



Intelligent Vehicle Automatic Identification System Based on YOLOv4 and ViSLAM

Chenzhi Nie, Wei Lin, Xiuwen Zheng

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract:

In this paper, we use intelligent vehicles as the platform and use convolutional neural networks for lane recognition and classification during driving. For the recognition of landmarks, we use YOLOv4, a popular YOLO series algorithm, as the model for recognition. At the same time, we study and explore intelligent vehicle mapping and positioning technology based on the SLAM framework in a laboratory working environment with weak signals.

Keywords: Intelligent Vehicle, YOLOv4, CNN model

1、 Plan explanation

With the development of the times, autonomous vehicle technology has flourished. It mainly relies on the intelligent autopilot in the car, which is mainly composed of computer systems, utilizing vehicle sensors to perceive the surrounding environment, and according to the data obtained by the perception, the judgment and analysis of the road condition are made to control the vehicle 's movement, so as to realize the unmanned driving.

SLAM technology can be applied in real time on mobile devices, but pure visual SLAM methods can not work in the area with less image text. In this paper, we use intelligent vehicles as the platform and use convolutional neural networks for lane recognition and classification during driving. For the recognition of landmarks, we use YOLOv4, a popular YOLO series algorithm, as the model for recognition. At the same time, we study and explore intelligent vehicle mapping and positioning technology based on the SLAM framework in a laboratory working environment with weak signals, and the image is fuzzy when moving fast. Based on the viSLAM structure, we use convolutional neural network CNN for lane recognition and classification during driving, while for the recognition of landmarks we use YOLOv4 in the popular YOLO series algorithm as a model for identification.

2、 Problem description

2.1 Road recognition clone

The smart car needs to analyze the pictures taken by its camera during the driving process of the site, and finally make judgments such as acceleration, deceleration and turning. The focus is on the classification of the

collected image samples to achieve its control, and for the image classification algorithm, we use the CNN model classification recognition of the deep learning algorithm, and we find that its accuracy rate greatly exceeds the traditional machine learning method.

In recent years, with the continuous popularization and development of the Internet, more and more people are constantly innovating on the CNN model invented by Alex Krizhevsk. The error rate of the Top-5 model has decreased to 3.5%, which is two percentage points lower than that of human eye recognition.

2.2 Target detection

Compared with the convolutional neural network used in road detection, we use YOLOv4 in marker recognition. YOLO is a real-time target detection algorithm and the first algorithm to balance the quality and speed of detection. Its detection speed is very fast, and its detection speed is 100 times that of 'Fast-CNN', and its accuracy is more than twice that of previous real-time systems.

3、 Technology program

3.1 Image classification by convolutional neural network

For this image recognition and classification, a combination of convolution and full connection is used.

3.1.1 Convolution layer

The purpose of convolution operation is to extract different features of the input, realize the iterative extraction of more important features from the low-level to the high-level features, and discover the local correlation and spatial invariant properties of the image.

The operation of the convolution layer has two stages: feature extraction and feature mapping. In the feature extraction stage, the input of each neuron is connected to the local receptive field of the previous layer, and the convolution filter is used for convolution operation to extract the local features. After extracting the local features in the feature extraction stage, the activation function is used to map them into a normalized value^[1].

3.1.2 Pooling layer

Pooling operation is an operating system that calculates the average or maximum value within a small domain.

3.1.3 Fully connected layer

After the convolution is realized, the feature vector obtained after the fully connected layer can be used for classification or retrieval.

3.2 Target detection through YOLOv4

On the basis of the original YOLO target detection architecture, the YOLOv4 model adopts the best optimization strategy in the field of CNN in recent years. It has different levels of optimization from data processing, backbone network, network training, activation function, loss function and other aspects. The overall model is as follows.

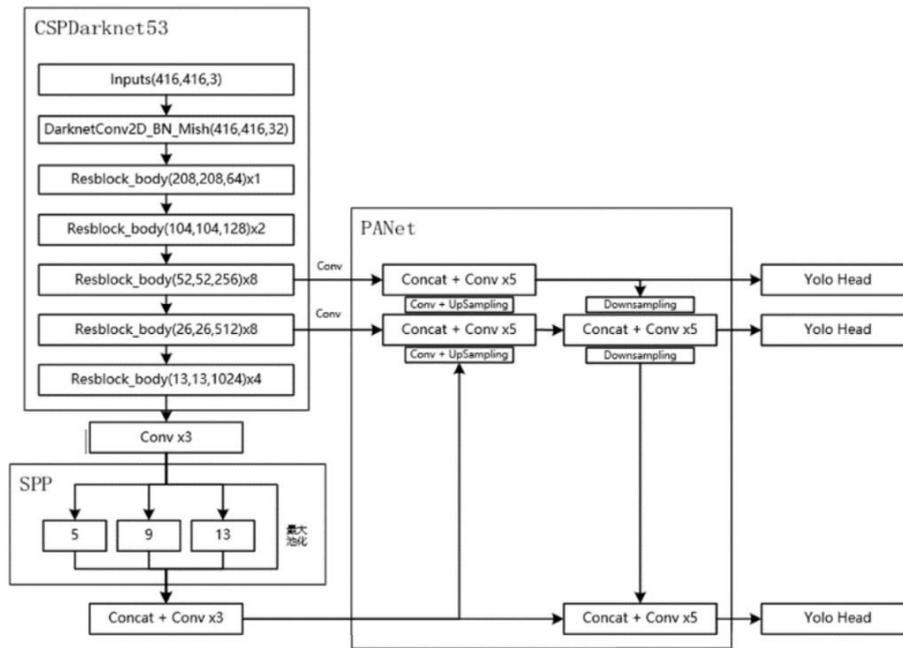


Fig.1 YOLOv4 Network model

From this, it can be seen that YOLOv4 performs better compared to most existing networks.

3.2.1 CSPDarkNet53

CSPDarkNet53 is an improved version of DarkNet53, and its improvements mainly include Mish activation function and the use of CSPnet structure.

1. Mish Activation function

The activation function is to improve the learning ability of the network and enhance the transfer efficiency of the gradient. The activation function commonly used by CNN range from ReLU and LeakyReLU in the early days to Swish and Mish in the later years. Their computational complexity is getting higher and higher. Mish activation function can be shown as follows in space:

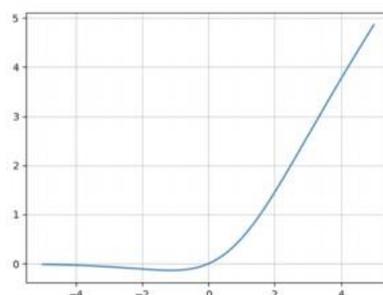


Fig.2 Mish Activation function diagram

If the figure is the output curve of the Mish activation function, we can see that its curve is relatively smooth and flat.

The slippery curve allows better information to penetrate the neural network for better accuracy and generalization. [2].

2.CSPnet Structure

The main idea of the Cross Stage Partial structure is to divide the input into two parts before inputting a block. One part is calculated through the block, and the other part is concatenated directly through a shortcut. Its advantage is that it enhances the learning ability of CNN and reduces memory consumption. The convolution operation in YOLOv4 uses CSPnet as the unit of the feature extraction network, and its structure is shown in the figure:

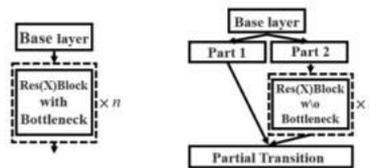


Fig.3 CSP net network structure

The figure (left) shows a typical stack using residual convolution. The figure (right) shows the CSPnet convolutional block, which is mainly divided into two parts. The main part is also the original stack of residual blocks, while the other part is the residual edges that have been slightly processed. Under this structure, the original features of the input can be retained to a certain extent.

3.2.2 Feature pyramid

CNN keeps the translation of the object unchanged, and the size transformation of the object cannot be processed, so the feature pyramid is used for processing. The feature pyramid is mainly divided into two parts, SPP structure and PANet structure.

3.2.3 YoloHead

After the feature pyramid structure, there will be three YoloHead feature layers, namely the middle layer, the middle and lower layers, and the bottom layer. This is the last part of the YOLOv4 network, which is used to predict the obtained features.

4、 Plan implementation

4.1 Image classification by convolutional neural network

4.1.1 Lane line data acquisition and preprocessing

1.Data acquisition

Start the car, connect it to the network and handle, and use remote sensing to control the car's left and right turns. Then, the two cameras on the car will collect the runway dataset information and turn angle information, and record the current speed control amount once. Run the car along the center of the lane line for 5 to 6 laps to end data collection. Press the end button 4 times to end data collection. Store the collected road data in the specified directory. The final collected images are as follows:



Fig.4 Lane line data

2. Pretreatment

Process the collected data to extract useful lane line information, and store the processed data in the designated directory.

4.1.2 Practice

There are two kinds of training, namely online training and local training. Online training uses Baidu AI Studio platform for online GPU training, while local training uses on-board computer for training. The first model trained a self-driving model, and the second trained an autonomous driving model that can detect markers.

1. Local training, which uses the CPU of the on-board computer for training, is much slower than online training.

2. For online training, you first need to register an account on the Baidu AI Studio platform to receive free GPU computing power.

4.2 Target detection through YOLOv4

4.2.1 Data acquisition and annotation

There are differences between target detection and runway classification. The dataset requires target annotation on the images before training, so we used labelImg software to annotate the data collected by the camera, as shown in the figure:



Fig.5 labeled dataset

The following is the save format for VOC:

```
|0 0.253125 0.3458333333333333 0.38125000000000003 0.475
```

Fist6 Data format

4.2.2 Practice

1.Mosaic Data augmentation

Due to the huge time cost of self-made data sets, Mosaic data enhancement is used to expand the data sets in the training process. Its main feature is to enrich the background of the detected object, to a certain extent, can prevent the situation of excessive fitting. ^[3].

The general process is to read four pictures, flip, zoom, color gamut change, etc., and finally put them in the same picture.

2.Label Smoothing

Sometimes too accurate model classification will lead to overfitting. In order to prevent this from happening, 'Label Smoothing' can smooth the data.

5、 Test analysis

5.1.1 Realize the automatic driving of the car

Download and extract the model to the local computer, place (replace) the corresponding model, lay out the track map, and combine the lane recognition content to automatically adjust the tire direction of the car, in order to achieve automatic driving of the car along the lane line. The following is a test diagram of the car starting to drive after successfully identifying the lane route:



Fig.6 Measuring view

After multiple experiments, we have calculated that the average line pressure of the car is very low, usually only once out of ten times.

5.1.2 Marker identity

The picture shows the photos taken during the operation using YOLOv4 to recognize the set markers, with a final recognition accuracy of up to 97%.



Fig.7 Marker recognition effect diagram

However, because the YOLOv4 network is more complex and has a greater advantage in the case of more data sets, we have relatively few data sets in the experiment, so the fitting degree of the prediction box is low.

6、 Conclusions

The most important task in the driving process of the unmanned car is the classification of the road and the recognition of the markers, but the recognition method between the two is actually different, so we use CNN convolutional neural network and YOLOv4 neural network as models for training and prediction.

In the road recognition used by the convolutional neural network, the effect is more obvious and the test accuracy is higher; in the process of YOLOv4 marker recognition, the recognition speed is also faster, and the accuracy rate can reach more than 95 %, which can meet the requirements of our project experiment.

What we hope is that in the future, more samples and model training can be used to achieve better results and application capabilities for the car, and it can be combined with GNSS positioning information to automatically and quickly draw maps.

Reference

- [1]Zhou Feiyan, Jin Linpeng, Dong Jun. A review of convolutional neural networks. *Journal of Computer Science*, 2017,40 (06) : 1229-1251.
- [2]Research on target detection of unmanned driving scene based on YOLO algorithm. Southwest University, 2021.DOI : 10.27684 / d.cnki.gxndx.2021.003227.
- [3]Cheng Ze, Lin Fusheng, Jin Chao, et al. Fatigue driving detection based on lightweight convolutional neural network. *Journal of Chongqing University of Technology (Natural Science)*, 2022,36 (02):142-150.