Deep Convolution Neural Network Based Research on Recognition of Mine Vehicle Head and Tail

Junqiang Li [1], ChaoWang [1,2], Lin Cui [1], ZhiweiZhang [1], Wenquan Tian [1], Zhenggao Pan [1], Wanli Zhang [1], Xiaoying Yang [1] and Guolong Chen [3]

1. School of Informatics and Engineering, Suzhou University, Suzhou 234000, PR China.

2. Institute of Machine Learning and Systems Biology, School of Electronics and Information Engineering, Tongji University, ShangHai 201804, PR China.

3. School of Computer Science and Information Engineering, Bengbu University, Bengbu 233000, PR China..

E-mail:szxycw@126.com

Abstract. The mine environmental monitoring system captures the photos of the head and the tail of the vehicle, and sometimes the system can not accurately distinct whether it is the head or the tail of the vehicle. When there are two trucks in the view of the surveillance camera, the captured image contains the head of one truck and the tail of another truck. What needs to be recognized is the head license plate number or the tail license plate number. However, because the system cannot distinguish the head and tail of the truck, it will cause more false alarms. In order to solve this problem, this paper proposes an end-to-end feature extraction and recognition model based on deep convolution neural network (Deep CNN). The Deep CNN model contains five stage CNN layer and each layer contains different kernel size to extract the features. The data set is provided by Huaibei Siyuan Technology Co., Ltd., which includes normal capture, escape and false alarm images of the trucks. The final prediction rate is 85% on the testing set, which occupied twenty percent of the whole image set. The prediction rate of our model has been higher than the prediction rate base on right-out-left-in principle, which is used in the mine environmental monitoring system. Finally, our model will be applied in the mine environmental monitoring system.

Keywords: head and tail recognition, deep learning, convolution neural network, mine environment monitoring system.

1 Introduction

China is rich in mineral resources, many of which play an important role in the economy and provide important support for the sustainable and healthy development of local national economy. With the continuous expansion of the mining scale of mineral resources, the effective supervision of mine resources development has become an increasingly important issue. In recent years, the state has issued a series of laws, regulations and policies on the management of mineral resources [1], which has promoted the healthy development of mining industry, and also promoted the improvement of the level of grass-roots Mining Administration. However, it is difficult to put an end to violations of laws and regulations in the development of mineral resources [2]. There are still a series of problems, such as unlicensed mining, illegal mining, destructive mining, mining not according to the approved development and utilization plan, low resource utilization rate, concealed production, environmental pollution, safety and geological hazards. It is very difficult to rely on the personnel supervision, and the cost of human, material and financial resources is also very large, and the effect is not ideal. How to achieve accurate supervision has become an urgent problem to be solved.

In recent years, vehicle identification technology is widely used in mine supervision, traffic control and other fields, and has been widely valued and concerned by researchers [3]. Therefore, the fine identification of the license plate, head and tail of the transport vehicle and the shape of the minerals truck load in the mining area is not only of great significance for the effective management of mine resources, but also can avoid the illegal behavior of tax evasion [4]. In the mining environment, there are many kinds of vehicles in and out. In addition to transport vehicles, there are trucks, SUVs, buses and other types of vehicles. In addition, there are also waste resources such as ballast. Therefore, it is of great significance for the construction of mine comprehensive supervision platform to effectively identify the transport vehicles and the goods they trucker and eliminate the interference of other types of vehicles.

When the mine environmental supervision system captures the photos of the head and tail of the truck, it is impossible to give the impact of tax evasion caused by the problem of the head and tail of the truck distinction. The purpose of this paper is to solve the problem that the head and tail of the mine transportation vehicle could not be identified accurately. The comparison between the head and tail images is shown in Figure1.

Fig 1. Vehicles' head and tail.

The research method of mining vehicle head and tail recognition based on deep convolution neural network proposed in this paper has the following contributions: First, the head and tail pictures are sorted, and the data that is difficult for people to distinguish between the head and the tail is deleted. Second, for the head and the tail have different size, the model uses different convolution kernel size, Third, The structure of the convolution tail is changed from the 1000 classes to the 2 classes that is the truck's head or tail, and then the Softmax activation function is used, and finally the Adam optimizer is used to update the parameters.

The organizational structure of this paper is as follows: in the introduction part, it introduces the problems in mining management and the shortcomings of the previous solutions, and puts forward our research on vehicle head and tail recognition based on deep learning. The related work parts introduces the previous work of vehicle recognition, including vehicle license plate recognition, vehicle logo recognition and vehicle head and tail recognition, two patents and one article are introduced as the reference of our work. The third part method part, mainly introduces the principle of deep convolution neural network, the basic network framework, the network hierarchy and loss function settings used in this paper. In the experiment parts, it mainly introduces the database, the parameters set in this experiment and the comparison of the experimental results under the corresponding settings. The last part is the summary and future work. The final part is the part of acknowledgement, which mainly lists the items that support the work.

2 Related Work

The visual recognition of vehicles mainly includes license plate recognition, logo recognition, head and tail recognition, etc. License plate recognition is an early research area in vehicle recognition. The main processes of license plate recognition include image preprocessing, license plate location, character segmentation and character recognition. The preprocessing mainly includes graying, contrast enhancement, binarization, filtering, smoothing and correction of the vehicle image taken by the camera. The typical methods in license plate location include color feature based [6-8], edge feature based [9-11], morphological feature based [12,13], support vector machine based [14], clustering method, neural network based, genetic algorithm based, hybrid feature based. The problem of character segmentation in license plate location is that the representative methods are Unicom segmentation, projection segmentation, and static boundary method. The last step is character recognition. The common methods are template matching, support vector machine, neural network. License plate recognition is a relatively mature recognition technology internationally recognized that the recognition rate of all vehicle models is over 85-95%, and the recognition time is within 200 ms.

Vehicle logo recognition is a more refined requirement for license plate recognition. This paper proposes a fast identification method for vehicle signs, which is mainly based on the filtering processing based on the characteristics of energy concentration in vertical direction of the position of the vehicle sign, and the accurate position of the vehicle sign by template matching in the rough positioning rectangle frame. According to the characteristics of high energy and concentration of the vehicle sign in the vertical direction, a new method of positioning the vehicle sign with high speed and robustness is proposed. Firstly, image filtering is truckried out by energy enhancement and adaptive morphological filtering, and then the candidate areas of the vehicle mark are segmented by adaptive threshold, Then, the vehicle logo is accurately positioned according to the characteristics of the vehicle and its relationship with the vehicle. In this paper, a fast identification method of vehicle signs based on edge histogram is proposed. The main technology is to identify the head signs by using correlation method and edge histogram based on template matching positioning.

There are not a lot of literature on vehicle head and tail identification. At present, the research on the recognition of the head and tail of large-scale transport vehicles is involved in a patent method of feature recognition of the residue truck based on convolution neural network. The patent mainly uses the deep convolution neural network to identify the head and tail of the residue truck to judge separately, If it is the tail information, it will be input into the trained traffic violation model to judge whether it violates the traffic law. The main basis is whether the tail cover of the muck truck is covered. If it is illegal to judge through the algorithm, the vehicle warning information will be notified to the supervisor. In the patent video-based accurate identification method for high-speed mobile vehicle logo, the pre-processing operation involves the positioning of the vehicle's head and tail, in which the image noise removal and contrast enhancement are mainly applied to the head, and homomorphic filtering technology is required for the tail image. In addition, vehicle logo recognition mainly uses the idea of relatively concentrated edge texture. This paper proposes that the vehicle is automatically identified by image segmentation and clustering technology in the camera assisted driving. The paper shows that this method has good recognition effect in foggy and rainy days. To sum up, it can be found in the current research that vehicle head and tail recognition is a processing link in the case that vehicle head and tail need to be distinguished in license plate recognition, and the research goal of this paper is to clearly distinguish the head and tail of the vehicle, let the surveillance camera capture the picture, and distinguish the head and tail of the vehicle through our algorithm, so as to solve the problem of background analysis. For this reason, our method is based on the deep convolution neural network (CNN) to extract features, while using the cross entropy loss function (binary entropy loss function) To measure the effect of classification.

3 Method

The method of mining vehicle head and tail recognition based on deep convolution neural network proposed in this paper, which has the following procedures: firstly, the pictures of the head and tail are sorted out, and the data that people can hardly distinguish the head and tail are filtered out. Secondly, considering the size and shape of the head and tail, the method contains total of 5 stages, each stage contains one convolution layer and one pooling layer in the experiment. Thirdly, two full connection layers and a prediction layer are used to process the convolution tail, and the softmax loss function is used to classify the head and tail of vehicles. Finally, the parameters are updated by Adam optimizer. Figure 2 shows the training process of the model.

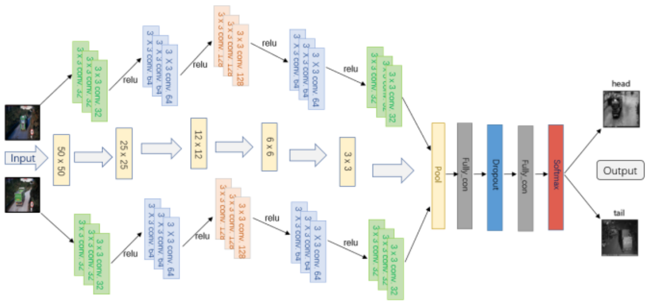


Fig 2. The structure of the Deep CNN model.

3.1 Image Processing

In order to reduce the search space and improve the accuracy of the network model, we normalized the input pictures before training. The size of all the input image is resized as 50×50, and each image contains the head or tail of the vehicle and its located in the center. Then, all the images are grayed before input in our designed deep CNN model.

3.2 Deep CNN

In this paper, considering the different sizes and shapes of the head and tail of the vehicle, if the large filter is used to extract the features of the head and tail of the small size vehicle, after multiple down sampling, the obtained feature map will only contain less feature information, even in extreme cases, reducing the classification accuracy. The same is true for the head and tail of large size vehicles. When using smaller filters to extract features, the overall information of the head and tail of vehicles could not be captured due to the smaller receiving area, resulting in low classification accuracy. Due to the uncertainty size of the head and tail of the vehicle, while a single size filter is used, the above problem will not be solved. Therefore, this paper designs a deep CNN to extract features of different scales. Each batch contains 32 pictures and with the picture size is 50×50. The final classification result is obtained by weighting and fusing the prediction results of 2D images. Deep CNN has three layers with the same structure but different convolution kernel sizes, and the convolution kernel sizes are 3×3, 5×5 and 7×7 . These convolution kernels are used to extract input features. The size of Max pooling kernel is the same as that of convolution kernel in each layer. Finally, the output of the three layers is combined, and then the global average pooling layer and softmax are used to obtain the prediction classification results, and finally the prediction is made.

3.3 Algorithm Equations

The formula for calculating the output feature map size is as follows:

$$\mathrm{out}_{\mathrm{width}} = \frac{\mathrm{Input}_{\mathrm{width}} + 2*\mathrm{Padding} - \mathrm{Kernel}_{\mathrm{width}}}{\mathrm{Stride}_{\mathrm{width}}} + 1 \tag{1}$$

$$\mathrm{out}_{\mathrm{length}} = \frac{\mathrm{Input}_{\mathrm{length}} + 2*\mathrm{Padding} - \mathrm{Kernel}_{\mathrm{length}}}{\mathrm{Stride}_{\mathrm{length}}} + 1 \tag{2}$$

In Eq. (1) and (2), outlength and outwidthare the lengthand widthof the output feature map respectively; Inputlength and Inputwidth are the lengthand widthof the input feature map; Padding is the number of edge filling pixels; Kernellength and Kernelwidth are convolution The size of the core; Stridelength and Stridewidth are the step size.

After the convolution operation of the convolution layer, the result needs to be processed with a non-linear activation function. The entire process expression is as follows:

$$S(i,j) = \sigma\big((I * K)(i,j)\big) = \sigma(\textstyle\sum_m \sum_n I(i+m, j+n)K(m,n)) \tag{3}$$

In Eq. (3), I represents the two-dimensional image tensor data; K represents the convolution kernel; m and n are the length and width of the convolution kernel respectively, usually 3x3 or 7x7; i and j are the values of horizontal and vertical coordinates respectively, and "*" represents multiplication; σ is the activation function, in the convolution neural network The ReLU activation function is commonly used, which can make the network sparse, facilitate the extraction of network features, and reduce the over-fitting phenomenon in the network training process. The calculation formula is formula (4), and the schematic diagram is shown in Figure3.

$$\sigma(x) = max(0,x) = \begin{cases} x, x > 0 \\ 0, x \le 0 \end{cases} \tag{4}$$

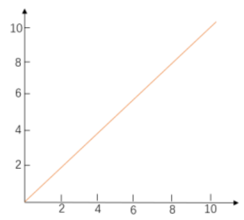


Fig 3 ReLU activation function.

The output layer of the convolution neural network is the Softmax layer, which is connected to the last fully connected layer in Figure 2. Softmax regression is a multi-classification algorithm, which belongs to supervised learning, which converts multiple outputs into probabilistic forms through normalization operations. For an n classification problem, given the input x belongs to an original metric h(x, yi) of the i-th category (yi), which the original metric is the input samples original class metric, the calculated probability is shown in Eq. (5).

$$P(x|y) = \frac{e^{h(x,y_i)}}{\sum_j^n e^{h(x,y_j)}} \quad (5)$$

The loss function of Softmax regression is shown in Eq. (6). The optimization of the Loss function is achieved by iterative solution of the gradient descent method.

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_j^k 1\{y^{(i)} = j\}\, log\, \frac{e^{h(x,y_i)}}{\sum_{j=1}^k e^{h(x,y_j)}}\right] \quad (6)$$

And the θ in Eq. (6) means the parameter of the model.

3.4 Algorithm Flow

CNN back propagation algorithm flow.

Input: M image samples, layer number L of CNN model and types of all hidden layers. For convolution layer, define the size k of convolution kernel, dimension F of convolution kernel matrix, filling size P and step s. For pooling layer, the size of pooling area K and pooling standard (max or average) are defined. For fully connected layer, the activation function (except output layer) and the number of neurons in each layer are defined. Gradient iteration parameter iteration step α， Max and threshold of stopping iteration ϵ。

Process:

1) Initialize W, B of each hidden layer and output layer

The value of is a random value.

2）for iter to 1 to MAX：

2-1) for i =1 to m：

a) Set CNN input A1 to the tensor corresponding to Xi

b) For L = 2 to L-1

b-1) if the current layer is fully connected: then ai,l=σ(zi,l)=σ(Wlai,l−1+bl)

b-2) if the current convolution layer is: then there is ai,l=σ(zi,l)=σ(Wl∗ai,l−1+bl)

b-3) if the current pooling layer is: ai,l=pool(ai,l−1), where pool refers to the process of reducing the input tensor according to the size of pooling area K and pooling criteria.

c) For output layer L: ai,L=softmax(zi,L)=softmax(WLai,L−1+bL)

c) The loss function is used to calculate the loss of the output layer δi,L

d) For L = L-1 to 2

d-1) if the current layer is fully connected: δi,l=(Wl+1)Tδi,l+1 ⊙σ′(zi,l)

d-2) if the current layer is convolution layer: δi,l=δi,l+1∗rot180(Wl+1)⊙σ′(zi,l)

d-3) if it is pool layer at present: δi,l=upsample(δi,l+1) ⊙σ′(zi,l)

2-2) for L = 2 to L, update the WL, BL of layer 1 according to the following two situations:

2-2-1) if the current layer is fully connected: Wl=Wl−α$\sum_{i=1}^{m} \delta^{i,l}$(ai,l−1)T，bl=bl−α$\sum_{i=1}^{m} \delta^{i,l}$

2-2-2) if the current convolution layer is a convolution layer, for each convolution kernel there is: Wl=Wl−α $\sum_{i=i}^{m} a^{i,l-1}$∗δi,l， bl=bl−α $\sum_{i=1}^{m} \sum_{u,v} (\delta^{i,l})_{u,v}$

2-3) if all the changes of W and B are less than the stop iteration threshold ϵ，Jump out of the iteration cycle and go to step 3.

3) The linear relation coefficient matrix W and bias vector B of each hidden layer and output layer are output.

Output: W, B of hidden layer and output layer of CNN model.

4 Experiment

This article tested the network structure and network performance through several experiments, and determined the best network structure and corresponding methods. These parameters are obtained through comparative experiments, so that the best truck head and truck tail classification results can be obtained. This article also discusses the comparison with other complex networks such as VGG16 and ResNet34 to illustrate the significant improvement of our network model.

4.1 Data Source and Composition

The data set used in the experiment is provided by the database of Huaibei Siyuan Technology Co., Ltd., who is responsible for regularly providing experimental data for our project. The data set is mainly based on image types, including 20,000 examples of the head and tail of the truck, and the division and preprocessing of the data set need to be in two stages. In the first stage, we randomly selected 16,000 pictures for training and another 4,000 pictures for testing. The rate of training set and test set is 4:1, and we manually mark the pictures in the training set as the head of the truck and the tail of the truck. In the second stage, we need to filter out non-jpg format pictures, unify the picture size and the grayscale of the picture, and finally scramble the picture. Figure 4 below shows the distribution of the number of training sets and test sets in the data set.

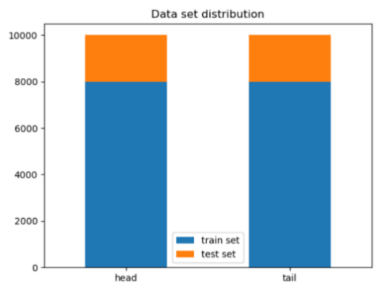


Fig 4 The distribution of training sets and testing sets.

4.2 Parameters Setting

The vehicle head and tail algorithm designed in this paper is implemented based on the structure of deep convolution neural networks. The training of deep convolution neural networks requires a lot of calculations. In order to speed up the network training speed and parameter tuning and optimization process, this paper first compares the configuration The high workstation, namely the PC, completes the network training, and then transplants it to the server platform. The specific experimental operating environment is shown in Table 1:

Table 1. Network model training environment.

CPU: i7-7700HQ	OS: Window10
GPU: Nvidia Geforce 1050	Deep Learning Framework:TensorFlow1.8
Memory: 16G	Cuda: 9.1
Python: 3.6	Cudnn: 7.4
OpenCV: 4.4	TFlearn: 0.3
Compiler: vs2017	GeForce Graphics Grive: 457.51

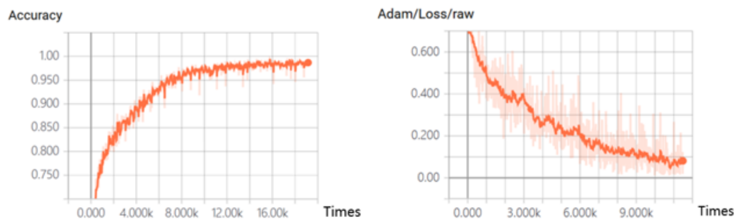
The hyper parameter settings during the network training process are shown in Table 2:

Table 2. Network training parameters.

Category	Set value
batch_size	32
learning_rate_begin	0.001
img_size	50
epoch_n	40
max_iter	20000

A larger learning rate will accelerate network learning in the early stage of training, making it easier for the model to converge to the optimal solution. However, there will be large fluctuations in the later stage, and even the value of the loss function may fluctuate around the minimum value. The fluctuation is large and it is always difficult to reach the optimum. Therefore, this chapter uses the learning rate decay method for training. At the beginning of training, a larger learning rate is used to speed up the convergence of the network. As the number of training increases, the learning rate will gradually decrease to ensure that the model will not fluctuate too much in the later stage of training, thus getting closer optimal solution. Specifically, the learning rate is attenuated by manual setting, that is, when the number of training steps reaches the set value, the learning rate is multiplied by the attenuation coefficient to obtain the new learning rate. The learning rate set in the pre-training stage and the fine-tuning training stage is 0.01, 0.005, 0.001.

After setting up the training environment on the PC side and setting the parameters, perform network training. The accuracy change curve during the training process is shown in Figure 5. The figure on the left is the accuracy change curve in the pre-training stage. When the number of iterations reaches 30, the training accuracy curve basically converges to about 0.95. It can be seen that the network model has room for further convergence. The figure on the right is the loss change curve in the fine-tuning training stage. Since the fine-tuning training is performed on the basis of the model obtained by the pre-training, the loss value at the beginning of the training is also small. It can be seen from the figure that the error jumps during the training process. In the end, the error of the entire model converges to about 0.1, which basically meets the convergence requirements of the network.



(Left) Accuracy curve of training process. (Right) Loss curve of training process.

Fig 5. CNN Accuracy rate and error loss change curve during training.

4.3 Comparison of Results

In order to explore the performance of the network, we use different patches and epochs to train the network to obtain the best performance of the network. From the parameter comparison table 3, it can be seen that when the values of patch and epoch are too large or too small, the accuracy of prediction is about 0.7, so when we fine-tune these two parameters, we find that the accuracy of prediction is gradually increasing. According to table 3, we finally receive the conclusion that when the patch is 32 and the epoch is 30, the performance of the network is the best 85.26%. The patch and epoch parameter fine-tuning comparison is shown in Table 3:

Table 3. Comparison of patch and epoch parameter.

patch	epoch	val_accuracy
16	5/10/30	70.23/76.35/79.12

32	10/30/40	82.43/85.26/84.35
64	15/30/50	79.65/82.56/80.82
128	30/50/80	70.32/75.65/72.36

4.4 Compare with VGG16 and ResNet34

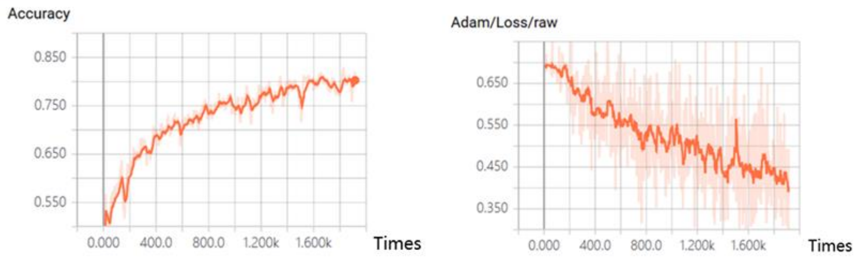
Table 4. Comparison with VGG16 and ResNet34.

methods	val_accuracy
VGG16	77.36
ResNet34	81.03
DeepCNN(ours)	**85.26**

(Left) Accuracy curve of training process. (Right) Loss curve of training process.

Fig 6. VGG16 Accuracy rate and error loss changing curve during training.

In figure 6, the figure on the left is the curve of accuracy in the pre training stage, and the curve of training accuracy basically converges to about 0.77. The figure on the right is the loss curve in the training stage. It can be seen from the figure that with the increase of training times, the value of loss function is gradually decreasing, and the error of the whole model finally converges to about 0.45.



(Left) Accuracy curve of training process. (Right) Loss curve of training process.

Fig 7. ResNet34 Accuracy rate and error loss changing curve during training.

In figure 7, the left figure is the curve of accuracy change in the pre training stage, and the accuracy curve of training is almost converging to about 0.81. The right figure is the curve of loss change in the training stage. It can be seen from the figure that with the increase of training times, the value of loss function is gradually decreasing, and the error of the whole model converges to about 0.40.

5 Summary and Future Work

In this paper, we proposed a method of mine vehicle head and tail recognition based on deep convolution neural network is proposed. The accuracy rate in the training set is more than 0.95, and the accuracy rate in the test set is about 0.85, which can better classify the head and tail. In the following work, we want to further improve the existing network based on the idea of attention module and knowledge distillation, hoping it can have better performance.

Acknowledgement. This work was supported by The Key Research and Technology Development Projects of Anhui Province (No.202004a0502043; No.202004a06020045; No.202004b11020023) and supported by the Open Project of Suzhou University Scientific Research platform Grant No.2017ykf12 and supported by the Natural Science Foundation of Anhui Province No.1908085QF283 and supported by the Doctoral Start-up Research Foundation No. 2019jb08 and supported by Overseas Visiting and Study Program for Outstanding Youth No. gxgwfx2020063. This work was also supported in part bythe Key Natural Science Project of Anhui Provincial Education Department under Grant KJ2019A0668 and supported by Anhui province's key R&D projects include Dabie Mountain and other old revolutionary base areas, Northern Anhui and poverty-stricken counties in 2019 No.201904f06020051 and supported by Natural science research project of Anhui Provincial Education Department No. kj20200a0733.

Reference

[1] Xi Yanyan, Li Hui. Prospect of the construction of China's mineral resources legal system [c] annual meeting of China Association of Geology and mineral economics. China Association of Geology and mineral economics, 2012

[2] Cai Guohua. On the shortcomings and Countermeasures of current mineral resources management [j]. Modern commerce, 2011, 000 (008): 115-116

[3] Wang Jiawei. "Design and implementation of mine vehicle monitoring system based on Android," Huaibei Normal University

[4] Li Laihu. Business tax planning in commercial mineral resources exploration project [j]. China land and resources economy, 2012

[5] Zhang Songlan, algorithm review of license plate recognition system [j] electronic technology and software engineering. 2021, (04):128-130

[6] Zhang Lijing, Sun Jie, yinxiaoyu. License plate location method based on HSV color space [j]. Microcomputer information, 2008 (7): 247-248

[7] Chang Qiaohong, gaomantun. License plate location research based on HSV color space and mathematical morphology [j]. Journal of graphics, 2013,34 (4): 159-162

[8] Tansiting, Hu Zhikun. A comprehensive method of license plate location based on HSV color space [j]. Computer and applied chemistry, 2011,28 (7): 903-906

[9] He Guijiao. Analysis and comparison of several classic image edge detection operators [j]. Computer optical disk software and application, 2014,17 (9): 182-183

[10] Liang Juan. An improved algorithm based on Sobel image edge detection [j]. Software introduction, 2014,13 (12): 78-82

[11] Luo fan, Chen Sheng, Wang Min, et al. An algorithm for locating license plate based on edge characteristics [j]. Journal of Huazhong University of science and Technology (NATURAL SCIENCE EDITION), 2004 (S1): 108-110

[12] Yang Liping. The study of license plate location based on mathematical morphology [j]. Information communication, 2016158 (2): 64-66

[13] Luyaqin, Wu LingChao. License plate location method based on mathematical morphology [j]. Computer Engineering, 2005,31 (3): 224-226