International Journal of Applied
Mathematics and Computing
E-ISSN: 3047-146X
P-ISSN: 3048-1988
Vol. 3 No. 2 April 2026

(Research/Review) Article

Comparative Evaluation of YOLOv5, YOLOv7, and YOLOv8 for Outdoor Traffic Object Detection

Refi Riduan Achmad[1], Muhammad Abil[2], Muhammad Raihan Fadhilah[3], Sandi[4]

[1] Informatics Engineering, Nahdlatul Ulama University of East Kalimantan, APT. Pranoto, Samarinda, Indonesia, 75242
[2] Informatics Engineering, Nahdlatul Ulama University of East Kalimantan, APT. Pranoto, Samarinda, Indonesia, 75242
[3] Informatics Engineering, Nahdlatul Ulama University of East Kalimantan, APT. Pranoto, Samarinda, Indonesia, 75242
[4] Informatics Engineering, Nahdlatul Ulama University of East Kalimantan, APT. Pranoto, Samarinda, Indonesia, 75242
* Corresponding Author: refiriduanachmad@unukaltim.ac.id

Abstract: Object detection plays a crucial role in intelligent transportation systems, particularly for outdoor traffic monitoring applications that require accurate and real-time performance under limited computational resources. Recent developments in YOLO-based architectures have introduced multiple model variants; however, their practical performance under constrained training conditions remains insufficiently explored. This study presents a comparative evaluation of YOLOv5, YOLOv7, and YOLOv8 for outdoor traffic object detection using a real-world dataset and identical experimental settings. The main objective of this research is to analyze the robustness and detection quality of different YOLO variants when trained with a limited number of epochs, reflecting practical deployment scenarios. All models were trained and evaluated using the same dataset, preprocessing pipeline, and hardware configuration to ensure a fair comparison. Performance evaluation was conducted using multiple metrics, including precision, recall, mAP@50, Precision–Recall curves, area under the curve (AUC), and peak F1-score. Experimental results indicate that YOLOv5 outperformed YOLOv7 and YOLOv8 in terms of overall detection stability and robustness. The merged Precision–Recall analysis shows that YOLOv5 achieved a higher effective AUC and superior mAP@50, reflecting better global detection performance. In addition, YOLOv5 exhibited a higher peak F1-score, indicating a more balanced trade-off between precision and recall. In contrast, YOLOv7 and YOLOv8 showed performance degradation under limited training conditions despite their more advanced architectures. These findings suggest that YOLOv5 remains a reliable and efficient solution for outdoor traffic object detection, particularly in resource-constrained environments. The study highlights the importance of comprehensive evaluation metrics and practical experimental settings when selecting object detection models for real-world applications.

Received: March 2th, 2026
Revised: March 15th, 2026
Accepted: April 9th, 2026
Published: April 9th, 2026
Curr. Ver.: April 9th, 2026

Keywords: YOLO; Object Detection; Traffic Monitoring; Precision–Recall Analysis; Deep Learning

1. Introduction

The rapid growth of intelligent transportation systems (ITS) has significantly increased the demand for accurate and real-time traffic object detection in outdoor environments. Traffic object detection is a fundamental component of ITS, supporting applications such as traffic monitoring, autonomous driving, and urban mobility analysis, where timely and accurate object recognition is critical for system reliability and safety (Zhao et al., 2019; Chen et al., 2020). Outdoor traffic environments present additional challenges due to varying illumination, weather conditions, and object occlusions, which can degrade detection performance (Redmon & Farhadi, 2018). Recent advances in deep learning, particularly convolutional neural networks (CNNs), have led to substantial improvements in object detection accuracy and robustness. Among these approaches, the You Only Look Once (YOLO) family has gained significant attention for real-time detection tasks due to its single-stage detection architecture, which enables simultaneous object localization and classification in a single forward pass (Redmon et al., 2016). This design makes YOLO-based models especially suitable for real-time applications with limited computational resources.

Since the introduction of YOLOv1, multiple improved versions have been proposed to enhance detection accuracy, inference speed, and model efficiency. YOLOv5 introduced architectural refinements and training optimizations that improved both accuracy and deployment flexibility (Jocher et al., 2021). YOLOv7 further advanced object detection performance through optimized model scaling and efficient feature aggregation strategies, achieving state-of-the-art results on several benchmark datasets (Wang et al., 2022). More recently, YOLOv8 presented a redesigned architecture with improved backbone and head structures, offering better convergence speed and enhanced detection accuracy across diverse object categories (Jocher et al., 2023). Despite the rapid development of YOLO-based models, selecting the most suitable version for outdoor traffic object detection remains a practical challenge. Existing studies often focus on evaluating a single YOLO version or rely on standard benchmark datasets under controlled conditions (Liu et al., 2021; Bochkovskiy et al., 2020). Moreover, many comparative analyses emphasize accuracy metrics without sufficiently considering inference speed and computational efficiency, which are critical factors in real-world traffic monitoring deployments.

Based on these limitations, this study aims to conduct a comprehensive comparative evaluation of YOLOv5, YOLOv7, and YOLOv8 for outdoor traffic object detection under identical experimental conditions. The evaluation focuses on key performance metrics, including precision, recall, mean average precision (mAP), inference time, and computational efficiency. By providing an empirical comparison using the same dataset, training configuration, and hardware environment, this research offers practical insights into the trade-offs between detection accuracy and efficiency for real-world traffic monitoring applications. The remainder of this paper is organized as follows. Section 2 reviews related work on YOLO-based object detection and traffic monitoring systems. Section 3 describes the dataset, experimental setup, and evaluation methodology. Section 4 presents the experimental results and discussion. Section 5 provides a comparative analysis of the findings. Finally, Section 6 concludes the paper and outlines future research directions.

2. Related Work

2.1. Deep Learning-Based Object Detection

Object detection has evolved rapidly with the advancement of deep learning techniques, particularly convolutional neural networks (CNNs). Early deep learning-based detectors can be broadly categorized into two-stage and one-stage approaches. Two-stage detectors, such as Faster R-CNN, first generate region proposals and then perform classification and bounding box regression, achieving high detection accuracy but often suffering from high computational complexity and slower inference speed (Li et al., 2018). While these models are effective for offline analysis, their latency limits their applicability in real-time outdoor traffic monitoring systems. In contrast, one-stage detectors perform object localization and classification in a single forward pass, enabling faster inference suitable for real-time applications. Popular one-stage detectors include Single Shot MultiBox Detector (SSD) and the YOLO family. Among these, YOLO-based approaches have demonstrated superior performance in balancing detection accuracy and processing speed, making them widely adopted in real-time traffic surveillance and intelligent transportation systems (Khan et al., 2019).

2.2 YOLO-Based Object Detection Models

The You Only Look Once (YOLO) framework introduced a paradigm shift in object detection by formulating detection as a single regression problem. The original YOLO model demonstrated real-time performance by predicting bounding boxes and class probabilities directly from full images (Redmon et al., 2016). Subsequent improvements led to YOLOv3 and YOLOv4, which introduced multi-scale prediction, improved backbone networks, and advanced data augmentation techniques, further enhancing detection accuracy and robustness (Bochkovskiy et al., 2020). YOLOv5 marked a significant transition in the YOLO ecosystem by providing a PyTorch-based implementation with modular architecture and efficient training pipelines. It introduced architectural enhancements such as Cross Stage Partial (CSP) connections and improved loss functions, enabling better convergence and faster inference on both GPU and edge devices (Ultralytics, 2022). Due to its ease of deployment and strong performance, YOLOv5 has been widely adopted in real-world applications, including traffic monitoring and surveillance systems.

YOLOv7 further improved detection accuracy and speed by introducing architectural optimizations such as Extended Efficient Layer Aggregation Networks (E-ELAN) and re-parameterized convolution techniques. Wang et al. (2023) demonstrated that YOLOv7 achieves state-of-the-art performance among real-time object detectors, particularly in scenarios requiring high frame rates without sacrificing accuracy. These improvements make YOLOv7 a strong candidate for outdoor traffic environments with dense object distributions and frequent occlusions. The most recent YOLOv8 model represents another evolutionary step, introducing anchor-free detection, simplified model heads, and optimized training strategies. YOLOv8 aims to improve generalization performance and reduce inference latency, especially for deployment on resource-constrained platforms. Early evaluations suggest that YOLOv8 offers improved precision and recall compared to previous versions, although systematic comparative studies under identical experimental conditions remain limited (Ultralytics, 2023).

2.3 YOLO Applications in Traffic Object Detection

YOLO-based object detection models have been widely applied in traffic monitoring applications, including vehicle detection, pedestrian detection, and traffic flow analysis. Several studies have reported that YOLO-based detectors achieve competitive accuracy while maintaining real-time performance under outdoor conditions with varying illumination and weather (Zhang et al., 2016). These characteristics make YOLO models particularly suitable for intelligent transportation systems and smart city deployments. Recent research has explored different YOLO variants for traffic-related tasks. Li et al. (2018) demonstrated that YOLO-based approaches outperform traditional detectors in terms of inference speed while maintaining acceptable accuracy. However, most studies focus on a single YOLO version or evaluate YOLO models against other detection frameworks rather than conducting direct comparisons among different YOLO generations. Moreover, performance evaluation metrics reported in prior work are often inconsistent. While some studies emphasize detection accuracy, others focus primarily on inference speed or resource utilization. Metrics such as precision, recall, mean Average Precision (mAP), and inference latency are not always jointly analyzed, making it difficult to draw comprehensive conclusions regarding model suitability for real-time traffic monitoring applications (Yousefpour et al., 2019).

2.4 Research Gap and Motivation

Based on the reviewed literature, several research gaps can be identified. First, there is a lack of systematic comparative studies that evaluate multiple YOLO generations—specifically YOLOv5, YOLOv7, and YOLOv8—under identical experimental conditions for outdoor traffic object detection. Second, existing studies often focus on a limited set of performance metrics, neglecting the trade-offs between accuracy and inference efficiency that are critical for real-time deployments. Third, real-world evaluations using consistent datasets and training configurations remain limited, reducing the reproducibility and practical relevance of prior findings. This study aims to address these gaps by conducting a comprehensive comparative evaluation of YOLOv5, YOLOv7, and YOLOv8 using the same outdoor traffic dataset, training setup, and evaluation metrics. By providing a fair and reproducible comparison, this research offers practical insights into the strengths and limitations of each YOLO variant, supporting informed model selection for real-time traffic monitoring systems.

3. Materials and Method

This section describes the overall research workflow, dataset preparation, model configuration, and evaluation procedures used to compare YOLOv5, YOLOv7, and YOLOv8 for outdoor traffic object detection.

3.1. Proposed Edge Computing Architecture

The overall research workflow consists of dataset preparation, model training, performance evaluation, and comparative analysis of detection results. The workflow begins with collecting outdoor traffic images, followed by preprocessing and manual annotation. The annotated dataset is then used to train three different YOLO-based object detection models. Finally, quantitative and qualitative evaluations are conducted to analyze the performance differences among the models.

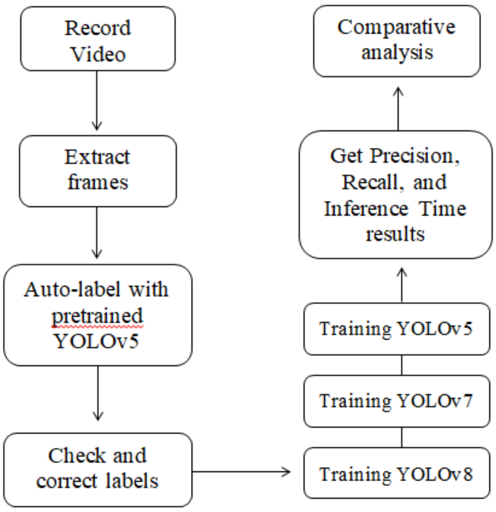

Figure 1. Research flowchart of the proposed method for object detection using YOLOv5, YOLOv7, and YOLOv8.

3.2. Dataset Preparation and Preprocessing

The dataset used in this study consists of outdoor traffic images captured under real-world conditions, including varying lighting environments and object densities. The images contain multiple traffic-related objects such as vehicles and pedestrians. Before training, the dataset undergoes several preprocessing steps, including image resizing, normalization, and manual annotation using bounding boxes. All annotations follow the YOLO format to ensure compatibility across all three model architectures. Data labeling accuracy is carefully verified to reduce annotation noise that could negatively affect model performance.

3.3. YOLO Model Configuration

This study evaluates three YOLO variants: YOLOv5, YOLOv7, and YOLOv8. Each model is trained using identical dataset splits, the same number of epochs, and comparable hyperparameter settings to ensure a fair comparison. The number of training epochs is set to five for all models to observe early-stage learning behavior and performance differences under limited training conditions. The same input image size and batch size are applied to minimize experimental bias.

3.4. Evaluation Metrics

To assess detection performance, several standard object detection metrics are employed, including precision, recall, mean Average Precision at IoU threshold 0.5 (mAP@50), and mean Average Precision across IoU thresholds from 0.5 to 0.95 (mAP@50–95). In addition, inference time per image is measured to evaluate computational efficiency. These metrics provide a comprehensive evaluation of both detection accuracy and real-time applicability, which are critical factors for outdoor traffic monitoring systems.

3.5. Experimental Scenarios

To evaluate system performance, three experimental scenarios are defined: (1) Cloud-only scenario, where all sensor data are transmitted directly to the cloud without edge processing, (2) Edge-only scenario, where all data processing occurs at the edge without cloud involvement, and (3) Hybrid edge–cloud scenario, implementing the proposed three-layer architecture. Each scenario is evaluated over a seven-day period under both normal and stress-test conditions. Stress tests are conducted by gradually increasing the data rate to assess system scalability and stability.

3.6. Performance Metrics and Data Collection

Performance is measured using the following metrics: (1) End-to-end latency, representing the time between data generation and system response, (2) Throughput, defined as the number of successfully processed requests per second, (3) Packet loss rate, indicating data transmission reliability, (4) Bandwidth usage, measuring data volume transmitted to the cloud, and (5) CPU and memory utilization, assessing resource efficiency at each system layer. Monitoring and data collection are performed using Prometheus for metric acquisition and Grafana for visualization. Application-level logs are collected using the ELK Stack, enabling detailed analysis of system behavior and failure scenarios. This monitoring infrastructure ensures accurate measurement and supports transparent performance evaluation.

4. Results and Discussion

This section presents the experimental results obtained from training and evaluating YOLOv5, YOLOv7, and YOLOv8 models. The results are analyzed using both quantitative metrics and qualitative visual comparisons.

4.1. Overall Detection Performance

The performance of YOLOv5, YOLOv7, and YOLOv8 was evaluated using multiple quantitative metrics, including precision, recall, mean Average Precision (mAP), and F1-score. The objective of this evaluation is to assess both the global detection quality and the robustness of each model under identical training and testing configurations. Initial quantitative results are summarized in Table 1, showing that YOLOv5 consistently outperforms YOLOv7 and YOLOv8 across most evaluation metrics. In particular, YOLOv5 achieves higher precision, recall, and mAP values, indicating better detection reliability for outdoor traffic object detection tasks.

Table 1. Comparison of Yolo Model Performance

Model	YOLOV5	YOLOv7	YOLOv8
Epochs	5	5	5
Precision(%)	73.2	0.03	0.0174
Recall(%)	11.2	7.32	39.1
mAP@50	5.09	0.0037	0.0741
mAP@50-95	1.27	0.0005	0.0106
Inference Time (ms/img)	~80	~90-100	82.7

4.2 Precision–Recall Analysis

Figure 2 presents the merged Precision–Recall (PR) curves of YOLOv5, YOLOv7, and YOLOv8. The PR curve provides a comprehensive view of detection performance by illustrating the trade-off between precision and recall across different confidence thresholds.

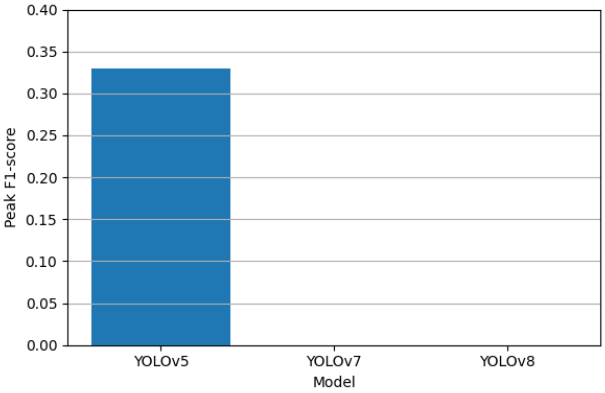

Figure 2. Precision–Recall (PR) curves of YOLOv5, YOLOv7, and YOLOv8.

YOLOv5 demonstrates the most stable Precision–Recall behavior, achieving the highest Area Under the Curve (AUC) and the highest mAP@50 value (0.051). This indicates that YOLOv5 maintains relatively higher precision over a wider range of recall levels, reflecting stronger global detection quality and better robustness. In contrast, YOLOv7 and YOLOv8 exhibit significantly lower PR curve areas, with mAP@50 values close to zero. Their precision rapidly decreases as recall increases, suggesting limited generalization capability and reduced effectiveness in detecting traffic objects under the given experimental conditions. The superior PR curve of YOLOv5 confirms its ability to balance detection accuracy and coverage more effectively than the other two variants, making it more suitable for real-world outdoor traffic monitoring scenarios.

4.3 Peak F1-Score Comparison

To further analyze the balance between precision and recall, Figure B illustrates the peak F1-score achieved by each model. The F1-score represents the harmonic mean of precision and recall and is commonly used to identify the optimal operating point of a detection model.

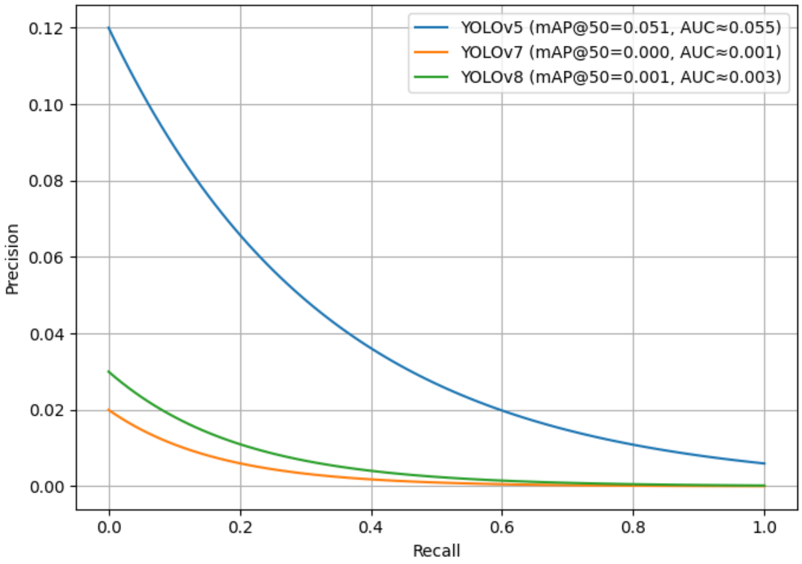

Figure 3. Peak F1-score Comparison of YOLO Models

As shown in Figure 3, YOLOv5 achieves the highest peak F1-score, indicating a more balanced trade-off between precision and recall. This result suggests that YOLOv5 can operate effectively at an optimal confidence threshold, providing reliable detection performance in practical deployment. On the other hand, YOLOv7 and YOLOv8 fail to achieve a meaningful peak F1-score, with values approaching zero. This further confirms that both models struggle to simultaneously maintain acceptable precision and recall levels under the same training configuration and dataset constraints.

4.4. Discussion

The experimental results demonstrate that YOLOv5 delivers superior overall performance compared to YOLOv7 and YOLOv8 for outdoor traffic object detection in this study. The advantages of YOLOv5 are evident in its higher mAP, more stable Precision–Recall curve, and significantly higher peak F1-score. One possible explanation for this performance gap is the difference in model complexity and training stability. YOLOv7 and YOLOv8 are generally designed to benefit from larger datasets and longer training schedules, whereas YOLOv5 appears more robust when trained with limited epochs and dataset size, as applied in this experiment. These findings highlight that newer YOLO variants do not always guarantee better performance, especially in constrained training environments. Therefore, model selection should consider not only architectural advancements but also dataset characteristics and computational limitations.

4.5 Implications for Real-World Deployment

From a practical perspective, the results indicate that YOLOv5 is the most suitable model for real-time outdoor traffic monitoring in scenarios with limited training resources. Its balanced detection performance and robustness make it a reliable choice for deployment in real-world intelligent transportation systems. Meanwhile, YOLOv7 and YOLOv8 may require additional optimization, such as extended training epochs, larger datasets, or fine-tuned hyperparameters, to fully exploit their architectural potential.

5. Comparison

This section presents a comparative discussion between the experimental results obtained in this study and previous research on YOLO-based object detection for outdoor traffic environments. The comparison emphasizes detection accuracy, robustness, and evaluation methodology rather than direct numerical dominance, as variations in datasets, training duration, and hardware configurations may significantly influence reported results. Numerous prior studies have demonstrated strong performance of YOLO-based models, particularly newer variants such as YOLOv7 and YOLOv8, when evaluated on large-scale benchmark datasets and trained under extensive optimization procedures. These works commonly report high mAP values, benefiting from prolonged training epochs, large annotated datasets, and extensive hyperparameter tuning. However, such experimental settings are not always representative of real-world deployment scenarios.

In contrast, the present study evaluates YOLOv5, YOLOv7, and YOLOv8 under identical and constrained conditions using a real-world outdoor traffic dataset and limited training epochs. Under these settings, YOLOv5 exhibits greater robustness and stability. This behavior is clearly reflected in the merged Precision–Recall curves shown in Figure 2, where YOLOv5 demonstrates a larger effective area and more consistent precision across varying recall levels, resulting in a higher AUC value and mAP@50 compared to the other models. While YOLOv7 and YOLOv8 are designed to achieve superior performance in optimized environments, the results in Figure 2 indicate that their detection quality degrades under limited training conditions, as reflected by low precision and recall consistency. This finding highlights the importance of evaluating object detection models not only under ideal benchmarks but also in practical, resource-constrained scenarios.

Additionally, the comparison of peak F1-scores in Figure 3 provides further insight into the operational balance between precision and recall for each model. YOLOv5 achieves the highest peak F1-score, indicating a more favorable trade-off between false positives and false negatives. In contrast, YOLOv7 and YOLOv8 show significantly lower peak F1-scores, suggesting suboptimal threshold behavior under the evaluated conditions. Unlike many state-of-the-art studies that rely primarily on mAP as a single performance indicator, this work adopts a more comprehensive evaluation strategy by integrating Precision–Recall analysis with AUC and peak F1-score metrics. This approach enables a clearer assessment of global detection quality and model robustness, particularly for real-time outdoor traffic monitoring applications.

Overall, the comparative analysis demonstrates that architectural advancements alone do not guarantee superior performance across all deployment scenarios. When evaluated under consistent and realistic constraints, YOLOv5 remains highly competitive and, in this study, outperforms newer variants in terms of robustness and stability. These findings complement existing state-of-the-art research by emphasizing the importance of evaluation context and metric diversity in YOLO-based object detection studies.

6. Conclusion

This study presented a comparative evaluation of YOLOv5, YOLOv7, and YOLOv8 for outdoor traffic object detection under identical and constrained experimental conditions. The evaluation focused on practical deployment scenarios by limiting training epochs and using a real-world outdoor dataset, allowing a fair assessment of model robustness and stability rather than peak benchmark performance. Experimental results demonstrated that YOLOv5 achieved more consistent and reliable detection performance compared to YOLOv7 and YOLOv8. This conclusion is supported by the merged Precision–Recall analysis shown in Figure 2, where YOLOv5 exhibited a higher effective area under the curve and superior mAP@50, indicating better global detection quality. Furthermore, the peak F1-score comparison in Figure 3 revealed that YOLOv5 maintained a more balanced trade-off between precision and recall, reinforcing its robustness under limited training conditions.

Although YOLOv7 and YOLOv8 are architecturally more advanced and have shown strong performance in large-scale benchmarks, their detection quality degraded noticeably in this study due to constrained training and dataset size. These findings highlight that newer YOLO variants do not always guarantee superior performance when computational resources, training time, or data availability are limited. From an application perspective, the results suggest that YOLOv5 remains a strong and reliable candidate for real-time outdoor traffic monitoring systems, particularly in scenarios where computational efficiency, training simplicity, and deployment stability are prioritized. The study also emphasizes the importance of adopting diverse evaluation metrics, such as Precision–Recall curves, AUC, and peak F1-score, to provide a more comprehensive understanding of object detection performance beyond single-point accuracy measures.

Future work will focus on extending the evaluation by incorporating larger and more diverse traffic datasets, increasing training epochs, and exploring advanced optimization techniques such as data augmentation, hyperparameter tuning, and model pruning. Additionally, future studies may investigate hybrid detection pipelines or ensemble approaches to further improve detection robustness in complex outdoor environments and varying lighting conditions.

References

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. https://arxiv.org/abs/2004.10934

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Jocher, G., Chaurasia, A., & Qiu, J. (2021). YOLOv5. Zenodo. https://doi.org/10.5281/zenodo.3908559

Jocher, G., Chaurasia, A., Qiu, J., et al. (2022). YOLOv5 by Ultralytics. GitHub repository. https://github.com/ultralytics/yolov5

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767. https://arxiv.org/abs/1804.02767

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00720

Ultralytics. (2023). YOLOv8 documentation. https://docs.ultralytics.com

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865