# Highlights in Science, Engineering and Technology EMIS 2024

## Abstract

Object detection is a crucial aspect of computer vision, focusing on identifying and localizing objects within a scene. Recent advances in deep learning have significantly enhanced the speed and accuracy of object detection methods. This paper discusses the latest developments in this field from various algorithmic perspectives and compares their performance on standard datasets. Challenges and future research directions are also highlighted.

## Introduction

The evolution of artificial intelligence has profoundly influenced autonomous vehicles, revolutionizing not only the automotive industry but also urban development and transportation. Key to these advancements is object detection, an essential component of environmental perception systems critical for the safety and reliability of autonomous vehicles. This section explores the foundational and transformative role of object detection in autonomous driving.

## Methodology

### Architecture Overview

The object detection task is tackled using different architectures such as dual-stage, single-stage, Transformer-based, and keypoint-based approaches, each with distinct advantages and tailored for specific scenarios.

### Algorithms and Equations

Object detection algorithms like Faster R-CNN and YOLO have been pivotal, with Faster R-CNN employing a two-stage process for heightened accuracy and YOLO optimizing for speed with a direct regression approach from images to bounding boxes.

$$L\_{BCE} = -[y \log(p) + (1-y)\log(1-p)] \tag{1}$$

### Training Details

Training methodologies involve various datasets and metrics to evaluate the performance of object detection algorithms, focusing on aspects such as precision, recall, and speed.

## Experiments & Results

### Datasets and Metrics

Popular datasets like Pascal VOC and COCO are utilized, with metrics such as average precision (AP) and mean average precision (mAP) serving as benchmarks.

### Results

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

| Method | Category | VOC2007 | COCO |  
|:--------|:---------|-------------:|----:|  
| Fast RCNN | Two-stage method | 70.0 | 19.7 |  
| Faster RCNN | | 73.2 | 21.9 |  
| YOLOv4 | Single-stage method | 81.8 | 43.5 |  
| Retina-Net | | 79.5 | 36.2 |  
| Deformable DETR | Transformer-based method | 79.5 | 52.3 |  
| Swin Transformer | | 81.0 | 57.7 |

### Ablation Studies

Detailed comparisons and ablation studies highlight the impact of different architectural decisions on the performance metrics.

## Discussion / Limitations

Discusses current limitations in object detection technologies such as the detection in multiple domains, small object detection, and the need for high-precision lightweight network architectures for execution in real-time on limited-resource platforms.

## Conclusion

The paper encapsulates the swift advancements in object detection facilitated by deep learning, illustrates the performance enhancement through algorithmic diversity, and points to future directions that could mitigate current challenges.

## References

1. Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. 1440–1448.  
2. Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv: 1506.01497, 2015.  
3. NEUBECK A, VAN GOOL L. Efficient non-maximum suppression. Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, China, Aug 20- 24, 2006. Piscataway: IEEE, 2006: 850-855.  
4. QIN Y, HE S, ZHAO Y, et al. RoI pooling based fast multi domain convolutional neural networks for visual tracking. Proceedings of the 2016 International Conference on Artificial Intelligence and Industrial Engineering, Beijing, Nov 20-21, 2016: 198-202.  
5. He KM, Gkioxari G, Dollár P, et al. Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. 2980–2988.