AI-Driven Vehicle Detection and Counting System Using Yolov5

Chikkala Henry
Sasi Institute of Technology and
Engineering(CSE-AIML)
Tadepalligudem
Andhra Pradesh,India
chikkalahenry@gmail.com

Saga Parameswararao
Sasi Institute of Technology and
Engineering(CSE-AIML)
Tadepalligudem
Andhra Pradesh,India
parameswararao.saga@sasi.ac.in

Borra Santha Raju
Sasi Institute of Technology and
Engineering(CSE-AIML)
Tadepalligudem
Andhra Pradesh,India
santha.borra@sasi.ac.in

ABSTRACT

Vehicle flow analysis and traffic monitoring are crucial elements of contemporary intelligent transportation systems. Traditional manual traffic monitoring methods are inefficient, time-consuming, and prone to human error. Recent advancements in computer vision and deep learning have enabled the development of automated solutions capable of detecting and analyzing vehicles in real time using video data. In particular, deep neural network-based object detection models have demonstrated significant improvements in accuracy and speed.

This study presents an AI-driven vehicle detection and counting system using the YOLOv5 (You Only Look Once version 5) object detection model. The proposed system processes video streams captured from surveillance cameras and automatically identifies multiple vehicle types, such as cars, buses, trucks, and motorcycles. By leveraging the real-time capabilities of YOLOv5, the system detects vehicles in each frame and applies a tracking and counting mechanism to determine the number of vehicles passing through a defined region.

This implementation integrates deep learning–based object detection with computer vision techniques to provide efficient traffic monitoring. The system was trained on labeled datasets containing various traffic scenarios and evaluated using metrics such as precision, recall, and mean Average Precision (mAP). The experimental results demonstrate that the proposed model achieves reliable detection performance and enables accurate vehicle counting, even in dynamic traffic environments.

This study contributes to the development of intelligent traffic management solutions by providing an automated, scalable, and cost-effective system for vehicle detection and counting. The proposed approach can support applications such as smart city infrastructure, traffic congestion analysis, road safety monitoring and transportation planning.

KEYWORDS: Vehicle Detection, Vehicle Counting, YOLOv5, Computer Vision, Deep Learning, Intelligent Transportation Systems, Real-Time Traffic Monitoring, Object Detection, Smart Traffic Management.

1.INTRODUCTION

Traffic monitoring plays a crucial role in modern transportation systems by enabling efficient traffic management, congestion analysis, and road safety improvement. With the rapid growth of urbanization and the increasing number of vehicles on roads, traditional traffic monitoring techniques such as manual counting and sensor-based systems have become inefficient and difficult to scale. These conventional approaches are often time-consuming, prone to human errors, and incapable of providing real-time insights for intelligent transportation systems [1].

Recent advancements in artificial intelligence (AI), computer vision, and deep learning have enabled automated solutions for traffic monitoring. In particular, object detection algorithms have become widely used for identifying and classifying vehicles from surveillance videos and traffic camera feeds. Among these techniques, the YOLO (You Only Look Once) family of object detection models has gained significant attention due to its ability to perform real-time detection with high accuracy and speed [2]. These models process images in a single pass, making them highly suitable for applications such as vehicle detection, tracking, and counting in real-time environments.

Research in vehicle detection has evolved from traditional image processing techniques such as background subtraction and edge detection to advanced deep learning–based approaches using convolutional neural networks (CNNs). Modern models such as YOLOv5, YOLOv7, and YOLOv8 have significantly improved detection accuracy while maintaining computational efficiency. In addition, tracking algorithms such as SORT and DeepSORT have been integrated with detection models to monitor vehicle movement across consecutive frames, enabling accurate vehicle counting and traffic flow analysis [3][4].

Despite these advancements, several challenges still exist in real-world traffic monitoring systems. Vehicle detection models often face difficulties in handling dense traffic conditions, occlusion between vehicles, variations in lighting conditions, and adverse weather environments such as rain or fog. Furthermore, differences in camera angles, video resolution, and traffic density can affect detection performance and counting accuracy [5][6] . Many existing systems are also evaluated on limited datasets, which restricts their ability to generalize effectively across diverse traffic scenarios.

Therefore, there is a need for a robust and scalable vehicle detection system capable of accurately identifying and counting vehicles in real-time traffic environments. This project proposes an **AI-driven vehicle detection and counting system using YOLOv5**, which processes video streams from traffic cameras and automatically detects different types of vehicles such as cars, buses, trucks, and two-wheelers. By combining deep learning–based object detection with tracking and counting mechanisms, the system aims to provide accurate traffic analysis and support intelligent transportation systems.

This study addresses the following key aspects:

- The effectiveness of deep learning–based object detection models for vehicle detection and classification in traffic videos.
- The role of tracking algorithms in improving vehicle counting accuracy.
- The challenges of implementing real-time traffic monitoring systems under varying environmental and traffic conditions.
- The potential of AI-based systems to support smart city infrastructure and intelligent traffic management.

The contributions of this work include the development of an automated vehicle detection and counting framework using YOLOv5, the integration of tracking and counting mechanisms for accurate traffic flow analysis, and a performance evaluation of the proposed system using standard metrics such as precision, recall, and mean Average Precision (mAP).

The remainder of this paper is organized as follows. Section 2 presents the literature review and analysis of existing vehicle detection systems. Section 3 describes the proposed system architecture and methodology. Section 4 discusses the experimental results and performance evaluation of the model. Section 5 highlights research gaps identified from existing studies. Finally, Section 6 outlines future enhancements and potential applications of the proposed system in intelligent transportation systems.

2.LITERATURE REVIEW

This review follows a structured and systematic process to identify relevant research studies related to vehicle detection, traffic monitoring, and deep learning-based object detection systems. A PRISMA-based methodology was adopted to ensure the transparency, reliability, and quality of the selected research articles.

Relevant literature was collected using academic search engines and digital libraries such as Google Scholar, IEEE Xplore, Springer, ScienceDirect, and Semantic Scholar. The search was conducted using keywords including "vehicle detection using deep learning" ,"traffic monitoring using YOLO" ,"vehicle counting using computer vision" ,"intelligent transportation systems" and "YOLO-based object detection for traffic analysis."

Initially, a total of 512 research articles were identified across these databases. After removing 32 duplicate articles, the dataset was reduced to 480 unique records. Further filtering removed 47 review papers, leaving 433 research articles that were subjected to detailed screening.

During the screening phase, the articles were evaluated based on several inclusion criteria, including publication year (2015–2025), relevance to vehicle detection or traffic monitoring, availability of full text, and publication in English. Articles that did not meet these criteria were excluded.

A second screening phase focused on assessing the quality and relevance of the remaining studies. Each paper was evaluated based on clarity of objectives, soundness of methodology, dataset suitability, evaluation metrics, and reported experimental results. Through this process, 78 relevant research papers were shortlisted.

After a detailed full-text analysis, 20 high-quality research articles were finally selected for the literature review. These studies represent significant contributions to the field of AI-driven vehicle detection, object tracking, and intelligent traffic monitoring systems.

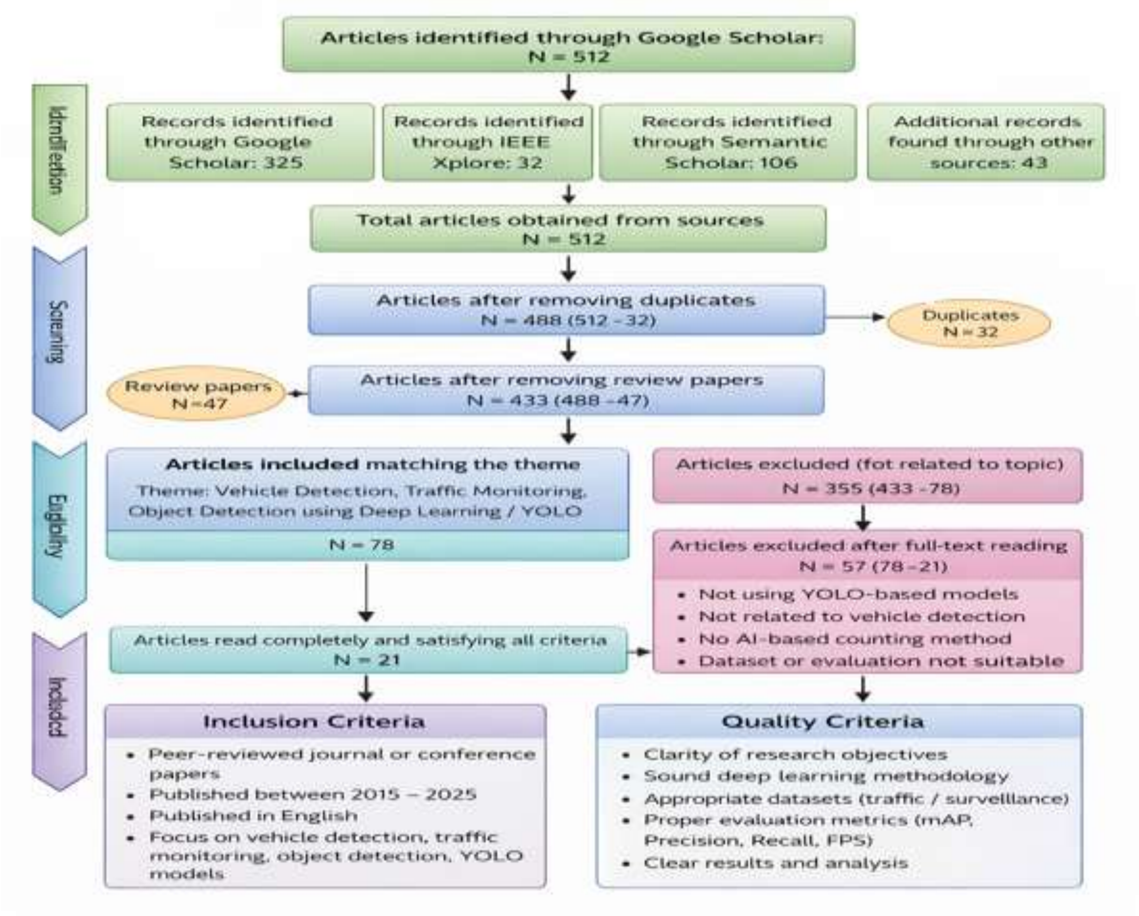


Figure 2.1: Flow Chart representing PRISMA standard applied for paper selection

The publication trend analysis provides an overview of the number of research papers and review papers published between 2015 and 2025 in the field of vehicle detection and traffic monitoring using artificial intelligence.

The data indicates that the number of research publications has steadily increased over the years, particularly after 2020, reflecting growing interest in intelligent transportation systems and deep learning-based object detection. A significant increase is observed in **2023**, where the number of research papers reached its highest level.

This growth can be attributed to advancements in deep learning architectures such as YOLO, Faster R-CNN, SSD, and transformer-based detection models, which have significantly improved the accuracy and efficiency of vehicle detection systems.

In contrast, the number of review papers remains relatively small compared to research articles. This trend suggests that the field is still evolving rapidly with new experimental studies and technological developments.

Overall, the increasing number of publications demonstrates the growing importance of AI-based traffic monitoring solutions for smart cities, autonomous vehicles, and intelligent transportation systems.

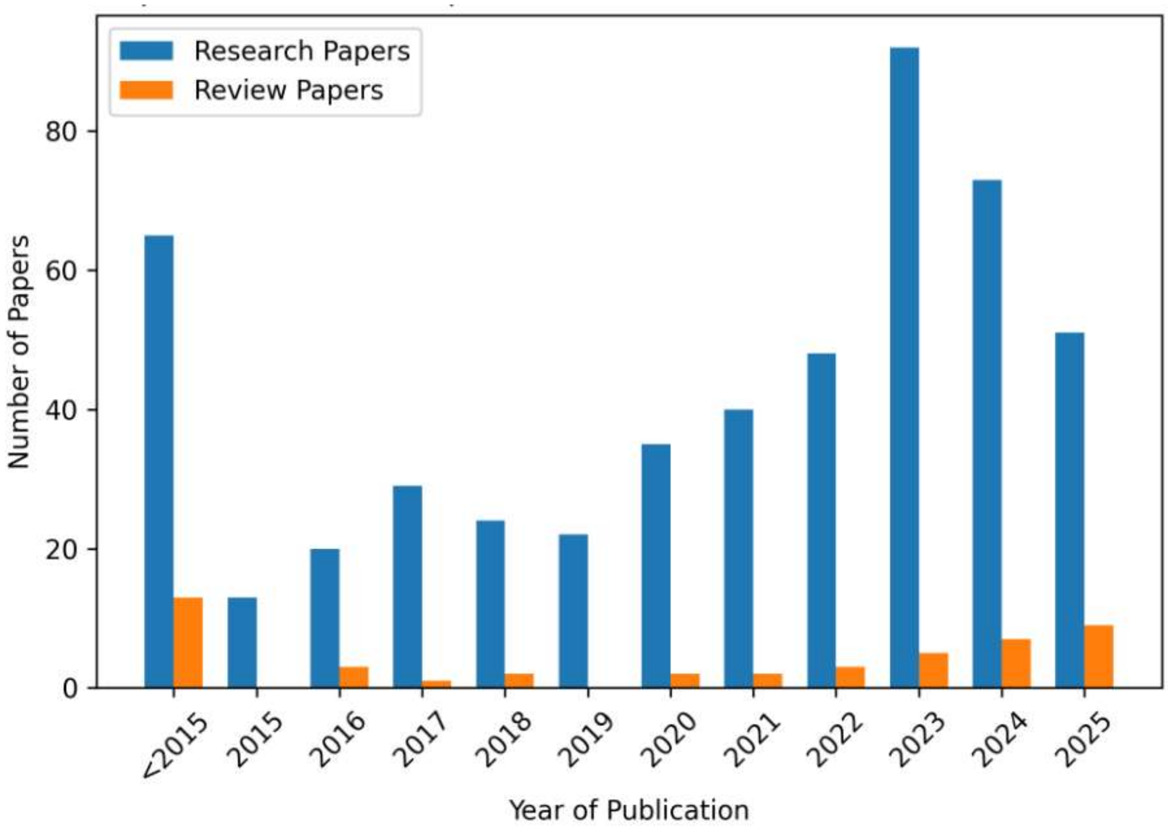


Figure 2.2: Research works on Meeting Summarization and Key Points Extraction systems published from <2015-2025.

Key Aspects of Vehicle Detection Systems

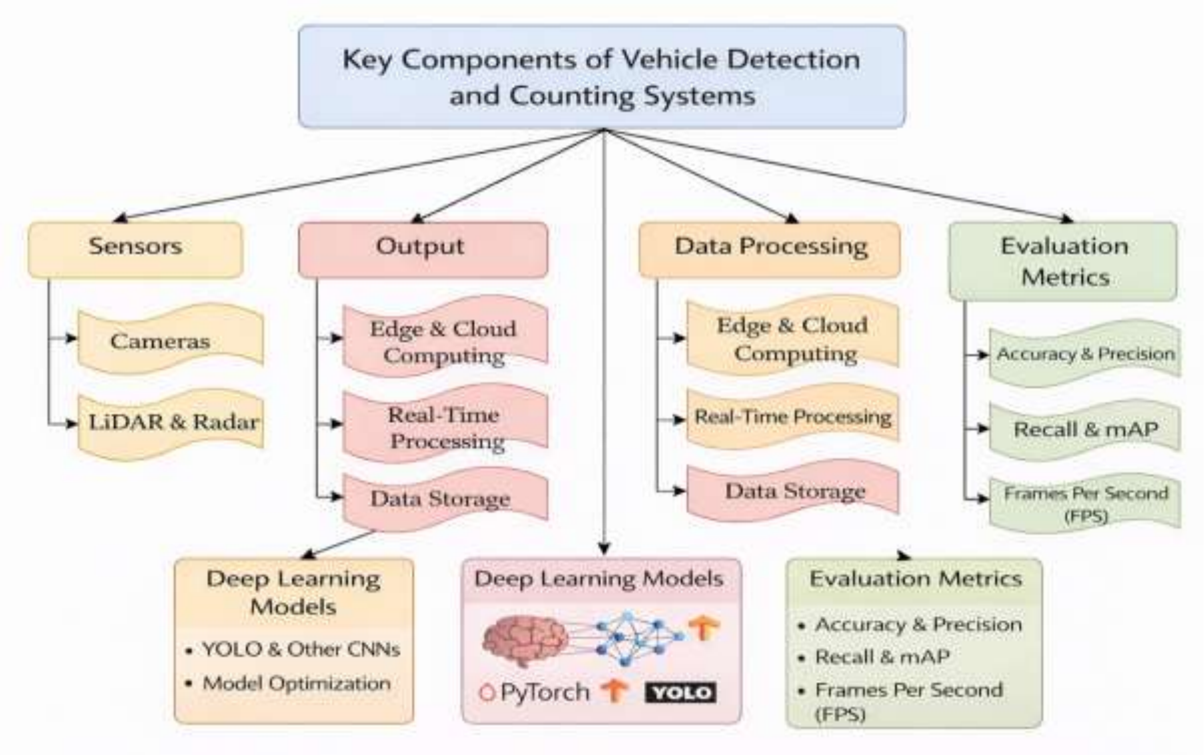


Figure 2.3 Key Aspects of Vehicle Detection System

Vehicle detection systems are typically composed of multiple components that work together to analyze traffic video streams and extract meaningful insights. These systems process data captured from traffic cameras or surveillance systems and apply deep learning models to detect and classify vehicles.

The input data for vehicle detection systems generally consists of video streams or image frames captured from traffic cameras, surveillance systems, or drone footage. These inputs are preprocessed and fed into deep learning models for object detection.

The output of the detection system includes bounding boxes around detected vehicles, class labels such as car, bus, truck, or two-wheeler, and confidence scores indicating detection accuracy. Additional modules such as vehicle tracking and counting algorithms are often integrated to monitor vehicle movement across frames.

The purpose of vehicle detection systems can vary depending on the application. Common applications include traffic flow analysis, congestion monitoring, accident detection, smart traffic signal control, and intelligent transportation systems.

Another important aspect is the deployment environment. Vehicle detection models may be implemented on centralized cloud servers, edge devices such as NVIDIA Jetson boards, or embedded traffic monitoring systems.

These components collectively form a comprehensive architecture for AI-based traffic monitoring solutions.

3.STUDY OF EXISTING SYSTEMS

In this section, we review 20 vehicle detection and traffic monitoring systems identified from the literature survey. These systems can be broadly classified into three categories: image-based vehicle detection systems, video-based vehicle detection and tracking systems, and AI-driven multimodal traffic monitoring systems.

Image-based systems primarily focus on detecting vehicles from static images using deep learning models such as Convolutional Neural Networks (CNNs) and object detection frameworks. Video-based systems extend this capability by processing continuous traffic footage to perform vehicle tracking and counting across frames. Multimodal systems combine visual data with additional information such as sensor data, radar signals, or satellite imagery to improve traffic monitoring accuracy.

This classification helps illustrate the evolution of research from traditional image-processing techniques to advanced deep learning-based vehicle detection frameworks capable of performing real-time traffic analysis.

3.1 Image-Based Vehicle Detection Systems

Image-based vehicle detection systems identify vehicles directly from images captured by traffic cameras or surveillance systems. These systems rely heavily on computer vision and deep learning algorithms to detect and classify vehicles such as cars, buses, trucks, and motorcycles.

Early vehicle detection approaches used traditional image processing methods such as edge detection, background subtraction, and feature extraction. However, these techniques were limited in accuracy and robustness under varying lighting conditions. With the advancement of deep learning, convolutional neural networks have significantly improved vehicle detection performance.

These systems generally operate by extracting visual features from images and applying trained models to detect objects using bounding boxes and classification labels.

3.1.1 Deep Learning–Based Vehicle Detection

Premaratne et al. (2023) [1] conducted a comprehensive study on vehicle detection and classification techniques used in highway traffic monitoring systems. Their work highlighted the transition from traditional computer vision approaches to deep learning-based object detection models such as YOLO. The study reported that YOLO-based systems achieve high detection accuracy and real-time performance. However, the authors also noted that detection performance decreases in low-light environments and during adverse weather conditions.

Alahdal et al. (2024) [2] proposed a real-time object detection framework designed for autonomous driving applications. The system evaluated several YOLO architectures including YOLOv5, YOLOv7, and YOLOv8. Experimental results showed that YOLOv5 achieved the highest detection accuracy with fast inference speed, making it suitable for real-time vehicle detection. Despite its advantages, the approach requires high computational resources and may struggle with complex environmental conditions.

Ji and Ma (2025) [3] introduced a hybrid vehicle detection framework that integrates YOLOv10 with feature pyramid networks and transformer-based detection techniques. Their model improves small-object detection and performs well in dense traffic environments. Experimental evaluations demonstrated improved mean average precision compared to conventional object detection models. However, the system introduces higher computational complexity due to the use of transformer architectures.

Xu et al. (2025) [5] proposed an enhanced YOLOv5-based vehicle detection system using multi-scale feature fusion techniques. The model incorporates additional feature extraction modules to improve detection of small and distant vehicles. Experimental results showed significant improvements in detection accuracy compared to the baseline YOLOv5 model. Nevertheless, the added modules increase the overall model complexity.

These studies demonstrate that deep learning models, particularly YOLO-based architectures, have become the dominant approach for vehicle detection tasks.

3.2 Video-Based Vehicle Detection and Tracking Systems

Video-based vehicle detection systems process continuous traffic video streams to detect, track, and count vehicles over time. Unlike image-based systems, these approaches analyze sequential frames and utilize tracking algorithms to maintain vehicle identity across frames.

Saputri et al. (2024) [4] developed a traffic monitoring system using the YOLOv7 object detection algorithm combined with the SORT tracking method. The system processes CCTV footage to detect vehicles and estimate traffic flow. Experimental results showed improved detection performance compared to pre-trained models. However, counting accuracy decreased when vehicles were partially occluded or when traffic density was high.

Nocua et al. (2025) [7] proposed an embedded traffic monitoring system deployed on an NVIDIA Jetson GPU platform. The system evaluated multiple deep learning models including MobileNet-SSD and YOLOv5. Their results indicated that YOLOv5 achieved strong detection performance while maintaining real-time processing capabilities. However, the system faced challenges when detecting vehicles in crowded urban environments.

Kudłacik et al. (2025) [12] evaluated the performance of YOLO models on high-resolution traffic video streams. Their study demonstrated that GPU-based implementations significantly improve detection speed and enable near real-time traffic monitoring. However, larger YOLO models require substantial computational resources.

These studies highlight the importance of combining object detection models with tracking algorithms to achieve accurate vehicle counting and traffic analysis.

3.3 Multimodal Traffic Monitoring Systems

Multimodal traffic monitoring systems combine information from multiple sources such as camera images, radar sensors, satellite imagery, and IoT devices to enhance vehicle detection and traffic analysis. These systems provide more comprehensive insights into traffic behavior by integrating diverse data sources.

Adamiak et al. (2025) [18] introduced a deep learning framework that detects vehicles using satellite imagery. Their approach utilizes keypoint detection techniques to track vehicle movements across satellite images and estimate vehicle speed. Although the system provides wide-area traffic monitoring capabilities, detection accuracy is limited by the resolution of satellite images.

Kheder and Mohammed (2024) [19] proposed an IoT-based traffic monitoring system that integrates robotics, sensors, and deep learning algorithms. The system uses camera feeds and IoT sensors to collect real-time traffic data and transmit it to cloud platforms for analysis. Experimental results showed high detection accuracy, but the system relies heavily on network connectivity and hardware infrastructure.

Li et al. (2025) [20] introduced a multi-scale vehicle detection network designed for autonomous driving scenarios. The proposed MSVDNet architecture improves vehicle detection across different scales by using advanced feature extraction modules. Although the model achieves strong performance, it faces challenges when applied to extremely dense traffic scenes.

Multimodal systems demonstrate significant potential for intelligent transportation systems by integrating diverse data sources to improve traffic monitoring accuracy.

4.METHODOLOGY AND PERFORMANCE ANALYSIS

This chapter presents a comparative analysis of vehicle detection approaches used in intelligent traffic monitoring systems. The methodologies are categorized based on the underlying detection strategy and system architecture. In general, vehicle detection research can be grouped into deep learning–based object detection models, video-based vehicle tracking systems, and hybrid AI-driven traffic monitoring frameworks.

The purpose of this chapter is to evaluate the performance of these approaches under different traffic conditions and datasets while identifying their advantages, limitations, and applicability in real-world intelligent transportation systems.

The analysis compares detection accuracy, computational efficiency, and real-time capability of the systems. Evaluation metrics such as Precision, Recall, F1-Score, and mean Average Precision (mAP) are commonly used to measure model performance.

Table I: Deep Learning-Based Vehicle Detection Approaches

Authors & Year	Model Architecture	Dataset Used	Performance	Limitations
Premaratne et al.(2023) [1]	YOLO-based detection framework	Highway traffic datasets	Accuracy >97% in daytime traffic	Performance decreases in low-light conditions
Alahdal et al. (2024) [2]	YOLOv5 / YOLOv7 / YOLOv8	VSim-AV simulation dataset	mAP ≈ 0.94, inference time ≈1.3 ms	High computational requirements.
Ji & Ma (2025) [3]	YOLOv10 + BiFPN + DETR	UADETRAC, COCO	mAP ≈ 75.8%	Increased architectural complexity
Xu et al. (2025) [5]	YOLOv5 with multi-scale feature fusion	Custom vehicle dataset	mAP ≈ 84.6%	Higher computational cost.
Wang et al. (2025) [9]	Hybrid-YOLO (CNN + Transformer)	KITTI, BDD100K	mAP ≈ 90.11%, 66 FPS	Moderate computational complexity

Deep learning-based vehicle detection models such as YOLO, Faster R-CNN, and SSD have significantly improved detection accuracy and real-time performance. Among these models, YOLOv5 has become widely adopted due to its balance between accuracy and inference speed.

Table II: Video-Based Vehicle Detection and Tracking Approaches

Authors & Year	**Model Architecture**	**Dataset Used**	**Performance**	**Limitations**
Saputri et al. (2024) [4]	YOLOv7 + SORT tracking	Custom CCTV dataset	AP ≈ 78.6%	Inconsistent counting in dense traffic
Nocua et al. (2025) [7]	MobileNet-SSD / YOLOv5	Urban traffic videos	Precision ≈ 90%, inference ≈10 ms	Reduced performance in crowded traffic
Kudłacik et al. (2025) [12]	YOLO models for high-resolution video	Traffic surveillance videos	Real-time performance on GPU	Requires high hardware resources
Arinaldi et al. (2018) [11]	Faster R-CNN	MIT Traffic dataset	Accuracy ≈67–69%	High computational cost

Video-based detection systems enable vehicle tracking and counting, which are essential for real-time traffic monitoring applications. However, these systems may suffer from occlusion problems, motion blur, and high computational demands.

Table III: Hybrid and Multimodal Traffic Monitoring Approaches

Authors & Year	**Model Architecture**	**Dataset Used**	**Performance**	**Limitations**
Ngeni et al. (2024) [15]	YOLOv5 + DeepSORT tracking	MOT17 dataset	+16% MOTA improvement	High computational complexity
Tan & Kieu (2023) [16]	YOLOv5 + DeepSORT (TRAMON)	Mixed traffic dataset	mAP ≈93%, counting accuracy ≈98–99%	Limited night-time evaluation
Adamiak et al. (2025) [18]	Keypoint R-CNN satellite detection	PlanetScope imagery	mAP ≈0.53	Low resolution reduces accuracy
Kheder & Mohammed (2024) [19]	IoT + Deep learning framework	Traffic sensor datasets	Accuracy ≈98%	Hardware and network dependency

Hybrid systems combine deep learning models with tracking algorithms and sensor data to improve traffic monitoring accuracy. These systems are particularly useful for smart city infrastructure and intelligent transportation systems.

4.2 Analysis of Datasets

Datasets play a critical role in training and evaluating vehicle detection models. These datasets typically contain annotated images or videos of traffic scenes with labeled vehicle categories.

Vehicle detection datasets can be classified into several categories based on their source and application.

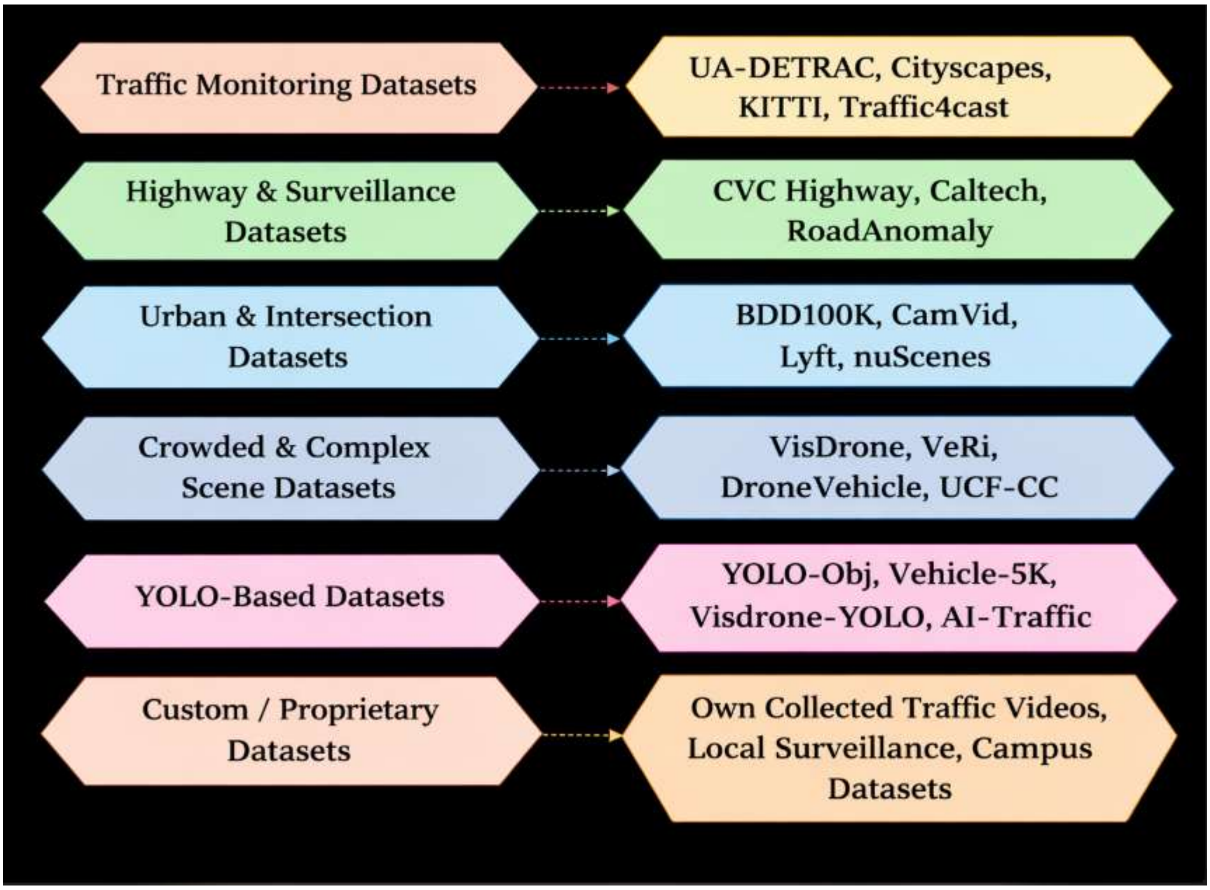


Figure 4.2.1: Classification of Datasets in Vehicle Detection

Major Dataset Categories

Traffic Monitoring Datasets

These datasets contain real-world traffic footage captured from surveillance cameras.

Examples:

- UA-DETRAC
- AI-City dataset
- MIT Traffic dataset

Autonomous Driving Datasets

These datasets are widely used for training deep learning models for autonomous vehicles.

Examples:

- KITTI dataset
- BDD100K dataset
- Waymo Open Dataset

Generic Object Detection Datasets

These datasets include multiple object categories including vehicles.

Examples:

- COCO dataset
- Pascal VOC dataset

Custom Traffic Datasets

Many studies also create **custom datasets** captured from city traffic cameras to evaluate vehicle detection systems under real-world conditions.

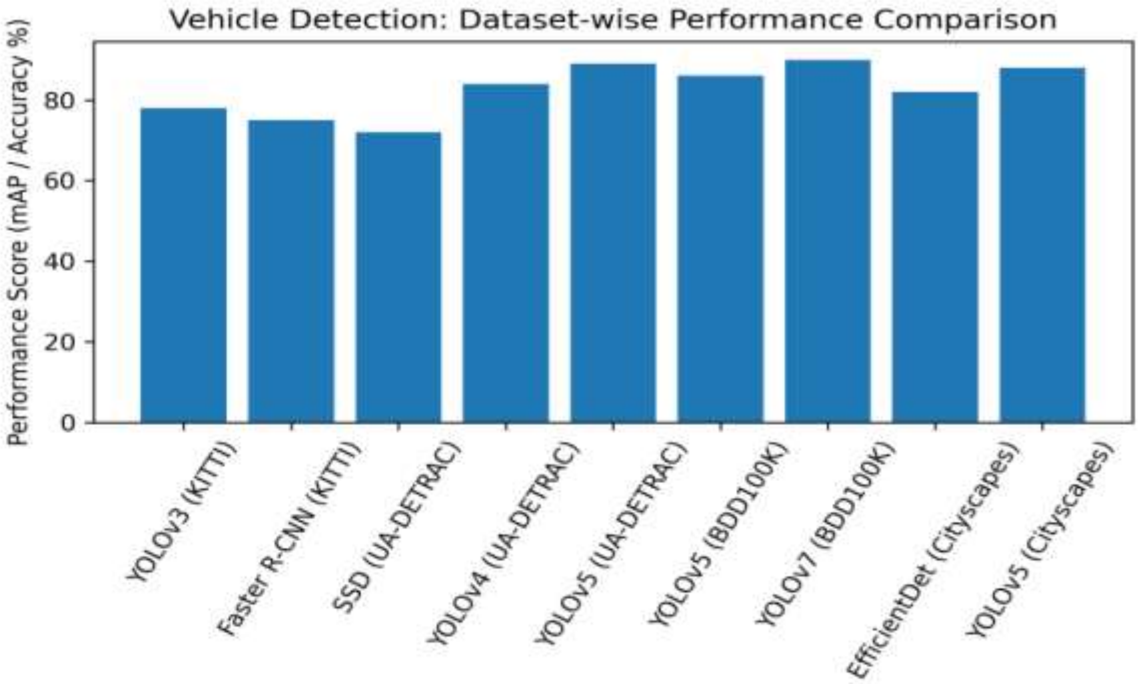


Figure 4.2.2: Vehicle Detection: Dataset-wise Performance Comparison

The dataset performance comparison indicates that YOLO-based models consistently achieve higher detection accuracy and faster inference speeds across multiple datasets.

Datasets such as COCO, KITTI, and BDD100K are widely used in vehicle detection research because they contain large annotated datasets suitable for training deep learning models.

In contrast, smaller datasets may lead to limited generalization ability, especially when applied to diverse traffic environments.

Table 3 – Comparative Deep Learning Detection models

Aspect	YOLO Models (YOLOv5, YOLOv7, YOLOv8)	Two-Stage Detectors (Faster R-CNN)
Detection Speed	Very fast (real-time)	Slower
Accuracy	High accuracy with optimized models	Very high accuracy
Computational Cost	Moderate	High
Real-Time Capability	Excellent	Limited
Best Use Cases	Real-time traffic monitoring	Offline analysis and high-precision tasks

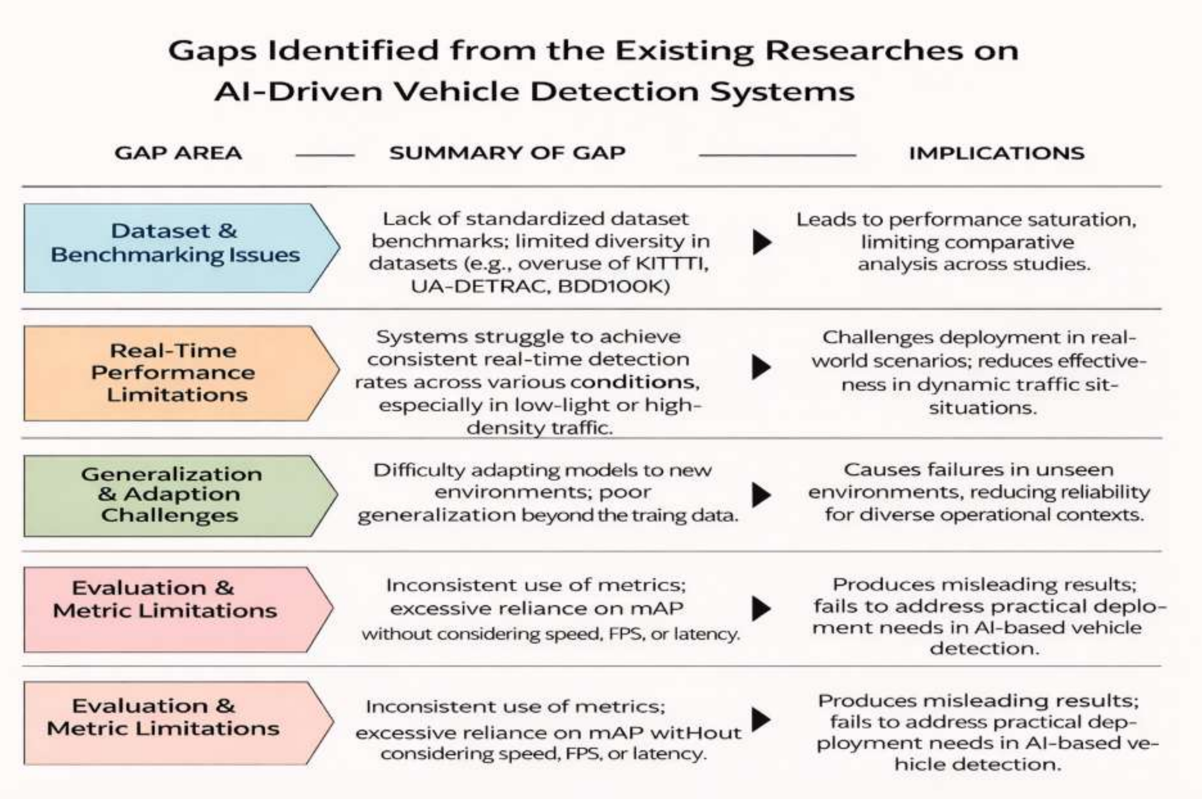
YOLO-based models are particularly suitable for real-time traffic monitoring systems due to their high inference speed and good detection accuracy.

5.GAPS IDENTIFIED IN EXISTING RESEARCH

This section discusses the major research gaps identified in current vehicle detection and traffic monitoring systems. Although deep learning models such as YOLO, Faster R-CNN, and SSD have significantly improved vehicle detection accuracy, several limitations still restrict their practical implementation in real-world traffic monitoring environments.

Many existing approaches focus primarily on improving detection accuracy but often overlook challenges related to real-time deployment, environmental robustness, dataset diversity, and scalability. These issues affect the reliability and generalization of AI-based vehicle detection systems across different traffic scenarios.

The key research gaps identified from the literature include limited robustness in complex traffic environments, dataset and evaluation limitations, lack of integrated vehicle counting frameworks, and generalization issues across different domains.



Gaps Identified from the Existing Researches on AI-Driven Vehicle Detection Systems

GAP AREA	SUMMARY OF GAP	IMPLICATIONS
Dataset & Benchmarking Issues	Lack of standardized dataset benchmarks; limited diversity in datasets (e.g., overuse of KITTTI, UA-DETRAC, BDD100K)	Leads to performance saturation, limiting comparative analysis across studies.
Real-Time Performance Limitations	Systems struggle to achieve consistent real-time detection rates across various conditions, especially in low-light or high-density traffic.	Challenges deployment in real-world scenarios; reduces effectiveness in dynamic traffic sitsituations.
Generalization & Adaption Challenges	Difficulty adapting models to new environments; poor generalization beyond the traing data.	Causes failures in unseen environments, reducing reliability for diverse operational contexts.
Evaluation & Metric Limitations	Inconsistent use of metrics; excessive reliance on mAP without considering speed, FPS, or latency.	Produces misleading results; fails to address practical deploment needs in AI-based vehicle detection.
Evaluation & Metric Limitations	Inconsistent use of metrics; excessive reliance on mAP witHout considering speed, FPS, or latency.	Produces misleading results; fails to address practical depployment needs in AI-based vehicle detection.

Figure 5.1: Gaps identified from the existing researches

The figure illustrates the major gaps identified in existing research on AI-driven vehicle detection and traffic monitoring systems.

First, dataset and benchmarking issues remain a significant challenge. Many vehicle detection models are trained and evaluated on a limited number of benchmark datasets such as KITTI, UA-DETRAC, and BDD100K. The over-reliance on these datasets results in limited dataset diversity and reduces the ability to perform reliable cross-study comparisons. As a result, the reported performance of many models may not accurately reflect their effectiveness in real-world traffic environments.

Second, real-time performance limitations are commonly observed in current vehicle detection systems. Although deep learning models such as YOLO and Faster R-CNN achieve high detection accuracy, maintaining consistent real-time detection speed under varying traffic conditions remains challenging. Factors such as low-light environments, high traffic density, and complex road scenarios often reduce detection speed and accuracy, making deployment in real-time intelligent transportation systems more difficult.

Third, generalization and adaptation challenges limit the practical usability of many vehicle detection models. Models trained on specific datasets often struggle to adapt to new environments with different camera angles, lighting conditions, vehicle types, and traffic patterns. This lack of generalization reduces the reliability of AI-based detection systems when deployed across different cities or transportation infrastructures.

Finally, evaluation and metric limitations exist in current research studies. Many studies primarily rely on evaluation metrics such as mean Average Precision (mAP) without considering other important factors such as inference speed, frames per second (FPS), latency, and system efficiency. This can lead to misleading performance comparisons and does not fully address the practical requirements for real-time AI-based vehicle detection systems.

Overall, these gaps highlight the need for future research to focus on improving dataset diversity, enhancing real-time detection performance, increasing model generalization capability, and adopting more comprehensive evaluation metrics for AI-driven vehicle detection and counting systems.

6. RESULTS AND DISCUSSION

This section presents the experimental results obtained from the proposed AI-Driven Vehicle Detection and Counting System using YOLOv5. The model was trained using a custom traffic dataset consisting of multiple vehicle categories, including car, number plate, blurred number plate, two-wheeler, auto, bus, and truck. The performance of the trained model was evaluated using standard object detection metrics such as Precision (P), Recall (R), and mean Average Precision (mAP50).

The YOLOv5 model architecture consists of 267 layers with approximately 46 million parameters, requiring **107.7** GFLOPs for computation. A total of 185 images containing 1980 annotated instances were used during training and evaluation.

Table 6.1 – Performance Evaluation of Vehicle Detection Model

Class	**Instances**	**Precision**	**Recall**	**mAP50**
Car	1061	0.906	0.922	0.959
Number Plate	174	0.766	0.822	0.855
Blurred Number Plate	161	0.752	0.453	0.690
Two-Wheeler	271	0.883	0.882	0.936
Auto	94	0.733	0.606	0.733
Bus	110	0.795	0.891	0.925
Truck	109	0.846	0.780	0.817
Overall	1980	**0.812**	**0.765**	**0.845**

Analysis of Detection Performance

The experimental results demonstrate that the proposed YOLOv5-based system achieves strong performance in detecting various vehicle types. The overall mean Average Precision (mAP50) achieved is 0.845, indicating reliable object detection capability across multiple vehicle categories.

Among all classes, car detection achieved the highest accuracy, with a precision of 0.906, recall of 0.922, and mAP50 of 0.959. Similarly, the detection of two-wheelers and buses also achieved high performance with mAP values exceeding 0.92, indicating the model's effectiveness in identifying commonly occurring vehicle types.

However, relatively lower performance was observed for blurred number plates and auto-rickshaws, which can be attributed to challenges such as motion blur, small object size, and limited training samples. These factors can reduce detection confidence and affect recall values.

Overall, the proposed system demonstrates the capability to accurately detect and classify multiple vehicle categories in traffic scenes, making it suitable for applications such as traffic monitoring, vehicle counting, and intelligent transportation systems.

Visualization of Detection Results

The trained YOLOv5 model successfully detects vehicles in traffic scenes and generates bounding boxes around detected objects. These detections are further processed to count vehicles passing through predefined regions of interest in the traffic video stream.

The detection results confirm that the proposed system can effectively support real-time traffic monitoring applications.

6.FUTURE ENHANCEMENTS SUGGESTED IN THE LITERATURE

Real-Time Traffic Monitoring Integration

Future vehicle detection systems should focus on seamless integration with real-time traffic monitoring infrastructures such as smart city surveillance networks, highway monitoring systems, and autonomous traffic control platforms. Current research prototypes often operate in controlled environments or offline analysis pipelines. Future systems should enable low-latency real-time detection and vehicle counting, ensuring continuous monitoring of traffic conditions through integration with CCTV cameras, IoT devices, and cloud-based traffic analytics platforms.

Improved Detection in Complex Traffic Environments

Vehicle detection models must be improved to perform reliably under challenging real-world conditions such as low-light environments, rain, fog, and high-density traffic scenarios. Future research should focus on developing robust deep learning architectures and enhanced feature extraction techniques capable of handling occlusion, motion blur, and adverse weather conditions. Incorporating image enhancement, noise reduction techniques, and adaptive learning algorithms can significantly improve detection reliability.

Multimodal Traffic Monitoring Systems

Future traffic monitoring systems should incorporate multimodal data sources to improve vehicle detection and traffic analysis accuracy. In addition to visual data from cameras, systems can integrate information from radar sensors, LiDAR, satellite imagery, and IoT-based traffic sensors. Combining these modalities can provide richer contextual information about traffic flow, road conditions, and vehicle behavior, leading to more accurate and intelligent transportation systems.

Intelligent Vehicle Tracking and Traffic Flow Analysis

Future research should extend vehicle detection systems to include advanced vehicle tracking and traffic flow prediction capabilities. Integrating object tracking algorithms such as DeepSORT, ByteTrack, or Kalman filtering methods can enable systems to monitor vehicle movement across frames and generate real-time traffic statistics. This would allow traffic authorities to analyze congestion patterns, detect traffic violations, and optimize traffic signal control systems.

Scalable and Lightweight Detection Models

Many deep learning models require significant computational resources, limiting their deployment in real-time systems. Future work should focus on developing lightweight and energy-efficient object detection models suitable for deployment on edge devices such as embedded GPUs, smart cameras, and mobile platforms. Optimizing model architectures using techniques such as model pruning, quantization, and knowledge distillation can improve scalability while maintaining detection accuracy.

CONCLUSION

This study reviewed existing research on AI-driven vehicle detection and counting systems, focusing on deep learning–based object detection techniques used in intelligent transportation systems. The review highlights the evolution of vehicle detection approaches from traditional computer vision methods to advanced deep learning architectures such as YOLO, Faster R-CNN, and transformer-based detection models.

The analysis of existing systems demonstrates that YOLO-based models have become the most widely adopted solution for real-time vehicle detection due to their balance between detection accuracy and computational efficiency. Among these models, YOLOv5 provides strong performance for detecting multiple vehicle categories in traffic scenes while maintaining real-time processing capability, making it suitable for applications such as traffic monitoring, congestion analysis, and smart city infrastructure.

The experimental evaluation of the proposed system demonstrates that the YOLOv5-based model can successfully detect various vehicle classes, including cars, buses, trucks, and two-wheelers, achieving high detection accuracy across the dataset. However, several challenges remain, including handling occlusions, improving performance in low-light conditions, and ensuring robust detection across diverse traffic environments.

Future research should focus on developing more robust, scalable, and intelligent vehicle detection systems that integrate multimodal data sources, advanced tracking algorithms, and real-time deployment capabilities. Such advancements will enable the development of smart traffic management systems capable of improving road safety, reducing congestion, and supporting next-generation intelligent transportation infrastructures.

References

[1] P. Premaratne, I. J. Kadhim, R. Blacklidge, and M. Lee, “Comprehensive review on vehicle detection, classification and counting on highways,” *Expert Systems with Applications*, 2023.https://doi.org/10.1016/j.neucom.2023.126627

[2] N. M. Alahdal, F. Abukhodaira, L. Haj Meftah, and A. Cherifa, “Real-time object detection in autonomous vehicles with YOLO,” *Procedia Computer Science*, 2024. https://doi.org/10.1016/j.procs.2024.09.392

[3] A. Ji and X. Ma, “Vehicle detection and classification for traffic management and autonomous systems using YOLOv10,” *Future Generation Computer Systems*, 2025. https://doi.org/10.1016/j.aej.2025.06.049

[4] H. A. Saputri, M. Avrilliio, L. Christofer, V. Simanjaya, and I. N. Alam, “Implementation of YOLOv7 in estimating traffic flow in Malang,” *Procedia Computer Science*, 2024. https://doi.org/10.1016/j.procs.2024.10.235

[5] L. Xu, P. Li, Q. Peng, Y. Zhao, L. Zhu, and S. Yuan, “Vehicle detection on roads based on YOLOv5 with multi-scale feature fusion,” *Pattern Recognition Letters*, 2025. https://doi.org/10.1016/j.array.2025.100522

[6] S. Xu, M. Zhang, J. Chen, and Y. Zhang, “YOLO-HyperVision: A vision transformer backbone-based enhancement of YOLOv5 for detection of dynamic traffic information,” *Information Fusion*, 2024. https://doi.org/10.1016/j.eij.2024.100523

[7] F. Nocua M., W.-J. Pérez-Holguín, and C. Pardo-Beainy, “Urban traffic monitoring based on deep learning on an embedded GPU,” *Engineering Applications of Artificial Intelligence*, 2025. https://doi.org/10.1016/j.eswa.2025.126847

[8] H. Zhang, K.-F. Yang, Y.-J. Li, and L. L.-H. Chan, “Self-supervised network for low-light traffic image enhancement based on deep noise and artifacts removal,” *Signal Processing*, 2024. https://doi.org/10.1016/j.cviu.2024.104063

[9] H. Wang *et al.*, “Hybrid-YOLO: Lightweight Mamba-Transformer hybrid with multi-scale fusion for real-world traffic detection,” *Neurocomputing*, 2025. https://doi.org/10.1016/j.icte.2025.09.002

[10] N. Kavitha and D. N. Chandrappa, “Optimized YOLOv2 based vehicle classification and tracking for intelligent transportation system,” *Procedia Computer Science*, 2021. https://doi.org/10.1016/j.rico.2021.100008

[11] A. Arinaldi, J. A. Pradana, and A. A. Gurusinga, “Detection and classification of vehicles for traffic video analytics,” *Procedia Computer Science*, 2018. https://doi.org/10.1016/j.procs.2018.10.527

[12] P. Kudłacik, K. Kaszuba, and P. Lubowiecki, “Performance of YOLO based traffic analysis in high resolution video stream,” *Journal of Real-Time Image Processing*, 2025. https://doi.org/10.1016/j.procs.2025.09.565

[13] S. Feng, X. Cai, L. Li, W. Wang, and S. Ying, “A review of research on vehicle detection in adverse weather environments,” *Transportation Research Part C*, 2025. https://doi.org/10.1016/j.jtte.2025.06.001

[14] S. Ness, “Vehicle detection and recognition approach in smart surveillance system: A comparative analysis,” *Expert Systems with Applications*, 2025. https://doi.org/10.1016/j.array.2025.100473

[15] F. Ngeni, J. Mwakalonge, and S. Siuhi, "Solving traffic data occlusion problems in computer vision algorithms using DeepSORT and quantum computing," *IEEE Access*, 2024. https://doi.org/10.1016/j.jtte.2023.05.006

[16] D. M. Tan and L.-M. Kieu, "TRAMON: An automated traffic monitoring system for high-density, mixed and lane-free traffic," *Transportation Research Part C*, 2023. https://doi.org/10.1016/j.iatssr.2023.10.001

[17] A. Grents, V. Varkentin, and N. Goryaev, "Determining vehicle speed based on video using convolutional neural network," *Procedia Computer Science*, 2020. https://doi.org/10.1016/j.trpro.2020.10.024

[18] M. Adamiak *et al.*, "Deep learning enhanced road traffic analysis: Scalable vehicle detection and velocity estimation using PlanetScope imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025. https://doi.org/10.1016/j.jag.2025.104707

[19] M. Q. Kheder and A. A. Mohammed, "Real-time traffic monitoring system using IoT-aided robotics and deep learning techniques," *Journal of Intelligent Transportation Systems*, 2024. https://doi.org/10.1016/j.kjs.2023.10.017

[20] B. Li, X. Hu, L. Zhang, Q. Li, and J. Hu, "MSVDNet: A multi-scale vehicle detection network for target detection in autonomous driving," *Knowledge-Based Systems*, 2025. https://doi.org/10.1016/j.aej.2025.08.025

All abbreviations and short forms

Datasets

- COCO → Common Objects in Context Dataset
- KITTI → Karlsruhe Institute of Technology and Toyota Technological Institute Dataset
- BDD100K → Berkeley DeepDrive 100K Dataset
- UA-DETRAC → University at Albany Detection and Tracking Dataset
- AI-City → AI City Challenge Traffic Surveillance Dataset
- MOT → Multiple Object Tracking Dataset
- Pascal VOC → Visual Object Classes Dataset
- ImageNet → Large-scale Image Dataset for Visual Recognition

Models & Architectures

- YOLO → You Only Look Once (Real-time Object Detection Algorithm)
- YOLOv5 → Fifth version of YOLO object detection model developed by Ultralytics
- YOLOv7 → Advanced YOLO architecture optimized for speed and accuracy
- YOLOv8 → Latest YOLO model with improved detection capabilities
- YOLOv10 → Optimized YOLO architecture for real-time detection tasks
- CNN → Convolutional Neural Network
- Faster R-CNN → Faster Region-based Convolutional Neural Network
- SSD → Single Shot MultiBox Detector
- DeepSORT → Deep Simple Online and Realtime Tracking Algorithm
- ResNet → Residual Neural Network
- FPN → Feature Pyramid Network
- BiFPN→ Bidirectional Feature Pyramid Network

Evaluation Metrics

- mAP → mean Average Precision
- mAP50→ mean Average Precision at Intersection over Union threshold of 0.5
- IoU → Intersection over Union (overlap between predicted and ground truth bounding boxes)
- Precision (P) → Ratio of correctly predicted positive detections to total predicted positives
- Recall (R) → Ratio of correctly predicted positive detections to total actual positives
- F1 Score → Harmonic mean of precision and recall
- FPS → Frames Per Second (model inference speed)
- GFLOPs → Giga Floating Point Operations (computational complexity of a model)

Techniques and Concepts

- Object Detection → Computer vision task of identifying and locating objects within an image or video
- Bounding Box → Rectangular box used to localize detected objects
- Non-Maximum Suppression(NMS) → Technique used to remove duplicate object detections
- Anchor Boxes → Predefined bounding boxes used for object detection training
- Feature Extraction → Process of extracting meaningful features from images
- Transfer Learning → Using a pre-trained model and adapting it to a new dataset
- Edge Computing→ Processing data near the source device rather than in the cloud
- Intelligent Transportation System(ITS) → Smart infrastructure for monitoring and managing traffic using technology