ISSN: 2664-9640

YOLOV3 traffic sign recognition and detection based on FPN improvement

Ao Qin, Yu Jun Zhang

University of Science and Technology Liaoning, Liaoning Anshan, China

Abstract

This paper applies the Tensorflow deep learning framework and adopts the target detection algorithm in deep learning to solve the problem of road traffic sign identification. Improve FPN on the basis of YOLOV3, change the fusion method from concat to ADD, and finally get the mAP value of 94.2.

Keywords

Tensorflow, YOLOV3, Improve the FPN, Fusion method, mAP.

1. Introduction

Target detection algorithms based on deep learning can be divided into two categories: target detection framework based on classification represented by R-CNN series; The target detection framework based on regression represented by YOLO and SSD algorithm.

In 2017, in terms of multi-feature fusion, Goxia[1] linearly fused the image orientation gradient histogram features and local binary mode features extracted to obtain new features. In 2018, Yao Hanli[2]proposed a method based on feature fusion and dictionary learning. In Literature 3, a transfer learning method based on CNN was proposed. Reference 4 proposes a multi-resolution feature fusion network structure, which can learn features more effectively for small size objects.

In June 2019, Tian Feng[5]chose a new loss function focal loss to replace the cross-entropy loss function in the original model.The improved YOLOV3 detection accuracy in the German traffic sign detection data set can reach 71%.In December 2019, Yang Jinsheng[6]proposed a lightweight traffic sign detection network based on YOLOV3-Tiny's depth separable convolution, the average accuracy of small and medium-sized traffic signs increased by 14.01%. In July 2020, Deng Tianmin[7]improved Dartnet53 network structure and introduced GIoU, an evaluation index, to guide orientation tasks, The improved YOLOV3 represents an 8% increase in average accuracy on the standard Lisa dataset. In March 2020, Bai Shilei proposed a lightweight YOLOV3 traffic sign detection algorithm, which compressed the model through pruning algorithm, reducing the weight of the model by 70% and saving the detection time by 90%[8]. In August 2020, Zhao Kun proposed a real-time adaptive image enhancement and optimization YOLOV3 network combined traffic sign detection and recognition method[9] , which improved recall rate and accuracy by 0.96% and 0.48%, respectively.

2. Deep leanring framework

2.1. YOLOV3

YOLOV3 used 5 residual pieces to form Darknet-53 and used the thought of residual neural network for reference. YOLOV3 adopts up-sampling and fusion method to predict the category results, and uses three different scales of fusion to detect the target. The detection effect of objects of different sizes and occluded objects is enhanced, and the jump layer connection is introduced to enhance the convergence effect[10] .

ISSN: 2664-9640

2.2. Analysis of the network structure of YOLOV3

In figure 1 on the left side of the blue red numbers in the first line of each module respectively the number of residual block of the network, in the blue box conv2D block for convolution module, upSamping2D for sampling, the characteristics of the green box for dartnet output figure and the characteristic of the sampling figure to concat feature fusion, finally yellow box for convolution, the output of the final three characteristic figure, including the size of the convolution kernels for 1 * 1 and 3 * 3, YOLOV3 overall network as shown in figure 1, the residual network structure as shown in figure 2.

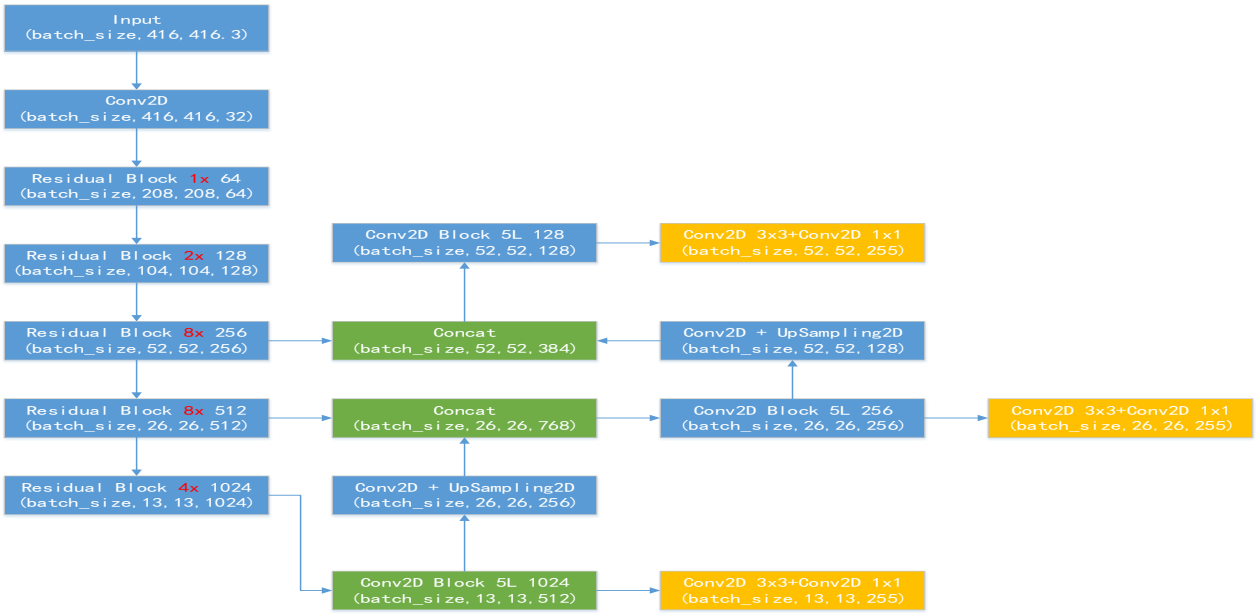

Figure 1: YOLOV3 network architecture

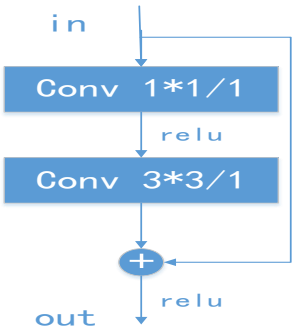

Figure 2: Residual network

2.3. Improvement method based on FPN

2.3.1. FPN

The purpose of FPN is to combine the high-level feature map with the low-level feature map in a certain way to obtain the feature map with the balance of resolution and semantic information to achieve better detection effect. The FPN proposed by Lin FPN integrates the high-level features with low resolution but rich semantic information and the low-level features with less semantic information but high resolution through lateral connection[11].

FPN is mainly divided into two processes: 1. 2. Top-down process and side connection.

Bottom-up process: as shown in figure 3 on the left side of the structure, the bottom-up process is prior to transmission of network, the forward calculation of convolution neural network channels, the process of using multiple pooling tong to extract features so as to get a different

ISSN: 2664-9640

size chart, in the process of the propagation characteristics of figure after some layer becomes smaller, and some layers will not decrease the size of feature maps.

Top-down process and lateral connections: as shown in figure 3 FPN pyramid structure, first of all, will be on the right side of the upper figure on sampling operation, then left with a size 1 x 1 convolution kernels from the bottom up to generate the feature of the map after the convolution results with the results of the sampling on the characteristics of the fusion, the fusion specific operation FPN lateral connection structure diagram as shown in figure 4. After fusion, the convolution of 1×1 and 3×3 will be used to check each fusion result for convolution, in order to eliminate the aliasing effect of up-sampling[12].

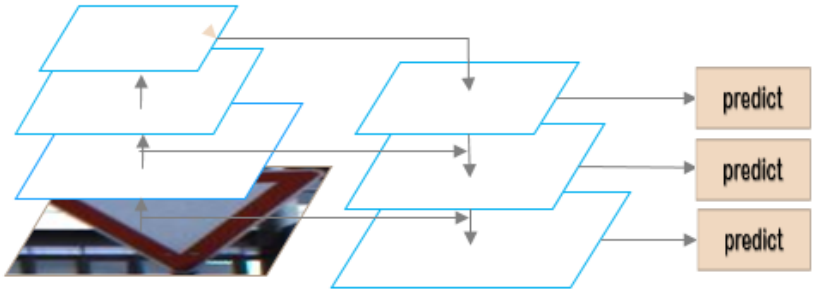

Figure 3: FPN structure

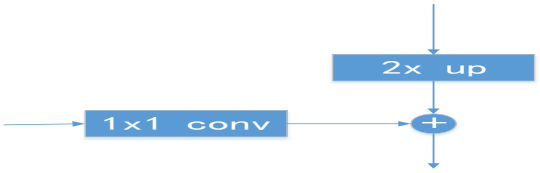

Figure 4: FPN connection structure

2.3.2. The FPN YOLOV3

The FPN in YOLOV3 is different from the FPN in Aiming he's paper. The FPN proposed by Kaiming He in his paper[13]is the eltwise operation between high-order features and low-order features after upsample, i.e. the fusion operation of addition. After upsample, high-order features in YOLOV3 are concat operation, namely the splice operation of channel direction.

2.3.3. Improvements to FPN

FPN pyramid structure is to enhance the resolution of the deep characteristic figure, enrich the semantic information of feature maps better forecast target, use the concat method can make the figure characteristics of the channel number increase, increase the amount of calculation, and the add increase the amount of information features figure, not increase the features of the channel number, the add method is a better choice, as shown in figure 5 and figure 6, the add method and concat method features fusion of 3D structure diagram.

Figure 5: Concat method

Figure 6. The Add method

ISSN: 2664-9640

The two figures above better describe the different features of the add fusion method and concat fusion method, while the formula below can more directly understand the difference between the two methods. The convolution kernel of each output channel is relatively independent, so we can only see the output of a single channel.Suppose the two input channels are X_1, X_2, ..., X_C and Y_1, Y_2, ..., Y_C. Then the single output channel of concat is (* denotes convolution) :

$$Z_{concat} = \sum_{i=1}^{c} X_i * K_i + \sum_{i=1}^{c} Y_i * K_{i+c} \quad (1)$$

The single output channel for Add is:

$$Z_{add} = \sum_{i=1}^{c} (X_i * Y_i) * K_i = \sum_{i=1}^{c} X_i * K_i + \sum_{i=1}^{c} Y_i * K_i \quad (2)$$

It can be seen from the above formula that :(1) the add method has more advantages for the detection and classification of the final target, because the amount of information under the feature description of the image increases, but the dimension of image description itself does not increase, only the amount of information under each dimension increases. (2) Concat is the combination of the number of channels, that is, the features describing the image itself are increased, while the information under each feature is not increased. (3) Therefore, ADD is equivalent to adding a prior. When the two input channels have similar feature of feature graph semantics of the corresponding channel, ADD can be used to replace concat, which saves more parameters and calculation. Concat's parameters and calculation amount are nearly twice as much as ADD.

Figure 7. dataset pictures

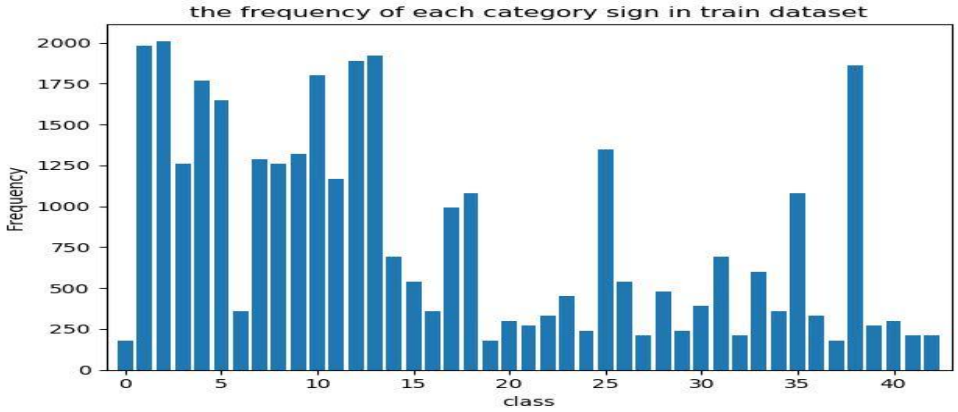

Figure 8. Number distribution of 43

2.4. German traffic dataset

The training set contains 39,209 pictures of traffic signs labeled as 43 categories. The specific pictures of traffic signs are shown in Figure 7. The number distribution of pictures of each category is shown in Figure 8, which shows that the data distribution of the data set of German

ISSN: 2664-9640

traffic signs is not balance. Some of the data is overexposed, too dark,unclear and other problems.

3. The Experiment

The experimental equipment uses the deep learning computer in the laboratory. The specific parameters of the computer are shown in Table 1.

Table 1: Deep learning device diagram

equipment	The parameter value
The graphics card	GeForce RTX 2070 SUPER(8G)
The processor	i9-9900k CPU @3.60GHZ 3.60GHZ
memory	64 GB
system	Windows 10 64位

3.1. Experimental results and analysis

The general parameters of YOLOV3 in this experiment are shown in Table 2. First, the Faster-RCNN was used to train again after adjusting the parameters, then the original structure of YOLOV3 was trained, and finally the training was modified according to the method in this paper.

Table 2: The YOLOV3 generic parameter

Parameter names	The parameter value
BATCH_SIZE	4
TRAIN.INPUT_SIZE	416
TRAIN.LR_INIT	1e-4
TRAIN.LR_END	1e-6
TRAIN.WARMUP_EPOCHS	3
TRAIN.EPOCHS	30

Evaluation standard:mAP

First, the Average Precision (AP) of each category is calculated, and then the Average Precision of all categories is obtained.

A picture a category target number a (Total objects), the number of correctly predicted for b (True Positives), then the model for the accurate rate of the class P (Precision) is b/a, such as formula (1) according to the number of data sets, each image has the accurate rate of this category, the category of the accurate rate averaged all images, the average is the average accuracy of the class, such as formula (2). Using these Average Precision values to judge the performance of the model for any given category, the data set generally has multiple categories, and the Average Precision of each category can be calculated, and the Mean Average Precision of each category can be calculated by adding the Average Precision of each category and calculated its average, which is the evaluation standard mAP (as shown in Formula (3)).

$$\mathrm{p} = \frac{\text{True positives}}{\text{Total objects}} \tag{1}$$

$$\mathrm{AP} = \frac{\sum p}{\text{Total images}} \tag{2}$$

ISSN: 2664-9640

$$\text{mAP} = \frac{\sum AP}{\text{Total classcs}} \tag{3}$$

Experimental analysis: The concat method of FPN's feature fusion is replaced by ADD, which increases the amount of information under the features of the description image and reduces the computation. The increase of mAP also reduces the computational complexity of the model and improves the detection speed. The original YOLOV3 model mAP on the left is 93.49%, the improved FPN is on the right, and the improved YOLOV3 mAP is 94.24%. The specific detection results of the improved YOLOV3 network are shown in Figure 9, and the average accuracy of each frame is shown in Table 3.

Table 3: Different framework mAP

Experiment	mAP
Faster-RCNN vgg16	81.8
Improvement of Faster-RCNN vgg16	90.8
YOLOV3+FPN	93.49
Improvement of YOLOV3+FPN	94.24

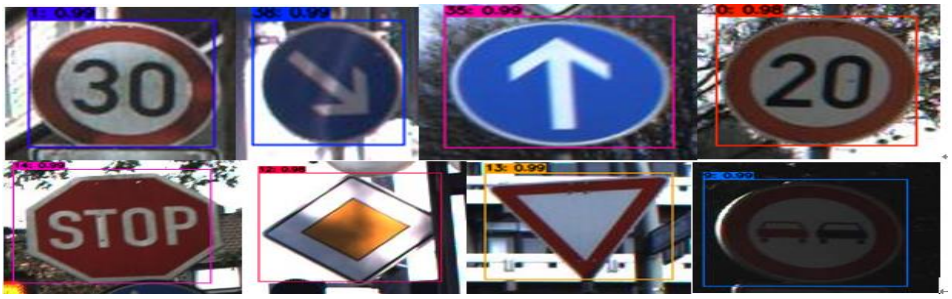

Figure 9. detection renderings

4. Conclusion

Study of traffic signs is currently the important direction of unmanned, this article is based on the present mainstream framework to study German traffic sign detection and classification, proposed the YOLOV3 + FPN change, the method of comparison YOLOV3 + FPN change original YOLOV3 improve the mAP value 0.75, compared to 12.44% higher than that of Faster-RCNN mAP value, compared with Faster-RCNN increased by 3.44%, after RCNN adjustable parameter in this paper, the methods for traffic signs based on YOLOV3 future research have good reference, but due to the diversity of samples, There are still deficiencies in some areas that need further study and improvement.

Acknowledgements

I would like to extend my deep gratitude to Professor Yu Jun Zhang who have offered me a lot of help and support in the process of my thesis writing.

References

[1] Ge Xia, YU Fengqin, Chen Ying. Traffic Sign Recognition based on Block Adaptive Fusion Feature [J]. Computer engineering and applications, 2017,53 (3) : 188-192.

[2] Yao Hanli, ZHAO Jin-jin, Bao Wen-xia. Traffic Sign Recognition based on Feature Fusion and Dictionary Learning [J]. Computer Technology and Development, 2012, 28(1) : 51-55.

ISSN: 2664-9640

[3]LIU W, LUR, LIU XL. Traffic sign detection and recognition via transfer learning [C]. 2018 Chinese Con-trol and Decision Conference, 2018: 5884 −5887.
[4]YUAN Y, XIONG Z, WANG Q. VSSA −−NET: vertical spatial sequence attention network for traffic sign detec-tion[J]. IEEE Transactions on Image Processing, 2019, 28(7) : 3423 −3434.
[5] Tian Feng, LEI Yinjie, DENG Qi. Research on natural road condition information recognition based on YOLOV3 [J]. Computer application research,2020,37(S1):391-393.
[6] Yang Jinsheng, Yang Yannan, Li Tianjiao. Traffic sign recognition algorithm based on deep separable convolution [J]. Liquid crystal & display,2019,34(12):1191-1201. [7] zhang xiuling, zhang ya-fu, zhou kaixuan. Cnn-squeeze traffic sign image recognition based on region of interest [J]. Traffic and Transport Systems Engineering and Information,19,19(03):48-53.
[7] Deng Tianmin, ZHOU Zhenhao, FANG Fang, WANG Lin. Study on improved Traffic Sign Detection Method of YOLOV3 [J/OL]. Computer Engineering and Application :1-11[2020-10-12]
[8] Bai Shilei. Research on Traffic Sign Detection and Recognition Algorithm Based on Deep learning [D]. Changchun University of Technology,2020.
[9] Zhao Kun, LIU Li, MENG Yu, SUN Ruo-can. Traffic sign detection and identification under low light conditions [J]. Journal of engineering science,2020,42(08):1074-1084.
[10] Yue Xiaoxin, JIA Junxia, Chen Xidong, LI Guangan. Improved YOLO V3 Road small target detection [J/OL]. Computer Engineering and Application :1-9[2020-09-30]
[11]LINTY, DOLLAR P, GIRSHICK R, et al. Feature pyra-mid networks for object detection [C]// Proceedings of the2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017: 936-944.
[12] Cao Yan, LI Huan, WANG Tianbao. A Review of Target detection Algorithms based on Deep learning [J]. Computer and Modernization,2020(05):63-69.
[13] Feature Pyramid Networks for Object Detection.Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge BelongiearXiv : 1612.03144 [cs.CV]