A Survey of Target Detection Based on Deep Learning

Hucheng Wang*, Fengzhi Dai, Min Zhao

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*E-mail: *18322744268@163.com*
www.tust.edu.cn

Abstract

Object detection is a hot topic in the field of visual detection. Deep learning can greatly compensate for the defect that traditional methods sacrifice real-time for improving accuracy. This paper mainly introduces the main networks and methods of two-stage deep learning algorithm and single-stage deep learning algorithm in the field of target detection. The advantages and disadvantages, usage scenarios and development of each network are described in detail. Finally, the follow-up development in this field is prospected..

Keywords: Machine earning, Deep learning, Object detection, Convolutional neural network

1. Introduction

Object detection is one of the most important topics in the field of computer vision. Its task is to find all interested objects in the detected image and classify them. Because the image quality is affected by weather conditions, lighting, occlusion and other factors, and different objects have different attitudes, shapes and surface roughness, target detection has always been a very hot and difficult problem in the field of computer vision.

Although the traditional method (machine learning) has been very mature in the field of object detection, it still has many problems, such as: (1) How to select the sliding window. If it is a complex practical project application, this problem will consume a lot of time and energy. (2) The robustness of manually calibrated features is poor. How to overcome the above two problems became a hot issue in the field of target detection at that time.

Depth convolution neural network can not only overcome the typical problems of target detection with traditional methods, but also greatly shorten the detection time. this model makes depth learning receive great attention, and gradually become the mainstream of researchers.

2. Two Stage Deep Learning Algorithm

The so-called "two stages" refer to: 1. Extract the object area first. 2. Then classify and identify the area by CNN. The common methods for extracting object regions are: selective search, boundary box, etc. The main feature of this algorithm is high accuracy, but slow speed. Typical algorithms include R-CNN, SPP-Net, fast R-CNN and fast R-CNN.

2.1. Region Convolution Neural Network

The Girshick team [1] first proposed this method in 2014. Its algorithm idea is very simple. First, select several candidate boxes from the original image based on the Selective Search method. Second, scale the image in each candidate box to a fixed scale and send it to the convolutional neural network for feature map. Finally, it is judged by the support vector machine (SVM), Determine whether the images in these candidate boxes are the image type we want to extract or the background.

Compared with traditional methods, the efficiency of this method is greatly improved. The disadvantages are as

follows: 1. When the candidate regions are normalized, it is easy to cause image loss. 2. CNN model parameters and classification regression parameters cannot be modified at the same time. 3. It takes a lot of disk space, making the detection time long.

2.2. Spatial Pyramid Pooling Network

In 2014, He Kaiming [2] and others put forward SPP-Net on the basis of R-CNN, which requires input convolutional neural network to extract features of the sub image to be tested with fixed size. SPP Net cancels this restriction. SPP Net has a pyramid space pyramid pool layer, which makes it possible to plan the sub image into a uniform size regardless of the size of the input sub image. Each sub image with the same size is sent to the subsequent network for special extraction, and the extracted features also have the same dimension.

Compared with CNN, SPP Net is more efficient in target recognition.

2.3. Fast Area Convolution Neural Network

In 2015, Girshick team [3] further optimized and improved on the basis of CNN and SPP Net, and proposed convolutional neural network (fast R-CNN), which can simultaneously predict the classification probability and position offset of targets in the network. The network solves two main problems of the two networks mentioned above: 1. There are many training steps and slow training speed. 2. It takes a lot of disk space, resulting in a long training time. Fig.1 is the schematic diagram of fast R-CNN algorithm.

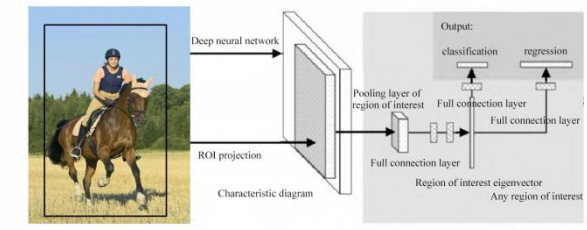

Fig.1. the schematic diagram of fast R-CNN algorithm

When using the same scale network and the same data set, fast R-CNN has faster detection speed and higher detection accuracy than R-CNN. Table 1 shows the performance comparison of R-CNN and fast R-CNN training and testing based on VGG-16 convolutional network model on the VOC2007 dataset.

Table 1. Performance comparison between VGG-16 based R-CNN and fast R-CNN algorithm

Algorithm	Training Time (h)	Test time (seconds/frame)	mAP(%)
R-CNN	84	47	66
Fast R-CNN	95	0.32	66.9

2.4. Fast Area Convolution Neural Network

This network was proposed by Ren S Q [4] team in 2015. The biggest feature of this network is that it integrates the four target detection steps into a deep network. The algorithm process can be divided into the following steps: 1. Input the tested image into the network. 2. Use regional bidding network (RPN) (Fig.2 shows the RPN algorithm) and discriminant function to obtain accurate candidate regions. 3. Unify the images to be input to the full connection layer into the same size through pooling layer. 4. Identify the category and position of the tested object in the tested image through the training of the full connection layer.

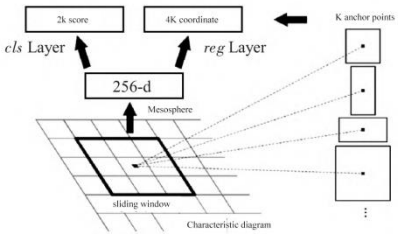

Fig.2. RPN algorithm

3. Single Stage Deep Learning Aigorithm

The single-stage deep learning algorithm refers to the end-to-end method. For a detected image, only one network is used to identify the category and location of the object to be detected, simplifying the complex steps of the two-stage deep learning algorithm, and greatly improving the efficiency. This method has excellent real-time performance. Typical single-stage deep learning algorithms include yolo and SSD.

3.1. YOLO

The two-stage detection algorithm needs to go through two steps: border regression and classification. This will generate a large number of candidate boxes, making the detection time very long despite the high accuracy. It can not meet the requirements when executing some tasks that require high real-time performance. In view of this, Joseph Redmon proposed the yolov1 model. The model cancels the candidate box and directly partitions the image to be measured. One area corresponds to one grid. A neural network is used to traverse the image and predict the border and category information of objects in each grid. Since the prediction of each border is based on the characteristics of the whole picture, the predicted object border combines the information of the whole picture, which not only improves the prediction efficiency but also has better accuracy, However, the prediction of small objects is not effective. Fig.3 is the schematic diagram of YOLOV1 network.

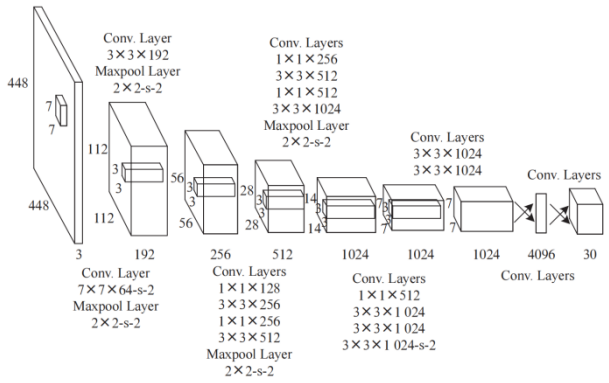

Fig.3. The schematic diagram of YOLOV1 network

In order to solve the problems such as the poor detection effect of yolov1, the yolov2 version was subsequently iterated. This version introduced the BN algorithm to improve the convergence speed of the model, and took Darknet-19 as the backbone network, instead of using the full connection layer, which simplified the network structure and improved the efficiency.

Yolo algorithm, as a classical algorithm for real-time target detection, has been continuously optimized and iterated since its appearance. So far, the yolov7 version has exceeded all known target detection algorithms both in real-time and accuracy. And it reaches 56.8% AP on COCO dataset. Fig.4 shows the performance comparison of yolo versions.

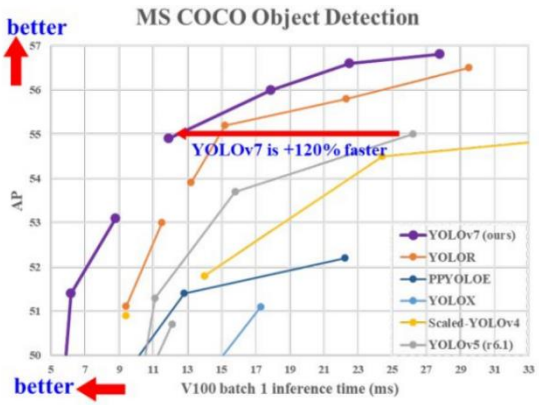

Fig.4. Comparison Chart of Yolo Performance of Different Versions

3.2. SSD

The emergence of SSD is mainly to solve the problems of low positioning accuracy and poor detection effect for small objects existing in the previous version of yolo network. In 2017, Fu C W team [5] further improved the algorithm and proposed the DSSD algorithm, which further improved the detection accuracy of small objects, but also increased the network complexity. Fig.5 is the schematic diagram of SSD network structure.

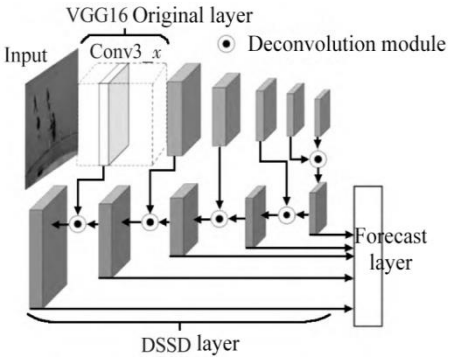

Fig.5. The schematic diagram of SSD network structure

4. Conclusion

Now, target detection is still a hot research direction in the field of artificial intelligence, and it still has huge development potential and broad application prospects. By using DHT11 module, the indoor temperature and humidity data can be real-time transmitted to the remote client.

The current target detection field still has very challenging problems, such as the lack of lightweight network model for mobile terminals; Lack of network model that can accurately detect small objects; Methods that can extract more levels and dimensions.

Acknowledgments

This paper is partly supported by the Education Reform Project (2021-JG-03) from the Teaching Guidance Committee of Electronic Information in Higher Education of the Ministry of Education, China in 2021, and the Graduate Education Reform and Innovation Project (2021YJCB02) from Tianjin University of Science and Technology, China in 2021.

References

1. GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580－587.
2. HE K M, ZHANG X Y, RRN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. European Conference on Computer Vision, Springer, Cham, 2014: 346－361.
3. GIRSHICK R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440－1448.
4. REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 2015: 91－99.
5. FU C Y, LIU W, RANG A, et a. DSSD: Deconvolutional single shot detecto. arXiv preprint, arXiv: 1701.06659, 2017.

Authors Introduction

Mr. Hucheng Wang

He is a second-year master candidate in Tianjin University of Science and Technology, majoring in machine learning.

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 College of Electronic Information and Automation, Tianjin University of Science and Technology, China.

Ms. Min Zhao

She is a master student in Tianjin University of Science and Technology. Her research is about automatic and adaptive control.