



Comparative Analysis of Deep Neural Networks YOLOv11 and YOLOv12 for Real-Time Vehicle Detection in Autonomous Vehicles

Mohammed Chaman^{*}, Anas El Maliki, Hamza El Yanboiy, Hamad Dahou, Hlou Laâmari,
Abdelkader Hadjoudja

Laboratory of Electronic Systems, Information Processing, Mechanics and Energetics. Faculty of Sciences, Ibn Tofail University, Kenitra 14000, Morocco

Corresponding Author Email: mohammed.chaman@uit.ac.ma

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijtdi.090104>

ABSTRACT

Received: 7 February 2025

Revised: 14 March 2025

Accepted: 25 March 2025

Available online: 31 March 2025

Keywords:

real-time object detection, YOLOv11, YOLOv12, autonomous vehicles, vehicle detection, deep learning, ADAS

Accurate, real-time vehicle detection is crucial for autonomous vehicles navigating dynamic traffic environments. This study compares YOLOv11 and the newly released YOLOv12, two state-of-the-art deep learning models for object detection, to assess enhancements in speed, accuracy, and robustness. YOLOv12 has improved upon YOLOv11's architecture with an attention mechanism and Residual Efficient Layer Aggregation Networks (R-ELAN). The improvements for YOLOv12 are designed to obtain better accuracy and improved computational performance as compared to YOLOv11. YOLOv11 and YOLOv12 were trained and tested on a newly developed dataset with 38,500 fully annotated images of seven classes of vehicles taken in different environmental conditions. Results show YOLOv12 achieves higher recall (95.0%), F1-score (96.03%), and mAP@50–95 (88.6%), while both maintain real-time inference speeds. YOLOv12 also demonstrated an improved capacity to detect small or partially occluded objects in challenging scenes. Overall, these findings establish YOLOv12 as a better solution for perceiving real-time data while autonomous driving, with a real prospect for implementation in intelligent transportation systems and edge-computing.

1. INTRODUCTION

The rapid advancement of autonomous vehicles relies heavily on real-time object detection to identify surrounding vehicles, pedestrians, and road signs under various and complex traffic environments [1]. Luckily, this list has been compiled to present some of the most exciting developments that have made the most out of these new-found algorithms and methodologies, specifically the YOLO (You Only Look Once) family that has generated for itself a sunshed amount of popularity surely due to its relatively high-speed and accuracy when it comes to solving standards object detection problems [2, 3].

The evolution of the YOLO object detectors from YOLOv1 (2016) to YOLOv12 (2025) depicts continuous enhancement in speed, performance, and computational efficiency, as shown in Figure 1. Over time, significant architectural modifications such as fusion schemes in features, attention units, and optimizations in the detection head have been introduced to improve object detection in autonomous vehicles, surveillance systems, and various real-world deployments [4]. This evolution outlines an unmistakable trajectory toward better generalizability, reduced latency, and improved detection performance, positioning YOLOv12 as the most advanced iteration to date [5, 6].

More recently, YOLOv11 and even more recently,

YOLOv12 in early 2025, which implemented attention mechanisms, R-ELAN modules and improved detection heads. It significantly increases accuracy while also boosting inference speed, further solidifying its position as a valuable option to be incorporated into Advanced Driver Assistance Systems (ADAS) and intelligent transportation systems (ITS) [5].

However, despite the architectural innovations introduced in YOLOv12, a comprehensive comparative evaluation with its immediate predecessor, YOLOv11, remains lacking, particularly in the context of real-time vehicle detection for autonomous driving.

The YOLO object detection framework has undergone continuous evolution since its initial introduction. Over the years, multiple versions have been developed, each bringing improvements in speed, accuracy, and computational efficiency. Early versions such as YOLOv3, YOLOv4, and YOLOv5 introduced enhanced feature extraction and detection head refinements, significantly improving real-time object detection performance [7].

This study addresses this gap by evaluating and comparing YOLOv12 with YOLOv11 regarding detection accuracy, inference speed, and computational efficiency in a driving environment. The outcomes will inform us on the best fit for deployment in ADAS in real-time.

Subsequent models, including YOLOv6 and YOLOv7

added more advanced mechanisms, including Swin Transformers [8, 9] and Multi-Stage Feature Fusion (MSFF) modules [10], further improving detection performance, especially in the case of occlusion and challenging

environments. Recent generations, such as YOLOv8 and YOLOv9, improved detection accuracy and efficiency, enabling them to be used in intelligent transportation systems and traffic monitoring applications [11-13].

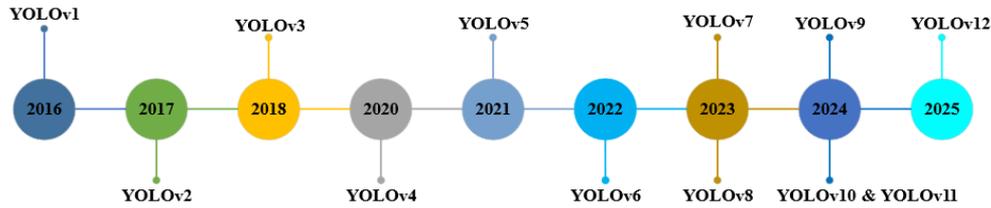


Figure 1. Evolution of the YOLO architecture

As discussed by Sundaresan Geetha et al. [14], YOLOv10 and YOLOv11 continued this trend, improving the detection of small objects and crowded traffic scenarios. Sharma et al. [15] found that YOLOv11 notably outperformed all other detectors for occluded vehicle detection, even in challenging environments, showing that YOLOv11 is a great candidate for real-time traffic surveillance [16, 17].

The most recent development in this lineage YOLOv12 shows lackluster progress, moving towards an attention-centric feature world, aiming etc., In terms of accuracy and real-time inference speeds. Studies by Tian et al. [18], YOLOv12 outperforms its predecessors, as well as the recent real-time Detection Transformer (RT-DETR) in terms of mean Average Precision (mAP) while consuming lower computational resources. Despite these advancements, a direct comparison between YOLOv11 and YOLOv12 specifically for autonomous vehicle perception remains an open research question, which this study aims to address.

To address this research gap, this study proposes a detailed comparative analysis of YOLOv11 and YOLOv12, focusing on three key performance metrics: detection accuracy, inference speed, and computational efficiency in real-world autonomous driving conditions. In contrast to all previous studies, which have considered YOLO models on separate benchmarks, this work is intended to evaluate their performance when it comes to their direct use in real-world environments, where real-time is paramount for ADAS. Through a performance analysis over a diverse set of monitored traffic scenarios, this work aims to evaluate if YOLOv12's architectural benefits outweigh its predecessor YOLOv11, thus making a case towards its adoption as a coverage tool for safety-critical environments.

The rest of the paper is organized as follows: Section 2 describes the YOLOv12 architecture, dataset preparation, and evaluation metrics. In section 3, we present the experimental results and comparisons of YOLOv12 against YOLOv11, along with some of the key observations from these experiments with an emphasis on the trade-offs between accuracy, speed, and computational efficiency. Finally, Section 4 concludes the paper with a summary of important contributions and directions for further research in online object detection models for autonomous driving.

2. MATERIAL AND METHOD

This study employs a systematic approach to evaluate and compare YOLOv11 and YOLOv12 in real-time vehicle detection for autonomous driving. YOLOv12 represents the latest milestone in real-time object detection, achieving state-

of-the-art performance through advanced attention mechanisms and Residual Efficient Layer Aggregation Networks (R-ELAN). By focusing on detection accuracy, inference speed, and computational overhead, this work highlights YOLOv12's architectural innovations and demonstrates how it surpasses its predecessor, YOLOv11, without sacrificing efficiency. The findings offer valuable insights into the trade-offs between precision and speed, serving as a crucial benchmark for researchers and practitioners shaping the future of autonomous vehicle applications.

2.1 YOLO neural network models

The YOLO (You Only Look Once) neural network family, developed by Ultralytics, has established a strong reputation for real-time object detection by balancing high accuracy with computational efficiency. In this study, we examine two of its recent iterations YOLOv11 and YOLOv12, which have gained traction in autonomous vehicle applications, where rapid and precise detection of objects is paramount for operational safety. Both models stem from a lineage of continual refinements, targeting challenges such as detecting smaller objects, optimizing attention mechanisms, and reducing computational overhead [15, 17, 19].

YOLOv11 introduced several notable architectural updates compared to its predecessors. As shown in Figure 2(a), a key change involved replacing the C3k2 module with the more flexible C2f module, enabling the network to adapt more effectively to diverse detection scenarios. Additionally, YOLOv11 incorporated a C2PSA block that improved the attention mechanism, enhancing the extraction of contextual features in complex environments. Another significant improvement was the use of depthwise separable convolutions in the detection head, which streamlined computations while minimizing any impact on overall accuracy. Despite these advances, YOLOv11 exhibited limitations in detecting smaller objects, spurring further research into refined feature pyramids or enhanced upsampling strategies [20-22].

Building on YOLOv11's framework, YOLOv12 broadened performance improvements through several innovations. As illustrated in Figure 2(b), YOLOv12 integrates A2 (area-attention) modules for more fine-grained focus on salient regions, especially in cluttered scenes or varied lighting conditions. The network also leverages R-ELAN to optimize gradient flow and multi-scale feature aggregation, and continues using depthwise separable convolutions refined for even lower computational overhead. In addition, Figure 2(c) compares the attention modules C3K2 and the novel R-ELAN introduced in YOLOv12, highlighting how R-ELAN

significantly improves residual connections and enhances feature aggregation. These advancements contribute to YOLOv12's improved ability to detect objects of varying sizes, even on resource-constrained platforms such as embedded systems and edge devices [6, 18].

Several new features (key improvements) further distinguish YOLOv12 from YOLOv11. Zone-based attention mechanisms efficiently handle large receptive fields while maintaining a balanced computational load across upstream layers. The enhanced feature aggregation in R-ELAN

produces more robust multi-scale representations, and the integration of residual connections with scaling improves training stability, especially in larger models. Flash Attention minimizes memory overhead by optimizing access patterns, and the simplified attention implementation eliminates the need for positional encoding, reducing model complexity. Furthermore, the optimization of MLP ratios ensures more efficient allocation of computing resources, allowing YOLOv12 to achieve high accuracy with fewer parameters. [23].

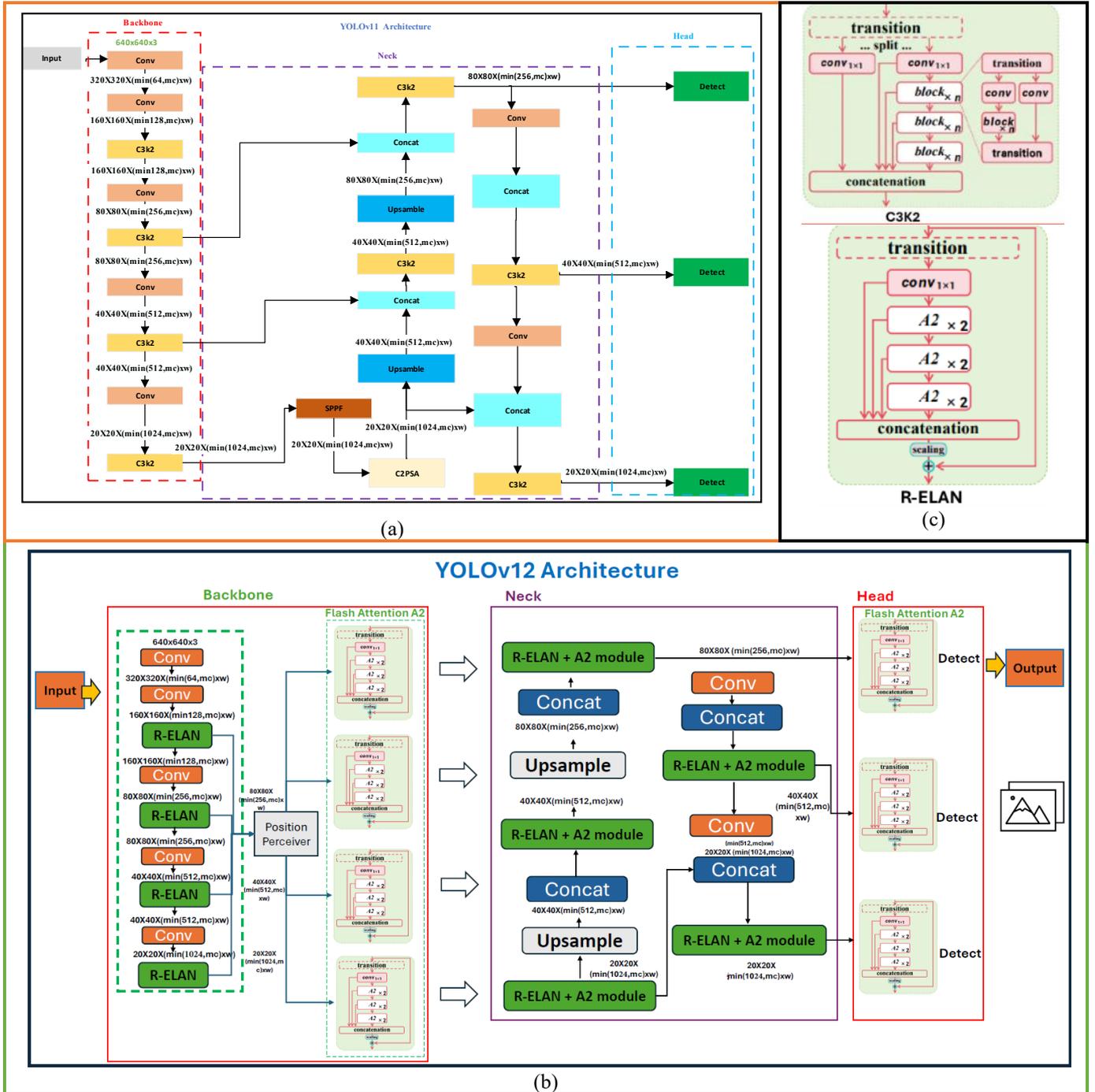


Figure 2. YOLO neural network architecture and module comparison: (a) YOLOv11 architecture; (b) YOLOv12 architecture, highlighting the integration of A2 modules; (c) Comparison of attention modules: C3K2 and the novel R-ELAN introduced in YOLOv12

A comparative evaluation of YOLOv11, YOLOv12, and other object detection frameworks is presented in Figure 3, underscoring YOLOv12's advantages in speed, accuracy, and

overall computational load. Small-object detection, traditionally a challenge for real-time systems, benefits from refined feature scaling and hardware acceleration options,

enabling YOLOv12 to handle intricate scenarios more effectively. Researchers can consult supplementary materials and referenced literature for deeper insight into these architectures, including specific implementation strategies,

hyperparameters, and ablation studies. By uniting advanced attention modules, efficient convolutional operations, and carefully tuned resource allocations, YOLOv12 offers a robust, forward-looking solution for modern object detection needs.

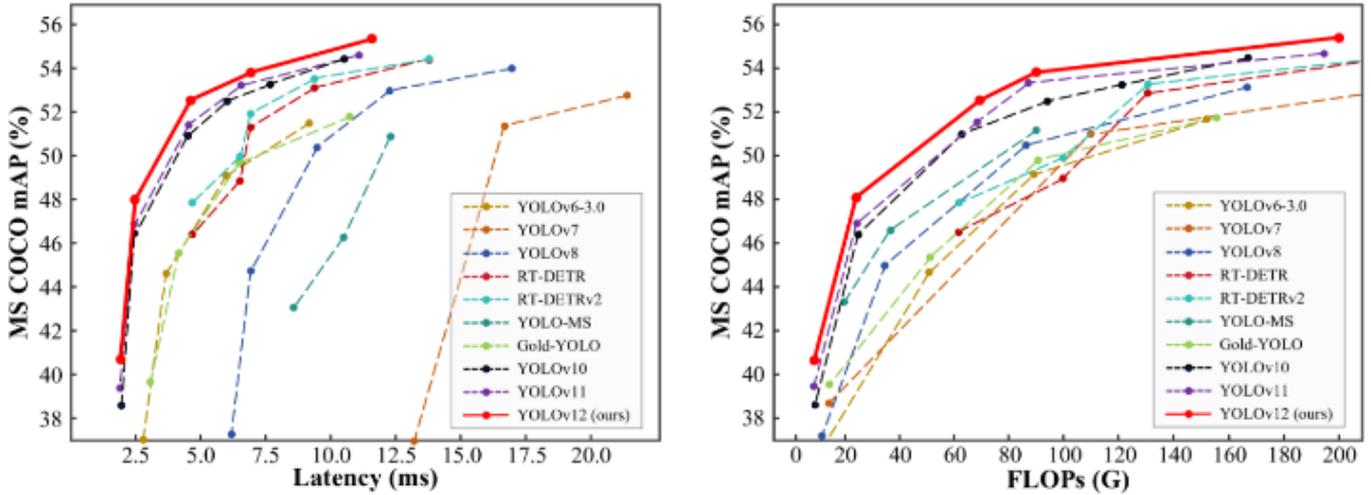


Figure 3. Performance comparison of YOLOv12, YOLOv11, and other object detection models

2.2 Dataset and resources for training and deployment

In this study, we curated a diverse and meticulously annotated dataset of 38,500 images to train and evaluate both YOLOv11 and YOLOv12 for vehicle detection. The images were derived from video frames recorded in various traffic environments urban streets, highways, intersections, and parking lots to ensure broad coverage of real-world scenarios. To further enhance representativeness, we included images captured under different weather conditions (daylight, nighttime, fog, rain) and in the presence of occlusions. The dataset encompasses seven categories of vehicles E-Scooter, Bicycle, Bus, Car, Motorcycle, Truck, and Emergency Vehicle reflecting the spectrum of traffic objects typically encountered by autonomous driving systems, as shown in Table 1. High-resolution images were collected from both in-vehicle cameras and roadside surveillance systems, supplemented by samples from open-source repositories.

Each image underwent manual annotation using Roboflow, with bounding boxes drawn around every target object. Annotations were then exported in YOLO format for streamlined integration with the training pipeline. Subsequently, images were resized to 640×640 pixels to strike a balance between computational efficiency and preservation of visual detail [24].

To enhance the dataset’s robustness, we applied data augmentation techniques such as flipping, rotation, noise injection, and exposure adjustments. These transformations effectively increased the variety of training samples, mitigating overfitting and improving generalization. In addition, saturation adjustment was introduced to account for varying lighting and color conditions. Specifically, saturation was modified within a range of -25% to $+25\%$, simulating both muted and vibrant scenarios. The transformation can be described as:

$$I' = \text{adjust_saturation}(I, \alpha), \alpha \in [-0.25, 0.25] \quad (1)$$

where, I is the original image and I' is the saturation-adjusted image.

Noise augmentation was implemented by adding small random pixel perturbations to simulate real-world image imperfections such as sensor noise or compression artifacts. This process is mathematically expressed as:

$$I' = I + N \quad (2)$$

where, N represents the noise applied to the image, affecting up to 0.1% of the pixels.

For exposure adjustment, brightness was modified uniformly by approximately 10% across the image to simulate varying lighting conditions. The formula used is:

$$I' = \text{clip}(I + \beta \times I, 0, 255), \beta \approx 0.1 \quad (3)$$

This adjustment enhances the model’s resilience to extreme lighting environments, ensuring better generalization across day, night, or glare-affected scenes.

Table 1. Dataset image categories

Id	Class	Image
0	E-Scooter	
1	Emergency vehicle	
2	Bicycle	
3	Bus	
4	Car	
5	Motorcycle	
6	Truck	

Random rotations were also introduced, wherein images were rotated between -15° and $+15^\circ$, simulating camera tilt and slight perspective changes. This transformation can be written as

$$I' = \text{rotate}(I, \theta), \theta \in [-15^\circ, 15^\circ] \quad (4)$$

and the bounding boxes were dynamically adjusted to ensure vehicles remained properly enclosed after rotation. Geometrically, each pixel's coordinates underwent a 2D rotation matrix transformation,

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (5)$$

where, $R(\theta)$ denotes the rotation matrix, which is especially relevant for rotating or mobile cameras.

In addition, horizontal flipping was applied with a 50% probability to simulate mirrored perspectives and enhance spatial diversity in vehicle orientation.

These combined augmentation methods not only broaden the range of environmental variations but also help address class imbalance by expanding underrepresented categories.

Finally, the dataset was divided into training (70%), validation (20%), and test (10%) sets to enable systematic hyperparameter tuning and unbiased performance evaluation. Figure 4 illustrates the workflow from image collection and annotation to structured dataset partitioning. By combining a thoroughly annotated dataset with systematic preprocessing and augmentation, this approach provides a robust foundation for assessing the accuracy, robustness, and real-time performance of YOLOv11 and YOLOv12 under challenging traffic conditions.

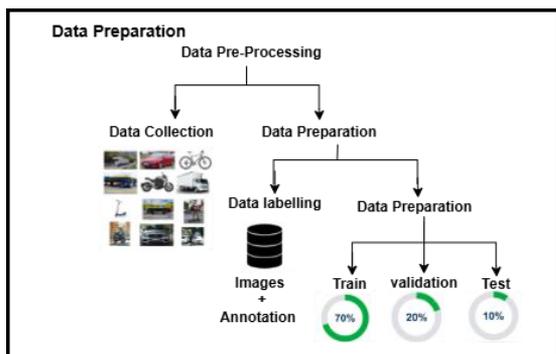


Figure 4. Dataset preparation workflow

2.3 Experimental environment and parameter settings

To ensure a fair and efficient evaluation of YOLOv11 and YOLOv12, all experiments were conducted in a high-performance computing environment, optimized for deep learning workloads. The training setup was carefully designed to provide consistent computational power and efficient GPU acceleration, ensuring stable model convergence and reliable performance assessment. The details of the hardware and software configurations, along with the hyperparameter settings, are summarized in Table 2.

The system was equipped with an AMD Ryzen 9 7940HX CPU, an NVIDIA GeForce RTX 4070 GPU with 8 GB of VRAM, and 32 GB of DDR5 memory, running on Windows 11. The software environment was built on Python 3.12.4, using PyTorch 2.5.1 with CUDA 11.8, ensuring compatibility

with GPU acceleration for optimized deep learning computations.

To achieve efficient convergence and optimal detection accuracy, key hyperparameters were configured as follows: 100 epochs, a batch size of 16, an image resolution of 640×640 pixels, and the Stochastic Gradient Descent (SGD) optimizer. Additional settings included a momentum of 0.937, a weight decay of 0.0005, an initial learning rate of 0.01, and a final learning rate of 0.01. These configurations, as detailed in Table 2, were selected to balance computational efficiency and model accuracy, ensuring robust real-time vehicle detection capabilities.

Table 2. Hardware and software configurations with hyperparameter settings

Hardware and Software Environment		Hyperparameters	
Name	Version	Parameters	Details
CPU	AMD Ryzen 9 7940HX	Epochs	100
GPU	NVIDIA GeForce RTX4070	Batch size	16
VRAM	8 GB	Image size (Pixels)	640×640
Memory	32 GB DDR5	Optimizer algorithm	SGD
Operating System	Windows 11	Momentum	0.937
Python Version	3.12.4	Weight Decay	0.0005
PyTorch Version	2.5.1	Initial Learning Rate	0.01
CUDA Version	11.8	Final Learning Rate	0.01

2.4 Models evaluation metrics

Precision, Recall, mAP, and F1-score are employed as the primary evaluation metrics for the thorough evaluation of YOLOv11 and YOLOv12 performance in vehicle detection in this study. These metrics measure the models' accuracy, stability, and generalization capacity for different categories of vehicles.

Precision, Eq. (6), determines the proportion of true positive detections correctly identified, defining the ability of the model to suppress false positives. Recall, Eq. (7), determines the proportion of true objects detected correctly, representing the model's sensitivity to detect all relevant instances. Average Precision (AP), Eq. (8), taken across all categories, gives the mean Average Precision (mAP), Eq. (9), as the global performance indicator combining precision-recall values for all detected classes. The F1-score, Eq. (10), balances precision and recall and gives the global view of the model detection reliability in realistic situations [25, 26].

With these criteria, this study ensures that there is a standardized and rigorous measurement of the detection performance, which simplifies comparison of YOLOv11 and YOLOv12 for real-time autonomous vehicle detection. The criteria enable a balance between accuracy, computational overhead, and the capability to manage variability so that the models are viable for use in dynamic traffic conditions.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

$$\text{AP} = \int_0^1 P(R) dR \quad (8)$$

$$mAP = \frac{\sum_{j=1}^c (AP)_j}{c} \quad (9)$$

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

3. RESULTS AND DISCUSSION

The experimental results from training and validation of the

YOLOv11 and YOLOv12 models demonstrate clear distinctions in their detection performance, which is critical for real-time autonomous driving applications. Both models were trained using NVIDIA GPUs in a high-performance computing environment, utilizing the PyTorch framework. The training setup incorporated an SGD optimizer with a learning rate of 0.01, momentum of 0.9, and a decay strategy. A batch size of 16 was used, and early stopping was activated after 20 epochs without validation improvement to mitigate overfitting.

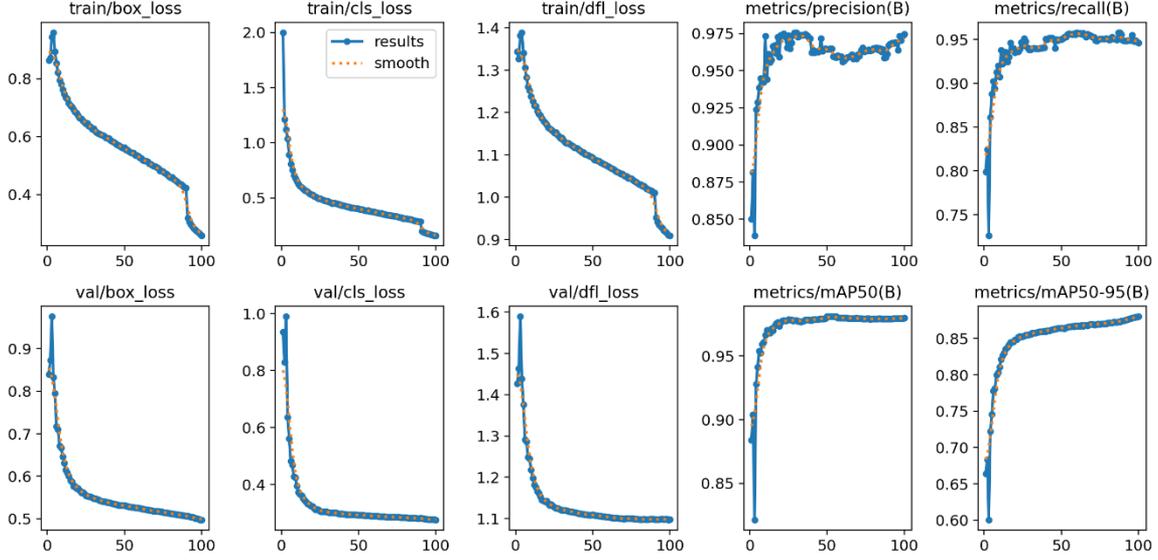


Figure 5. Training results of YOLOv11

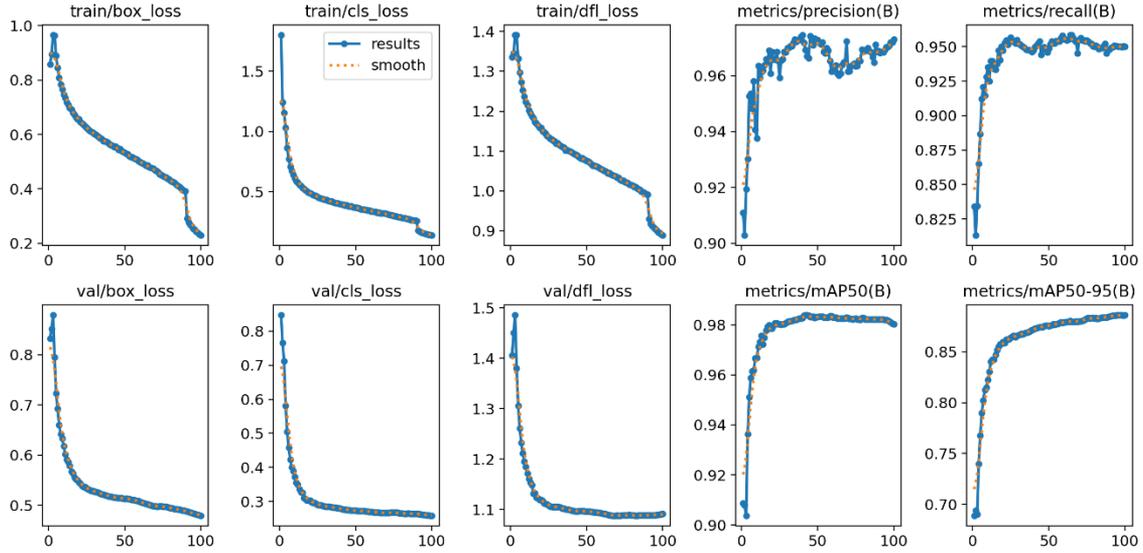


Figure 6. Training results of YOLOv12

Table 3. Metrics of the proposed models

Model	Metrics of the Models				
	Precision	Recall	F1-score	mAP@50	mAP@50-95
YOLOv11	97.7%	94.3%	95.96%	98%	88.1%
YOLOv12	97.1%	95%	96.03%	98.2%	88.6%

As shown in Figures 5 and 6, both models exhibited steady convergence, with training and validation losses decreasing

smoothly over epochs. The quantitative performance comparison (Table 3) shows that while YOLOv11 achieved a slightly higher precision (97.7%), YOLOv12 demonstrated superior recall (95.0%), F1-score (96.03%), and mAP@50-95 (88.6%). These improvements can be attributed to YOLOv12's refined architecture, which incorporates advanced attention mechanisms and R-ELAN. These enhancements provide more effective feature extraction and

object localization, particularly in cluttered or occluded scenes.

Nevertheless, YOLOv12 is not without limitations. Class-specific analysis reveals that it still struggles with detecting certain categories such as bicycles and buses. These challenges are likely due to several factors: the relatively small size of these objects in the input images, frequent occlusions in urban environments, and high intra-class variability in shape, orientation, and color. Such characteristics reduce the effectiveness of feature extraction layers, especially when spatial resolution is limited. This indicates that while YOLOv12 improves overall performance, further refinement is needed to ensure reliable detection of all vehicle types. Future enhancements could include integrating multi-scale feature fusion to better preserve small object features and applying sensor fusion strategies such as incorporating LiDAR or radar data to complement visual inputs.

Figures 7 and 8 illustrate the Precision-Recall (PR) confidence curves for both models. YOLOv12 consistently outperforms YOLOv11 in recall, demonstrating fewer missed detections, while maintaining high precision. Its mAP@50 increased slightly from 0.980 (YOLOv11) to 0.982 (YOLOv12), reinforcing its enhanced classification accuracy. Vehicle classes such as E-Scooters, motorcycles, and trucks show high and stable performance in both models. However, bicycles and buses display noticeable variability in recall, emphasizing the need for targeted improvements in these categories.

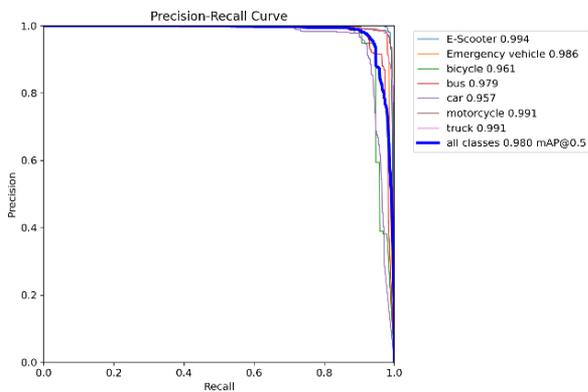


Figure 7. Precision-Recall confidence curve of YOLOv11

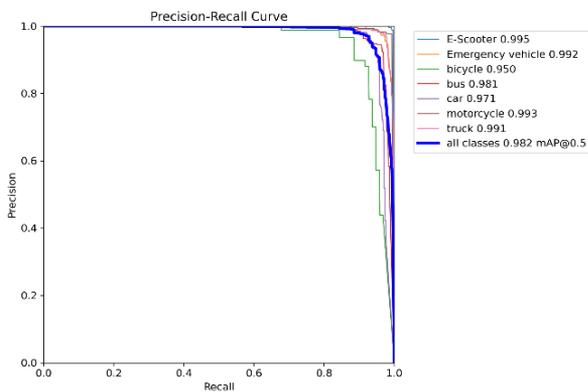


Figure 8. Precision-Recall confidence curves of YOLOv12

The F1-confidence curves for YOLOv11 and YOLOv12, provide a detailed evaluation of their precision-recall trade-offs at varying confidence thresholds. The F1 score, which

balances precision and recall, is a key metric for determining the overall detection effectiveness of the models.

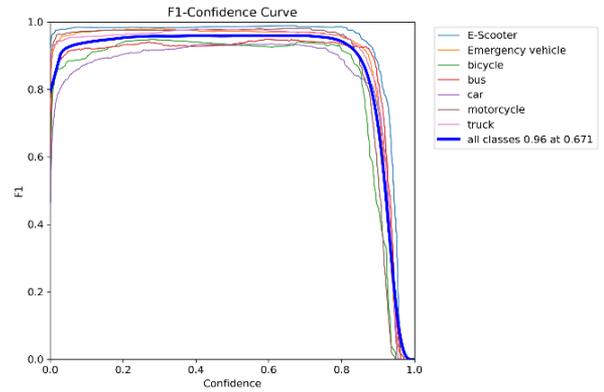


Figure 9. F1 and confidence curves of YOLOv11

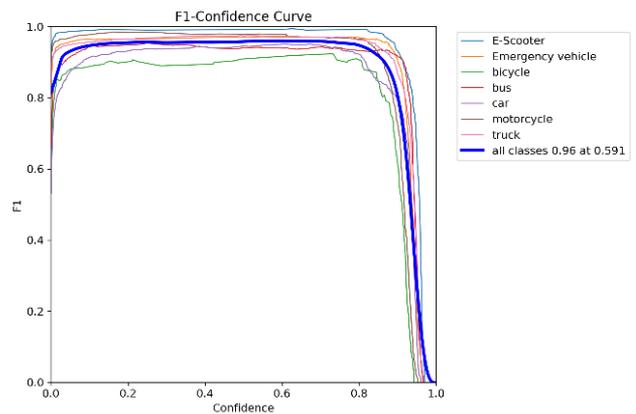


Figure 10. F1 and confidence curves of YOLOv12

As depicted in Figures 9 and 10, both models achieve consistently high F1 scores, indicating strong detection performance across multiple vehicle categories. However, YOLOv12 exhibits a superior F1 score at lower confidence thresholds (0.591) compared to YOLOv11 (0.671), confirming its enhanced detection capability and reduced false negatives. This improvement reflects YOLOv12's refined feature extraction, optimized detection head, and better object classification under varying conditions.

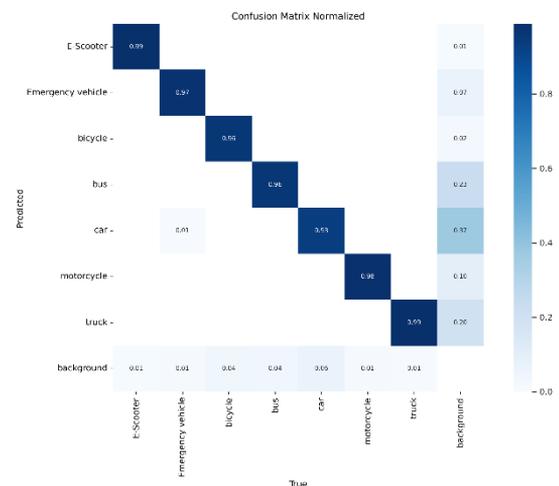


Figure 11. Normalized confusion matrix for YOLOv11

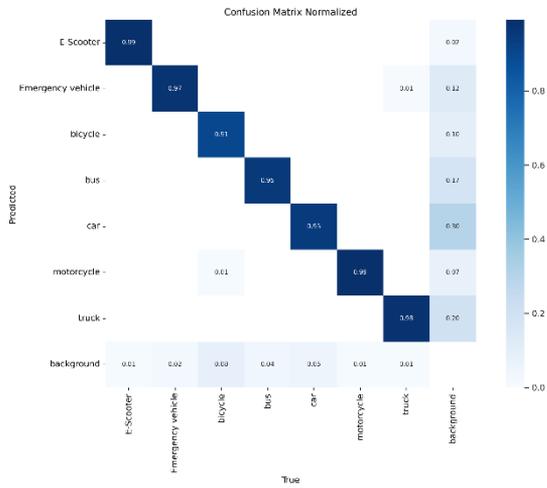


Figure 12. Normalized confusion matrix for YOLOv12

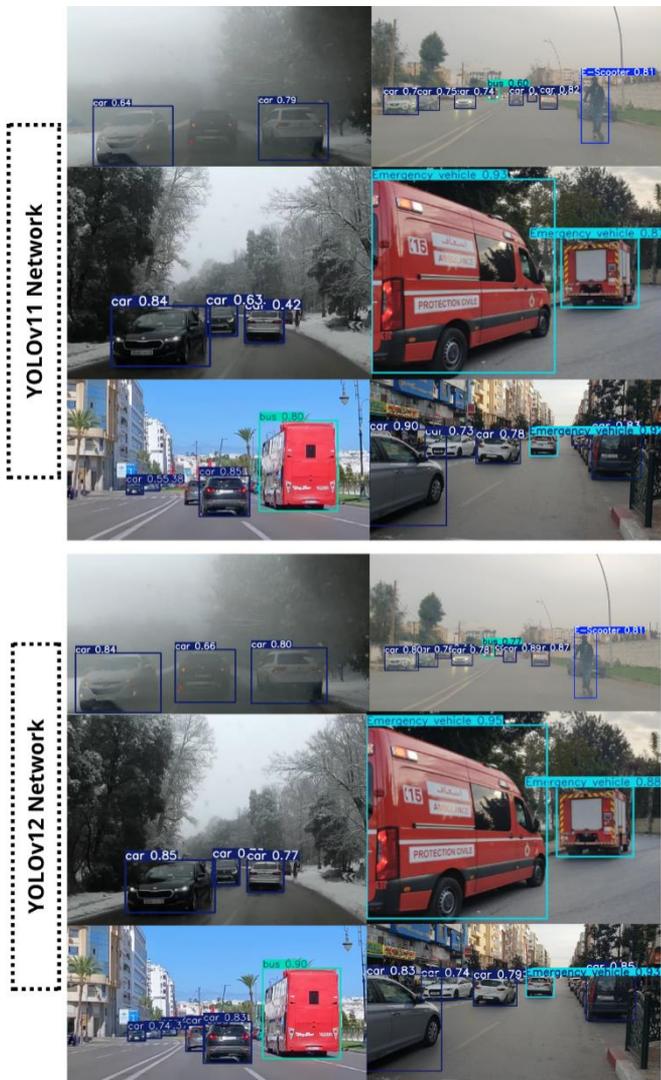


Figure 13. Comparison of YOLOv11 and YOLOv12 in diverse driving conditions

Among different vehicle classes, bicycles and buses show slight dips in F1 scores, likely due to occlusion challenges and size variations in the dataset. Conversely, E-Scooters, motorcycles, and trucks maintain stable F1 scores across both models, highlighting their robust detection accuracy. The improvements in YOLOv12’s architectural design, including

R-ELAN and attention mechanisms, contribute to its higher efficiency in real-time autonomous vehicle perception.

Figures 11 and 12 show the normalized confusion matrices, which offer deeper insights into classification performance. YOLOv11 displayed misclassifications, particularly for bicycles and cars, likely due to similarities in visual features and environmental clutter. YOLOv12, by contrast, achieved more accurate class distinctions and a reduced false positive rate, especially for visually similar objects. These improvements stem from the attention-based architecture and enhanced feature learning provided by the R-ELAN modules.

Figure 13 presents qualitative comparisons in diverse environmental conditions, including fog, grayscale, occlusions, and dense traffic. YOLOv12 showed improved detection consistency and bounding box accuracy in all scenarios, especially under adverse conditions like low visibility and partial object obstruction. These results validate its superior real-time adaptability in complex driving environments.

From a real-world perspective, the improved detection capabilities of YOLOv12 have several important implications for autonomous driving systems. Higher recall reduces the risk of missing critical objects, particularly in dense urban or high-speed environments, enhancing situational awareness and safety. The model’s robustness in detecting occluded or small objects supports more reliable performance in complex real-world traffic scenes, directly benefiting ADAS. However, these advantages must be balanced against deployment constraints, especially in edge-computing environments. YOLOv12’s enhanced architecture introduces additional computational demands, making model optimization necessary for use in real-time applications. Techniques such as quantization, pruning, or deploying lightweight variants may be required to ensure responsiveness on embedded systems.

YOLOv12 outperforms YOLOv11 in most evaluation metrics and real-world detection scenarios, confirming its readiness for advanced perception tasks in intelligent transportation systems. Yet, continued efforts are needed to address class-specific weaknesses and enhance its deployability across embedded systems. The findings in this study offer a strong foundation for future enhancements in object detection models tailored for autonomous driving.

4. CONCLUSIONS

The comparison of YOLOv11 and YOLOv12 demonstrates the continuous progress in real-time object detection, particularly for autonomous vehicle perception. YOLOv12 exhibits significant enhancements in regard to the accuracy, recall, and robustness of detections built on architectural advances, including the design of attention modules and R-ELANs. Experimental evaluation demonstrates stable performance of YOLOv12 in difficult conditions such as low visibility, occlusion, and high traffic density, suggesting its possibilities for real-world applications.

In addition, the model has better capability to identify partially occluded or smaller cars, which is one of the needed capabilities for enhancing safety and situation perception in complex city driving environments. YOLOv12, though, still has difficulty with certain categories of objects, such as buses and bicycles, especially under cluttered or overlapping conditions. These issues indicate the need for further optimization in handling visually ambiguous or scale-invariant objects.

In the future, sensor fusion techniques that combine YOLOv12 with LiDAR or radar data need to be investigated for better performance in adverse weather or low-light environments. Multiscale feature learning improvements would address the model's issue with small and occluded objects. Model compression techniques such as quantization and pruning might also be necessary to enable real-time deployment on low-power embedded systems. Domain adaptation and expanding training datasets, especially for underrepresented classes, would further increase the detections' reliability. These areas will propel the further development of safe, efficient, and scalable object detection architectures for autonomous vehicles and intelligent transportation systems.

ACKNOWLEDGMENT

This research was carried out at the Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco, within the Laboratory of Electronic Systems, Information Processing, Mechanics, and Energetics, in collaboration with the National School of Applied Sciences.

REFERENCES

[1] Tahir, N.U.A., Zhang, Z., Asim, M., Chen, J., ELAffendi, M. (2024). Object detection in autonomous vehicles under adverse weather: A review of traditional and deep learning approaches. *Algorithms*, 17(3): 103. <https://doi.org/10.3390/a17030103>

[2] Gallagher, J.E., Oughton, E.J. (2025). Surveying You Only Look Once (YOLO) multispectral object detection advancements, applications and challenges. *IEEE Access*, 13: 7366-7395. <https://doi.org/10.1109/ACCESS.2025.3526458>

[3] Sani, V., Kantipudi, M.V.V., Meduri, P. (2023). Enhanced SSD algorithm-based object detection and depth estimation for autonomous vehicle navigation. *International Journal of Transport Development and Integration*, 7(4): 341-351. <https://doi.org/10.18280/ijtdi.070408>

[4] Ultralytics. YOLO12: Détection d'objets centrée sur l'attention. <https://docs.ultralytics.com/fr/models/yolo12>, accessed on Mar. 02, 2025.

[5] Alif, M.A.R., Hussain, M. (2024). YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain. *arXiv preprint arXiv:2406.10139*. <https://doi.org/10.48550/arXiv.2406.10139>

[6] Alif, M.A.R., Hussain, M. (2025). YOLOv12: A breakdown of the key architectural features. *arXiv preprint arXiv:2502.14740*. <https://doi.org/10.48550/arXiv.2502.14740>

[7] Mohammed, G.S.A., Diah, N.M., Ibrahim, Z., Jamil, N. (2023). Vehicle detection and classification using three variations of you only look once algorithm. *International Journal of Reconfigurable and Embedded Systems*, 12(3): 442-452. <https://doi.org/10.11591/ijres.v12.i3.pp442-452>

[8] Zhang, Y., Sun, Y., Wang, Z., Jiang, Y. (2023). YOLOv7-RAR for urban vehicle detection. *Sensors*, 23(4): 1801. <https://doi.org/10.3390/s23041801>

[9] Ling, H., Zhao, T., Zhang, Y., Lei, M. (2024). Engineering vehicle detection based on improved YOLOv6. *Applied Sciences*, 14(17): 8054. <https://doi.org/10.3390/app14178054>

[10] Liang, Z., Wang, W., Meng, R., Yang, H., Wang, J., Gao, H., Fan, J. (2024). Vehicle and pedestrian detection based on improved YOLOv7-tiny. *Electronics*, 13(20): 4010. <https://doi.org/10.3390/electronics13204010>

[11] Bakirci, M. (2024). Real-time vehicle detection using YOLOv8-nano for intelligent transportation systems. *Traitement du Signal*, 41(4): 1727-1740. <https://doi.org/10.18280/ts.410407>

[12] Soylyu, E., Soylyu, T. (2024). A performance comparison of YOLOv8 models for traffic sign detection in the Robotaxi-full scale autonomous vehicle competition. *Multimedia Tools and Applications*, 83(8): 25005-25035. <https://doi.org/10.1007/s11042-023-16451-1>

[13] Yaamini, H.G., Swathi, K.J., Manohar, N., Kumar, A. (2025). Lane and traffic sign detection for autonomous vehicles: Addressing challenges on Indian road conditions. *MethodsX*, 14: 103178. <https://doi.org/10.1016/j.mex.2025.103178>

[14] Sundaresan Geetha, A., Alif, M.A.R., Hussain, M., Allen, P. (2024). Comparative analysis of YOLOv8 and YOLOv10 in vehicle detection: Performance metrics and model efficacy. *Vehicles*, 6(3): 1364-1382. <https://doi.org/10.3390/vehicles6030065>

[15] Sharma, A., Kumar, V., Longchamps, L. (2024). Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and faster R-CNN models for detection of multiple weed species. *Smart Agricultural Technology*, 9: 100648. <https://doi.org/10.1016/j.atech.2024.100648>

[16] Alif, M.A.R. (2024). Yolov11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems. *arXiv preprint arXiv:2410.22898*. <https://doi.org/10.48550/arXiv.2410.22898>

[17] Alkhamash, E.H. (2025). A comparative analysis of YOLOv9, YOLOv10, YOLOv11 for smoke and fire detection. *Fire*, 8(1): 26. <https://doi.org/10.3390/fire8010026>

[18] Tian, Y., Ye, Q., Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*. <https://doi.org/10.48550/arXiv.2502.12524>

[19] Banduka, N., Tomić, K., Živadinović, J., Mladineo, M. (2024). Automated dual-side leather defect detection and classification using YOLOv11: A case study in the finished leather industry. *Processes*, 12(12): 2892. <https://doi.org/10.3390/pr12122892>

[20] Rodríguez-Lira, D.C., Córdova-Esparza, D.M., Álvarez-Alvarado, J.M., Romero-González, J.A., Terven, J., Rodríguez-Reséndiz, J. (2024). Comparative analysis of YOLO models for bean leaf disease detection in natural environments. *AgriEngineering*, 6(4): 4585-4603. <https://doi.org/10.3390/agriengineering6040262>

[21] Wang, Y., Jiang, Y., Xu, H., Xiao, C., Zhao, K. (2025). Detection method of key ship parts based on YOLOv11. *Processes*, 13(1): 201. <https://doi.org/10.3390/pr13010201>

[22] Tian, S., Lu, Y., Jiang, F., Zhan, C., Huang, C. (2024). Improved campus vehicle detection method based on YOLOv11 and grayscale projection-based electronic

- image stabilization algorithm. *Traitement du Signal*, 41(6): 3335-3341. <https://doi.org/10.18280/ts.410649>
- [23] Sharma, B. (2025). YOLOv12: Object detection with attention. <https://learnopencv.com/yolov12/>.
- [24] Roboflow annotate: Label images faster than ever. <https://roboflow.com/annotate>, accessed on Mar. 2, 2025.
- [25] Flores-Calero, M., Astudillo, C.A., Guevara, D., Maza, J., Lita, B.S., Defaz, B., Armingol Moreno, J.M. (2024). Traffic sign detection and recognition using YOLO object detection algorithm: A systematic review. *Mathematics*, 12(2): 297. <https://doi.org/10.3390/math12020297>
- [26] Parambil, M.M.A., Ali, L., Swavaf, M., Bouktif, S., Gochoo, M., Aljassmi, H., Alnajjar, F. (2024). Navigating the YOLO landscape: A comparative study of object detection models for emotion recognition. *IEEE Access*, 12: 109427-109442. <https://doi.org/10.1109/ACCESS.2024.3439346>