Research on Visual Recognition Method of Unmanned Driving Based on Deep Learning

Liu Qingyong\textsuperscript{1,a}

\textsuperscript{1}School of computer science, Sichuan University, Chengdu, 610207, China
\textsuperscript{a}liuqingyong@stu.scu.edu.cn

Abstract. Visual recognition is an important part of unmanned driving technology. With the development of computer technology, the application of deep learning in the field of unmanned driving has been promoted. This article uses Raspberry Pi to build a car with a camera, uses deep learning technology to recognize the acquired image information, and optimizes the structure and parameters of the convolutional neural networks, and finally makes the efficiency of car recognition as high as possible.

1. Introduction

With the continuous development of society, the number of cars is increasing, and the problems of traffic safety and road congestion have become more and more prominent. Unmanned driving technology can effectively improve this problem. This is because unmanned driving can avoid traffic accidents caused by human control, and at the same time can choose the optimal road conditions autonomously, thereby greatly reducing traffic pressure\cite{1}. Through the application of deep learning technology, the system's environmental perception and road condition analysis capabilities can be further improved, and the safety and reliability of unmanned driving can be significantly enhanced. These advantages greatly prove that deep learning is of great significance to the improvement of unmanned driving.

As early as the 1970s, the United States, the United Kingdom and other countries began to study driverless technology, and Google was the first company to develop driverless vehicles\cite{2}. In 2009, Google converted a Toyota car into Google's first-generation driverless car, which can independently realize road conditions. In 2011, Google developed the second generation of driverless cars, combined with deep learning technology to strengthen the car's ability to perceive the environment. In 2011, Google obtained a license for testing on the road, and in the following two years, Google released the third generation of driverless cars. By 2016, Google's driverless project was divided into waymo\cite{3}. In addition to Google, companies such as Delphi in the UK, INRIA in France, and IBEO in Germany have also carried out research on driverless technology.

China began to study driverless technology since 1980s. At present, Baidu company in China has launched a number of driverless platforms, and Ali DAMO Academy has also made a breakthrough in the field of driverless, pushing the world's first hybrid simulation test platform for driverless. In addition to Baidu and Ali, Tencent, Didi and other companies have also started the research on driverless. These show that driverless vehicles have been fully valued in China, however, the development of driverless technology is not perfect, there are still many problems to be solved. Among them, the image acquisition and processing of the surrounding environment of the driverless vehicle is one of the key points. How to realize the visual recognition efficiently has become a major bottleneck restricting the development of driverless technology.
Because of the great practical significance of driverless, and the importance of visual recognition in driverless, we study the method of driverless visual recognition based on deep learning: we use Raspberry Pi to build a car, and use deep learning to analyze and process the images collected from the camera, and optimized the structure and parameters of the neural network to make the efficiency of image recognition as high as possible.

2. Deep learning theory

2.1. Convolutional neural networks
Convolutional neural network is one of the most representative algorithms in deep learning technology. It is a kind of neural network with convolution operation and depth structure. The network structure is widely used in many fields, such as computer vision, voice recognition, natural language processing, etc[4]. It can autonomously learn and extract features according to the image information input into the network. Therefore, more details can be extracted from the image, and the judgment of image features will be more accurate.

2.2. Convolution layer
Convolution layer is the core of CNN. It extracts image features by sliding the input image. In the sliding process, it mainly depends on convolution kernel. Each convolution core in the convolution layer will convolute the image matrix of its coverage area. The convolution results of several convolution cores at the same position are added, and the bias is added to get the output results of the input image at this position. This process is called feature mapping. The process of convolution calculation is shown in the figure below:

![Figure 1 convolution process](image1)

In Figure 1, the white background on the left is the input, and the white background on the right is the corresponding output. The convolution kernel used is shown in Figure 2:

![Figure 2 convolution kernel](image2)

2.3. Activation function
Usually, activation function is a kind of nonlinear and differentiable function, which can map the upper input data nonlinearly, and then enhance the fitting ability of neural network to complex functions. The common activation functions are Sigmoid function, Tanh function and ReLu function. Because of the gradient disappearance of Sigmoid and Tanh in the process of neural network reverse conduction, they are not frequently used in practice. Using ReLu as the activation function not only has faster training speed, but also solves the problem of gradient disappearance, so it has become one of the most widely used activation functions in CNN. The image of ReLu activation function is as follows:
The ReLu function takes 0 in the negative interval and itself in the positive interval. In other words, it takes the maximum value between 0 and itself, so it only needs to judge whether the input is greater than 0, which reduces the calculation amount of the neural network to a certain extent and speeds up the operation speed.

2.4. Pooling layer
The pooling layer is usually placed behind the convolution layer and the activation function, which is used to downsample the output feature map of the previous convolution layer. The number of parameters is reduced by reducing the known tensor size. There are two common pooling layers: maximum pooling and average pooling. Maximum pooling is to extract the maximum value in the local receiving domain, and average pooling is to average all the values in the local receiving domain. The two methods are shown in the following figure:

![Figure 4 maximum pooling diagram](image)

![Figure 5 average pooling diagram](image)

Using pooling layer can reduce the amount of calculation, improve the calculation speed of neural network, and enhance the robustness of neural network. In addition, the pooling layer can reduce the feature dimension and keep the feature scale unchanged, which can preserve the main information after compressing the image.

2.5. Fully connected layer
In CNN, the fully connected layer is usually placed after multiple convolution layers and pooling layers. Each neuron in the full connectivity layer is fully connected with all neurons in the upper layer, and its function is to integrate the class distinguishing feature representation of the front layer. The full connection layer generally has two or more layers, which can be classified by convolution operation. Because of its full connection characteristics, the parameters of this layer generally account for about 80% of the whole network structure parameters, which makes the whole model very bulky. At present, some network models use global average pooling instead of full connection layer[5], which not only has better prediction performance, but also greatly reduces the amount of network parameters.
3. Experimental design

3.1. Hardware system composition
The driverless car uses Raspberry Pi 4B platform. Raspberry Pi integrates all kinds of basic components of modern computer, but it has the advantages of small volume, low power consumption and low price. However, the computing power of Raspberry Pi is limited, so the network model needs to be trained in PC. After the model is trained, it will be migrated to Raspberry Pi. Raspberry Pi collects and predicts the image data, and sends instructions to control the operation of the car, including start, left turn, right turn, stop and other actions. The driverless car is equipped with an 8-megapixel camera to collect road data.

3.2. Network design
In order to compare the effect of GAP layer instead of full connection layer, two neural network models are designed. The input layer of the two models is the image collected by the car camera, and the convolution layer is two layers. After the convolution layer, one model uses two full connection layers, while the other model uses one GAP layer.

In general, the size and step size of each convolution kernel are customized. In order to reduce the amount of computation, the first convolution layer uses a larger convolution kernel, which has a size of $5 \times 5$ and a transverse and longitudinal step size of 2. When filling, it fills in 0 around it. The convolution kernel size of the second convolution layer is $3 \times 3$, the horizontal and vertical steps are 1, and 0 is still filled around the input image. Using small convolution kernel and small step length in the second convolution layer is conducive to better learning effect. The activation function is the commonly used ReLu function. The second layer of the full connection layer is the classification output layer, which uses softmax classification.

3.3. Model training and testing
About 7000 pictures were collected through the car camera, and the training set and validation set were divided according to the ratio of 4:1. The training set was 5600 pictures, and the validation set was 1400 pictures. The image is grayed and binarized in OpenCV, and the size is reduced to $32 \times 32$.

Training is divided into two stages, the first stage of learning rate is 0.01, the second stage of learning rate is 0.001. The first stage has fast convergence speed, which can reduce the training time. Loading the model parameters after the first stage training into the model for the second stage learning can further improve the learning effect of the model.

4. Experimental result
In the experiment, according to the pre divided data set, the model is trained on the training set, and then the accuracy is verified on the validation set. After training, it is observed that the model with GAP layer is similar to the model with full connection layer in training accuracy, but it takes less time to train the model with GAP layer and converges faster. The training process of the model with GAP layer is shown in the figure below:
Figure 7 the training process of replacing the fully connected layer with the GAP layer

The left figure is the first stage of training, and the learning rate is 0.01. It can be observed that in the first round of training, the accuracy of training set is only 0.14, after the 5th iteration, the accuracy reaches 0.4, and after the 10th iteration, the accuracy reaches 0.5; In the first round of training, the accuracy on the validation set is only 0.1, after the 5th iteration, it reaches 0.2, and after the 10th iteration, it reaches 0.41.

The right figure is the second stage of training. At this time, the learning rate is 0.001, which is further optimized on the basis of the existing model. It can be observed that at the end of the 5th iteration, the accuracy on the training set has reached 0.62, the accuracy on the validation set has reached 0.58, and at the end of the training, the accuracy on the training set and the validation set has reached 0.75 and 0.65 respectively.

5. Conclusion

In this paper, the research and experiment design of unmanned driving technology based on deep learning are mainly carried out. The important role of image processing in unmanned driving is analyzed, and the two neural network models with GAP layer and with fully connected layer are compared. The results of the two models in image recognition accuracy are similar. But the neural network with GAP layer has fewer parameters and the model converges faster. Therefore, using the GAP layer to replace the fully connected layer network can have better performance.

References

[1] Ma chentian,chentian Ma. Application of Deep Learning in Vehicle Driverless Technology[J]. Journal of Physics: Conference Series,2020,1682(1).
[2] Yingjie, Wang. Development and Application of Artificial Intelligence[C]. 2015 4th International Conference on Mechatronics,Materials,Chemistry and Computer Engineering(ICMMCCE 2015). 0.
[3] Google Company.
https://www.egouz.com/topics/16023.html
[4] Gu, Jixiang, Wang, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition the Journal of the Pattern Recognition Society, 2018.
[5] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.