Research on Image Recognition of Convolutional Neural Network under Different Computer Data Set Capacities

Yuanyi Chen1,*

1 College of Internet of Things, Jiangxi Teachers College, JTC, Yingtan, Jiangxi, China, 335000

*Corresponding author e-mail: chenyuanyi@jxsfgz.edu.com

Abstract. Target detection is chinafrica growing to a task, computer vision in intelligent video surveillance, automatic monitoring and are widely used in industrial detection, over the years, with the deepening of the study and deepened, target detection algorithm based on depth of convolution neural network is gradually replacing the traditional target detection algorithm, become algorithms, which are frequently used in the field in this article, the focus on the target detection algorithm of commonly used data sets and the performance evaluation index, this paper introduces the development of convolution neural network, two stages of target detection algorithm and the unit target detection algorithm is focused on the content analysis, The future development of target detection algorithm based on deep convolutional neural network is prospected.

Keywords: Target Detection, Deep Learning, Computer Vision, Convolutional Neural Network

1. Introduction

Target detection is a basic problem in the field of computer vision and image processing. It is a hotspot of theoretical and applied research in recent years. Its main goal is to accurately locate the categories and position information of various targets in images or image sequences. There are many types of target detection tasks, such as single-class target detection, multi-class general target detection, static image target detection, video target detection and so on. The task of target detection, according to the needs of follow-up tasks, has three main levels: classification, detection, segmentation. However, such factors as target type, quantity, scale change and external environment interference pose challenges to the target detection task. In order to overcome these difficulties, many scholars have devoted themselves to the research in this field[1].

The flow of the traditional target detection algorithm includes the following stages: (1) Select the candidate box of the input image by means of sliding window; (2) The method of scale invariant feature transformation (SITF) and directional gradient histogram (HOG) was used to extract features of the image information in each window; (3) The features are classified by the classifier such as deformed part model (DPM) and support vector machine (SVM); (4) NMS algorithm is used to merge candidate boxes to realize object detection. However, target detection algorithm is rely on traditional
manual design, the characteristics of the it with the method of sliding window to select candidate box, lead to the window and the serious problem of redundant and its generalization performance is poorer, feature extraction method leads to lower precision and algorithm steps of the speed of the low detection, real-time performance, so the need for the traditional target detection algorithm for a series of improvement, in order to achieve better performance index.

In recent years, with the rapid development of deep learning, deep convolutional neural network has been more and more applied in the field of computer vision and has made great progress in this field[2].has led to the development in the field of target detection in 2012, Hinton et AlexNet deep convolution neural network for the first time application in large scale image classification task, and winning the ILSVRC image classification with RCC champion, from the convolution neural network got the favour of many scholars in computer vision. The target detection algorithm based on deep convolutional neural network USES large-scale data sets such as ImageNet, PASCAL VOC and COCO to train the convolutional neural network (CNN) for feature extraction, and finally greatly improves the performance index of the target detection algorithm. For example, in 2013, Girshick et al. proposed R-CNN[61, as the pioneering work of using deep convolutional neural network for target detection, using AlexNet network for feature extraction, and the final detection accuracy was nearly 30% higher than the traditional method[3].

Target detection algorithms based on deep convolutional neural network are mainly divided into two categories: class II is target detection algorithm based on regression (namely single-stage target detector). The two-stage target detector first selects the candidate region of the input image. Then the candidate regions are classified and regressed to achieve target detection. The single-stage target detector omitted the step of candidate region generation, and directly integrated the processes of feature extraction, target classification and position regression into a convolutional neural network, which reduced the target detection process to an end-to-end regression problem.

2. Data set and target detection algorithm performance evaluation index

2.1. The data set
Data sets are very important for supervised learning algorithms. For target detection tasks, data sets with strong applicability can not only test and evaluate algorithm performance, but also promote the development of research fields related to target detection. In the target detection task, PASCAL VOC and Microsoft COCO data sets are the two most frequently used data sets[4].

The PASCAL VOC dataset was originally released in 2005 and has been updated in several versions until 2012. VOC2007 and VOC2012 are the two most commonly used versions in the academic world, both with 20 target categories. Among them, VOC2007 data set contains 9 963 pictures, which is composed of three parts of train/val/test, while VOC2012 data set contains 11 1530 pictures, which is composed of two parts of train/test. There are two methods to use this data set for target detection tasks in academia: train2007+ Train2012 as the training set and Test2007 as the test set, and train2007+ Train2012. The main evaluation index of this dataset is the mean accuracy rate mean (mAP).

Originally released in 2014, the Microsoft COCO dataset consists of 91 object categories, 328,000 images and 25,000 label labels. COCO data set aims at scene understanding and can be used in image classification, image segmentation and target detection. The research focuses on the three questions of non-iconic Views, Contextual Reasoning between Objects and PRECISE 2D localization. The release of this data set has promoted the development of image segmentation in recent years and is the largest data set in the field of image segmentation so far.

2.2. Performance evaluation index of target detection algorithm
Common performance evaluation indexes of target detection algorithm include: detection speed, intersection ratio, accuracy rate, etc.

Among them, detection speed (Frames Per Second (FPS) means the number of images detected by
the algorithm model Per Second. FPS can evaluate the real-time performance of a model. The greater the value, the better the real-time performance of the model. Occurring simultaneously ratio (Intersection computes Over the Union, IOU), expressed as the forecast box and original algorithm model Intersection and and set the ratio of the box, it describes the two areas of coincidence degree, the higher value on behalf of the algorithm is the more accurate positioning accuracy of the model (Precision, P), expressed as classification after the correct number of positive samples and classification criterion is the ratio of number of samples, points out the measure is a classifier is, indeed, is the probability that the sample is sample. Recall (R) is the ratio between the correct number of positive samples classified and the true number of positive samples. It measures the ability of a classifier to find out all positive samples. In general, the higher the precision rate is, the lower the Recall rate will be. The greater the value, the better the detection effect of the classifier for a certain category. Mean Average Presion (mAP), expressed as the mean of the mean accuracy of all categories, measures the detection effect of the classifier on all categories, and is the most important index used in the target detection algorithm.

3. Convolutional neural network
Image is the foundation of target detection, target detection is on the basis of classification, image to realize the target orientation, target specific spatial location and boundary information. Target detection based on convolutional neural network network convolutional neural network for feature extraction are needed to fulfill the task of target detection, so the convolution of the neural network performance is directly related to the target detection performance of the network is good or bad. In 2012, The AlexNet proposed by Hinton et al. surpassed many traditional image classification algorithm models and won the champion of ILSVRC image classification competition with a top-5 error rate of 16.4%, it made the application more and more, marking the advent of the era of deep learning. In 2013, Lin M the NiN put forward network layer 1 x1 convolution is introduced and the global pool, effectively reduces the network number, after ResNet19 and Inception borrowed this design such as network also. Simonyan K et al pr.of parameter of the model, makes the model in the image classification task on Top - 5 error rate down to 6.8%, This proves that increasing network depth can also improve model accuracy.

4. Target detection algorithm based on candidate regions
The classifies and regresses the suggestion box to obtain the final detection result. Typical examples of such algorithms are: R-CNN series, R-FCN, Mask R-CNN, Cascade R-CNN, Trident Net, etc.

[Figure 1. R-CNN network structure.]
5. Summary and prospect
On the whole, the detection process of target detection network based on candidate regions is getting simpler and more accurate\textsuperscript{[8]}, but the speed is generally slow, which is difficult to meet the real-time demand of practical application, and it faces difficulties in application promotion. Although the target detection network based on regression is generally inferior to the former in accuracy, it has more advantages in detection speed. With the development in recent years, its detection accuracy has been gradually improved, and it has more potential in practical application. The future research direction will focus on the combination of precision and speed, draw on the advantages of high precision based on candidate area network and high speed based on regression network, improve the precision and speed of target detection network, and achieve trade-off of precision and speed, so as to meet the actual requirements.

6. Conclusion
The rapid development of neural network and the continuous improvement of hardware facilities provide a very good development opportunity for the field of target detection. Although the performance of the existing target detection algorithms based on convolutional neural network has far exceeded that of the traditional target detection algorithms, there is still much room for improvement. How to implement the target detection under small data sets the network training, how to better implement multi-objective and small target detection under complex scene, how to make better use of scene information and semantic information, how to meet the intelligent transportation system, intelligent video surveillance, automatic monitoring, industrial detection accuracy and real-time performance in the field of practical application requirements, etc., these will be the next important research direction in the detection.

References
[1] Sun Guopeng, Hao Xiangyang, Zhang Zhenjie, et al. Cooperative Target Recognition Method based on Multi-Feature Judgment [J] Journal of System Simulation, 2012, 30(6): 2377-2383.
[2] Journal of shaanxi normal university (natural science edition), 2019, 47(5): 1-9. (in Chinese).
[3] White ice, Zhong Bing, Ouyang Valley. Fusion features of Hierarchical convolution and Target Tracking of Scale-Adaptive Kernel Correlation Filter [J] Minicomputers, 2017. 38(9): 2062-206.
[4] Krizhevsky A, Sutskever I, Hinton G. Image Net classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
[5] Neural Computation, 1989. 1(4): 541-551. [5] Lecun Y.B. Oser B. Denker J S, et al. Back propagation Applied to handwritten zip code Recognition [J].
[6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection
and semantic segmentation [C]/Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[7] Zhang Zemiao, HUO Huan, Zhao Fengyu. A review of target detection algorithms for deep convolutional neural networks [J]. Minicommicrocomputer systems, 2019, 40(9): 1825-1831.

[8] Lin M. Chen Q. yan S, et al. Network in network [J]. Ar Xiv: 1312. 4400, 2013.