ORIGINAL RESEARCH PAPER
Volume-7 | Issue-2 | February-2018 | PRINT ISSN No 2277 - 8179
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

A FAST OBJECT DETECTION METHOD BASED ON DEEP RESIDUAL NETWORK

Computer Science

Xuzheng Hao School of Computer Science and Software Engineering, Tianjin Polytechnic University, China.

ABSTRACT

Real-time Convolution Neural Network object detection methods with limited computing devices still have in less accuracy and low detection speed. Considering the above problem, a new method based on Deep Residual Network and tiny-YOLO is proposed. The new model integrates deep residual network with pre-activation, which has more layers of convolution neural networks but less weight parameters. The experiment has been conducted on the authoritative PASCAL VOC dataset, the experimental results show that the average intersection over union is improved by 0.29% and the recall rate by 2.44% with quick detection speed, which verifies the effectiveness of the method.

KEYWORDS

Object Detection; Deep Residual Network; YOLOv2; Convolution Neural Network

INTRODUCTION

Object Detection as the primary function of the interactive system is used in autonomous driving and AI robots. Traditional object detection methods are hardly applied to multi-object detection because of complex priori characteristics. Convolution neural network can extract features for multi-object detection automatically[1]. The representative CNN models include OverFeat[2], Faster R-CNN[3,4,5], SPP-net[6], YOLOv2[7,8] and SSD[9].

Object detection method pay more attention to accuracy while ignoring object detection speed, especially the frame rate of object detection models based on limited computing devices. In this paper we leverage Residual Network[10,11](Resnet) and tiny-YOLO[12] to evolve a new real-time object detection method for the limited devices. Experiments show that the proposed method has better average intersection over union(IoU) and recall with fast frame rate, the model is more flexible and practical.

RES-YOLO METHOD

YOLOv2 has a excellent performance in object detection and Resnet deals effectively with classification tasks. In our paper we attempt to improve a fast YOLO method which combines with Resnet v2 and tiny-YOLO, the YOLOv2 simplified version, Res-YOLO object detection method. Bottleneck residual blocks which deepen the network hierarchy and reduce the weight parameters moderately is used in Res-YOLO. The blocks help our model to learn higher-level abstract features to improve the detection accuracy.

TABLE – 1
TINY-YOLO AND RES-YOLO

layer	tiny - YOLO	Res - YOLO
1_x	3×3×16c 2×2/2m	3×3×16c 2×2/2m
2_x	3×3×32c 2×2/2m	3×3×32c 2×2/2m
3_x	3×3×64c 2×2/2m	1×1×32c 3×3×32c 1×1×128c
4_x	3×3×128c 2×2/2m	(1×1×64c 3×3×64c 1×1×256c) ×2
5_x	3×3×256c 2×2/2 m	(1×1×128c 3×3×128c 1×1×512c) ×2
6_x	3×3×512c 2×2/2m	1×1×256c 3×3×256c 1×1×1024c
7_x	3×3×1024c	

The classification network of our method refers to the Resnet. Bottleneck residual block is used to extract higher hierarchical features with fast speed, instead of a very deep but very slow network with depth residual network. Our model is designed as 6 depth residual blocks with preactivation[11](Figure 1). In order to reduce the weight parameters, the head of classification network is designed as a 3×3×16 convolution layer, a 2×2 maximum pooling layer, a 3×3×32 convolutional layer and a 2×2 maximum pooling layer. Our model is compared with tiny-YOLO in Table 1(convolution layer is abbreviated as c, and maximum pooling layer is abbreviated as m).

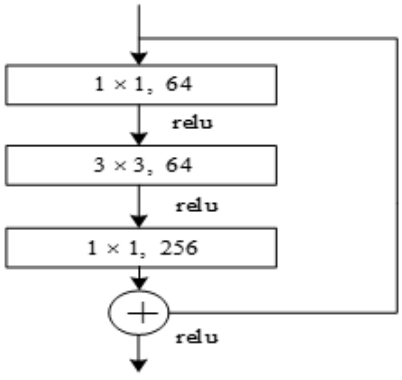

Figure 1: depth residual block

Our network uses detection part of tiny-YOLO. Detection part of our method includes a 3×3×1024 Convolution layer, a 13×13×125 intermediate layer and a object prediction layer. It can effectively improve the model performance. The above method of combining Resnet and tiny-YOLO detection network is named as Res-YOLO object detection method, the structure of our network shown in Figure 2:

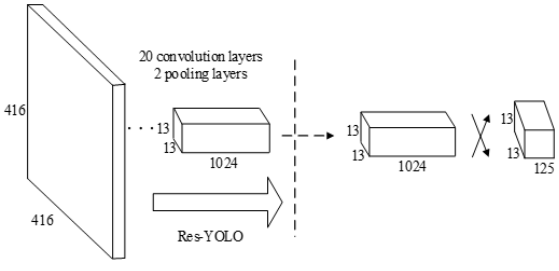

Figure 2: Res-YOLO model

EXPERIMENT AND RESULT

Res-YOLO network is based on Darknet, a open source deep learning framework. The learning rate decay strategy is a polynomial decay with the power of 4, the initial learning rate of 0.01; the momentum parameter is set to 0.9, a weight decay parameter is set to 0.0005.

In this paper, PASCAL VOC dataset was chosen as experimental dataset. The PASCAL VOC 2007 training dataset and the PASCAL VOC 2012 dataset were used to the experiment training dataset. The VOC 2007 test dataset was used as the test dataset of this experiment.

EVALUATION OF RESULTS:

In the experiment the recall rate and average IoU of our model and tiny-YOLO are evaluated. Recall and IoU are the most important indicators in evaluation of the object detection. Recall rate is a category accuracy index, the greater the recall rate is, the more correct predictions are. The average IoU is the index of positioning accuracy. The larger average IoU means the more accurate position of the system.

As can be seen from Table 2, the average IoU of our method is 62.20%, 0.29% higher than that of tiny-YOLO method, and the recall rate is 79.33%, which was 2.44% higher than tiny-YOLO. Because the tiny-YOLO classification network only has 7 layers, while Res-YOLO network has 20 layers of convolution neural networks. Res-YOLO can extract high-level features better, so that the average IoU and the recall rate are all improved.

TABLE – 2
COMPARISON OF RECALL AND AVG_IOU

model	Recall	Avg_IoU
Tiny-YOLO	76.89%	61.91%
Our method	79.33%	62.20%

EVALUATION OF REAL-TIME:

For the object detection method, real-time speed has practical significance in the low-latency scenarios. Therefore, this paper evaluates the real-time performance between our model and tiny-YOLO in Table 3. Our method can reach 44.2fps, the same with 44.5fps of tiny-YOLO (based on 1080p video and GTX1070). Although the weight parameters of our method are obviously reduced, the frame rate of the method has not changed significantly. Because the increase of convolution layers leads to the frame rate remain unchanged.

TABLE – 3
FPS COMPARISON

model	layers	weight	frame rate
Tiny-YOLO	7	63.5MB	44.5fps
Our method	20	45.0MB	44.2fps

CONCLUSIONS

In this paper, a fast object detection method based on deep residual network and tiny-YOLO is proposed. The average IoU and the recall rate are all improved with the real-time performance, especially the recall rate promoted obviously. And our method can also be applied to video object detection. Next, we will do further research in improving the speed of the Res-YOLO method.

ACKNOWLEDGMENTS

Supported by the National Natural Science Foundation of China (U1504613).

REFERENCES

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
[2] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & Lecun, Y. (2013). Overfeat: integrated recognition, localization and detection using convolutional networks. Eprint Arxiv.
[3] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2016). Region-based convolutional networks for accurate object detection and segmentation. IEEE transactions on pattern analysis and machine intelligence, 38(1), 142-158.
[4] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).
[5] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).
[6] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9), 1904-1916.
[7] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779-788).
[8] Redmon, J., & Farhadi, A. (2016). YOLO9000: better, faster, stronger. arXiv preprint arXiv:1612.08242.
[9] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.
[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In European Conference on Computer Vision (pp. 630-645). Springer International Publishing.
[12] YOLO: Real-Time Object Detection. (2018, January 25). Retrieved January 25, 2018, from pjreddie: https://pjreddie.com/darknet/yolo/.