Vehicle detection algorithm based on LW-SSD

Yunxiang Liu¹,a, Guoqing Zhang¹,b
¹School of Computer Science and Information Engineering Shanghai Institute of Technology Shanghai, China
²zgqpop1001@163.com, b749229869@qq.com

Abstract. Due to the complex structure of the traditional SSD network model and the large amount of parameters, it is not suitable for porting to mobile or embedded platforms, so the LW-SSD vehicle detection algorithm is proposed, this paper redesigns the basic feature extraction network of SSD, using MobileNet feature extraction network instead of VGG16, for the original MobileNet to improve, remove the original MobileNet network average pooling layer and Softmax layer, while reducing the MobileNet network from the original 28 layers to 13 layers. Then, on the basis of the subsequent network, feature maps of different scales are scaled, residual mapping and deep feature fusion operations, and the low-level position information and high-level semantic information are effectively combined to obtain new feature maps of different scales. The designed network model is trained using the Tensorflow framework, and the test effect of the network model is evaluated. The experimental results show that the accuracy of the network model designed in this paper is 78.76%, which is 9.27% lower than the original SSD network model. However, the amount of parameters is only 1/4 of the SSD network model. At the expense of partial accuracy, the design of a lightweight network model is exchanged, which lays the foundation for subsequent development on mobile or embedded platforms.

1. Introduction
Convolutional neural network has been widely used in the field of target detection and has achieved good results[1-5]. However, with the improvement of model accuracy, the depth and complexity of the model become deeper and deeper. However, in real life scenarios, it is difficult to move too big model deployed to end or embedded devices, not only faced with the problem of insufficient memory, and these scenarios require low latency, real-time and fast response speed, considering the existing chip technology has not developed to layer can be too deep, too much network model parameters ported to mobile terminal development, so focus on optimizing the lightweight design of the network model.

2. Methods and Innovations
2.1. Basic network improvement
The SSD network model extracts the initial feature graph by improving the traditional VGG16[6] feature extraction network. Although it shows good detection effect in the target detection, it is not suitable to transplant the model to the mobile end for development due to the excessive parameters required in the detection process and the large amount of calculation. MobileNet[7] network is released by Google in 2017, a lightweight depth of neural network, it is the basic unit of the depth of the separable convolution (depthwise separable convolution), depth of separable convolution is one kind of decomposable convolution (factorized convolutions), it contains the depthwise convolution and pointwise convolution two sub operations. The convolution kernel of standard convolution operation ACTS on all input
channels. Different from standard convolution operations, depthwise convolution (convolution) should be performed on one input channel for operation, and then all channel operations should be integrated. Common convolution operations of Pointwise convolution are similar, but the convolution kernel of 1x1 is adopted. Table 1 shows the experimental results of MobileNet compared with GoogleNet and VGG16. The accuracy of MobileNet is slightly lower than that of VGG16, but better than that of GoogleNet. However, in terms of model computation and model parameters, MobileNet has an absolute advantage.

| Model         | ImageNet | Million | Million |
|---------------|----------|---------|---------|
| 1.0 MobileNet-224 | 70.6%    | 569     | 4.2     |
| GoogleNet     | 69.8%    | 1550    | 6.8     |
| VGG 16        | 71.5%    | 15300   | 138     |

Due to the MobileNet feature extraction network compared to VGG16, the accuracy of the model slightly decreased, so this paper for the original MobileNet to improve, the improved MobileNet network remove the original MobileNet network average pooling layer and Softmax layer, while the MobileNet network from the original 28 to 13 layers.

As shown in Figure 1, take images with an input of 300*300 as an example. First, the convolution is performed on 1 layer of 3*3 standard convolution, and then, after 13 layers of improved depth, the convolution can be separated. The feature output of each channel is obtained by using 3*3 Depthwise convolution for convolution, and then the feature output of each channel is merged by 1*1 Pointwise convolution.

2.2. Depth feature fusion and feature scaling module

Through the improvement of the basic network, the final output characteristic graph is obtained. After a series of standard convolution operations, the output feature map was obtained with feature layers of different scales, as shown in Figure 2, Conv13, Conv14, Conv15, Conv16, and Conv17. In this paper, on the basis of the original network, characteristics of the different scales of figure for feature scaling, operating features mapping and depth fusion, will lower the location of the information and high-level semantic information for effective combination of the characteristics of the new scales, specific operation for Conv11 input to a three layer of residual block get Conv11_1 for residual map operation, the three layers of residual block, 1 * 1 convolution have played an important role in dimension reduction and nonlinear transformation, and the introduction of more significantly increased the depth of the residual block, improves its expressing ability of residual network, The problem of gradient disappearance and degradation caused by too deep network is avoided. So the same thing with Conv13_1, Conv14_1, Conv15_1, Conv16_1. At each point of these feature maps of different sizes, preset some...
target precheck boxes, and finally put these target precheck boxes into the NMS module, remove some overlapping or incorrect target precheck boxes, and get the final target precheck boxes.

3. Analysis of experimental results

In this chapter, the data set adopted by the algorithm includes bit-Vehicle data set and data set collected and shot manually. The model evaluation index includes Average Precision, multi-category Average Precision mAP and the number of model parameters. In this chapter, the algorithm sets the initial learning rate of the model to 0.01, the learning rate decline strategy to multistep, and the maximum number of iterations to 100000. The resulting loss reduction curve and mAP value curve of vehicle training are shown in FIG 3 and 4.
For the same vehicle test set, YOLOV3 model, SSD model, and the improved SSD model were used for comparative experiments. The experimental results are shown in Table 2, and some of the improved SSD network model detection results are shown in Figure 5.

| Model    | mAP  | car  | bus  | SUV  | truck | minibus | minivan | Number of model parameters (M) |
|----------|------|------|------|------|-------|---------|---------|-------------------------------|
| SSD      | 88.03 | 92.63 | 90.03 | 89.55 | 88.65 | 86.20   | 81.12   | 100                           |
| YOLOV3   | 87.63 | 91.22 | 90.05 | 85.40 | 88.77 | 84.22   | 86.11   | 236                           |
| LW-SSD   | 78.76 | 79.93 | 80.32 | 80.31 | 76.25 | 80.15   | 75.58   | 25                            |

Experiments show that the iterative 100000 times, The accuracy of LW-SSD model designed in this paper is 78.76%, compared with the original SSD network model, a fall of 9.27% accuracy, but it only equivalent to the SSD network model and the number of a quarter, at the expense of the part accuracy, the design of the network model for lightweight, helps in the mobile terminal or embedded platform development, conform to the designed in this paper.
4. Conclusion

This paper redesigns the SSD basic feature extraction network, using MobileNet feature extraction network instead of VGG16, for the original MobileNet to improve. The improved MobileNet network removed the original MobileNet network average pooling layer and Softmax layer, at the same time, the MobileNet network from the original 28 to 13 layer. On the basis of the subsequent network, feature scaling, residual mapping and depth feature fusion were carried out for feature maps of different scales. By effectively combining low-level position information and high-level semantic information, new feature maps of different scales were obtained. Experimental results show that the design of the network model accuracy of 78.76%, compared with the original SSD network model, a 9.27% drop in accuracy, but it only equivalent to the SSD network model and the number of a quarter, at the expense of the part accuracy, the design of the network model for lightweight, for the follow-up on the mobile end or embedded platform development laid a solid foundation.

References

[1] Zhou Feiyan, Jin Linpeng, Dong Jun. A Review of research on convolutional neural networks [J]. Chinese Journal of Computing.2017, (6):1229-1251.
[2] Chen Chao, Qi Feng. Development of convolutional neural network and its application in computer vision [J]. Computer Science. 2019, (3):63-73.
[3] Lu Hongtao, Zhang Qinchuan. A Review of the application of deep convolutional neural networks in computer Vision [J]. Data Acquisition and Processing. 2016, (1):1-17.
[4] Jing Chenkai, Song Tao, Zhuang Lei, et al. Review of Facial recognition Technology based on deep convolutional neural network [J]. Computer Applications and Software, 2018, (1):223-231.
[5] Zhou Junyu, Zhao Yanming. Overview of application of convolutional neural network in Image classification and Target Detection [J]. Computer Engineering and Applications. 2017, (13):34-41.
[6] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprintarXiv:1409.1556, 2014.
[7] Andrew G, Howard, Menglong Zhu, et al. Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprintarXiv: 1704.04861, 2017.