A survey of object detection algorithms based on machine learning

Li Mingbo

Changchun University of Technology，

School of computer science and technology

Abstract： Target detection is a popular technology in the field of computer vision and machine learning. The classical target recognition direction has the primary feature extraction methods, such as extracting the target's hog features, training SVM classifier for recognition, etc. the accuracy of these methods has been exceeded by machine learning for a long time. With the application of machine learning in target detection technology, a variety of excellent target detection algorithms appear, but lack of a systematic summary and comparison of these algorithms. This paper compares R-CNN, Fast RCNN, Fast RCNN, end-to-end detection algorithms such as Yolo series, SSD algorithm, and recently proposed Cascaded-RCNN algorithm, RefineDet algorithm, SNIP algorithm, R-FCN-3000 algorithm, DES algorithm, STDN algorithm. This paper analyzes the recognition accuracy and speed of each target recognition algorithm in each open data set, as well as the advantages and disadvantages of each algorithm. And a summary of target detection

algorithm is given.

Key Words： Machine Learning, Target Detection, Area Proposal

With the rapid development of machine learning technology, machine learning is applied in more and more fields. Especially in the direction of image processing, more and more people begin to use machine learning for target detection. The accuracy and speed of the emerging algorithms are constantly improving, so it is necessary to organize and compare each algorithm systematically. The purpose of target detection technology is to detect the interested target in the given picture, and lock the rectangular area and type of the target. The challenging problems of target detection technology are the different shapes, postures, the influence of light, the interference of occlusion and so on. These problems make the task of target detection more difficult. Traditional detection methods tend to use sliding windows, which mainly include: using windows of different sizes to select some areas in the image; analyzing the features of the selected areas, such as the hog feature with better pedestrian attitude effect, and the Haar feature with better facial feature recognition effect; using SVM classifier, AdaBoost classifier to classify the target. DPM algorithm of multi-scale deformation component model is the best one. DPM regards the object as many parts,

and describes the whole object with the relationship between the parts. This property is in good accordance with the non rigid nature of most objects, and has achieved good results. Then the R-CNN algorithm proposed by Ross girshick raised the accuracy of the deep learning target detection algorithm to a level. MsrA he Kaiming and others put forward SPP_net algorithm to eliminate the cross / warp operation in r-cnn, and use spatial pyramid pooling layer (SPP) to solve the problem that CNN's full connection layer requires the input image size to be consistent. Then fast RCNN algorithm and fast RCNN algorithm appeared. Fast RCNN algorithm solves the problem that RCNN algorithm needs a lot of repeated frame operations to be detected in the region proposal stage, which greatly improves the recognition speed of the algorithm. The fast RCNN algorithm directly places the region proposal in the network model, called region proposal networks, to calculate the frame to be detected. The fast RCNN algorithm abandons the original independent region proposal steps, so that it can return to the common convolution characteristics of the network with the classified network, and the detection speed can be further improved. Later, the r-fcn algorithm uses the position sensitive convolution network to replace the last full connection layer, further sharing the computation and improving the efficiency.

Based on the end-to-end method, there are a series of Yolo

algorithms famous for their speed, The original Yolo algorithm combines the recognition behavior and location behavior of the target, simplifies the detection process, realizes the end-to-end detection, and the speed can reach 45fps. SSD algorithm improves the weakness of Yolo algorithm in detecting small targets. It uses the anchor mechanism and multi-scale to improve the accuracy of the algorithm. After that, the improved Yolov2 algorithm appeared, adjusted the multi-objective framework of the Yolo algorithm, and improved the accuracy. Yolov3 algorithm [1] improves recognition accuracy, especially for small targets.

The Cascade RCNN algorithm is different from the ordinary cascade. The detection network of Cascade R-CNN is trained on the positive and negative samples determined by different IOU thresholds. The accuracy of the algorithm is excellent. Refinedet algorithm [2] combines the advantages of SSD algorithm and RPN algorithm, and improves the detection accuracy on the premise of ensuring the algorithm speed. The R-FCN-3000 algorithm introduces the location sensitive score graph to solve the repeated calculation of the original R-FCN algorithm, and improves the algorithm speed. DES algorithm is improved based on SSD algorithm, which solves the problem that the original algorithm is not sensitive to small target objects, and improves the accuracy of the algorithm. STDN algorithm [3] uses DenseNet network as feature extraction network, which improves the detection effect of target

detection algorithm on different scale targets.

1、 Accuracy comparison

The comparison of the accuracy of each target detection algorithm on each big open data set is shown in Table 1. Pascal VOC challenge data set and coco data set contain a large number of complex scenes containing targets and multiple categories of targets, which are relatively authoritative public data sets to measure the performance of the target detection algorithm.

Table 1: comparison of accuracy of target detection algorithm

算法名称	VOC2007mAP(%)	VOC2012mAP(%)	COCOmAP(%)
Fast RCNN	70.0	68.4	-
Faster-RCNN	73.2	70.4	34.9
R-FCN	79.5	77.6	29.9
YOLOV1	66.4	57.9	-
SSD	76.8	74.9	30.3
YOLOV2	78.6	73.4	-
YOLOV3	79.4	-	33.0
Cascade RCNN	-	-	42.8
RefineDet	83.8	83.5	41.8
R-FCN-3000	83.6	-	34.9
DES	81.7	80.3	32.8
STDN	80.9	-	31.8

The refinedet algorithm of Shifeng Zhang et al. Is the best in the overall accuracy of the algorithm, which is the best in voc207 and voc2012, second only to cascade RCNN algorithm in coco data set. On

the one hand, RefineDet algorithm introduces the idea of regression from coarse to fine in the target detection algorithm, that is, first through the RPN network coarse-grained rectangular box information, through the general regression branch regression to get more accurate rectangular box information. On the other hand, the method of feature fusion is introduced for network detection, which effectively improves the detection effect of small targets.

The best performing Cascade RCNN algorithm in coco data set is composed of multiple detection models, which are obtained by training positive and negative samples based on different IOU thresholds. The output of the previous model is used as the input of the later model, and the IOU threshold of the later defined positive and negative samples is rising. Using Cascade RCNN, the detector of each step can focus on detecting the proposed area of IOU in a certain range. Because the output IOU is generally larger than the input IOU, the detection effect will be better and better.

The R-FCN-3000 algorithm improves the R-FCN algorithm to adapt to more categories of detection. The limitation of the R-FCN algorithm is that once there are many categories, the calculation of the location sensitive score graph of the classification branch is relatively large. Therefore, the R-FCN-3000 decouples the generation process of the classification branch and turns the original method into two branches.

This improves the accuracy and speed. The performance of recognition accuracy on voc207 is ideal, and it is also good on coco data set.

The DES algorithm which performs well on voc207 and voc2012 is an improvement based on SSD algorithm, which solves the problem that the detection effect of SSD algorithm is not ideal for small targets. Des increases the high-level semantic features of small target recognition, and introduces segmentation module to improve the semantic information of low-level feature map. The introduction of global activation module improves the semantic information of feature map at high level. The DenseNet network used by the main network of STDN algorithm, which also performs well, introduces the scale transfer layer, and generates large-scale feature map under the premise of hardly increasing the number of parameters and the amount of calculation, which also improves the recognition accuracy.

Yorov3 algorithm can guarantee high detection speed and achieve almost 80% accuracy in the open data set. Yolov3 algorithm is the latest algorithm in the series of Yolo. It uses the end-to-end method to train, and a loss function solves the training problem, only needs to pay attention to the input end to the output end. At the same time, Yolov3 has carried out multi-scale training, and the detection accuracy for small targets is higher.

Other algorithms, such as fast RCNN algorithm, Faster algorithm and SSD algorithm, have been surpassed by the latest algorithms in accuracy

and have no advantage in speed.

2、 Speed comparison

The speed of the algorithm is also an important index to measure the practicability of the algorithm. The detection speed is too slow, which has no practical value in many application scenarios. The speed comparison of the target detection algorithm is shown in Table 2:

Table 2：comparison of speed of target detection algorithm

算法名称	速度 (FPS)
Fast RCNN	3
Faster-RCNN(low)	5
Faster-RCNN(high)	17
R-FCN	6
YOLOV1	45
SSD	19
YOLOV2	40
YOLOV3-320	45.4
YOLOV3-416	34
Cascade RCNN	8.6
RefineDet	24.1
R-FCN-3000	30
DES	31.7
STDN321	40.1
STDN513	28.6

In Table 2, on the premise of high accuracy, the fastest algorithm is the Yolo series algorithm famous for its speed. With the input of 320 × 320 resolution, Yolov3 can achieve 45.4fps speed, which is very high speed in the target detection algorithm. At this speed, it can still have high

recognition accuracy and high practicability. Both the speed of Yolov2 and Yolov1 are satisfactory, but the accuracy of them has been greatly exceeded by Yolov3.

Secondly, the STDN algorithm with 321 × 321 input can achieve 40.1 FPS, that is to say, adjusting the input to 513 × 513 can improve the accuracy and still ensure the real-time performance to reach 28.6 FPS. In the case of speed and Yolov3 algorithm, the accuracy is better than Yolov3 algorithm.

The detection speed of DES algorithm with high accuracy is also up to 31.7fps. The R-FCN-3000 with higher accuracy also achieves 30fps, and the RefineDet algorithm with the highest accuracy can achieve a good detection speed of 24.1fps. The three algorithms with the highest accuracy have achieved real-time performance and high practicability.

In contrast, SSD algorithm and fast-rcnn algorithm, which are barely real-time, have no advantages in accuracy. Like Fast-RCNN, Yolov1, Yolov2 and other algorithms, they are gradually eliminated by powerful new algorithms.

3、 Conclusion

In this paper, we make a series of comparison on the algorithm of target detection based on machine learning. In terms of algorithm

accuracy, the accuracy of today's target detection algorithm is getting higher and higher. The accuracy of classic Fast-RCNN algorithm, Yolov2 algorithm, SSD and other algorithms has been exceeded. In terms of algorithm speed, the overall algorithm speed of machine learning target detection has become faster and faster. The real-time performance is almost the price adjustment that the target detection algorithm must meet, and both guarantee the premise of high accuracy. At the same time, many algorithms can adjust the accuracy and speed by adjusting the input resolution and network structure, and can make a better balance between the accuracy and speed according to the specific use requirements.

When we pay more attention to the real-time performance of the algorithm, STDN algorithm is the most recommended algorithm, followed by DES algorithm, R-FCN-3000 algorithm, Yolov3 algorithm, and RefineDet algorithm, which has the highest accuracy. If the required target detection algorithm is more accurate, then the RefineDet algorithm and Cascade RCNN algorithm are relatively satisfactory.

4、 Reference

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

[2]Zhang S, Wen L, Bian X, et al.Single-shot refinement neural network

for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2018:4203-4212.

[3]Zhou P, Ni B, Geng C, et al.Scale-transferrable object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2018:528-537.