2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE) | 978-1-6654-2314-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMCCE51767.2020.00486

Smoking target detection based on Yolo V3

Jianhao Liao[1st*]
School of Automation, Wuhan University of Technology, Wuhan 430063, China
e-mail:289963@whut.edu.cn

Jiayu Zou
School of Information Engineering, Wuhan University of Technology, Wuhan 430063, China
e-mail: Zoujy@whut.edn.cn

***Abstract*—The current smoking recognition and detection system in my country is mainly based on an intelligent analysis of the video, that is, target detection and recognition of smoking behavior in the monitoring area, and early warning in the fastest way when abnormal conditions are found, therefore, the advantages and disadvantages of the detection algorithm directly affect the safety factor.The Yolo V3 model used in this article is based on DarkNet53 as the backbone. After the feature map is output, it is decoded by the Bounding Box, and finally the prediction result is obtained. The prediction results can be used as the evaluation results of the model.The prediction results show that the smoke detection effect based on Yolo v3 is good, reaching a mAP of 0.76.**

Keywords-component;Object detection;Computer vision; Image recognition;

I. INTRODUCTION

With the continuous development and maturity of target detection algorithms, the applications that have been born are also constantly emerging. For example, the aerospace industry, vehicle driving, underwater detection, etc. also apply target detection algorithms to very mature applications. The research in this paper is the application innovation of smoking detection on the target based on the Yolo V3 algorithm on smoking behavior of smokers in public places (specifically to cigarette butts), which is beneficial to the detection of smoking behavior in public places and facilitates subsequent processing. The reason why the Yolo V3 algorithm is used is because in the target detection scenarios we explored, cigarette butts are always small targets for the target detection algorithm most of the time, and Yolo V3 uses the FPN structure to improve the accuracy of the corresponding multiple scales. Small target detection can detect target objects of different sizes more accurately than the previous Yolo V1[1] and Yolo V2[2]. Therefore,the characteristics of Yolo V3 to meet the requirements of high efficiency and high accuracy in the detection of small targets such as cigarette butts in the Specific scene smoking scene we studied.

II. YOLO V3 MODEL BUILDING DESIGN

A. *Yolo V3 Network*

The overall Yolo V3 network architecture is shown in Figure 1. We only introduce the 3 core parts of the network.

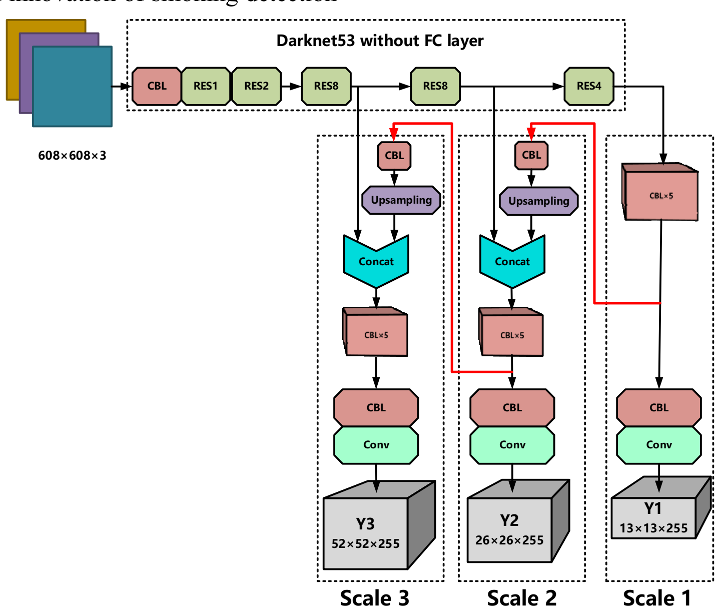

Figure 1 Yolo V3 network architecture

978-1-6654-2314-4/20/$31.00 ©2020 IEEE
DOI 10.1109/ICMCCE51767.2020.00486

1) CBL network: The basic components of Yolo V3

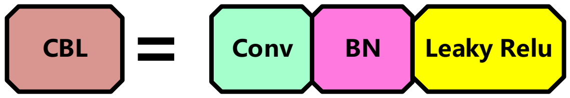

Figure 2 CBL network component

The CBL network component in the figure 2 is the basic component of the Yolo V3 Network. The specific workflow is that the input is convolved through the convolutional neural network and then Batch Normalization is performed. Before the data is input to the neuron, it is translated, stretched and transformed, and the data The distribution of is normalized to the standard distribution in a fixed interval, and finally activated by the Leaky Relu function, which reduces the probability of neuron inactivation and greatly improves the performance of the model.

2) BackBone：Darknet53

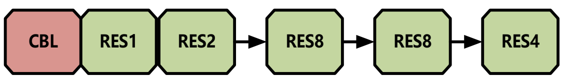

Figure 3 The Darknet53 without FC layer

Because in the process of superimposing the network, when the deep network starts to converge, there will be a degradation problem which is as the number of network layers increases, the accuracy drops immediately from reaching saturation. The reason is not because of overfitting but because of adding too much the number of layers will cause the training error to become higher. The figure 3 is the backbone network Darknet53 for extracting features. It draws on the idea of residual network Resnet[3], and its principle is shown in the figure 4 below.

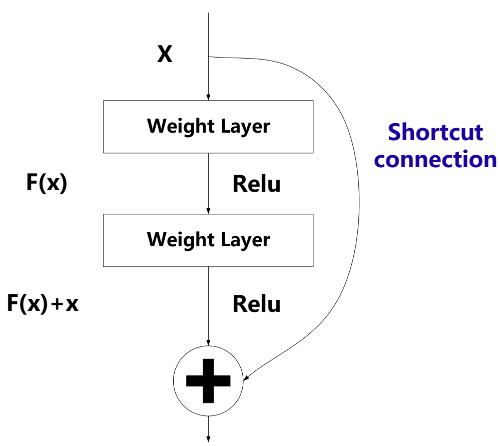

Figure 4 Resnet component

RESn is the residual component, n is the number of residual components, each residual component has two Conv and one Shortcut connection. This residual structure can just solve the aforementioned network degradation problem.

3) Three scales: Adaptive detection based on target size

Since the deeper the network, the smaller the feature map, the smaller the object, the more difficult it is to detect. Therefore, Yolo V3 draws on the idea of feature pyramid maps,and the FPN network structure is shown in the figure 5.Small-size feature maps are used to detect large-size objects, and large-size feature maps are used to detect small-size objects. Back to the figure 1, the Y1 in scale1 is a small-size feature map for detecting large-size objects, the Y2 in scale2 is a medium-size feature map for detecting medium-size objects, and the last Y3 in scale 3 is a small-size feature map for detecting small-sized objects.

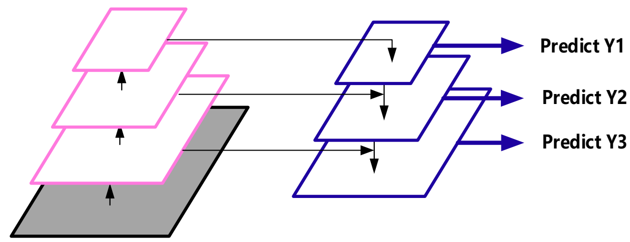

Figure 5 Feature Pyramid Network

B. *Bounding Box*

Bounding Box is shown in the red frame in the figure 6.

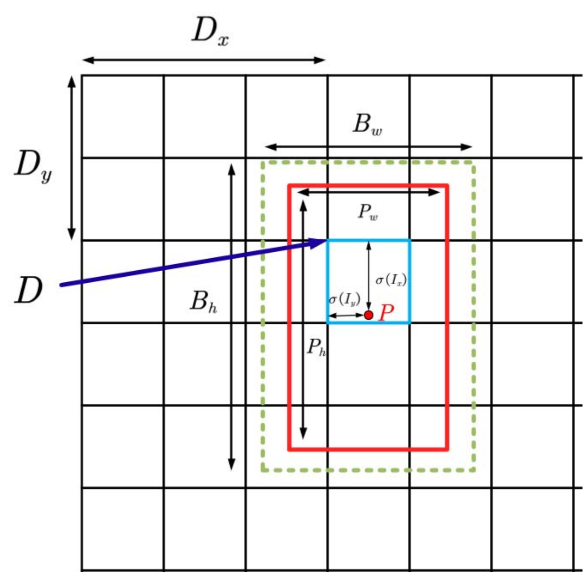

Figure 6 Bounding Box prediction schematic

Yolo V3 first divides the picture into N×N blocks, and then uses the result of the border clustering of the data set as the prior border. The dotted box in the figure 6 is the prior box Anchor Box, Bw obtained by border clustering. And Bh are the height and width of the Anchor Box respectively. The target center coordinates of the Anchor Box are the coordinates of point D in the upper left corner of the blue square pointed by the arrow in the figure 6. The Bounding Box is based on the Anchor Box, the feature vector is calculated through the neural network, combined with the regression method, and the box is corrected. The network first predicts 4 coordinates for each bounding box: Ix, Iy, Iw, and Ih. If the offset between the grid of the target center of the Anchor Box and the upper left corner of the image is (Dx, Dy), and the width and height of the corresponding prior bounding box Anchor Box are Bw, Bh, then the prediction of the Bounding Box is defined The box information will be calculated by the following formulas:

$$P_x = \sigma(I_x) + D_x \quad (1)$$

$$P_y = \sigma(I_y) + D_y \quad (2)$$

$$P_w = B_w \cdot e^{I_w} \quad (3)$$

$$P_h = B_h \cdot e^{I_h} \quad (4)$$

The σ in the above formula is the sigmoid activation function, and exp(Iw) and exp(Ih) are the scaling ratios.

III. The Result and Discussion of Yolo V3 Training

A. Implementation platform and environment

Table 1 Implementation platform and environment

Processor	**Intel（R）Core（TM）i5-8300H CPU @2.30GHz**
System	**Windows 10（64 bit）**
Programming Language	**Python3.7.4**
Environment	**Anaconda3（64 bit）**
Framework Construction	**Paddlepaddle**
Network compression method	**Pruning and quantification**

B. Experimental result graph

Smoke target detection results based on Yolo V3 are shown in Fig. 7, FIG. 8 and Fig. 9.

Figure 7 Left view

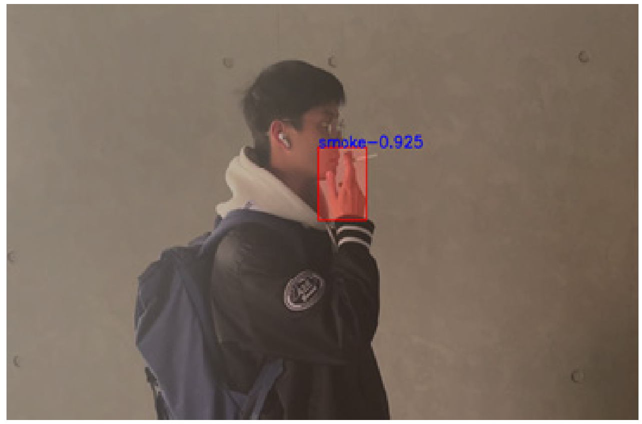

Figure 8 Right view

Figure 9 Main view

C. Experimental results data

Table 2 Results data of Smoke target detection

mAP	**0.76**
FPS	**100Hz**

IV. Conclusion

The Yolo V3 network used in this article can perform target detection for smaller targets, that is, the cigarette butts in the context of this article. Because DarkNet53 in Yolo V3 includes an up-sampling step, the speed is also faster, and the performance is stronger than the previous Yolo algorithm. , When we adjust the ratio of the training set and the validation set to 9:1, the mAP of the Yolo V3 using the DarkNet53 model for smoking detection reached 0.76, and the reasoning speed reached 100HZ. In short, Yolo V3 has good results in the detection of small targets. It can be considered that after extensive application and development, the algorithm can identify smoking behavior in specific occasions earlier, thereby reducing the probability of safety accidents. On the other hand, it is also conducive to the diversified development of the application of various target detection algorithms in the future.

Acknowledgment

The resecarch presented here was supported by "National innovation and entrepreneurship program for college students" (Item Number No. S202010497230).

References

[1] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[2] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[3] 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.

[4] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[5] Lin W H , Zhong J X , Liu S , et al. RoIMix: Proposal-Fusion among Multiple Images for Underwater Object Detection[J]. 2019.

[6] Plastiras G , Kyrkou C , Theocharides T . EdgeNet: Balancing Accuracy and Performance for Edge-based Convolutional Neural Network Object Detectors[C]// the 13th International Conference. 2019.

[7] Nie J , Anwer R M , Cholakkal H , et al. Enriched Feature Guided Refinement Network for Object Detection[C]// International Conference on Computer Vision. 0.

[8] Tan M , Pang R , Le Q V . EfficientDet: Scalable and Efficient Object Detection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[9] Liu S , Huang D , Wang Y . Learning Spatial Fusion for Single-Shot Object Detection[J]. 2019.

[10] Wang J , Zhang W , Cao Y , et al. Side-Aware Boundary Localization for More Precise Object Detection[J]. 2019.

[11] Ma W , Tian T , Xu H , et al. AABO: Adaptive Anchor Box Optimization for Object Detection via Bayesian Sub-sampling[J]. 2020.

[12] Zhu C , Chen F , Shen Z , et al. Soft Anchor-Point Object Detection[J]. 2019.

[13] Wei F , Sun X , Li H , et al. Point-Set Anchors for Object Detection, Instance Segmentation and Pose Estimation[J]. 2020.

[14] Yang X , Yan J . Arbitrary-Oriented Object Detection with Circular Smooth Label[J]. 2020.

[15] Sindagi V A , Oza P , Yasarla R , et al. Prior-based Domain Adaptive Object Detection for Hazy and Rainy Conditions[J]. 2019.