IJLAI transactions on Science and Engineering

Quarterly Research Journal

ISSN: Applied

https://ijlaitse.com/index.php/site

Published by:

Ali Institute of Research & Skills Development (AIRSD)

Office No 1, Moiz Clinic Building, Khan Village Road,

New Gulgasht, Multan Pakistan.

Email: admin@ijlaitse.com

IJLAI transactions on Science and Engineering, vol. 1, no. 1, pp. 1-6

Object Detection Based on Deep Learning: A Brief Review

Shoulin Yin[1]

1. **Software College, Shenyang Normal University, Shenyang 110034, China**
2. **Email:yslinhit@163.com**

Abstract

Object Detection is one of the basic tasks in the field of computer vision. In recent years, with the hot development of deep learning technology, object detection algorithm has changed from the traditional algorithm based on manual features to the detection technology based on deep neural network. From the first proposed R-CNN, OverFeat, to the later Fast/Faster RCNN, SSD, YOLO series, and most recently Pelee. In less than five years, the object detection technology based on deep learning has improved the network structure from two stages to one stage, from bottom-up only to Top-Down. From single-scale network to feature pyramid network, from PC side to mobile side, many good algorithm technologies have emerged, and these algorithms have excellent detection effect and performance on open target detection data sets. This paper first reviews the traditional target detection algorithms, then introduces several mainstream two-stage target detection algorithms and single-stage target detection algorithms, analyzes the structure, advantages and disadvantages of these algorithms, and finally forecasts the future research direction of target detection algorithms.

Keywords: Deep learning; Object detection; Single-stage; Two-stage

1. Introduction

With the continuous development of artificial intelligence technology, deep learning technology has been widely used in the field of computer vision. Vehicle target detection is always a challenging problem in the field of computer vision and has a wide range of application prospects. Traditional methods often need to manually extract features and build classifiers to achieve vehicle target detection, which is susceptible to environmental changes and requires a lot of adjustment and optimization. In recent years, with the continuous development of deep learning technology, vehicle object detection algorithms based on deep learning have achieved great improvement in accuracy and processing speed, and gradually become a research hotspot. Figure 1 shows the development of object detection algorithm.

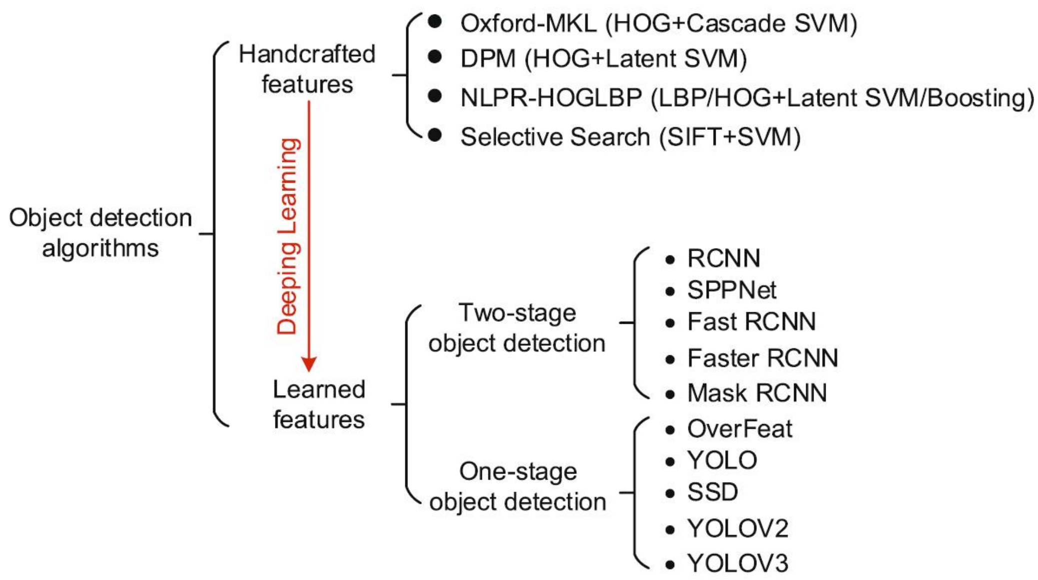

Figure 1. Object detection process

This paper aims to review the vehicle object detection algorithms based on deep learning, review the traditional object detection algorithms, and then introduce several current mainstream two-stage vehicle object detection algorithms and single-stage vehicle object detection algorithms, and analyze the structure, advantages and disadvantages of these algorithms, in order to provide reference and inspiration for related research. At the same time, this paper will also look forward to the future research direction [1-5].

2. Traditional target detection algorithm

The traditional target detection algorithm generally consists of three steps: region selection, feature extraction and feature classification. Region selection refers to finding areas in the input image that may contain objects to be detected. A common method of area selection is to use the sliding window technique, where Windows of different sizes and proportions are slid over the image, and each window is evaluated by a classifier to get the area that is likely to contain the object. Feature extraction refers to the use of artificially designed feature extraction methods from selected regions to extract features from images. Common manual design features include SIFT, HOGP, etc. These feature extraction methods can effectively capture local texture and shape information in images.

And it plays a key role in target classification. However, this way of designing features manually requires a lot of time and effort, which limits the scalability and generalization ability of the algorithm. Feature classification refers to the classification of extracted features, to determine whether the region contains objects to be detected, and to identify them. The commonly used classification algorithms are support vector machine (SVM) and AdaBoost4. However, in recent years, the development of deep learning techniques has begun to change this situation, especially the widespread use of convolutional neural networks (CNNS) has made end-to-end training possible to learn feature representations directly from raw images, avoiding the process of manually designing features.

3. Object detection algorithm based on deep learning

The vehicle object detection algorithm based on deep learning refers to the automatic identification and location of cars in the road scene by using deep neural network model. By learning a large number of vehicle image data, the algorithm can not only complete the target detection task, but also have a certain degree of robustness, that is, it can have the ability to adapt to changes in lighting, weather and other factors. Common deep learning-based vehicle target detection algorithms include two-stage detection algorithm represented by R-CNN series network and single-stage detection algorithm represented by SSD and YOLO series [6-10]. These algorithms all use convolutional neural networks as feature extractors to realize object detection through different network structures and techniques. Compared with the traditional method based on manual features, the algorithm based on deep learning has higher accuracy and faster running speed, and has been widely used in automatic driving, traffic monitoring and other fields.

3.1 Two-stage target detection algorithm

The two-stage vehicle object detection algorithm is to generate the candidate frame, and then classify and regression the candidate frame, so as to realize the detection and positioning of the target object in the image. Next, we mainly introduce R-CNN, Fast R-CNN and Faster R-CNN.

3.1.1 R-CNN

R-CNN was proposed by Ross Girshick et al., in 2014. It is a deep learning-based object detection algorithm that is able to identify different categories of objects in pictures and label their locations.

The R-CNN algorithm is mainly divided into three steps: First, the algorithm generates some candidate boxes that may contain objects by Selective Search (SS) method. The method is based on the image features such as color, texture, size and shape to get some areas that may contain objects. Then, for each candidate box, the algorithm uses convolutional neural networks to extract features. Finally, support vector machine classifier is used to classify each candidate box to determine whether there is a target object in the region. At the same time, regressors are used to fine-tune the coordinates of the bounding box to position the object more precisely.

In general, R-CNN divides the object detection task into three sub-tasks: candidate frame generation, feature extraction and classification localization. However, it is slow because of the need to perform CNN calculations and SVM classification for each candidate box, which makes it difficult to apply to real-time object detection scenarios. Subsequent improved models, such as Fast R-CNN and Faster R-CNN, have made some improvements on the basis of R-CNN to improve the detection speed and accuracy.

3.1.2 Fast R-CNN

Ross Girshick et al. improved R-CNN and proposed Fast R-CNN. Firstly, the algorithm uses the selective search algorithm to generate about 2000 candidate regions on the image, then extracts the features of the whole image through the convolutional neural network, projects the candidate regions generated by the SS algorithm onto the feature map to obtain the corresponding feature matrix, then uses RolPooling to extract the fixed-size feature vector for each feature matrix, and finally uses the full connection layer Classification and regression prediction were performed. Fast R-CNN still needs to use a selective search algorithm to generate candidate regions, which is time-consuming and limits the speed of the algorithm.

3.1.3 Faster R-CNN

Shaoqing Ren et al., aiming at the shortcomings of R-CNN and Fast R-CNN, made improvements again in 2015 and proposed the Faster R-CNN algorithm. It is the fastest of the R-CNN series of algorithms. Compared with R-CNN and FastR-CNN, Faster R-CNN introduces a new neural Network structure called Region Proposal Network (RPN), which can extract candidate regions with potential detection targets on the image. This structure can share the convolution layer features with the classifier, thus avoiding double computation. The network structure of Faster R-CNN mainly consists of three parts: Convolutional neural network (CNN), RPN network and Fast R-CNN detection network. First, feature extraction of input images is carried out in convolutional neural network. The feature map is then fed into the RPN network to generate a candidate region containing potential targets. These candidate regions are further processed to remove the irrelevant regions, and the remaining regions are fed into the Fast R-CNN detection network for target classification and bounding box regression.

3.1.4 YOLOv3

YOLOv3 is an end-to-end target detection algorithm, mainly composed of backbone feature extraction network DarkNet-53 and multi-scale feature fusion prediction network. YOLOv3 continuously stacked Residual blocks to form the backbone network DarkNet-53 for feature extraction. The overall network structure is constructed with FPN2 type feature pyramid structure, which realizes the fusion detection of three scale features. The feature maps of each scale are allocated with three different sizes of anchor frames, and 9 different sizes of anchor frames are set on the feature maps of the three scales for detection. Compared to its predecessor YOLOv2, YOLOv3 has a great improvement in accuracy and speed.

4. Conclusions

This paper mainly expounds the object detection algorithm, first reviews the traditional object detection algorithm, and then introduces the representative two-stage and single-stage vehicle object detection algorithm based on deep learning. To detect vehicle targets in complex road scenarios, it is necessary to develop more efficient deep learning algorithms to further improve the accuracy and real-time detection of vehicle targets. In the future, we can explore the combination of intelligent algorithms and hardware to achieve a more intelligent and efficient vehicle target detection system, and can also continue to integrate technologies in various fields, such as the fusion of camera and lidar and other sensor data, to further improve the accuracy and reliability of vehicle target detection.

Acknowledgements

None

Conflict of Interest

No potential conflict of interest was reported by the authors.

References

[1] Amit Y, Felzenszwalb P, Girshick R. Object detection[J]. Computer Vision: A Reference Guide, 2020: 1-9.

[2] Papageorgiou C, Poggio T. A trainable system for object detection[J]. International journal of computer vision, 2000, 38: 15-33.

[3] Zou Z, Chen K, Shi Z, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023.

[4] Padilla R, Netto S L, Da Silva E A B. A survey on performance metrics for object-detection algorithms[C]//2020 international conference on systems, signals and image processing (IWSSIP). IEEE, 2020: 237-242.

[5] Hu H, Gu J, Zhang Z, et al. Relation networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3588-3597.

[6] Zhao Z Q, Zheng P, Xu S, et al. Object detection with deep learning: A review[J]. IEEE transactions on neural networks and learning systems, 2019, 30(11): 3212-3232.

[7] Wu X, Sahoo D, Hoi S C H. Recent advances in deep learning for object detection[J]. Neurocomputing, 2020, 396: 39-64.

[8] Xiao Y, Tian Z, Yu J, et al. A review of object detection based on deep learning[J]. Multimedia Tools and Applications, 2020, 79: 23729-23791.

[9] Zhou X, Gong W, Fu W L, et al. Application of deep learning in object detection[C]//2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). IEEE, 2017: 631-634.

[10] Wang W, Lai Q, Fu H, et al. Salient object detection in the deep learning era: An in-depth survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 3239-3259.