Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 4, No. 3, 2023

Review of Target Detection Algorithms

Xu Zhou, Guojun Lin *

School of Automation and information engineering, Sichuan University of Technology, Zigong 643000, China
*** Corresponding author:** Guojun Lin (Email: 386988463@qq.com)

Abstract: Object detection is a popular direction of computer vision and digital image processing, which is widely used in robot navigation, intelligent video surveillance, industrial detection, aerospace and other fields, using computer vision to reduce human capital consumption has important practical significance. Because of the wide application of deep learning, the algorithm of target detection has been developed rapidly. This paper mainly introduces the traditional target detection algorithm and two kinds of target detection algorithm based on depth learning and the data set commonly used in target detection.

Keywords: Object Detection; Deep Learning; Computer Vision.

1. Introduction

Object detection is one of the important branches of computer vision, whose task is to find out some specific objects in images, and give their categories and positions. In recent years, object detection algorithm has been a hot research field in computer vision and image processing. Firstly, it is the foundation of more complex visual task processing such as image semantic segmentation [1] and instance segmentation [2], secondly, it has great application prospect in many fields such as robot navigation, intelligent monitoring and industrial detection.

The development of target detection can be divided into two stages: the traditional stage based on artificial feature extraction and the new stage based on deep learning. Traditional target detection usually adopts the method of sliding window combined with feature extraction by hand, and finally combines with special classifier to classify. Typical algorithms are Viola Jones [3], HOG [4], DPM [5] and so on. Until 2014, R. Girshick et al applied CNN (Convolutional Neural Networks) [6] to target detection, trying to extract features using Convolutional Neural network, improving the average detection accuracy by about 30% compared with traditional methods, to the detection effect has brought a qualitative leap.

This paper mainly introduces the traditional target detection algorithm, the target detection algorithm based on depth learning and the commonly used data set of target detection.

2. Traditional Target Detection Algorithm

Taking 2012 as a watershed, target detection in the past two decades can be roughly divided into two periods: the "Traditional target detection period" before 2012 and the "Depth-learning-based target detection period" after that. Most of the early target detection algorithms are based on manual features. Due to the lack of effective image representation, people had no choice but to design complex feature representation and various acceleration techniques to make the most of the limited computing resources.

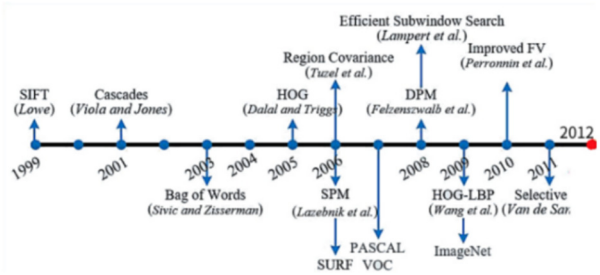

Figure 1. Traditional target detection model and its development

As shown in Figure 1 above, the traditional target detection algorithm mainly relies on the traditional feature extractor to extract the image features, and uses the sliding window to generate a large number of target candidate regions, at the same time, there are some problems such as low detection speed and low detection precision.

The basic operational steps of target detection include the following steps:

1. Data collection and tagging: collect image or video data containing the target, and tagging, tagging the target's boundary box and category information.

2. Feature extraction: to extract features from an image to represent a target, common methods include feature extraction using convolutional neural network (CNN).

3. Candidate region generation: a candidate region generation algorithm (such as Selective Search, EdgeBoxes, etc.) is used to generate candidate regions that may contain the target.

4. Feature Matching: to match the extracted features with the features in the candidate area, the common method is to calculate the similarity between the two.

5. Target classification: according to the features obtained by matching, targets are classified by using classifiers, such as Support vector machine support vector machine (SVM) and Logistic Regression.

6. Bounding box regression: a regression algorithm is used to fine-tune the bounding box of a candidate area to frame the location of the target more accurately.

3. Target Detection Algorithm based on Depth Learning

3.1 One-stage Target Detection Algorithm

One-Stage target detection algorithm can directly generate the class probability and position coordinate values of objects in one stage, and the overall process is simpler than the Two-Stage target detection algorithm, which does not need the Region Proposal stage.

The following is a model of one-stage target detection algorithm.

SSD: the Backbone of the SSD detection algorithm is a VGG network, using feature maps with different resolutions at different stages to make predictions.

Yolo: Yolo [7] is an end-to-end neural network. Yolo is an algorithm for object prediction based on global image information.

Yolov2: YOLOV2 [8] improves the precision of Yolov1. By adding the BN layer, small objects can be detected better, and following the practice of Faster R-CNN, YOLOV2 removes the fully connected layer in YOLOV1 and adopts the convolutional kernel anchor boxes to predict the boundary boxes, the detection accuracy is improved.

Yolov3: Yolov3 [9] improves the previous Yolo algorithm, improves the backbone of the network, uses the multi-scale feature graph to detect the target, uses the single neural network to process the image, and divides the image into several regions, the probability of each region and the information of boundary box are predicted, so the global perception and local precision of target detection are combined, and good detection results are obtained.

Yolov4: developed by Alexey Bochkovskiy, Chien Yao Wang, and Hong-yuan Mark Liao and released in 2020, the algorithm uses a number of optimization strategies, such as rapid testing, cross-layer connectivity, and convolution operations, so that it can achieve faster detection speed while maintaining high accuracy.

Yolov5: June 2020, Jocher presents Yolov5 [10]. In the input side, the adaptive anchor frame calculation and image zooming technology are adopted to obtain the suitable anchor frame and reduce the computation of model, and improve the performance and effect quality of the whole target detection algorithm.

Yolov7: YOLOV7 [11] target detection algorithm has higher detection accuracy and faster detection speed than previous models. And for different target detection tasks, there are 7 different size models, such as YOLOv7-w6, YOLOv7-d6, etc.

Yolov8 is an improved version of YOLO（You Only Look Once), which is known for fast and accurate real-time object detection capabilities. Yolov8 incorporates several improvements over its predecessors, such as a powerful backbone network (Dark-53), advanced data augmentation techniques, and better anchor box clustering. These improvements help to enhance the accuracy and efficiency of object detection tasks in various applications.

3.2 Two-stage Target Detection Algorithm

The Two-Stage object detection algorithm can be regarded as two One-Stage detection, the first Stage can detect the location of objects, the second Stage can refine the results of the first Stage, one-Stage detection is performed for each candidate region. The overall process is shown in Figure 2 below. While testing, the input image is convolutional neural network to produce the first stage output. The output is decoded to generate candidate regions, then the feature representation (Rois) of the corresponding candidate region is obtained, and the second stage output is generated by refining ROIs, and the final result is generated by decoding (post-processing), and the corresponding detection box is generated by decoding While Training, you need to encode Ground Truth into a format corresponding to the CNN output in order to calculate the corresponding loss.

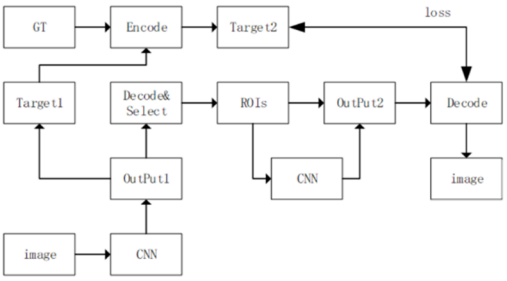

Figure 2. Two-stage detection algorithm schematic

R-CNN: R-CNN [12] (Regions with CNN features) uses DCNN as the backbone network for feature extraction, but the feature-sharing ability of DCNN is not utilized when extracting the features of each candidate region separately, resulting in a large amount of waste of computing resources, and it consumes a lot of storage space.

Fast R-CNN: Fast R-CNN is improved on the basis of R-CNN algorithm. Compared with R-CNN/SPPNet, Fast R-CNN has less memory space and higher accuracy.

Faster R-CNN: Faster R-CNN proposes a candidate area network (RPN) to replace the selective search algorithm for generating candidate areas, which improves both accuracy and speed.

Mask R-CNN: Mask R-CNN is composed of Faster R-CNN and semantic segmentation algorithm FCN, which performs target classification and bounding box regression parallel RoI prediction segmentation, while completing target detection and instance segmentation.

3.3 Common Data Sets for Target Detection

Pascal VOC: The Pascal VOC (Visual Object Classes) challenge is one of the early and important competitions in the field of computer vision, including image classification, Object detection, scene segmentation, event detection and other tasks. Voc2007 and Voc2012 datasets are commonly used for target detection. VOC2007 included 5,000 training images and 12,000 labeled targets, and VOC2012 included 11,000 training images and 27,000 labeled targets.

Imagenet: Imagenet is a large labeled image data set. The dataset contains more than 14 million images and about 22,000 categories. More than a million of these images are labeled with specific categories and positions of objects in the image, making them a suitable target detection dataset.

MS-COCO: the MS-COCO dataset is a data set released by the Microsoft team that can be used for image recognition, segmentation and annotation. It consists of 91 classes of targets and 328,000 images. It features images that are mostly from life, more complex backgrounds, and more small objects. On the other hand, it also notes the segmentation information of each instance in addition to the border of each object. Therefore, Coco has gradually become a mainstream target detection evaluation set.

Open Images: Open Images comes from the Google team

and includes 9 million Images and more than 6,000 categories. Data sets have more diverse objects, often containing complex scenarios of multiple objects. Visual relationship annotations are also provided. There is more information than in the previous data set.

4 Conclusion

In this paper, the traditional target detection methods and target detection methods based on depth learning are reviewed. Combining two-stage target detection method and single-stage target detection method, the paper summarizes their advantages and disadvantages, introduces the data set, and summarizes the important applications of target detection. With the gradual development of target detection technology, the detection accuracy has been gradually improved at present. However, with the development of diversified application scenarios, there are still many challenges and problems to be solved in improving model algorithms, Data pre-processing, deep learning network design and model optimization.

Based on the current research status of target detection, the future research prospects are as follows: (1) for multi-scale target detection under complex background, such as dense small target detection is still insufficient. (2) with the increasing complexity of the application scene, how to identify the target quickly and accurately in the complex environment, and how to use the context information of the scene and the relationship between objects are the future research directions in the field of target detection.

Acknowledgments

This work was supported in part by the 2022 Graduate Innovation Fund Project of Sichuan University of Science and Engineering (Y2022162). The authors express their acknowledgement for the anonymous review.

References

[1] LIU X, DENG Z, YANG Y. Recent progress in semantic image segmentation[J]. Artificial Intelligence Review, 2019, 52(2): 1089-1106.

[2] HAFIZ A M, BHAT G M. A survey on instance segmentation: state of the art [J]. International journal of multimedia information retrieval, 2020, 9(3): 171-189.

[3] VIOLA P, JONES M J. Robust real-time face detection[J]. International journal of computer vision,2004,57(2):137-154.

[4] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE,2005.

[5] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE transactions on pattern analysis and machine intelligence, 2009, 32(9): 1627-1645.

[6] GU J, WANG Z, KUEN J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77: 354-377.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al.You only look once: unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:779–788.

[8] REDMON J, FARHADI A.YOLO9000: better, faster, stronger [C]// Proceedings of the IEEE conference on computer vision and pattern recognition,2017:6517–6525.

[9] REDMON J, FARHADI A .YOLOv3: An Incremental Improvement[J].arXiv e-prints, 2018. DOI: 10.48550/ arXiv. 1804.02767.

[10] JOCHER G, STOKEN A, BOROVEC J, et al.YOLOv5:V3.1 - bug fixes and performance improvements [EB/OL].2020. doi: 10.5281/zenodo.4154370, 2020.

[11] WANG C Y, BOCHKOVSKIY A, LIAO H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for realtime object detectors[J]. arXiv preprints arXiv: 2207, 02696, 2022.

[12] 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.