WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 1, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n1.11
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Overview of Object Detection Methods

Jingbin Yang [1, 2, 3], Xiumin Shi [1, 2, 3], Sanpeng Deng [1, 2, 3, *], Yuming Qi [1, 2, 3]

[1] Institute of Robotics and Intelligent Equipment, Tianjin University of Technology and Education, Tianjin 300222, China

[2] Tianjin Key Laboratory of Intelligent Robot Technology and Application, Tianjin 300350, China

[3] Tianjin Bonus Robotics Technology Co. td, Tianjin 300350, China

*Corresponding Author: Xiumin Shi

ABSTRACT

Object detection is an important task in computer vision, aimed at detecting and recognizing the position and category of target objects from images or videos. With the rise of deep learning, the accuracy and efficiency of object detection have significantly improved, especially the application of convolutional neural networks (CNN) in this field, which has made significant breakthroughs in object detection methods. This article provides an overview of the development history of object detection, with a focus on classic object detection algorithms, deep learning methods, and their evolution. It explores the evaluation criteria and challenges faced by object detection, and looks forward to future development trends.

KEYWORDS

Object detection; Deep learning; Convolutional neural networks; YOLO; Faster R-CNN; Evaluation Criteria1

1. INTRODUCTION

Object detection is one of the core problems in computer vision, which aims to find all interested target objects from an image and determine their categories and positions (through bounding boxes). Unlike image classification tasks, object detection not only requires classifying objects in the image, but also accurately marking the spatial position of objects, making it a complex and challenging task in computer vision.

In recent years, with the development of deep learning, especially convolutional neural networks (CNN) [1-5], object detection technology has made significant progress. Deep learning methods have made significant breakthroughs, especially in problems such as multi-scale, complex backgrounds, occlusion, and real-time detection. However, despite significant progress, object detection still faces many challenges, such as detecting small objects, recognizing multiple types of targets, and class imbalance.

This review will explore in detail the background, development history, main algorithms, evaluation criteria, challenges faced, and future research directions of object detection, aiming to provide valuable references for scholars and engineers engaged in related research.

2. THE DEVELOPMENT HISTORY OF OBJECT DETECTION

2.1. Traditional Method

Before deep learning, object detection mainly relied on manual features and traditional machine learning methods. The core idea of these methods is to describe the objects in the image by designing some feature extraction algorithms, and to perform object recognition through classifiers such as support vector machines, decision trees, etc.

2.1.1. Haar features and AdaBoost

In the late 1990s, the combination of Haar features [6-7] and AdaBoost [8-9] algorithm became a classic object detection method, particularly achieving great success in face detection. The Haar feature calculates the brightness difference of an image through rectangular regions, which has the characteristics of simple calculation and fast speed. AdaBoost is used to select the optimal features from a large number of weak classifiers and perform weighted combinations. However, the main issue with this method is its poor robustness when the background is complex.

2.1.2. HOG features and SVM

In 2005, Dalal and Triggs proposed the HOG feature [10-11] and combined it with support vector machine [12-13] for pedestrian detection. HOG features can effectively capture the shape and contour of objects by calculating the local gradient direction information of the image. The advantage of this method lies in its good robustness, especially for pedestrian detection, which has a wide range of applications.

2.1.3. Sliding Window and Feature Pyramid

In traditional methods, sliding a window [14-15] is a common technique that detects targets in an image by sliding a fixed size window at different scales. In order to improve detection accuracy and speed, the feature pyramid [16] method is commonly used, which processes targets of different sizes through multi-scale feature representation. However, sliding windows have high computational overhead and poor detection performance for complex backgrounds and high-density targets.

2.2. The Rise of Deep Learning

The successful application of deep learning, especially convolutional neural networks (CNN), in image classification tasks has provided strong impetus for the advancement of object detection. CNN can automatically learn complex features from raw pixels to high-level semantics, greatly reducing the workload of manual feature extraction.

2.2.1. R-CNN

In 2014, R-CNN (Region based CNN) [17-18] was proposed, marking a breakthrough in the application of deep learning in object detection. R-CNN first uses selective search to generate candidate regions, then uses CNN to extract features from each candidate region, and finally uses SVM for classification. Although R-CNN has high detection accuracy, its computational efficiency is low because each candidate region requires separate forward propagation.

2.2.2. Fast R-CNN

In order to solve the problem of low computational efficiency in R-CNN, Girshick proposed Fast R-CNN [19-20] in 2015. Fast R-CNN extracts features by performing a convolution operation on the entire image, and uses RoI Pooling to pool each candidate region, thus avoiding the inefficient problem of repeated calculations in R-CNN. Fast R-CNN has made significant improvements in both accuracy and speed.

2.2.3. Faster R-CNN

In 2015, Faster R-CNN [21-22] further improved the efficiency of object detection by introducing Region Proposal Networks (RPNs). RPN achieves end-to-end training by sliding windows on convolutional feature maps to generate candidate regions. The biggest advantage of Faster R-CNN is that it simultaneously performs object detection and region proposal by sharing convolutional features, greatly improving detection speed.

2.3. Single Stage Detection Method

Although two-stage methods based on region proposals, such as R-CNN and Faster R-CNN, have shown outstanding accuracy, their computational complexity is relatively high. Therefore, researchers have proposed single-stage object detection methods. These methods avoid the region proposal stage and improve detection speed by directly predicting the target category and position on the entire image.

2.3.1. YOLO

YOLO [23-24] was proposed by Redmon et al. in 2016, using a single convolutional neural network for object detection. YOLO treats object detection tasks as regression problems, where the network outputs both the object category and bounding box coordinates during a forward propagation process. The advantage of YOLO is its extremely fast detection speed, making it suitable for real-time detection tasks. Although YOLO has high accuracy, it has certain limitations in small object detection.

2.3.2. SSD

SSD [25-26] proposes a multi-scale convolutional feature map that can simultaneously perform object detection on feature maps of different scales. SSD performs well in processing multi-scale targets by introducing multiple detectors. Similar to YOLO, SSD is also a single-stage detection method, but it has improved accuracy.

2.3.3. RetinaNet

RetinaNet [27-28] solves the problem of class imbalance by introducing focal loss. Focus loss can suppress the influence of background categories and improve the model's detection ability for rare targets. RetinaNet performs well in single-stage detection methods, especially when dealing with target categories with long tail distributions.

3. EVALUATION CRITERIA AND CHALLENGES FOR OBJECT DETECTION

3.1. Evaluation Criteria

The evaluation of object detection typically relies on the following criteria:

(1) Average precision (mAP)

Mean Average Precision (mAP) is the most commonly used evaluation metric in object detection. MAP calculates the average precision (AP) for each category, and then averages the AP for all categories. The calculation method of AP is based on different Intersection over Union (IoU) thresholds.

(2) Precision and Recall

Accuracy and recall are commonly used performance evaluation metrics, representing the proportion of correct targets in the detection results and the proportion of actual targets detected, respectively. The comprehensive performance of accuracy and recall is usually evaluated by F1 score.

3.2. Challenge

The main challenges faced by object detection include:

(1) Small object detection: Small objects have weaker features and are easily interfered by background, resulting in lower detection accuracy. The existing multi-scale detection methods have made some improvements in small object detection, but still find it difficult to completely solve this problem.

(2) Multi category and long tail problems: The distribution of categories in object detection is often uneven, with fewer samples in certain categories, resulting in lower detection accuracy of the model on these categories.

(3) Real time detection: In some applications that require high real-time performance, such as autonomous driving and monitoring, object detection models not only need to maintain high accuracy, but also need to have high processing speed

4. RESEARCH PROSPECTS

The field of object detection is continually evolving, with significant advancements driven by deep learning and neural network architectures. However, several key areas remain for further improvement and innovation.

(1) Improving Small Object Detection: Despite recent progress, detecting small objects continues to be a major challenge. Future research should focus on developing more effective multi-scale architectures, utilizing attention mechanisms or advanced feature fusion strategies to enhance the representation of small objects. Moreover, integrating temporal information in video-based object detection could further aid in identifying small moving objects.

(2) Handling Class Imbalance: As object detection often faces imbalanced datasets with many more background samples than foreground objects, addressing class imbalance remains critical. New loss functions, like focal loss (used in RetinaNet), and techniques such as hard negative mining could be further refined and applied to improve the model's performance in detecting underrepresented categories.

(3) Real-Time Detection: Achieving high accuracy while maintaining real-time processing is crucial, especially for applications in autonomous vehicles, security, and robotics. To address this, future research should explore novel network architectures optimized for speed without sacrificing detection precision. This may involve developing lightweight models, pruning techniques, or hardware acceleration methods tailored for mobile and embedded systems.

(4) 3D Object Detection and Scene Understanding: Expanding object detection from 2D images to 3D space is an emerging research frontier. By incorporating depth information from LiDAR or stereo cameras, researchers can improve the robustness of detection in environments with complex geometries, such as urban and indoor scenes. Future methods should focus on end-to-end learning systems that combine image and point cloud data for improved scene understanding.

(5) Cross-Domain and Unsupervised Learning: The effectiveness of object detection models often diminishes when trained on one dataset and applied to another (cross-domain problems). To mitigate this, unsupervised or semi-supervised learning techniques should be explored to reduce the dependency on large labeled datasets, especially in domains where annotation is expensive or impractical.

(6) Multimodal and Multitask Learning: Future research should focus on combining object detection with other tasks such as semantic segmentation, tracking, and scene recognition in a unified framework. Multimodal approaches that integrate data from various sensors (e.g., visual, auditory,

and tactile) could lead to more comprehensive models capable of making robust predictions in diverse environments.

5. SUMMARY

Object detection remains one of the most challenging and dynamic areas of computer vision, with a variety of applications across industries such as autonomous driving, surveillance, healthcare, and robotics. This review has provided an overview of the evolution of object detection, from traditional feature-based methods to the more recent deep learning-driven approaches.

Traditional methods, such as Haar features combined with AdaBoost, and HOG features with SVM, paved the way for early successes in face and pedestrian detection. However, these techniques were limited in handling complex scenarios, such as varied object scales and challenging backgrounds. The introduction of deep learning, particularly convolutional neural networks (CNN), marked a major breakthrough, allowing for end-to-end feature learning and improved detection performance.

The progression from R-CNN to Faster R-CNN and single-stage detectors like YOLO, SSD, and RetinaNet has significantly improved detection accuracy and speed. Despite these advancements, challenges such as small object detection, class imbalance, and real-time performance still present significant hurdles. Current and future research directions aim to address these challenges, exploring novel architectures, loss functions, and multimodal approaches to push the boundaries of what object detection models can achieve.

In conclusion, while object detection has made remarkable strides, the field is far from reaching its full potential. Continuous advancements in deep learning, as well as the integration of new techniques such as 3D detection and cross-domain learning, will shape the future of object detection systems, enabling more robust, efficient, and versatile solutions for real-world applications.

REFERENCES

[1] Luis R. Mercado Diaz, Derek Aguiar & Hugo F. Posada Quintero. (2025). Graph-based multi-modal MRI analysis with probabilistic attention for stroke lesion detection. Neurocomputing, 657, 131620-131620. https://doi.org/10.1016/J.NEUCOM.2025.131620

[2] Mingjian Yuan, Gongcheng Peng, Ruru Xiao, Bing Zeng, Wenxuan Xu, Guangyao Li... & Hao Jiang. (2025). Prediction model for 1 mm thickness A1060 aluminum and T2 copper magnetic pulse welding joints based on CNN-LSTM-AM deep learning algorithm and DeepSHAP interpretability analysis. Welding in the World, (prepublish), 1-19. https://doi.org/10.1007/S40194-025-02195-Z

[3] Hamed Sabahno & Davood Khodadad. (2025). A convolutional neural network-based joint detection and localization spatiotemporal scheme for process control through speckle pattern imaging. Computers & Industrial Engineering, 210, 111538-111538. https://doi.org/10.1016/J.CIE.2025.111538

[4] N. Regnier, V. Mungkung & L. Mezeix. (2025). A CNN-based statistical method for land cover classification to assess urban vulnerability to explosions: Case study of Paris, France. International Journal of Applied Earth Observation and Geoinformation, 144, 104878-104878. https://doi.org/10.1016/J.JAG.2025.104878

[5] Yichi Zhang, Yuxin Ma, Hui Fang & Hongqing Wang. (2025). Investigation on forecast of offshore wind power generation hybrid attention mechanism and bi-directional long short-term memory based on deep learning. Ocean and Coastal Management, 270, 107884-107884. https://doi.org/10.1016/J.OCECOAMAN.2025.107884

[6] Fujiao Ju, Shuhan Zhao & Shaotao Zhu. (2025). Enhanced pneumonia lesion segmentation using a hybrid CNN-BiFormer network with residual haar wavelet downsampling and shared attention. Multimedia Systems, 31 (6), 409-409. https://doi.org/10.1007/S00530-025-01990-4

[7] Jinshuai Xu & Juan Zhang. (2025). Infrared and visible image fusion based on Haar wavelet downsampling and multi-scale feature aggregation. Signal, Image and Video Processing, 19 (14), 1200-1200. https://doi.org/10.1007/S11760-025-04746-9

[8] Haixiang Yao & Chunzhuo Wan. (2025). Multi-factor portfolio optimization: A combined random Forest–AdaBoost model with cost-sensitive learning. Pacific-Basin Finance Journal, 94, 102946-102946. https://doi.org/10.1016/J.PACFIN.2025.102946

[9] Jidong Li, Jian Cui & Qian Su. (2025). Boosting fuzzy classification rules under footprint of uncertainty of type-2 fuzzy sets. Applied Soft Computing, 185 (PA), 113876-113876. https://doi.org/10.1016/J.ASOC.2025.113876

[10] Kumaresh Pal, Kumari Namrata, Ashok Kumar Akella, Akshit Samadhiya, Ahmad Taher Azar, Mohamed Tounsi... & Walid El Shafai. (2025). Intelligent islanding detection framework for smart grids using wavelet scalograms and HOG feature fusion. Scientific Reports, 15(1), 30351-30351. https://doi.org/10.1038/S41598-025-08391-7

[11] Peiran Zhang, Fuqiang Zhou, Xinghan Wang, Wentao Guo & Zhipeng Song. (2025). Monocular vision measurement for kinematic parameters of separated load in drop tower test. Optics and Lasers in Engineering, 193, 109074-109074. https://doi.org/10.1016/J.OPTLASENG.2025.109074

[12] H.P.D. Shiran Madhuranga, Wong Jee Keen Raymond, Hazlee Azil Illias & Nurulafiqah Nadzirah Binti Mansor. (2025). Robust open-set partial discharge diagnosis based on hybrid supervised contrastive learning and SVM framework. Ain Shams Engineering Journal, 16(12), 103762-103762. https://doi.org/10.1016/J.ASEJ.2025.103762

[13] Zhenman Gao, Jianyong Zhuang & Xiaoyong He. (2025). FALCON-Net: A hybrid fuzzy-attentive LSTM-convolutional neural network architecture for high-precision classification of steel alloys via femtosecond laser-ablation spark-induced breakdown spectroscopy. Analytica Chimica Acta, 1378, 344701-344701. https://doi.org/10.1016/J.ACA.2025.344701

[14] Tao Shen, Bo Li, Facai Ren, Enxiang Fan & Jianrui Zhang. (2025). A monitoring framework for predicting laser directed energy deposition property. International Journal of Mechanical Sciences, 307, 110861-110861. https://doi.org/10.1016/J.IJMECSCI.2025.110861

[15] Sirui Guo, Jinchang Li, Wei Jiang, Jun Yang, Yingying Du, Tao Luo... & Limei Qi. (2025). Chirality recognition of amino acid by combining machine learning method and sliding window technique. Optics and Laser Technology, 192 (PE), 113937-113937. https://doi.org/10.1016/J.OPTLASTEC.2025.113937

[16] Mbietieu Amos Mbietieu, Tapamo Kenfack Hippolyte Michel, Kouamou Georges Edouard & Norbert Tsopze. (2025). ShipFPN: New Feature Pyramid Network architecture for object detection in high-resolution satellite images: Application to ship detection. Science of Remote Sensing, 12, 100270-100270. https://doi.org/10.1016/J.SRS.2025.100270

[17] Hritu Raj & Gargi Srivastava. (2026). A novel data augmentation strategy for gas leak detection and segmentation using Mask R-CNN and bit plane slicing in chemical process environments. Computers and Chemical Engineering, 204, 109407-109407. https://doi.org/10.1016/J.COMPCHEMENG.2025.109407

[18] Seunghyeon Wang. (2025). Effectiveness of traditional augmentation methods for rebar counting using UAV imagery with Faster R-CNN and YOLOv10-based transformer architectures. Scientific Reports, 15 (1), 33702-33702. https://doi.org/10.1038/S41598-025-18964-1

[19] Seunghyeon Wang. (2025). Effectiveness of traditional augmentation methods for rebar counting using UAV imagery with Faster R-CNN and YOLOv10-based transformer architectures. Scientific Reports, 15(1), 33702-33702. https://doi.org/10.1038/S41598-025-18964-1

[20] Mehmet Özgür Özemre, Jale Bektaş, Hüseyin Yanik, Lütfiye Baysal & Hazal Karslioğlu. (2025). Enhanced diagnostic pipeline for maxillary sinus-maxillary molars relationships: a novel implementation of Detectron2 with faster R-CNN R50 FPN 3x on CBCT images. BMC Oral Health, 25 (1), 1473-1473. https://doi.org/10.1186/S12903-025-06337-Z

[21] Seunghyeon Wang. (2025). Effectiveness of traditional augmentation methods for rebar counting using UAV imagery with Faster R-CNN and YOLOv10-based transformer architectures. Scientific Reports, 15(1), 33702-33702. https://doi.org/10.1038/S41598-025-18964-1

[22] Mehmet Özgür Özemre, Jale Bektaş, Hüseyin Yanik, Lütfiye Baysal & Hazal Karslioğlu. (2025). Enhanced diagnostic pipeline for maxillary sinus-maxillary molars relationships: a novel implementation of Detectron2 with faster R-CNN R50 FPN 3x on CBCT images. BMC Oral Health, 25(1), 1473-1473. https://doi.org/10.1186/S12903-025-06337-Z

[23] Meiyun Chen, Jiacheng Tian, Xiuhua Cao, Zhenxiao Fu & Dawei Zhang. (2025). DM-YOLO for MLCCs' automatic defect detection. Optics and Laser Technology, 192 (PE), 113977-113977. https://doi.org/10.1016/J.OPTLASTEC.2025.113977

[24] Yingying Zhou, Chao Liu, Hao Tian, Xin Zhang & Nan Li. (2025). CSST Slitless Spectra: Target Detection and Classification with Yolo. The Astronomical Journal, 170(5), 256-256. https://doi.org/10.3847/1538-3881/ADEE1C

[25] Fan Zhang, Chi Fai Cheung, Yanbin Zhang & Chunjin Wang. (2025). Modelling and experimental analysis of subsurface damage in low-temperature nano-lubrication grinding. International Journal of Mechanical Sciences, 306, 110729-110729. https://doi.org/10.1016/J.IJMECSCI.2025.110729

[26] Priyanka P. Chavan, Rahul S. Redekar, Umesh D. Babar, Ashok D. Chougale, N.L. Tarwal & Pradip D. Kamble. (2025). Reaction time-driven structural and electrochemical analysis of MnFe2O4 for solid-state supercapacitor application. Journal of Power Sources, 659, 238459-238459. https://doi.org/10.1016/J.JPOWSOUR.2025.238459

[27] Shwetha V., Barnini Banerjee, Vijaya Laxmi & Priya Kamath. (2025). Optimizing TB Bacteria Detection Efficiency: Utilizing RetinaNet-Based Preprocessing Techniques for Small Image Patch Classification. International Journal of Biomedical Imaging, 2025(1), 3559598-3559598. https://doi.org/10.1155/IJBI/3559598

[28] S. Priya & K. Amshakala. (2025). An adaptive fall detection system based on ensemble learning using variants of YOLO V8 retinanet and DETR. Scientific Reports, 15(1), 33161-33161. https://doi.org/10.1038/S41598-025-97634-8