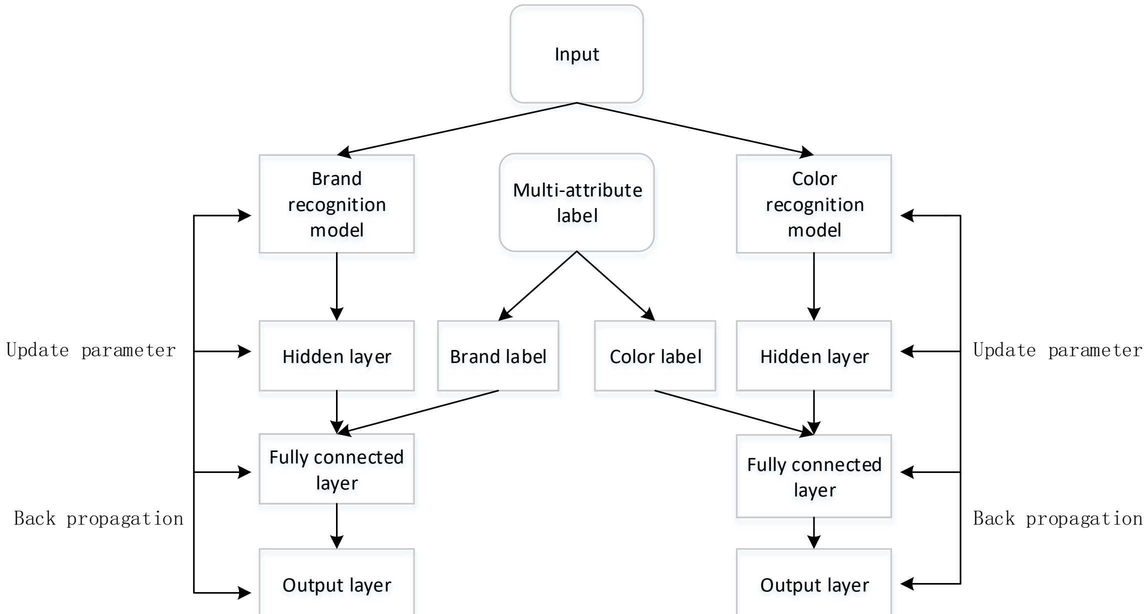

Fig. 4. The overall framework

B. *Network Model Construction*

For multitasking learning problems, each task has its own goals, and the final goal of the network is to maximize the accuracy of all task learning. For this paper, the task of learning is the identification of multiple attributes of the vehicle, In the upper half of the multitasking neural network structure in Figure 4, Defining a vehicle brand recognition network structure. In order to effectively deal with the problems in the vehicle brand recognition process, the network infrastructure uses the deep neural network DenseNet, and the DenseNet overall structure mainly contains Dense Block and Transition Layers. The Dense Block mainly implements dense connection of the dense block by using a loop. Each of the dense blocks contains a 1*1 convolution and a 3*3 convolution. The 1*1 convolution is to reduce the number of input feature map, reduce the computational amount by dimension reduction, and fusing the characteristics of each channel, in order to maintain the same size of the feature map inside the Dense Block, The input of each layer uses the Concatenate operation on the feature dimension when connecting across layers, rather than the Element-wise Addition operation on ResNet. Transition Layers are used to connect two Dense Blocks. there is no need to use a transition layer at the end of the last Dense Block. The transition layer consists of four parts: Batch Normalization, ReLU, 1*1 Conv, and 2*2 Max pooling. The Conv operation implements a 1*1 convolution operation and uses a compression rate to adjust the number of channels and add a pooling layer and two fully connected layers after the DenseNet structure, the pooling layer adopts the global average pooling, As another convolutional neural network with a deeper number of layers, DenseNet has fewer parameters than ResNet, and enhances the reuse of features. The network is easier to train and has certain regular effects, which alleviates the problem of gradient disappearance and model degradation. it is suitable for tasks such as vehicle brand recognition. The specific structure of the network is shown in Figure 5.

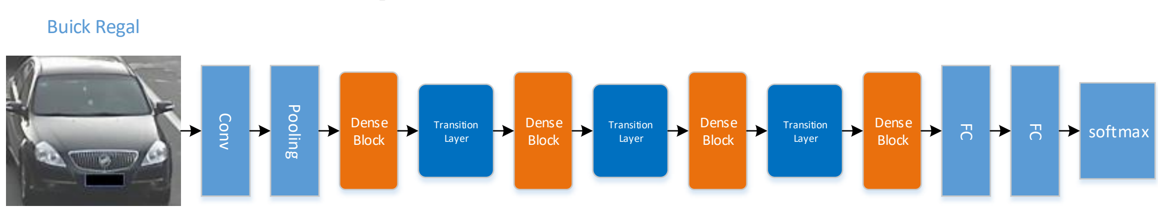

Fig. 5. Vehicle brand recognition model structure

In the other half of the multitasking neural network structure in Figure 4, Defining a vehicle color recognition network structure, mainly for the problems in the process of vehicle color recognition. The network structure adopts a bilinear CNN structure. the first two layers of each basic network

(Conv1+Conv2) is a translation layer, Convolution processing through normalization and pooling, the last process of the first two layers is pooling process, using max pooling with a size of 3*3, a step size of 2,The second layer (Conv2), the fourth layer (Conv4), and the fifth layer (Conv5) are divided into two groups, each group being independent of each other. The third layer (Conv3) and the fourth layer (Conv4) mainly perform convolution operations, and there is no pooling and normalization process. The fifth layer (Conv5) is a convolutional layer, only the pooling process. Before entering the fully connected layer, the pooling layer outputs of the fifth layer (Conv5) of the two basic networks are connected and dispersed into a long vector. The sixth layer (FC1), the seventh layer (FC2) and the eighth layer (FC3) are fully connected layers, and the sixth layer (FC1) and the seventh layer (FC2) adopt the Dropout regularization method to reduce overfitting, The last layer is the Softmax layer, and all layers (including the fully connected layer) use the ReLU activation function. The specific structure of the network is shown in Figure 6.

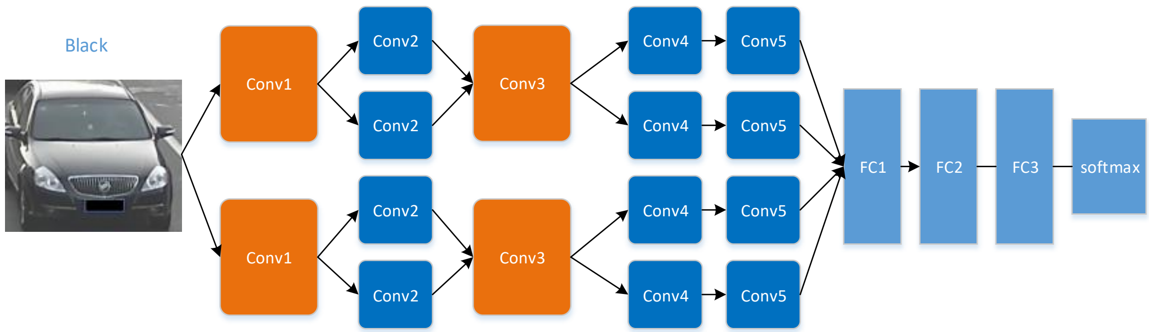

Fig. 6. Vehicle color recognition model structure

After defining the network structure for each task, Apply the multi-task learning mode to integrate the sub-task networks into an overall multi-tasking network framework. The overall network model framework is a soft parameter sharing mechanism in multi-task learning, Single input multiple output structure, The input is the vehicle image and the multi-attribute label. The two output structures are the vehicle brand and the vehicle color. The model loss function is divided into two parts. Vehicle brand recognition and vehicle color recognition are both cross entropy loss functions. The final optimization effect is to minimize L_{final_loss}, and the network model loss function is defined as follows:

$$L_{final\_loss} = \lambda_1 \cdot L_{color\_loss} + \lambda_2 \cdot L_{brand\_loss} \quad (4)$$

Among them, L_{final_loss} is the loss function of the network model, L_{color_loss} is the vehicle color part loss function, L_{brand_loss} is the vehicle brand part loss function, λ_1 and λ_2 are the two-part loss weights.

IV. Experiments

A. Environment Configuration

In order to evaluate the performance of the algorithm in vehicle multi-attribute identification, we trains on the public dataset PKU-VD Dataset, The experiments use the current popular deep learning tools, using TensorFlow as the experimental platform, TensorFlow is Google's open source deep learning framework, with rich applications in image classification, audio processing, recommendation systems, and natural language processing. TensorFlow is mainly used to build and train depth models. The experimental platform hardware configuration is processor Intel Core i7-6700K@4.00GHz*8, GeForce GTX 1080. The software uses the 64-bit version of Ubuntu 16.04.

B. Dataset

The experimental dataset used in this paper is from the PKU open source dataset. The PKU-VD dataset are constructed by National Engineering Laboratory for Video Technology (NELVT), Peking University, sponsored by the National Basic Research Program of China and Chinese National Natural Science Foundation. The dataset is based on two city real scene collections, built with VD1 and VD2,The image data in VD1 is obtained from a high-resolution traffic camera, and the image in VD2 is obtained from the surveillance video. The dataset is based on raw data and only contains one car per image. Each image in the dataset is provided with two different attribute annotations, including the vehicle brand and vehicle color. Because the PKU-VD dataset is relatively large, we first preprocesses the PKU-VD dataset. removes the identification number in the label, Screen out 201 precise vehicle branding labels and 11 common colors. we screened 500,000 images from the dataset as experimental data, of which 228,685 were used as training data and the rest were used as test and validation data. Table 1 shows the sample category attribute distribution of the PKU-VD dataset used in this experiment. Each vehicle image in the experimental data set contains two attributes. Figure 7 shows the partial vehicle dataset image and its multi-attribute label meaning.

TABLE I. PKU-VD DATASET SAMPLE ATTRIBUTE

Type	Class	Training	Testing/Validation
Brands	201	228685	225929
Color	11	228685	225929

Fig. 7. Vehicle data and multi-attribute labels

C. Model Training Process

1. Data processing: we first adjust the training samples according to the model input size requirements, Preprocessing data according to different identification features, such as image graying, denoising, etc. Data enhancement processing is performed on samples with insufficient sample numbers, such as random scaling, random clipping, random offset, random rotation, etc.

2. Training: The training process of the model is monitored by the callback function, and the result with the least loss on the test set is saved as the final weight file of the model. End training when the model training reaches the set training round or the test set loss lasts for 5 rounds without decreasing.

3. The two-part loss of the model training process is categorical_crossentropy loss function. Loss weights are set to 1.0 and 1.0 respectively, Use the Adam optimizer to set the base learning rate base_lr=0.001 and monitor the val_loss value, When the val_loss stops for 3 consecutive rounds, the learning rate is reduced by 0.5 times. The advantage is that the initial solution can be quickly approached with a large learning rate, and then gradually approach the optimal solution with a small learning rate to prevent the occurrence of oscillation. In order to make the model infinitely close to the optimal solution, select the training round epochs=10, Because the accuracy rate of the training rounds tends to be stable after 10 rounds of training, the basic learning rate is gradually reduced, So that the model gradually approaches the optimal solution at a slower speed. Model training loss, accuracy and iteration rounds are shown in Figure 8.

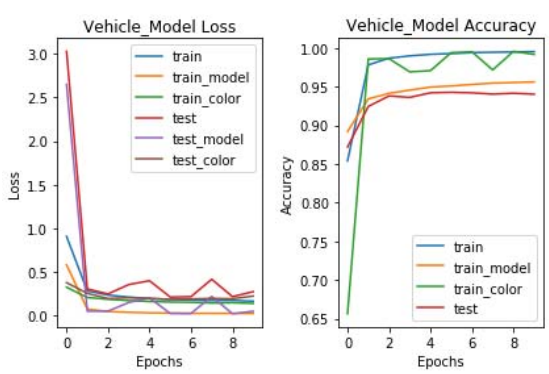

Fig. 8. Relationship between training loss, accuracy and iteration rounds

D. Analysis

In order to evaluate the performance of the vehicle multi-attribute identification model proposed in this paper, we judge the pros and cons of the model from the perspective of vehicle color recognition accuracy and vehicle brand recognition accuracy. We use the more popular neural networks InceptionV3 and ResNet50 in recent years for comparative analysis in this experimental environment. The results are shown in Table 2. Judging from the number of vehicle attribute recognition, the traditional deep learning models InceptionV3 and ResNet50 can only identify a single attribute of the vehicle at a time. The method used in this paper can identify both the brand and the color of the vehicle. From the perspective of vehicle brand and color recognition accuracy, ResNet50 is lower, InceptionV3 is second, and the method used in this article is the highest. In addition, we select the multi-attribute recognition algorithm MTL-CNN for comparative analysis. It can be seen from the verification results that the algorithm is slightly higher than the MTL algorithm in color recognition accuracy and vehicle recognition accuracy. the accuracy of vehicle brand recognition on the test set is 99.50%, and the vehicle color recognition rate is 94.27%. Therefore, we can see that the proposed algorithm model has great application prospects in vehicle multi-attribute identification.

TABLE II. COMPARISON OF RESULTS BY DIFFERENT METHODS

Methods		**Color recognition accuracy（%）**	**Brand recognition accuracy（%）**
Single attribute	InceptionV3	94.14	98.86
	ResNet50	94.12	97.03
Multiple attribute	MTL-CNN	94.21	99.40
	This paper	94.27	99.50

V. CONCLUSIONS

We proposes a vehicle multi-attribute identification network model based on multi-task learning framework. By constructing the vehicle brand recognition network and vehicle color recognition network, and introducing a soft parameter sharing mechanism based on multi-task learning, the two networks are merged into the same model structure to realize multiple attribute recognition of vehicles. Train the entire convolutional network end-to-end by using multiple attributes of the vehicle as an objective function of the network, The network model learns multiple attributes of the vehicle at the same time, which makes the output more rich, and can achieve good results with limited data support, and has greater expandability, the algorithm is directly efficient, and has great application value.

ACKNOWLEDGMENT

This paper is based on the project of urban intelligent traffic perception and decision-making system, and is supported by the School of Computer and Communication Engineering, University of Science and Technology Beijing.

REFERENCES

[1] B. Tian,B. Morris,M. Tang, “Hierarchical and networked vehicle surveillance in ITS,” IEEE Transactions on Intelligent Transportation Systems, vol. 16, pp. 557-580, April 2015.

[2] H. Bay,T. Tuytelaars,Van L. V. Gool, “Surf:Speeded up robust features,” European Conference on Computer Vision,vol. 3951, pp. 404-417, 2006.

[3] N. Dalal,B. Triggs, “Histograms of oriented gradients for human detection,” IEEE conference on Computer Vision and Pattern Recognition, vol. l, pp. 886-893, June 2005.

[4] J. Zhan,H. Zhang,X. Luo, “Fine-grained vehicle recognition via detection-classification-tracking in surveillance video,” International Conference on Digital Home. IEEE Computer Society,2014,pp. 14-19.

[5] J. Krause,M. Stark,J. Deng, “3D object representations for fine-grained categorization,” IEEE conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 554 -561.

[6] Y. Lecun,Y. Bengio,G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436-444, May 2015.

[7] R. Szeliski, “Computer vision:algorithms and applications, ” Springer Science&Business Media, 2010.

[8] A. Krizhevsky,I. Sutskever,G.E.Hmton, “Imagenet classification with deep convolutional neural networks, ” Advances in neural information processing systems,2012, pp.1097-1105.

[9] C. Szegedy,W. Liu,Y. Jia, “Going deeper with convolutions, ” IEEE conference on Computer Vision and Pattern Recognition,2015,pp.1-9.

[10] Y. Lin,M. Maire,S. Belongie, “Microsoft coco:Common objects in context, ” European Conference on Computer Vision, 2014, pp. 740-755.

[11] J. Deng,W. Dong, R. Socher, “Imagenet:A large scale hierarchical image database, ” IEEE conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255.

[12] C. Hu,X. Bai,L. Qi. “Vehicle color recognition with spatial pyramid deep learning, ” IEEE Transactions on Intelligent Transportation Systems, vol. 16, pp. 2925-2934, 2015

[13] Y. Huang,R. Wu,Y. Sun, “Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy, ” IEEE Transactions on Intelligent Transportation Systems, vol. 16, pp. 1951-1960,2015.

[14] R. Caruana, “Multitask learning, ” Machine Learning, vol. 28, pp.1-4, 1997.

[15] A. Abdulnabi,G. Wang,J. Lu, “Multi-task CNN models for attribute prediction, ” IEEE Transactions on Multimedia,vol. 17, pp. 1-1, 2015.

[16] G. Hinton,S. Osindero,Y. Teh, “A fast learning algorithm for deep belief nets, ” Neural Computation, vol. 18, pp. 1527-1554,2006.

[17] Pan S J, Q.Yang, “A Survey on Transfer Learning, ” IEEE Transactions on Knowledge &Data Engineering, vol. 22, pp. 1345-1359, 2010.

[18] N.Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting, ” Journal of Machine Learning Research, vol. 15 , pp: 1929-1958, June 2014.

[19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “ Caffe: Convolutional Architecture for Fast Feature Embedding, ” ACM International Conference on Multimedia 2014, pp. 675–678.

[20] Kaiming He, Xiangyu Zhang, “Deep residual learning for image recognition, ” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[21] A. Abdulnabi,G. Wang,J. Lu, “Multi-task CNN models for attribute prediction, ” IEEE Transactions on Multimedia,vol. 17, pp. 1-1, November 2015.

[22] Y. Wei,W. Xia,J. Huang, “CNN:Single-label to multilabel, ” arXiv preprint arXiv:1406.5726,2014.

[23] B. Bakker,T. Heskes, “Task clustering and gating for Bayesian multi-task learning, ” Journal of Machine Learning Research,vol. 4, pp. 83-99, January 2003.