Road identification system based on CNN

Xuebin Yang¹, Hongzhi Yu², Di Wang³ and Ding Wang¹*

¹Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730000, China
²Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730000, China
³Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730000, China
*Corresponding author’s e-mail:1023521598@qq.com

Abstract. In recent years, convolutional neural network (CNN) has been widely used in image recognition, but there is still a big gap in automatic driving and road recognition. Here we have designed a system that maps raw pixels directly from a single front-facing camera to a steering command. The system mainly uses the camera to recognize the auxiliary lines on the road to realize automatic driving, and the accuracy rate has reached 98%. Unfortunately, such data can only be used in the laboratory, and the recognition of complex road conditions needs further development of machine learning.

1. Introduction

CNN is a feed forward neural network. CNN was first proposed by Hubel and Wiesel in the 1960s when they studied the neurons used for local sensitivity and direction selection in the cat cortex and found that its unique network structure could effectively reduce the complexity of the feedback neural network. Therefore, CNN neural network was proposed.

CNN has revolutionized pattern recognition. Before CNN was widely adopted, most pattern recognition tasks were completed through the initial stage of manual feature extraction and classifier. The breakthrough of CNN is to automatically learn features through training examples, which greatly simplifies the steps of image recognition. Similarly, in recent years, CNN learning algorithm is also implemented on massively parallel graphics processing unit, which greatly accelerates the learning and reasoning speed.

In this paper, we describe a CNN that goes beyond pattern recognition, and it can achieve automatic road recognition part. The training set is all the images automatically collected by the front camera of the vehicle, and also includes some unexpected special cases. The whole process is completed in the laboratory. We will also set up some emergencies to determine the stability of the whole road identification system. In general, this system can be used in the laboratory environment.

2. System overview

Our training set contained all images collected by camera. As shown in figure 1, we installed the camera in the front of the vehicle to collect images. The time stamp video on the camera head and the steering Angle of the controller were taken at the same time. After the shooting, the camera was transferred to the storage device by the control board to save the training set. About every second of vehicle travels,
an image of the environment at that time should be taken. The pixels of the images should be the same as far as possible, so as to improve the accuracy of training in the later period. The training data should correspond to the corresponding control command. It means to add labels to the collected training set.

![Diagram](image1)

**Figure 1. Training set collection system**

In terms of controlling the motherboard, we compared Arduino and raspberry pie. Arduino was created in 2005 by teachers at a high-tech design school and Spanish-born chip engineers. Launched in 2012, the raspberry PI claimed to be the world's smallest computer, the size of a credit card, but with all the basic functions of a computer. The Arduino is really just a microcontroller, not a mini computer. The microcontroller is only one part of the computer, a subset of the raspberry pie, and offers limited functions. Arduino is still a little difficult to control the movement of the vehicle and the calculation of machine learning, but raspberry PI can well perform the task of controlling them. The Raspberry Pi foundation has released its fourth generation of Raspberry Pi, which greatly speed up the Pi's ability to process data. So we choose raspberry PI as our control board.

Our training system is shown in figure 2. When the image is input into the CNN training, CNN will give the next suggested command, and the system will compare the next suggested command with the collected tags. Then the weight of CNN will be adjusted by back propagation to make the CNN output closer to the expected output.

![Diagram](image2)

**Figure 2. Training system diagram**

### 3. Data collation

The general process of data consolidation is shown in figure 3. All the training sets are collected in the laboratory environment. During the process of collecting pictures, labels should be added to each picture to facilitate the identification of the training. When we get the training set, we have to filter the training set first. Some images will be very blurry when shooting. We have to delete these images manually. Since the size of the image acquired may be different each time, we also need to preprocess the image to compress it into the same size.

![Diagram](image3)

**Figure 3. Collection process**

In the whole process of data collection, a total of 26,408 images were collected. And the images quality was excellent.
4. Training data

4.1 Basic model introduction

The basic model we adopted is the five-layer CNN model proposed by NVIDIA in 2016, which can improve the accuracy of road recognition to 90%. The neural network contains 27 million connections and 250,000 parameters, which is very large.

The first normalization layer of the network adopts normalization operation, dividing each dimension of the image by 255, and normalizing all elements to -0.5 to 0.5. There is no learning process involved here. Normalization is converted into a pre-processing operation before the input convolution layer. Normalization operation can help adapt to different types of network structure and accelerate the training of the model on the GPU. The next five convolution layers are followed. The function of the convolution layer is to extract features for subsequent training. The first three convolution layers choose 5x5 kernel and 2x2 strides, and the last two convolution layers choose 3x3 kernel and no strides. The kernel is the convolution kernel, and strides can be understood as the move step length of the convolution kernel on the matrix. After the convolution layer, NVIDIA add three full connection layers, which are used to train the features extracted from the convolution layer and finally output the steering control signal.

The structure proposed by NVIDIA is end-to-end. The end-to-end means that the input of neural network is the original picture and the output of neural network is the instruction of direct control. The feature of end-to-end deep neural network is that the boundary between feature extraction layer and control output layer is not obvious, because every part of the network plays a role in feature extraction and control for the system.

4.2 Model improvement

4.2.1 Activation function selection. In image recognition, activation functions are generally selected as sigmoid function, the relu function, the tanh function and the elu function. The effects of these four functions are as follows:

1) The sigmoid function

Sigmoid function is a common logistic regression function, which is described as:

\[ g(z) = \frac{1}{1 + e^{-z}} \]  \hspace{1cm} (1)

As a nonlinear function, sigmoid function, when the z value is very large or small, the derivative of the function will approach 0, making the update gradient slow, which is called the disappearance of the gradient. After such training, the learning rate will be very low. The correct rate of 32 batches of training is shown in figure 4. It can be seen that after the model training with sigmoid function, the accuracy rate does not increase after the 7th batch. The accuracy rate is not very high, so sigmoid function is not an ideal activation function.

![Figure 4. Accuracy of sigmoid function](image)

2) The tanh function

![Figure 5. Accuracy of tanh function](image)
Tanh is a function that maps the values from \((-\infty, +\infty)\) to \((-1, 1)\). Its function is described as:

\[ g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \]  

(2)

Tanh function to some extent makes up for the shortcomings of sigmoid function, but in essence there will be the problem that the derivative of the function approaches 0, leading to the disappearance of the gradient. The correct rate of 32 batches about tanh function training is shown in figure 5.

Tanh function is very similar to sigmoid function, and it will not increase after training to the 8th batch. The accuracy rate is very low. So tanh is not the ideal activation function for our model.

(3) The relu function

Relu function is a piecewise linear function. It is also known as modified linear element, which is now the mainstream image recognition activation function. Its function is described as:

\[ g(z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases} \]  

(3)

The relu function to a large extent makes up for the disappearance of the gradient. Both forward propagation and back propagation are much faster than the sigmoid function and tanh function. However, when the input value is negative and the gradient is 0, there will still be a series of problems such as the disappearance of the gradient. The correct rate of relu function training in 32 batches is shown in figure 6.

Although relu has no gradient disappearance problem to a large extent, the training results show that relu function is the same as tanh function, with low accuracy. Therefore, relu function is not suitable for our road recognition model.

(4) The elu function

Elu function is an improvement of relu function. The most important is to improve the gradient. When the input value is negative, it can avoid the gradient disappearing. Its function is described as:

\[ g(z) = \begin{cases} z & z \geq 0 \\ a(e^z - 1) & z < 0 \end{cases} \]  

(4)

It can be said that elu function is a relatively successful improvement of relu function, which greatly avoids the disappearance of gradient. The correct rate of elu function training 32 batches is shown in figure 7.

It can be seen from this experiment that compared with the above three activation functions, elu function has a significant improvement in accuracy. The prediction of accuracy when NVIDIA road recognition model is put forward is also about 90%, so we can use elu function to meet the requirements.

Table 1 is the comparison table of accuracy of the selected activation function after training. From the accuracy after training, it can be seen that the effect of elu function is significantly better than the other three activation functions. We use the elu function in our model.

| The name of the function | Accuracy after training |
|-------------------------|------------------------|
| sigmoid                 | 44.38%                 |
4.2.2 Dropout choice. Dropout is to prevent over-fitting in the process of model training, leading to a decrease in the accuracy of the model. A certain proportion of neurons are randomly disconnected, leaving them in an inactive state. Since there are many choices of random disconnection, we chose the three proportions of 50%, 60% and 70% to conduct the experiment, and the correct rates are shown in table 2.

| Dropout rate | Accuracy after training |
|--------------|-------------------------|
| 50%          | 90.2%                   |
| 60%          | 89.5%                   |
| 70%          | 88.2%                   |

According to the above experiment, we can see that 50% Dropout ratio can make the training set the most accurate.

4.2.3 full connection layer design. According to the design of the basic model, the number of neurons in the three fully connected layers is 100, 50 and 10 respectively. After training, the accuracy was indeed up to 90% of what NVIDIA had proposed. However, after the development of CNN, we found that the number of neurons in the three fully connected layers seemed not reasonable. Therefore, after several experiments, we found an appropriate number of neurons allocation method.

We added a full connection layer on top of the original three. The number of neurons was allocated as 500, 250, 50, 5. The comparison chart of training accuracy is shown in figure 8.

![Data comparison chart](image)

Figure 8. Comparison chart of accuracy

We can see from the improved model training that the accuracy rate is increased from 90% to 98%, which greatly improves the expected accuracy, and also makes it possible for the road identification system to change from the laboratory stage to the normal road stage.

5. Simulation test
During the test phase, we designed a simulator, as shown in figure 9, which took a pre-recorded video from the front-facing camera of a human-driven data acquisition vehicle. The generated image was roughly the same as that of the trained model when operating the vehicle. The simulator sent the first frame of the selected test video to the model that has completed the training, and then the model returned the next command for that frame. This command and the recorded human driving command were entered into the dynamic model of the vehicle to update the position and direction of the simulated vehicle.
Record video and turn commands

move and turn

model

Update car position and direction

Composite road image

The neural network gives instructions

Figure 9. Working flow chart of the simulator

The simulator then modified the next frame in the test video to make the image appear to be where
the command issued by the model was. The image was then transmitted to CNN, and the process
repeated itself. The simulator recorded the distance away from the centre and reset the position and
direction of the virtual vehicle for easy reference.

6. Summary
CNN has been able to learn the whole task of road recognition now, and it is no longer necessary to
manually decompose it into road or lane marker detection. In this paper, based on the existing NVIDIA
road recognition model and the new experiments in recent years, the model has modified to improve the
accuracy rate from 90% to 98%. However, the automatic driving is still in the laboratory stage. If the
automatic driving is to be applied in life by modifying the CNN algorithm, it can make the road
recognition overcome the bad weather such as rainy days and Yangtze sand.

Acknowledgments
This work was supported by Northwest MinZu University the central college basic scientific research
operating expenses special fund self-help graduate student project (No. Yxm2019116).

References
[1] Yu Feiwu, Wu Xinxiao, Chen Jialu, Duan Lixin. (2019) Exploiting Images for Video Recognition:
Heterogeneous Feature Augmentation via Symmetric Adversarial Learning. In: IEEE
transactions on image processing. USA. pp. 5308-5321.
[2] Son Ngoc Truong, Khoa Van Pham, Wonsun Yang et al. (2017) Time-Shared Twin
Memristor Crossbar Reducing the Number of Arrays by Half for Pattern
Recognition. Nanoscale Research Letters, 12:1-6.
[3] Shitao Chen, Zhiqiang Jian, Yuhao Huang, Yu Chen, Zhuoli Zhou, Nanning Zheng. (2019)
Autonomous driving: cognitive construction and situation understanding. Science China,
62:42-49.
[4] Francisco Martinez, Ariel Carrasco. (2015) Pattern recognition applications in computer vision
and image analysis. Pattern Recognition, 48: 1025-1026.
[5] Changbin Shao, Xiaoning Song, Xin Shu, Xiao-Jun Wu. (2017) Converted-face
identification: using synthesized images to replace original images for recognition.
Multimedia Tools and Applications, 76: 6641-6661.