

Progress of Object Detection Algorithms Based on Deep Learning, Lightweight Optimization, and Cross-Domain Applications

Yayu Cai

*College of Science and Technology, Beijing Normal-Hong Kong Baptist University, Zhuhai, China
s230026004@mail.uic.edu.cn*

Abstract. As one of the core tasks in the field of computer vision, object detection has made significant progress over the last few years, driven by deep learning technology. This paper systematically reviews the development process of object detection algorithms, from traditional methods to modern deep learning methods, focusing on the technical characteristics and performance of representative algorithms such as the You Only Look Once (YOLO) series, Region-Based Convolutional Neural Networks (R-CNN) series, and the Single-Shot MultiBox Detector (SSD). In this paper, the lightweight optimization strategy is discussed in depth to deal with difficulties of computing resource limitations in practical applications of object detection, including network structure simplification, computing efficiency optimization, and hardware adaptation. In addition, this paper further combines typical cases, such as a two-stage efficient parking space detection method, to expound the application status and challenges of object detection technology in cross-field fields such as autonomous driving, intelligent transportation, and industrial testing. Finally, this paper looks forward to the future development direction of object detection technology, and puts forward potential research hotspots such as multimodal fusion and edge computing adaptation.

Keywords: Object Detection, Deep Learning, Computer Vision

1. Introduction

Object detection technology, designed to automatically identify and position specific objects in images or videos, is a fundamental problem in the field of computer vision. The rapid advancement of deep learning has led to groundbreaking improvements in both accuracy and efficiency for CNN-based object detection algorithms. After AlexNet made a breakthrough in the ImageNet competition in 2012, many innovative algorithm architectures have emerged in the domain of object detection, such as the R-CNN series, the YOLO series, SSD, etc., which have promoted technological innovation in this field. On this basis, the researchers explored the application of object detection in different scenarios. Computer vision and deep learning models, such as YOLO, offer an excellent solution for detecting, classifying, and counting insect populations in a shorter period of time, which has significant advantages over traditional sampling methods [1]. Onososen et al. provide a scheme

for fatigue detection of construction workers according to the YOLOv8 algorithm and computer vision technology, which achieved an average accuracy of 92% and an inference speed of 7.5 milliseconds on the test dataset [2].

However, in practical application scenarios, object detection systems often face challenges with limited computing resources and high real-time requirements. For example, parking space detection in autonomous driving systems relies on complex sensor arrays, resulting in bloated and costly systems. At the same time, its feature extraction mainly relies on manual design, and its generalization ability and robustness are insufficient. These limitations have prompted researchers to seek more efficient and lightweight object detection solutions.

This paper aims to comprehensively sort out the technical evolution path of object detection algorithms, deeply analyze the key technologies of lightweight optimization, and explore the application potential and implementation challenges of object detection technology in different fields, in combination with typical application cases such as parking space detection. Through systematic literature review and technical analysis, it provides a reference framework for researchers in related fields and looks forward to possible future technological development directions.

2. Evolution of object detection algorithms and technical analysis

2.1. Traditional object detection methods

Before the popularization of deep learning technology, object detection mainly relied on the extraction and classification of manually designed features. These traditional methods typically employ multi-stage processing, including candidate region generation, feature extraction, region classification, and border regression. Typical feature extraction methods include directional gradient histogram and scale invariant feature transformation. In general, in the field of parking space detection, traditional methods are mostly based on the technical route of computer vision. For example, Jung et al. identified parking lines based on the fixed-width parking space line hypothesis by using peak detection and clustering techniques in Hough space. A parking line detection method based on Radon space was proposed by Wang et al., which showed better noise immunity and robustness than the Hough transform [3]. However, these methods are sensitive to changes in parking line width and have limited adaptability.

Traditional object detection methods have several significant limitations, which are mainly reflected in three aspects: feature expression ability, processing speed, and generalization ability. In terms of feature expression, it is difficult to effectively capture the diversity of targets in complex scenes. In terms of processing processes, multi-stage independent optimization leads to inefficient systems and low overall integration. In addition, the characteristics of this method are designed for specific scenarios, making it lack the ability to generalize across scenarios. It is these inherent flaws that have driven researchers to turn to deep learning-based object detection methods.

2.2. Algorithm innovation in the era of deep learning

The rise of deep convolutional neural networks (DCNNs) has revolutionized object detection. Unlike traditional staged processing, DCNN can learn the mapping relationship from the unprocessed image to the target location and category end-to-end, greatly enhancing the composite performance of the system. DCNN automatically extracts hierarchical features through convolutional layers, and gradually builds the representation ability of the target from low-level edges and textures to high-level semantic features. Typical network architectures include

convolutional layers, pooled layers, and fully connected layers, which significantly reduce model complexity through parameter sharing mechanisms [3].

In the field of object detection, algorithm innovation develops along two main lines: two-stage detectors and single-stage detectors. The two-stage detector, represented by the R-CNN series, first generates candidate regions and then classifies and regresses each region. From R-CNN to Fast R-CNN, Faster R-CNN to Mask R-CNN, these methods continuously optimize the field generation and feature sharing mechanisms, gradually improving the detection accuracy and speed. In particular, Faster R-CNN introduces the regional proposal network (RPN) to achieve end-to-end training and set a new benchmark in accuracy. Viardin et al. used Faster R-CNN to automate the microstructure of dendrites, with recognition rates ranging from 98% to 100% in all cases [4]. This suggests that Faster R-CNN has advantages in detecting dendrites near equiaxial dense fronts as well as overlapping dendrites.

Single-stage detectors are represented by YOLO and SSD. They treat object detection as a regression problem and directly predict bounding boxes and class probabilities on the image grid. These methods eliminate the step of generating candidate regions, showing significant speed advantages and being more applicable for real-time systems. The series of YOLO continuously optimized the network structure and training strategies from the initial version to YOLOv11, improving detection accuracy while maintaining real-time performance. For example, YOLOv11 achieved a detection accuracy close to that of two-stage methods on standard datasets through architecture search and loss function improvements, while maintaining extremely high inference speed.

It is worth noting that in recent years, the Transformer architecture has also begun to penetrate the field of object detection. Detection transformer (DETR) was the first to introduce the Transformer into object detection, using the self-attention mechanism to globally model the image content and breaking away from the dependence of traditional methods on predefined anchor boxes. Although the computational cost is relatively high, the DETR series of algorithms has shown excellent detection performance in complex scenarios, opening up new ideas for object detection. Table 1 lists the types of current mainstream two-stage efficient parking detection methods and the corresponding time consumption, providing a more intuitive comparison.

Table 1. Comparison of computational efficiency of two-stage efficient parking detection method

Method	Time consumption (ms)
Faster-RCNN	63.7
SSD	27.1
DETR	106.4
YoloV11	9.3

Taking into account all the advantages and disadvantages, and based on the time consumption, it can be concluded that YoloV11 is the most appropriate option for detecting the essentials of parking spaces.

2.3. Automatic recognition methods and reduction of annotation costs

By integrating the latest computer vision technology with intelligent training methods, the field of industrial defect detection has achieved a dual breakthrough in efficiency and accuracy. Schaller et al. proposed an automatic recognition method for holes in aircraft engines, using the YOLOv8n

object detection algorithm. The model achieved the best detection effect under the metric of an average accuracy rate of 50, and proposed an improvement plan for optimizing rotated images and integrating synthetic data [5]. Yang et al. applied YOLOv10 to the real-time detection of polycrystalline diamond compact drill cutting teeth and built a lightweight wear-type and wear-grade classification model based on SqueezeNet. During the training process, data augmentation (such as rotation and scaling) was applied to solve the problem of sample imbalance, and semi-supervised iterative training (initial training → prediction generation → manual correction → re-training) was adopted to reduce annotation costs [6]. Based on the improved schemes of YOLOv8n and YOLOv10, combined with the semi-supervised learning strategy, not only has efficient automatic recognition of industrial defects been achieved, but also the cost of manual annotation has been significantly reduced, providing an expandable solution for quality inspection in the field of intelligent manufacturing.

3. Lightweight optimization strategies and technology implementation

3.1. Model compression and acceleration techniques

As the deployment requirements of object detection algorithms on mobile devices and embedded systems increase, model lightweighting has become a key research direction. Lightweight optimization mainly focuses on three aspects: network structure design, computational efficiency optimization, and hardware adaptation.

At the network structure level, techniques such as depthwise separable convolution, channel pruning, and layer fusion can effectively reduce both parameter count and computational cost. By factorizing standard convolution into depthwise and pointwise steps, depthwise separable convolution achieves a quadratic reduction in computations. Channel pruning assesses the importance of each channel and removes redundant channels that have little impact on the output. Layer fusion combines adjacent linear operations to reduce memory access overhead.

In the computational optimization aspect, quantization technology converts floating-point parameters into low-bit integer representations, which can both compress the model size and accelerate integer operations. Knowledge distillation transfers capabilities from a large teacher model to a compact student model by mimicking the teacher's output distributions, enabling lightweight models to achieve performance close to that of complex models. Additionally, neural network architecture search (NAS) can automatically explore the optimal lightweight structure. Liu et al. presented a NAS framework based on knowledge distillation for the method of driving distraction recognition, and for the experimental dataset, the parameter quantity was reduced by 55% and the accuracy was increased by 2.91% [7]. In summary, lightweight optimization techniques, through multi-level and multi-dimensional innovative methods, have achieved efficient deployment and application of object detection algorithms in resource-constrained environments.

3.2. Lightweight implementation of two-stage parking space detection

The two-stage parking space detection method presents a typical case of lightweight optimization. This method decomposes parking space detection into two stages, global keypoint detection and local direction judgment, and adopts the most suitable technical solutions for each stage. The first stage uses the YOLOv11 to capture the bounding boxes of parking space essentials from the surround view images, fully leveraging the advantages of deep learning in feature extraction. The second stage employs traditional computer vision methods, processing local images with specially

designed convolution kernels, efficiently obtaining parking space direction information. This hybrid strategy retains the representational ability of deep learning while achieving an overall efficiency improvement by simplifying some calculation processes.

The two-stage efficient parking space detection method proposed by Jiang et al. achieved a detection accuracy of 98.24% on the public dataset ps2.0, and the inference time for parking space detection on a single image was only 12.3 milliseconds [3]. This is due to the use of the computationally efficient YOLOv11 network in the keypoint detection stage and the design of dedicated convolution kernels in the direction judgment stage to replace complex neural networks, significantly reducing computational costs. This phased and differentiated optimization approach provides an important reference for lightweight object detection.

4. Future outlook and challenges

Combining object detection with multi-source data, such as radar, can provide richer environmental information and enhance the robustness of the detection system. Researchers in related fields will continue to probe the performance of deep learning in industrial automation, healthcare, environmental protection, and other fields by integrating interdisciplinary technologies [7]. Moreover, object detection is also widely applied in scenarios such as product recognition, smart construction site safety monitoring, intelligent search, and target tracking, demonstrating strong versatility [8].

However, small targets suffer from detail loss due to downsampling, and dense scenes encounter difficulties in detection due to occlusion and overlap. Solutions include multi-scale fusion, high-resolution input, and occlusion-aware modules (such as Occlusion-Net), but the computational complexity still needs to be optimized. In terms of multimodality and cross-domain generalization, a single sensor performs poorly under extreme weather conditions (rain, fog, low light), and it is necessary to combine multimodal data such as lidar, infrared, and depth cameras. Existing methods (such as Combined Spatial Attention and Improved Pyramid Pooling) improve robustness, but they have high computational complexity and deployment costs. In the future, more efficient fusion strategies are needed. Edge computing devices (such as drones, intelligent cameras) require lightweight models, which drives the development of technologies such as Anchor-Free, pruning, quantization, and knowledge distillation. A series of algorithms balance performance and real-timeity through structural optimization (such as lightweight backbone networks), but further optimization is still needed to adapt to low-power scenarios [9].

The breakthroughs of future single-stage detectors should focus on collaborative innovation at the data level (generative models + weak supervision to reduce annotation dependence), the architecture level (lightweight design such as without non-maximum suppression for multi-modal adaptation), the task level (improvement of small object detection and robustness in complex scenarios), and the deployment level (real-time optimization of edge devices such as Tiny-YOLO) [10]. Moreover, the core goal of future research lies in balancing accuracy and efficiency (reducing computational reliance through unsupervised/supervised learning), enhancing cross-domain generalization ability (improving robustness through multi-modal data fusion), and achieving low-cost engineering implementation (developing end-to-end deployment toolchains to support quantization and standardized formats) [11].

5. Conclusion

This article systematically reviews the object detection technology based on deep learning and computer vision, analyzes the algorithm evolution path and lightweight optimization strategies, and discusses cross-domain application practices. The two-stage parking space detection method, as a typical case, demonstrates the feasibility of integrating the strengths of deep learning and traditional computer vision, providing a reference solution for efficient object detection in resource-constrained scenarios. Cross-domain applications show that the successful deployment of object detection technology not only depends on the performance of the algorithm itself, but also needs to consider factors such as scene characteristics, hardware constraints, and user experience. With the continuous innovation of algorithms and the continuous evolution of hardware, object detection will play a key role in more fields and provide basic support for intelligent applications. Especially in professional fields such as autonomous driving and industrial inspection, algorithm optimization tailored to local conditions and system integration will be the key to successful deployment.

References

- [1] Lim, M. H., Chan, H. H., & Ong, S.-Q. (2025). An annotated image dataset of urban insects for the development of computer vision and deep learning models with detection tasks. *Data in Brief*, 60, 111673.
- [2] Onososen, A. O., Musonda, I., Onatayo, D., Saka, A. B., Adekunle, S. A., & Onatayo, E. (2025). Drowsiness Detection of Construction Workers: Accident Prevention Leveraging Yolov8 Deep Learning and Computer Vision Techniques. *Buildings*, 15(3), 500.
- [3] Jiang, J., Tang, R., Kang, W., Xu, Z., & Qian, C. (2025). Two-Stage Efficient Parking Space Detection Method Based on Deep Learning and Computer Vision. *Applied Sciences*, 15(3), 1004.
- [4] Viardin, A., Nöth, K., Pickmann, C. et al. Automatic Detection of Dendritic Microstructure Using Computer Vision Deep Learning Models Trained with Phase Field Simulations. *Integr Mater Manuf Innov* 14, 89–105 (2025).
- [5] Schaller, T., Li, J., & Jenkins, K. W. (2025). A Data-Driven Approach for Automatic Aircraft Engine Borescope Inspection Defect Detection Using Computer Vision and Deep Learning. *Journal of Experimental and Theoretical Analyses*, 3(1), 4.
- [6] Yang, X., Feng, X., Cheng, C., Yu, J., Zhang, Q., Gao, Z., Liu, Y., & Chen, B. (2025). Automatic detection and classification of drill bit damage using deep learning and computer vision algorithms. *Natural Gas Industry B*, 12(2), 195–206.
- [7] Dong, J., Sang, F., Guo, Q., et al. (2025). A review on lightweight research of object detection algorithms based on deep learning. *Computer Science and Exploration*, 1–34. Advance online publication. <https://doi.org/xxxx>
- [8] Hou, X. (2025). Object detection and applications based on deep learning. *Innovation and Application of Science and Technology*, 15(8), 21–26.
- [9] Liu, G., Wang, H., Ren, G., et al. (2025). A review on monocular vision object detection based on deep learning. *Computer Engineering and Applications*, 1–27. Advance online publication. <https://doi.org/xxxx>
- [10] Wang, N., & Zhi, M. (2025). A review on research on general object detection algorithms in deep learning. *Computer Science and Exploration*, 19(5), 1115–1140.
- [11] Wang, Y. (2025). A review of object detection algorithms based on deep learning. *Science and Technology Information*, 23(2), 64–66.