# Research Advances in Object Detection Based on Deep Learning

## Authors Boao Li1, \*, Wangzhixin Qi2, Xiaoning Zhang3 1North China University of Technology, Beijing, China 2International high school of Yuhuatai, Nanjing, China 3King’s College London, London, UK \*Corresponding Author Email: liboao@mail.ncut.edu.cn

## Abstract Object detection has always been a fundamental research topic in the computer vision community, focusing on predicting the category and location of all objects in a scene. In recent years, the rapid development of deep learning has significantly enhanced the speed and accuracy of general object detection methods. This paper reports the latest research progress in deep learning-based object detection and systematically introduces the progress from four aspects: dual-stage, single-stage, Transformer-based, and keypoint. It also compares the performance of different methods on common datasets and summarizes the ongoing challenges and future directions.

\*\*Keywords:\*\* computer vision, natural language processing, Object detection, anchor-free.

## Introduction With the fast development and widespread application of artificial intelligence, autonomous vehicles are becoming an integral part of future transportation. Autonomous driving technology aims to achieve navigation and control of vehicles using computer vision, sensor technology, and deep learning algorithms. The accuracy and real-time performance of object detection directly influence the decision-making and control capabilities of autonomous driving systems.

## Methodology ### 2.1 Two-stage Detection Approaches Two-stage detection methods, including algorithms like Faster R-CNN, decompose the object detection task into two parts: region proposal generation and target classification with bounding box regression. Faster R-CNN, for instance, uses a two-stage network structure with a region proposal network (RPN) and a classifier to achieve high-precision detection across multiple datasets. Despite its advantages, there are limitations such as the small resolution of features maps, which may affect the detection of small and multi-scale objects.

### 2.2 Single-stage Detection Approaches Single-stage detection methods such as YOLO directly predict bounding boxes and categories from the input images without an intermediate region proposal step. These methods are known for their speed, suitable for real-time applications. However, they may face challenges such as missed detections when multiple targets are present in the same grid.

### 2.3 Key Point-Based Detection Approaches Key point-based methods like CornerNet and CenterNet use anchor-free designs to detect objects by locating key points. This approach reduces the need for multiple anchor boxes, cutting down resource consumption and simplifying the model.

### 2.4 Transformer-Based Detection Approaches Transformer-based methods such as DETR use a combination of CNN and Transformer structures to model object detection. These methods leverage the attention mechanism to capture global dependencies, although they may require longer training times and struggle with detecting small objects.

## Experiments and Results ### 3.1 Datasets and Metrics Evaluation of detection algorithms is performed on datasets like Pascal VOC and COCO, using metrics such as Average Precision (AP), mean Average Precision (mAP), and detection speed (FPS).

### 3.2 Performance Comparison The performance of various detection methods is summarized in the following table:

\*\*Table 1. Accuracy (mAP) comparison between different representative methods\*\*  
| Method | Category | VOC2007 | COCO |  
|:------|----------|---------|------|  
| Fast R-CNN | Two-stage method | 70.0 | 19.7 |  
| Faster R-CNN | | 73.2 | 21.9 |  
| YOLOv4 | Single-stage method | 81.8 | 43.5 |  
| SSD | | 81.6 | 26.8 |  
| RetinaNet | | 79.5 | 36.2 |  
| CornerNet | Key point-based method | 79.5 | 42.5 |  
| CenterNet | | 78.0 | 41.8 |  
| Deformable DETR | Transformer-based method | 79.5 | 52.3 |  
| Swin Transformer | | 81.0 | 57.7 |

## Conclusion This review highlights significant advances in object detection technologies, particularly those based on deep learning. While existing methods have improved speed and accuracy, challenges remain, especially in handling small objects and diverse scenarios. Ongoing research focuses on developing more robust models to address these issues and improve the safety and efficiency of autonomous systems.