# Research Advances in Object Detection based on Deep Learning

## 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 the scene. Recent developments in deep learning have significantly enhanced the speed and accuracy of general object detection methods. This paper reports the latest research progress in object detection using deep learning and aims to inspire and promote further research.

## Introduction

With the rapid advancement of artificial intelligence technology, autonomous vehicles are becoming integral to future transportation. Object detection, a critical component of environmental perception in autonomous driving, impacts the decision-making and control capabilities of autonomous systems. This paper discusses the development of object detection technologies, identifies shortcomings, discusses societal implications, and outlines future research directions.

## Methodology

### Two-stage Detection Approaches

Two-stage detection methods, such as Faster R-CNN, involve a region proposal stage followed by classification and bounding box regression. These methods provide high accuracy, especially for small and multi-scale objects but have some drawbacks like slower speeds and high computational requirements.

### Single-stage Detection Approaches

Single-stage methods, like YOLO and SSD, directly predict bounding boxes and class labels from full images, offering faster detection rates suitable for real-time applications but sometimes at the cost of lower accuracy compared to two-stage methods.

### Transformer-based Detection Approaches

Transformers, utilized in DETR, offer a promising direction for object detection by modeling it as an ensemble prediction problem. Improvements such as Deformable DETR and UP-DETR have addressed issues of training efficiency and small object detection inherent in earlier transformer-based methods.

## Experiments & Results

### Datasets and Metrics

Commonly used datasets include Pascal VOC and COCO. Performance metrics such as average precision (AP) and mean average precision (mAP) were highlighted to gauge the effectiveness of different object detection methodologies.

### Performance Comparison

\*\*Table 1. Accuracy (mAP) comparison between different representative methods\*\*

| Method | Category | VOC2007 | COCO |  
|:--------|:---------|-------------:|----------:|  
| Fast RCNN | Two-stage method | 70.0 | 19.7 |  
| Faster RCNN | Two-stage method | 73.2 | 21.9 |  
| SSD | Single-stage method | 81.6 | 26.8 |  
| YOLOv4 | Single-stage method | 81.8 | 43.5 |  
| Retina-Net | Single-stage method | 79.5 | 36.2 |  
| RefineDet | Single-stage method | 79.2 | 39.1 |  
| CornerNet | Key point-based method | 79.5 | 42.5 |  
| CenterNet | Key point-based method | 78.0 | 41.8 |  
| FCOS | Key point-based method | 78.8 | 44.7 |  
| Deformable DETR | Transformer-based method | 79.5 | 52.3 |  
| Swin Transformer | Transformer-based method | 81.0 | 57.7 |

## Conclusion

Object detection technologies have significantly advanced, driven by deep learning innovations. This paper reviewed various methods, discussed their strengths and shortcomings, and highlighted emerging trends that could shape future research directions in object detection.