Dean&Francis

ISSN 2959-6157

The Advanced and Practical Application of Existing Object Detection Technologies

Kang Li

Department of Information Science and Engineering, Ocean University of China, Qingdao, China

likang4702@stu.ouc.edu.cn

Abstract:

Object detection is the core task in computer vision, widely used in industry, medicine, agriculture, and other fields. With the development of deep learning technology, object detection algorithms have made remarkable progress in improving accuracy, improving efficiency, and reducing cost. However, the existing technology still faces challenges such as insufficient accuracy, high computational complexity, and deployment flexibility. In this paper, the basic principle of object detection is summarized in detail, including two-stage and single-stage object detection algorithms, and the advantages and disadvantages of several classical algorithms and their application scenarios are analyzed. At the same time, this paper summarizes the specific cases of the application of object detection technology in industrial production, medical diagnosis, and smart agriculture, and discusses the limitations and future development direction of the existing object detection technology. The review in this paper provides a valuable reference for further enhancing the application potential of target detection technology, and it is expected that future research can overcome the current challenges and achieve a wider application.

Keywords: object detection; two-stage; single-stage; application.

1. Introduction

Object detection is one of the core tasks in the field of computer vision, whose main purpose is to identify and locate objects in images. As one of the most basic and challenging problems in computer vision, object detection has received great attention in recent years. Over the past two decades, the people have seen rapid advances in object detection technology and its profound impact on the entire field of computer vision [1]. With the rapid development of deep learning technology, object detection technology has been more and more widely used in many fields, mainly including industry, medicine, agriculture, and many other major fields. Because of its high precision and stability in complex environments, the two-stage target detection algorithm is suitable for solving the needs of small-volume target detection and accurate positioning. The single-precision target detection algorithm has the advantages of low cost and good

ISSN 2959-6157

real-time performance and has a place in the scene of efficiency priority. The widespread applications of scene understanding, video surveillance, robotics, and autonomous driving systems have sparked a great deal of research in the field of computer vision over the last decade. At the heart of all these applications, visual recognition systems that incorporate image classification, positioning, and detection have gained tremendous research momentum. Thanks to major developments in neural networks, especially deep learning, these visual recognition systems have achieved extraordinary performance. Object detection is one of the most successful fields in computer vision [2]. Although the research on target detection technology has made great progress, there are still many limitations and challenges.

This paper gives an overview of existing target detection algorithms, including two-stage target detection and single-stage target detection. At the same time, the practical application of target detection in industry, medicine, and agriculture is discussed, and the specific practice of target detection in industrial production automation and safety detection, medical image analysis, cell behavior tracking, and intelligent monitoring of agricultural diseases and pests is analyzed. At the same time, this paper lists the limitations of the current two object detection algorithms and summarizes the possible future development direction, such as multi-modal information fusion and an adaptive optimization algorithm to improve robustness and environmental adaptability. With the continuous iteration of algorithms and hardware technology, the application prospect of object detection technology will be very broad.

2. Object Detection

As one of the core problems in the field of computer vision, the challenge of object detection is divided into four main categories, namely, the determination of the category of the image in the picture, the determination of the location of the target, the determination of the size of the target, and the determination of the shape of the target. The current target detection algorithms are mainly divided into two categories: two-stage target detection and single-stage target detection. The two-stage target detection mainly focuses on the selective region proposal strategy through complex architecture. Single-stage object detection, however, focuses on all spatial region proposals at once through a relatively simple architecture so that objects may be detected. The performance of any object detection algorithm is evaluated by detection accuracy and inference time. Generally, the detection accuracy of two-stage target detection is better than that of single-stage target detection. However, the inference time of single-stage target detection is shorter than that of similar products [3].

2.1 Two-stage Object Detection

Two-stage object detection is widely used in various complex target detection tasks because of its high detection accuracy and reliability. The method is usually divided into two main stages: candidate region generation and object classification and location.

Candidate region generation stage: Firstly, a pre-trained convolutional neural network (such as ResNet or VGGNet) is used to extract the features of the input image and generate the feature map; A predefined set of anchor points is then generated on the feature map, which are bounding boxes of fixed size and scale. Each feature map point is associated with multiple anchor points to provide candidate boxes with different sizes and aspect ratios; Binary (target/background) predictions are made for each anchor point and the position of these anchors is adjusted by regression. The RPN outputs an adjusted set of candidate regions that may contain targets; Non-Maximum Suppression (NMS) is used to eliminate redundant candidate regions with high coincidence, leaving the regions most likely to contain the target.

Object classification and location stage: The candidate regions of different sizes are converted into fixed-size feature maps. This is typically achieved by an ROI Pooling or ROI Align operation, which preserves more granular spatial information; The feature map of the candidate region is processed by using the fully connected layer or other convolution layer to further extract the feature. Output a class probability distribution (classification) and bounding box position adjustment (regression) for each candidate region; The advantage of this process is that it can take advantage of the high-quality candidate boxes generated by the RPN, thus improving the detection accuracy.

There are two main algorithms for two-stage object detection, namely Faster R-CNN and Cascade R-CNN. Faster R-CNN integrates RPN and Fast R-CNN into a unified network, which has higher detection accuracy and significantly improves detection speed. Cascade R-CNN connects multiple detection heads based on Faster R-CNN to gradually improve detection accuracy, especially to improve detection performance under high IoU conditions.

The advantages and disadvantages of two-stage object detection are also very obvious. It has high accuracy, suitable for complex scenes, especially in the detection of small objects superior performance, and modular design for high flexibility. But the calculation amount is large, the reasoning speed is relatively slow; In addition, large hardware resources are required and cannot be deployed on resource-limited devices.

2.2 Single-stage Object Detection

The single-stage object detection method skips the candidate region generation stage, optimizes the computational efficiency, and reduces the processing time. Such methods are known for their high efficiency and real-time performance and are widely used in scenarios requiring rapid detection. Its working principle is simpler than two-stage object detection, a single-stage method directly on the feature map object detection, which usually does not need the candidate box generation step. After the feature extraction, the model takes all possible object bounding boxes and corresponding class probabilities as output.

There are also two main algorithms for single-stage target detection, namely the You Only Look Once (YOLO) algorithm and the Single Shot MultiBox Detector (SSD) algorithm. YOLO divides images into grids and simultaneously predicts bounding boxes and categories in each grid. It is known for its fast speed and is suitable for real-time detection. After iterative updates from YOLOv1 to YOLOv4 and later versions, the YOLO algorithm adds an S×S grid to divide input images, each grid cell predicts a boundary box and its corresponding category probability, introduces Anchor Boxes and better feature pyramid mechanism, and introduces multi-scale prediction. Target detection results were generated on different scales, and network structure, training strategy, and loss function continued to be optimized. More innovative algorithms such as CSPNet and Mish activation function were introduced to optimize records. Each generation of YOLO model upgrades further improves accuracy while maintaining high efficiency. Due to its ease of use and high object detection accuracy, YOLO is used as the most popular object detection software in many smart video applications. In addition, in recent years, various intelligent vision systems based on high-performance embedded systems have been developed. Despite this, YOLO still requires high-end hardware to successfully perform real-time object detection [4]. On the other hand, SSD enhances the detection ability of objects of different sizes by making predictions on feature maps of different scales, which can be roughly divided into using pre-trained convolutional neural networks (such as VGG16) to extract basic features and constructing several additional convolutional layers on top of basic features. There are four stages of generating feature maps of different sizes, using anchor boxes of different sizes and proportions on each feature map for target detection, and predicting a category probability and corresponding boundary box position for each anchor box.

Similarly, single-precision target detection also has its advantages and disadvantages, because only one forward propagation is required, the reasoning speed is fast, suitable for real-time applications, such as monitoring and unmanned driving, and the structure is simple, the hardware requirements are low, and the deployment is easy. However, the detection accuracy is low, and in the detection of complex scenes or small objects, the accuracy is usually not as good as that of the two-stage method. At the same time, the category is unbalanced, and the proportion of positive and negative samples in the single-stage detection is seriously unbalanced, which easily leads to a decline in classification performance.

3. The Application of Object Detection

3.1 Industrial Field

In the context of benefiting from mechanization and automation, modern industry is often in the pursuit of low cost, high efficiency, and high quality, so the industrial production process is moving toward digitalization and intelligence. Object recognition technology has a wide range of application scenarios in industrial applications, especially in the case of limited computing resources or high real-time requirements.

In the material production line, the target detection technology can be used to identify whether there is a gap on the surface of the metal material, to separate the defective products. In the assembly line, intelligent robot arms are combined to complete automatic assembly and welding. Now some studies have proposed that in the industrial miracle Man vision guidance system, if irregular industrial parts are encountered, the target detection algorithm based on template matching makes it difficult to achieve high recognition accuracy. In a structured environment, compared with traditional target detection algorithms, the visual target detection practice based on single-precision target detection model YOLOv3 has a higher level of positioning accuracy and recognition accuracy, reaching the standards for industrial field deployment [5].

At the same time, considering that the target recognition model has a higher accuracy in identifying the human body, the target detection technology can be effectively applied to complex industrial environments where safety devices need to be worn, such as the factory floor. The technology is used to detect whether workers are wearing safety ropes and helmets. With the continuous development of computer vision, object detection has been applied in more and more scenarios, and the industrial environment is no exception. Target detection based on deep learning has gradually become the main means of security management. Compared with traditional methods, it is more helpful to realize real-time supervision on site and create a safe production environment, to ensure the safety

ISSN 2959-6157

of workers [6].

On the other hand, in automated warehouses, object detection can identify items and classify them, while guiding robots to perform automatic screening, which can be widely used in the distribution centers of various express or logistics companies. To meet the needs of small and medium-sized logistics warehouses, a six-degree-of-freedom robot arm wheat-wheel robot based on object detection has been developed. By combining object detection technology with processing strategies, high-precision and real-time cargo identification is achieved, which significantly improves operational efficiency and reliability, and is especially suitable for meeting the growing logistics needs of small and medium-sized e-commerce [7].

3.2 Medical Field

As deep learning has made breakthrough progress in computer vision, speech processing, and other fields, more and more researchers have begun to apply deep learning to medical image detection, segmentation, diagnosis, and other tasks [8]. Object detection combines computer vision and artificial intelligence technology in the medical field, which can automatically identify, locate, segment, classify, and label target areas in medical images, to assist doctors in diagnosis and treatment. However, due to accuracy problems, it can only be used as an auxiliary means.

In the field of pathology, target detection technology has begun to be applied to automated analysis of pathological section images to help identify abnormal pathological features such as cancer cells, inflammatory areas, and necrotic tissue. For imaging doctors, the burden of analyzing a large number of case image data is gradually increasing. The application of target detection technology as an auxiliary means can reduce the burden on doctors and improve the accuracy of analysis.

In biological research, target detection technology can be used to study cell behavior and track cell movement. For example, when tracking the growth of cancer cells, the results obtained by a research team show that there are six stages in obtaining the growth information of ovarian clear cell carcinoids: obtaining images of ovarian clear cell carcinoma organoids; Getting the target model after training; The acquired images were input into the target model to obtain the target cells on the ovarian clear cell carcinoma organoids. Then each target cell was labeled to establish a unique identifier. The target cell map was obtained from multiple organoid images of ovarian clear cell carcinoma. According to the obtained target cell map, all kinds of information about the target cell can be obtained, such as circumference or area. Such application examples of target detection can automatically obtain all kinds of information needed for tracking growth conditions, thus effectively liberating manpower and improving research efficiency [9].

3.3 Agricultural Field

Through object detection technology, the growth of crops, health status, and the occurrence of pests and diseases can be monitored and diagnosed in real-time, to help farmers take timely measures to reduce crop losses. Using convolutional neural networks to train crop image data, combined with farmland monitoring equipment such as drones, healthy crops and crops infected with pests and diseases can be identified, and marked in thumbnails, so that farmers can get more intuitive and graphical information. The use of automated monitoring technology can reduce the burden on farmers and provide a more systematic overview. The research team of Nanjing Forestry University believes that the current agricultural pest detection mode is generally manual detection, and its detection efficiency and accuracy are completely dependent on the experience of technical personnel, resulting in low efficiency and high error rate of traditional agricultural pest detection. Given the current situation of agricultural pest detection, they proposed a monitoring model of agricultural pests based on a neural network. By extracting features from crop disease images and applying a convolutional neural network model, the types of pests and diseases in the images are identified. This machine learning-based pest and disease identification model improves the efficiency and accuracy of pest and disease identification [10].

4. Limitations and Prospects

At this stage, target detection technology is still faced with many challenges and limitations, among which, single-stage target detection technology in the pursuit of efficiency, but still faces the dilemma of low detection accuracy, especially in the processing of multi-scale targets, it is difficult to accurately capture and distinguish different sizes of target objects, in addition, a high false detection rate has become a bottleneck restricting its application. Relatively speaking, two-stage target detection technology has improved its accuracy, but its high computational complexity and strong dependence on high-end hardware configuration, as well as the limitation of candidate region generation strategy, make this technology subject to many limitations in actual deployment and difficult to be widely popularized.

In the future development, the development of target detection technology Target detection may move towards multi-modal information fusion. That is, the future target detection system will no longer be limited to relying on a single visual information, but will actively expand the field of vision, including the accurate distance and spatial structure information provided by LiDAR, the thermal radiation difference information captured by infrared sensors, and the surface depth information obtained by depth sensors. Through an intelligent information fusion algorithm, accurate identification and detection of target objects under complex environmental conditions can be achieved, significantly improving the accuracy and robustness of detection.

At the same time, the future target detection system will pay more attention to the optimization of self-adaptation and real-time. In diversified application scenarios with limited resources (such as limited computing power and low hardware configuration), the object detection model will have intelligent self-adjustment ability, and can flexibly adjust the complexity and calculation path of the model according to the performance and actual requirements of the current device, to maximize the real-time performance based on ensuring the detection accuracy. Ensure stable and efficient operation in various devices and scenarios, and provide strong technical support for the development of automatic driving, intelligent security, human-computer interaction, and other fields.

5. Conclusions

This paper introduces the principle of target detection technology and discusses the application of target detection technology in many fields, including the industrial field, medical field, and agricultural field, which can be summarized as industrial production line automation, safety monitoring, and storage logistics, automatic analysis of medical pathological sections, cell behavior tracking, and so on. Real-time monitoring of crop pests and diseases in agriculture and practical applications in three fields of smart agriculture. Although the target detection technology still has the limitations of low accuracy and large resource demand, after future model iteration and algorithm optimization, the accuracy and efficiency can be significantly improved, and the computing cost can be reduced. Meanwhile, after multi-modal information fusion, robustness and adaptability in complex environments will be further improved, and the universality will also have great room for improvement.

References

[1] Zou Z, Chen K, Shi Z, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.

[2] Pathak A R, Pandey M, Rautaray S. Application of deep learning for object detection[J]. Procedia computer science, 2018, 132: 1706-1717.

[3] Diwan T, Anirudh G, Tembhurne J V. Object detection using YOLO: Challenges, architectural successors, datasets and applications[J]. multimedia Tools and Applications, 2023, 82(6): 9243-9275.

[4] Lee J, Hwang K. YOLO with adaptive frame control for real-time object detection applications[J]. Multimedia tools and applications, 2022, 81(25): 36375-36396.

[5] Wu Wendi, Liu Yinzhu, Liu Jiwei, Chen Zhangbao. Target detection for industrial sorting based on YOLOv3 algorithm[J]. Industrial Control Computer, 2023, 36(10): 106-107+110.

[6] Pei Jiabin. Research on small target detection algorithm for complex industrial environments based on YOLO[D]. Qilu University of Technology, 2024. DOI:10.27278/d.cnki.gsdqc.2024.000257.

[7] Shen Jinchao, Feng Sikai, Peng Hui. Application of six-degree-of-freedom robotic arm Mecanum wheel robot in logistics handling based on target detection[J]. Electrical Times, 2024(05): 32-36.

[8] Xu Peiyuan. Research on target detection and segmentation of medical images[D]. Nanjing University of Aeronautics and Astronautics, 2019. DOI:10.27239/d.cnki.gnhhu.2019.000775.

[9] Chen Yiying, Guo Dan, Sun Jian, Zhang Sumei, Li Zhiyuan, He Jing, Li Mei, Li Wenbo, Wang Anqi, Zhang Yang. Method and device for acquiring growth information of ovarian clear cell carcinoma organoid[P]. Beijing: CN118570801A, 2024-08-30.

[10] Hu Shuangshuang. Research and implementation of agricultural pest detection system based on convolutional neural networks[D]. Nanjing Forestry University, 2023.