SUGISAKA MASANORI

Journal of Advances in Artificial Life Robotics
Vol. 2(2); September (2021), pp. 70–73
ON LINE ISSN 2435-8061; ISSN-L 2435-8061
https://alife-robotics.org/jallr.html

Research Article

A Study of YOLO Algorithm for Multi-target Detection

Haokang Wen[1], Fengzhi Dai[1,2]
[1]*Tianjin University of Science and Technology, China;*
[2]*Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China*

ARTICLE INFO

Article History

Received 25 October 2020
Accepted 29 July 2021

Keywords

Target detection
YOLOv5
Deep learning
Computer vision technology

ABSTRACT

With the development of deep learning, target detection has become one of the research directions of many scholars. As one of the more mature algorithms, the single-stage YOLO algorithms have been widely used in real life. Combining the development history of the YOLO algorithm, this article focuses on the main framework and main content of the current latest YOLOv5 algorithm, and uses the YOLOv5s model to identify and detect multi-target. The test results show that YOLOv5s algorithm has good detection effect and wide application meaning in real life.

1. Introduction

Computer vision includes target detection, target segmentation, target tracking, image description, event detection, and activity recognition. Target detection is the cornerstone of other more complex vision tasks. Its main task is to use computers to predict: a given image and video Object, what is it or where it is. Multi-target detection is to achieve the task of detecting multiple targets.

Currently, target recognition technology is widely used in the following fields:

Security field: fingerprint recognition, face recognition, etc.

Military field: terrain survey, flying object recognition, etc.

Traffic field: license plate number recognition, unmanned driving, traffic sign recognition, etc.

Medical field: electrocardiogram, B-ultrasound, health management, etc.

Life field: smart home, shopping, etc.

2. Object detection algorithm

With the rapid development of deep learning technology, since 2012 target detection algorithms have shifted from traditional target recognition algorithms based on manual features to target recognition technologies based on deep neural networks. Object detection is one of the most fundamental and challenging problems in computer vision in recent years. A road map of object detection [1] is shown in the Fig.1.

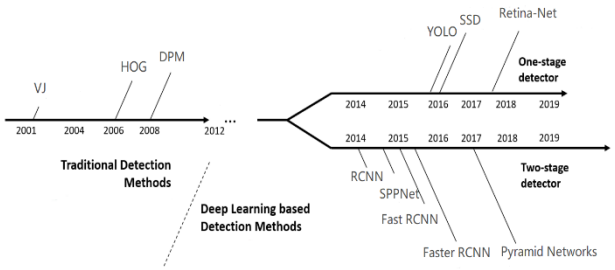


Fig.1 A road map of object detection

Single-stage algorithm and two-stage algorithm are currently two mainstream target detection algorithms based on deep neural networks.

The single-stage algorithm treats the target detection process as a regression problem, and uses a unified deep neural network for feature extraction, target classification

Corresponding author's E-mail: daifz@tust.edu.cn. URL: www.tust.edu.cn

and bounding box regression, achieving end-to-end reasoning. It has a faster detection speed, but its detection accuracy is lower than the two-stage algorithm. YOLO series algorithms and SSD series algorithms are the representative single-stage algorithm.

The two-stage target detection algorithm uses the region proposal network to extract the region of interest, that is, the region containing the target, and then uses a deep neural network to classify the region of interest and return the bounding box, which has higher detection accuracy. However, the detection speed of the two-stage target detection algorithm is too slow to meet the real-time requirements.

Comparing these two methods, the single-stage target detection algorithm is usually simpler, faster than the two-stage target detection algorithm, and has more advantages in real-time detection. Considering the real-time requirements, the YOLO algorithm is mainly introduced in this article.

3. YOLO algorithm

Since the first generation YOLOv1 algorithm was proposed in 2016, after recent years of development, there are now five main versions proposed. Next, the advantages of the various stages of the algorithm will be introduced.

3.1. *YOLO v1*

YOLOv1 proposes to detect the target through grid division, and detect the target through the position of the target center point on the grid, which significantly improves the detection speed.

Before the YOLO algorithm was proposed, object detection methods were based on the method of first generating candidate regions and then detecting. Although there is a relatively high detection accuracy rate, the running speed is slow.

YOLO creatively treats the object detection task directly as a regression problem, combining the candidate area and the detection phase into one. "You Only Look Once", YOLO really can let you know what objects are in each image and where the objects are at a glance [2].

3.2. *YOLO v2*

Compared with YOLOv1, YOLOv2 greatly improves the accuracy and speed of object detection [3]. YOLOv2 uses multi-scale feature maps to detect objects based on SSD, and proposes pass through layer to link high-resolution feature maps with low-resolution feature maps to achieve multi-scale detection.

YOLOv2 applies the anchor mechanism on the basis of grid constraints. By presetting a priori boxes of different scales, the detector focuses on detecting objects that are similar in shape to the a priori box. At the same time, the batch normalization method is used to accelerate convergence and avoid overfitting.

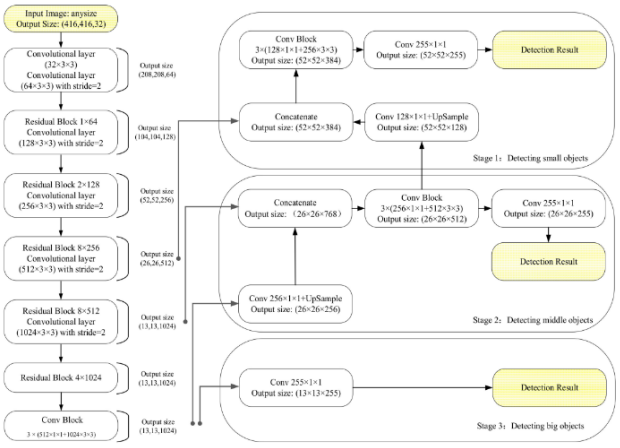
3.3. *YOLO v3*

YOLOv3 has made great progress because it has made two improvements on the basis of YOLOv2. One is to use the residual model to further deepen the network structure; the other is to use the FPN architecture to achieve multi-scale detection.

YOLOv3 uses the residual network idea in Res Net to design Darkent-53 as the backbone network for feature extraction [4]. On the basis of YOLOv2, it draws on the multi-scale idea of FPN and designs 3 different scales.

YOLOv3 predicts the position of the object through the method of frame regression prediction, which solves the problem of instability of the linear regression of the prior frame mechanism. In YOLOv3, each box uses multiple label classifications to predict a bounding box might contain which classes [5]. YOLOv3 architecture [6] is shown in the Fig.2.

11Fig.2 YOLOv3 architecture



3.4. *YOLO v4*

The backbone network CSPDarknet53 of YOLOv4 is the core of the algorithm and is used to extract target features. Drawing lessons from the CSPNet's experience in maintaining accuracy, reducing computational bottlenecks and memory costs, YOLOv4 adds CSP to each large

residual block of Darknet53, divides the feature map of the base layer into two parts, and merges them through a cross-stage hierarchy. This reduces the amount of calculation and ensures accuracy. The activation function of CSPDarknet53 uses the Mish activation function, and the following network uses the leaky_relu function, so that the setting is more accurate in target detection.

Unlike the YOLOv3 algorithm that uses FPN for upsampling, YOLOv4 draws on the idea of information circulation in the PANet network. First, the semantic information of high-level features is propagated to the low-level network through up-sampling, and then it is fused with the high-resolution information of the underlying features to improve the detection effect of small targets. Then increase the information transmission path from the bottom to the top, and enhance the feature pyramid through downsampling. Finally, the feature maps of different layers are fused to make predictions.

3.5. *YOLO v5*

Compared with YOLOv4, YOLOv5 has a higher accuracy rate and better ability to recognize small objects [7]. YOLOv5 is more flexible and faster than YOLOv4, and has great advantages in the rapid deployment of models. There are four network models in YOLOv5. Model design of different complexity can be realized by adjusting depth and width multiple parameters.

The YOLOv5 detection network uses CSPDarknet as the feature extraction network to extract rich information features from the input image. CSPNet solves the gradient information duplication problem of network optimization in other large-scale convolutional neural network frameworks, and integrates the gradient changes from the beginning to the end into the feature map, thus reducing the amount of model parameters and FLOPS values. This not only ensures the speed and accuracy of inference, but also reduces the scale of the model.

YOLOv5 proposes to use the Fcos algorithm to participate in the calculation of the frame selection area, which greatly improves the detection efficiency. And through image enhancement, new training samples are generated from the existing training data. Various advanced data enhancement techniques are used to maximize the use of data sets to achieve a breakthrough in the performance of the target detection framework. Through a series of image enhancement technology steps, the performance of the model can be improved without increasing the reasoning delay.

YOLOv5 uses the CSPDarknet feature extraction network to effectively extract image features, and uses BottleneckCSP instead of shortcut residual connection to strengthen the description of image features. The Neck module is mainly used to generate feature pyramids. The feature pyramid can enhance the model's detection of objects of different scales, thereby being able to identify the same object of different sizes and scales.

4. The test of YOLO v5

Based on the environment that is shown in Table 1, the YOLOv5s model was built based on the Pytorch framework.

Table 1. Computer environment

GPU	NVIDIA GeForce MX150
Video memory	2 GB
operating system	Windows 10
CUDA architecture	CUDA 10.2

In this article, 4000 and 1000 pictures in the PASCAL VOC data set are used as the training set and the test set. After training the neural network and using YOLOv5s algorithm, actual tests were performed on multi-target detection. The test result is shown in the Fig.3.

Fig.3 The test result

Experiments show that the YOLOv5 algorithm can detect different objects in complex scenes and has good results and can accurately identify objects. Overall, I think the algorithm is very user-friendly and can easily train our own data set.

5. Conclusion

After years of development, the YOLO algorithm has been continuously improving the network structure to

maintain the advantage of faster detection speed while maintaining high accuracy. As an excellent target detection algorithm, YOLOv5 algorithm has very considerable prospects in future detection work. In theory, this method has a wide range of application value in real life.

References

1. Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: A survey, *arXiv preprint arXiv*: 1905.05055, 2019.
2. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: pp.779-788.
3. Sang J, Wu Z, Guo P, et al. An improved YOLOv2 for vehicle detection. Sensors, 2018, 18(12): 4272.
4. Jing, Junfeng, et al. Fabric defect detection using the improved YOLOv3 model, *Journal of Engineered Fibers and Fabrics*, 2020(15): 1558925020908268.
5. Yuan Y, Dai F, Song Y, et al. On Fatigue Driving Detection System Based on Deep Learning, *Chinese Intelligent Systems Conference*. Springer, Singapore, 2020: pp.734-741.
6. Li Y, Zhao Z, Luo Y, et al. Real-Time Pattern-Recognition of GPR Images with YOLO v3 Implemented by Tensorflow, *Sensors*, 2020, 20(22): pp.6476.
7. Yifan Liu, BingHang Lu, Jingyu Peng, et al. Research on the Use of YOLOv5 Object Detection Algorithm in Mask Wearing Recognition, *World Scientific Research Journal*, 2020, 6(11).

Authors Introduction

Mr. Haokang Wen

He is the second-year graduate student of Tianjin University of Science and Technology. His major is information processing and Internet of Things technology. His main research field is digital image processing and computer vision. During his studies in school, he published several research papers.

Dr. Fengzhi Dai

He received an M.E. and Doctor of Engineering (PhD) from the Beijing Institute of Technology, China in 1998 and Oita University, Japan in 2004 respectively. His main research interests are artificial intelligence, pattern recognition and robotics. He worked in National Institute of Technology, Matsue College, Japan from 2003 to 2009. Since October 2009, he has been the staff in Tianjin University of Science and Technology, China, where he is currently an associate Professor of the College of Electronic Information and Automation.