Fine Tuned YOLOv11-Based Road Sign Detection

Safa Teboulbi
Laboratory of Systems Integration & Emerging Energies, National Engineering School of Sfax, Sfax University, Tunisia | Higher Institute of Computer Sciences and Mathematics of Monastir, University of Monastir, Tunisia
teboulbisafa@gmail.com (corresponding author)

Seifeddine Messaoud
Laboratory of Condensed Matter and Nanoscience (LR11ES40), Department of Physics, Faculty of Sciences of Monastir, University of Monastir, Tunisia
seifeddine.messaoud@fsm.rnu.tn

Mohamed Ali Hajjaji
Research Laboratory in Algebra Numbers theory and Intelligent Systems (RLANTIS), University of Monastir, Tunisia | Higher Institute of Applied Sciences and Technology of Sousse, University of Sousse, Tunisia
mohamedali.hajjaji@issatso.rnu.tn

Mohamed Atri
Computer Engineering Department, King Khalid University, Abha, Saudi Arabia
matri@kku.edu.sa

Abdellatif Mtibaa
Laboratory of Systems Integration & Emerging Energies, National Engineering School of Sfax, Sfax University, Sfax, Tunisia
abdellatif.mtibaa@enim.rnu.tn

Received: 16 April 2025 | Revised: 9 May 2025 and 28 May 2025 | Accepted: 31 May 2025

ABSTRACT

The rapid growth of autonomous driving and intelligent transportation systems has increased the need for accurate and efficient traffic sign detection. Recognizing traffic signs in real time plays a crucial role in enabling vehicles to understand and respond to dynamic road conditions, ensuring both safety and regulatory compliance. This study investigates the performance of three lightweight YOLOv11 variants, YOLOv11n, YOLOv11s, and YOLOv11m, for road sign detection, aiming to balance accuracy with computational efficiency for real-time deployment in resource-constrained environments. Each model was trained and evaluated using a consistent traffic sign dataset, with performance metrics including precision, recall, mean Average Precision (mAP), and F1 score. The YOLOv11n model demonstrated stable training behavior and achieved a peak mAP@0.5 of 0.52, with a mean F1 score of 0.47, indicating efficient detection of dominant classes but limited performance for underrepresented ones. The YOLOv11s showed improved generalization and localization abilities with a higher mAP@0.5 of 0.55 and a mean F1 score of 0.64, suggesting a balanced trade-off between speed and accuracy. The most advanced variant, YOLOv11m, achieved the highest mAP@0.5 of 0.70 and an F1 score of 0.63, demonstrating robust detection and convergence properties. However, all models exhibited difficulty in detecting rarely represented classes, such as "crosswalk," emphasizing the importance of dataset balancing. These findings confirm the suitability of these YOLOv11 variants for embedded traffic monitoring systems and highlight avenues for further improvement through data augmentation and fine-tuning.

Keywords-road sign detection; fine-tuned YOLOv11; computer vision

I. INTRODUCTION

Object detection on roads plays a crucial role in autonomous driving and Intelligent Transportation Systems (ITS), aiming to enhance road safety, optimize traffic flow, and support autonomous and semi-autonomous vehicles [1]. A fundamental component of ITS is the ability to accurately detect and recognize traffic signs, which provide essential regulatory, warning, and guidance information to drivers and onboard systems [2, 3]. Early approaches relied on handcrafted features, such as color histograms and edge descriptors, combined with classifiers such as SVM and KNN, but their performance degraded in complex environments due to lighting variations, occlusions, and cluttered backgrounds [4]. The advent of deep learning, particularly Convolutional Neural Networks (CNNs), enabled automatic feature extraction and significantly improved robustness and accuracy [5]. Notable early works used multilayered deep networks and benchmarks such as the German Traffic Sign Recognition Benchmark (GTSRB), establishing deep learning as the standard [4, 5].

Modern object detection frameworks, such as Faster R-CNN, SSD, and especially YOLO, have further advanced the field, offering real-time performance and unified architectures [6-9]. In [8], a multistage deep learning pipeline was proposed to improve recognition accuracy in challenging conditions such as partial occlusion and low resolution. This study used the publicly available Traffic Guider Dataset [10], which includes a wide range of annotated traffic signs captured under diverse conditions.

The YOLO series has evolved from YOLOv1 to YOLOv4, incorporating stronger backbones, feature pyramids, and improved training strategies [9, 11]. Region-based methods such as Fully Convolutional Networks (FCNN) and R-CNN variants have also offered pixel-level precision for better sign localization [12]. Data augmentation, transfer learning, and multiview approaches have further enhanced model generalization and robustness in diverse environments [13]. The most recent version, YOLOv11, integrates advanced feature aggregation and attention mechanisms that increase detection accuracy under difficult visual conditions [14].

This study presents a traffic sign detection system based on the YOLOv11 architecture. The proposed model is trained and evaluated using the publicly available Traffic Guider Dataset [10], which consists of annotated images captured from real-world road environments. This dataset features a wide range of traffic sign categories, including stop signs, pedestrian crossings, and bus stops, collected under diverse conditions such as variable lighting, weather, and occlusion levels. By leveraging the state-of-the-art capabilities of YOLOv11, the proposed system demonstrates high detection accuracy and low inference latency, making it well-suited for deployment in smart transportation infrastructure and autonomous vehicle platforms. The primary contributions of this study are summarized as follows:

- Design and implement an efficient traffic sign detection model using the advanced YOLOv11 architecture.
- Train and evaluate the model on the Traffic Guider Dataset, showcasing its robustness across different traffic sign classes and environmental conditions.
- Assess the real-time performance of the model and validate its applicability to ITS and autonomous driving through quantitative metrics.

II. PROPOSED FINE-TUNED YOLOV11 FOR ROAD SIGN DETECTION

This study presents a comprehensive road sign detection framework that leverages the powerful capabilities of the YOLOv11 architecture, adapted and fine-tuned in three versions, nano (n), small (s), and medium (m), to accommodate diverse operational requirements ranging from edge devices to high-performance servers. The proposed system, as shown in Figure 1, is specifically designed for the robust detection of various traffic-related objects, including bus stop signs, crosswalks, pedestrians, crossing signs, and stop signs, which are critical to the safe and intelligent navigation of modern vehicles and the deployment of smart city infrastructure. The detection pipeline begins with the utilization of the Traffic Guider Dataset. To increase the model's generalization ability and mitigate overfitting, a wide range of data augmentation strategies were employed, such as rotation, scaling, contrast adjustment, blurring, flipping, and color space manipulations.

At the core of the proposed framework is YOLOv11, an advanced object detection model enhanced with novel modules to boost speed and accuracy. Key components include the SPPF module for efficient multiscale feature fusion, critical for detecting small or distant signs, the C2PSA block, which combines partial connections with self-attention to highlight semantically important regions, and the C3K2 block, which deepens feature extraction using stacked convolutions with minimal overhead. Together, these modules improve object localization, classification, and robustness in complex traffic scenes.

The YOLOv11-based detection pipeline shows strong generalization across diverse traffic conditions, with a modular and scalable design that supports efficient deployment on various platforms. Through the integration of attention mechanisms, multiscale feature fusion, and efficient convolutional blocks, the system advances road sign detection and is well-suited for applications in intelligent transportation, autonomous driving, and smart city monitoring. The YOLOv11 architecture was selected due to its flexibility, performance, scalability, and real-time detection capabilities, which are critical for traffic sign recognition tasks in complex urban environments. YOLOv11 is available in multiple variants, namely YOLOv11n (nano), YOLOv11s (small), and YOLOv11m (medium), each offering different trade-offs between accuracy, speed, and computational demand, thus allowing fine-grained control over model deployment based on hardware constraints and application needs. Nano suits low-power devices, small offers a balance of speed and accuracy for real-time use, and medium prioritizes detection precision for performance-critical applications. All variants were trained and fine-tuned on the Traffic Guider Dataset using transfer learning and hyperparameter tuning to achieve high detection accuracy.

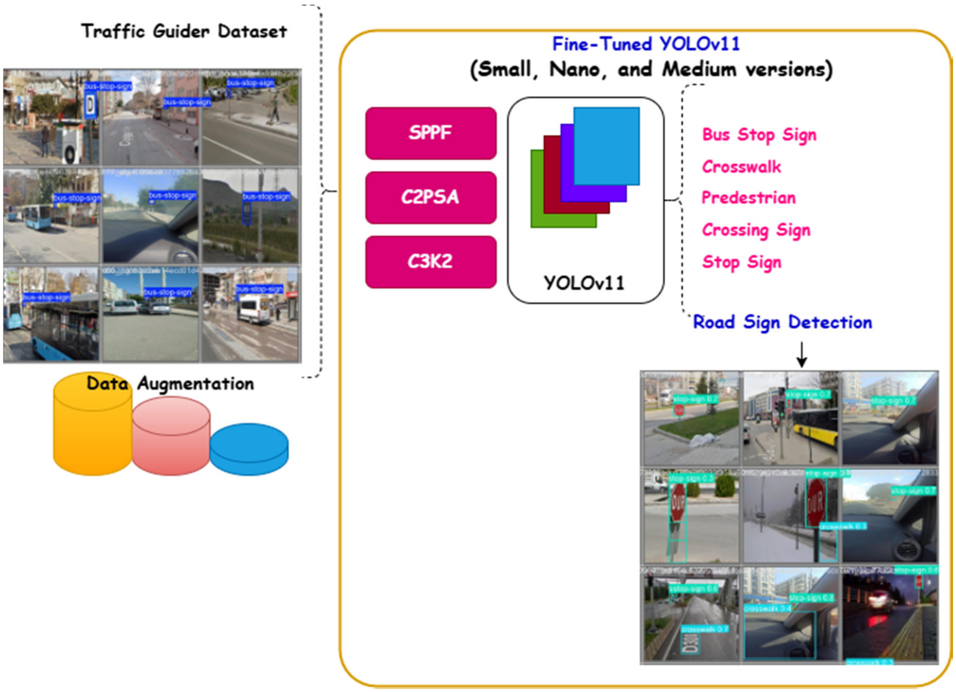

Fig. 1. Proposed system for road sign detection.

YOLOv11n represents the lightest version in the series. It is characterized by a depth coefficient of 0.33 and a width multiplier of 0.25. This compact design significantly reduces the number of parameters and floating-point operations, enabling real-time inference even on resource-constrained devices such as embedded systems and mobile platforms. Despite its small size, YOLOv11n retains the essential building blocks of the architecture, ensuring a baseline level of performance suitable for rapid prototyping and edge AI applications. YOLOv11s increases the model's representational capacity by maintaining the same depth (0.33) while doubling the width multiplier to 0.5. This results in a model that offers improved accuracy over YOLOv11n without incurring a substantial increase in inference time. Thus, YOLOv11s is often considered the best compromise between speed and detection precision, especially when deployed on mid-range GPUs or CPUs. YOLOv11m, the medium variant, further amplifies the complexity of the model by adopting a depth of 0.67 and a width of 0.75. This configuration supports deeper feature extraction and a finer semantic understanding of the input images, leading to improved accuracy in challenging scenarios such as low lighting, occlusions, and small object detection. However, this comes at the cost of increased memory consumption and slower inference times, making it more suitable for deployment on high-performance computing platforms.

All YOLOv11 variants share a modular architecture featuring the C2f block for improved feature propagation, the SPPF module for multiscale receptive field generation, and a PAN-based head for effective feature fusion. This design enhances the ability of the network to detect diverse traffic signs in complex and dynamic environments. Emphasizing adaptability and modularity, YOLOv11 proves to be a flexible and powerful solution for a wide range of intelligent transportation applications.

III. EXPERIMENTAL SETUP

A. *Traffic Guider Dataset*

The Traffic Guider dataset [10] comprises annotated images representing various road-related sign classes, specifically "bus-stop-sign," "crosswalk," "pedestrian-crossing-sign," and "stop-sign." As shown in Figure 2, the distribution of instances per class is imbalanced, with the "crosswalk" class being the most frequent, exceeding 340 instances, while the other classes such as "bus-stop-sign" and "stop-sign" have fewer than 100 instances each. The annotation bounding boxes, shown on the top-right of the figure, exhibit a consistent spatial distribution, suggesting reliable annotation quality. Furthermore, spatial analysis of the bounding boxes using scatter and heatmap plots (bottom row of Figure 2) reveals that most objects are centered horizontally and located slightly above the vertical midpoint of the images, indicating that signs tend to appear in predictable positions in the visual field.

The correlogram in Figure 3 provides a deeper look at feature relationships, such as between bounding box coordinates (x, y) and dimensions (width, height), demonstrating moderate correlations and certain distributional patterns. In particular, bounding boxes exhibit clustering around smaller widths and heights, confirming the prevalence of compact sign instances in the dataset. This comprehensive analysis underlines the dataset's suitability for training object detection models, while also highlighting potential challenges related to class imbalance and small object representation [13].

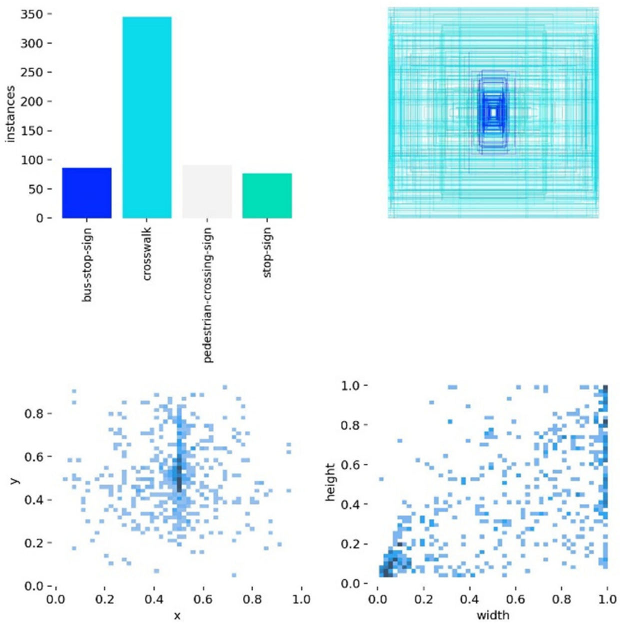

Fig. 2. Dataset labels.

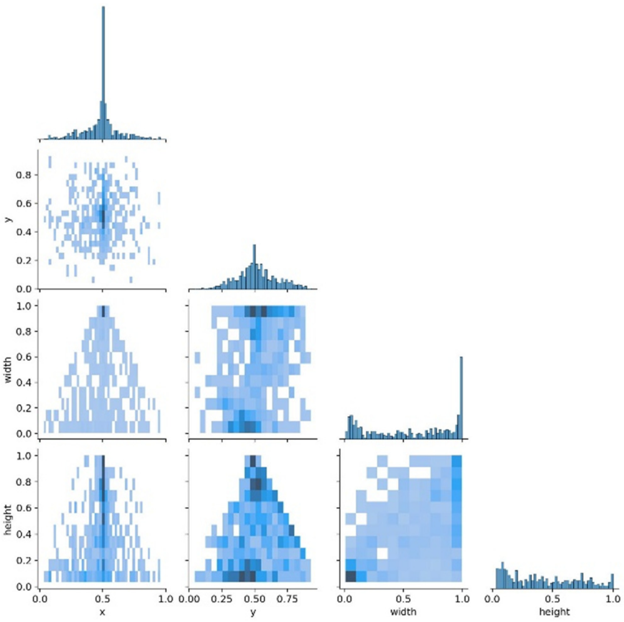

Fig. 3. Dataset correlogram.

B. Training Results of YOLOv11n

The training results of the YOLOv11n for the detection of road signs, illustrated in Figure 4, demonstrate a consistent and stable convergence of all loss components, including box loss, classification loss, and Distribution Focal Loss (DFL), with training losses decreasing smoothly and validation losses showing moderate divergence, indicating mild overfitting. Precision stabilizes around 0.75, while recall plateaus near 0.52, suggesting that the model is effective in reducing false positives but may miss some true detections. The mAP@0.5 metric reaches approximately 0.52, and the more stringent mAP@[.5:.95] metric peaks around 0.31, reflecting solid localization and classification performance given the model's lightweight architecture. These results confirm YOLOv11n's suitability for real-time, resource-constrained applications such as embedded traffic monitoring systems. Table I presents the validation performance of the YOLOv11n model for each class using three key metrics: precision, recall, and F1 score. The bus-stop-sign and stop-sign classes continue to exhibit strong detection performance, achieving F1 scores close to 0.92 and 0.91, respectively, indicative of well-balanced precision and recall. In contrast, the "crosswalk" and "pedestrian-crossing-sign" classes show extremely poor detection capabilities, with F1 scores close to zero, reflecting the model's difficulty in recognizing these underrepresented classes. Overall, the mean F1 score across all classes is 0.47 at a confidence threshold of 0.469, confirming the model's effective detection for dominant classes while underlining the need to enrich the dataset for minor categories.

TABLE I. EVALUATION RESULTS OF YOLOV11N

Class	Images	Instances	Precision (P)	Recall (R)	F1 score
All	50	47	0.515	0.478	0.47
Bus-stop-sign	23	23	0.957	0.961	~0.92
Crosswalk	2	2	0.000	0.000	~0.18
Pedestrian-crossing-sign	1	1	0.000	0.000	~0.00
Stop-sign	20	20	0.904	0.952	~0.91

C. Training Results of YOLOv11s

The training results of the YOLOv11s model for road sign detection, as shown in Figure 5, exhibit a well-structured convergence across training loss components, including box loss, classification loss, and DFL, with all showing a consistent downward trend that indicates effective learning. Validation losses also generally follow this trend with some early spikes, particularly in the first few epochs, suggesting initial instability that stabilizes over time. The precision metric fluctuates initially but eventually stabilizes around 0.68, reflecting the model's ability to minimize false positives. Meanwhile, recall increases steadily and levels at around 0.48, indicating moderate success in capturing true positives, although there is room for improvement. The mAP@0.5 reaches about 0.55, and the more comprehensive mAP@0.5:0.95 settles near 0.34, showcasing the model's strong performance in both detection and localization across various IoU thresholds. These results highlight YOLOv11s as a robust option for road sign detection tasks, especially in environments where a balance between inference speed and accuracy is critical. Further enhancements could involve hard example mining or fine-tuning on more diverse road sign datasets to boost recall and overall detection robustness. Table II shows the performance metrics of the YOLOv11s model, with the F1 score derived from the model's confidence curve. The model shows a high F1 score of 0.87 for the "bus-stop-sign" class, reflecting an excellent balance between precision and recall for this category. The overall F1 score across all classes is 0.64 at a confidence threshold of 0.214, indicating moderate performance. The "crosswalk" class records an F1 score of 0 due to the absence of correct detections, possibly caused by the minimal number of training instances. These results underline the importance of dataset balance in multi-class object detection.

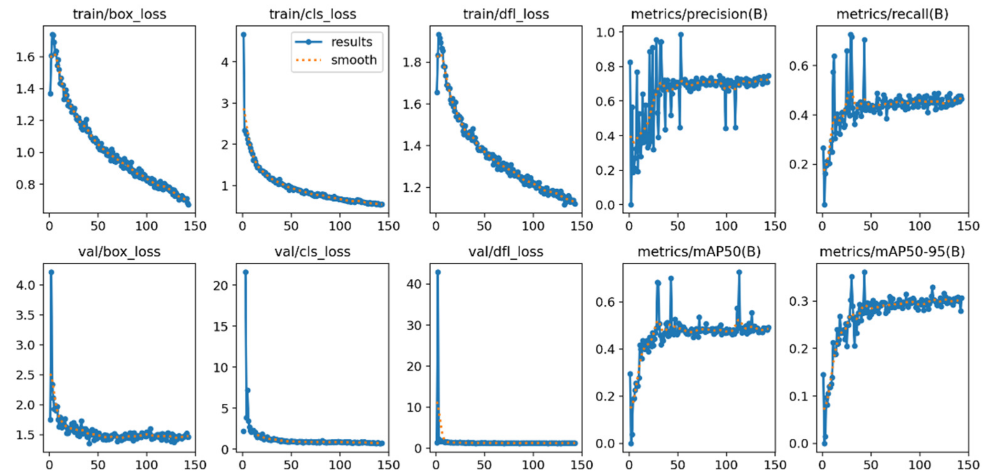

Fig. 4. Training results of YOLOv11n.

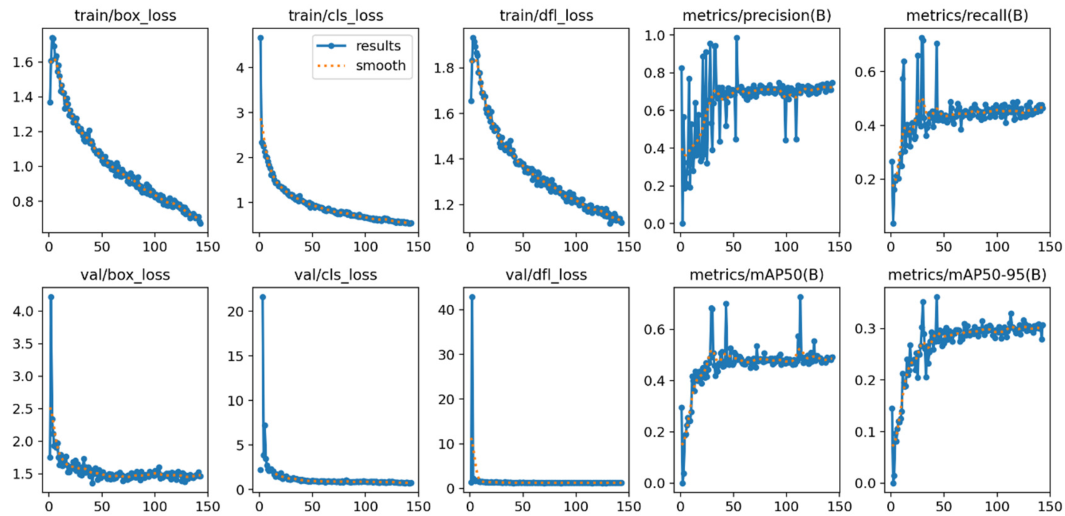

Fig. 5. Training results of YOLOv11s.

TABLE II. EVALUATION RESULTS OF YOLOV11S

Class	Images	Instances	Precision (P)	Recall (R)	F1 score
All	50	47	0.747	0.464	0.64
Bus-stop-sign	23	23	0.780	0.913	0.87
Crosswalk	2	2	0.000	0.000	0.00

D. Training Results of YOLOv11m

The training results of the YOLOv11m model demonstrate effective learning and convergence over 200 epochs. Training losses show consistent decreases, with the box loss reducing from approximately 2.0 to 0.55, the classification loss from 3.5 to around 0.5, and the DFL from 2.2 to about 1.1. Similarly, validation losses followed a downward trend: the validation box loss dropped from 4.0 to around 1.5, the classification loss decreased sharply from over 60 to about 2.0, and DFL fell from approximately 15 to 2. These patterns reflect the model's ability to generalize well across both training and validation data. In terms of performance metrics, precision improved steadily, peaking around 0.8 and stabilizing between 0.65 and 0.7, while recall increased to approximately 0.5. The mAP@0.5 reached 0.7 with some fluctuations, and mAP@0.5:0.95 gradually increased to about 0.35, indicating robust detection capabilities across different IoU thresholds. Overall, the YOLOv11m model exhibited strong performance, with stable learning behavior and improved detection accuracy throughout the training process. Table III presents the validation results of the YOLOv11m model. The model exhibits high detection performance for the "stop-sign" and "pedestrian-crossing-sign" classes, both showing F1 scores close to 1.0, indicating nearly perfect precision and recall. The model had a strong performance on the "bus-stop-sign class," with an approximate F1 score of 0.85. In contrast, the model fails to detect crosswalk instances, resulting in an F1 score of 0.00, which highlights a weakness in recognizing this specific class. The overall F1 score across all classes is approximately 0.63 at a confidence threshold of 0.043, reflecting a reasonably balanced detection ability that could benefit from further tuning or data augmentation.

TABLE III. EVALUATION RESULTS OF YOLOV11M

Class	Images	Instances	P (Precision)	R (Recall)	F1 score
bus-stop-sign	23	23	0.735	1.000	~0.85
crosswalk	2	2	0.000	0.000	~0.00
pedestrian-crossing-sign	6	6	1.000	1.000	~1.00
stop-sign	20	21	1.000	0.952	~0.97
Overall	50	47	0.716	0.475	0.63

E. Comparative Study

The comparative analysis is based on the training and evaluation of the YOLOv11n, YOLOv11s, and YOLOv11m models on the same dataset, ensuring a consistent and fair comparison. YOLOv11n, the most lightweight variant, exhibits smooth training convergence and effectively minimizes false positives, achieving a precision of approximately 0.75. However, its recall is relatively low (~0.52), and the mean F1 score (0.47 at a confidence threshold of 0.469) indicates limited effectiveness in detecting less represented classes. YOLOv11s demonstrates a more balanced performance, with a precision of about 0.68, a recall of 0.48, and an improved mean F1 score of 0.64 at a confidence threshold of 0.214, making it well-suited for real-time applications requiring both speed and accuracy. YOLOv11m, the most robust variant, achieves the highest mAP@0.5 (0.70) and strong F1 scores on dominant classes such as "stop-sign" and "pedestrian-crossing-sign" (both near 1.0). Despite a similar recall (~0.50), its overall F1 score (0.63 at a low confidence threshold of 0.043) reflects superior detection capabilities, particularly for complex or infrequent classes. All models exhibit challenges in detecting the underrepresented "crosswalk" class. Overall, YOLOv11m is the most reliable choice for high-accuracy detection in moderately constrained environments, while YOLOv11n is optimal for highly resource-limited deployments. The comparative results reported in Table IV are exclusively from the experiments in this study.

TABLE IV. COMPARATIVE STUDY

Model	Precision	Recall	mAP @0.5	mAP @0.5:0.95	Mean F1 score
YOLO v11n	~0.75	~0.52	0.52	0.31	0.47 (@ conf. 0.469)
YOLO v11s	~0.68	~0.48	0.55	0.34	0.64 (@ conf. 0.214)
YOLO v11m	~0.70	~0.50	0.70	0.35	0.63 (@ conf. 0.043)

IV. CONCLUSION

The evaluation of YOLOv11n, YOLOv11s, and YOLOv11m models for road sign detection demonstrates their effectiveness in real-time object detection tasks, particularly under resource constraints. YOLOv11n, the most lightweight variant, delivers rapid inference with acceptable accuracy for prominent classes, making it well-suited for ultra-low-power or embedded applications. YOLOv11s offers a favorable trade-off between detection precision and recall, ideal for scenarios where moderate accuracy and robustness are essential. YOLOv11m, with its superior detection and localization capabilities, emerges as the most accurate and reliable option for high-performance use cases, while still maintaining efficient computational costs. It should be noted that all models exhibit performance drops in minority classes, such as "crosswalk," underscoring the persistent challenge of class imbalance. This highlights the need for dataset enrichment and targeted augmentation techniques to improve generalization across underrepresented classes. The comparison of the three YOLOv11 variants on the same dataset shows that they offer competitive or improved accuracy, especially in real-time contexts, while achieving better deployment feasibility across different hardware platforms. Future work will focus on integrating adaptive loss functions, attention mechanisms, and semi-supervised training to further enhance minority class detection without compromising speed. In general, these YOLOv11 variants are scalable and practical solutions for deployment in intelligent transportation systems and smart city infrastructures, where both accuracy and resource efficiency are critical.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through a Large Group Project under grant no. RGP2/444/45.

REFERENCES

[1] T. Saidani, "Deep Learning Approach: YOLOv5-based Custom Object Detection," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12158–12163, Dec. 2023, https://doi.org/10.48084/etasr.6397.

[2] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, Sep. 2012, https://doi.org/10.1109/TITS.2012.2209421.

[3] S. Saxena, S. Dey, M. Shah, and S. Gupta, "Traffic sign detection in unconstrained environment using improved YOLOv4," *Expert Systems with Applications*, vol. 238, Mar. 2024, Art. no. 121836, https://doi.org/10.1016/j.eswa.2023.121836.

[4] Y. Zhu, C. Zhang, D. Zhou, X. Wang, X. Bai, and W. Liu, "Traffic sign detection and recognition using fully convolutional network guided proposals," *Neurocomputing*, vol. 214, pp. 758–766, Nov. 2016, https://doi.org/10.1016/j.neucom.2016.07.009.

[5] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," *Machine Vision and Applications*, vol. 25, no. 3, pp. 633–647, Apr. 2014, https://doi.org/10.1007/s00138-011-0391-3.

[6] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 3642–3649, https://doi.org/10.1109/CVPR.2012.6248110.

[7] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Networks*, vol. 32, pp. 323–332, Aug. 2012, https://doi.org/10.1016/j.neunet.2012.02.016.

[8] P. Sanjeewani, B. Verma, and J. Affum, "Multi-stage Deep Learning Technique for Improving Traffic Sign Recognition," in *2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Tauranga, New Zealand, Dec. 2021, pp. 1–6, https://doi.org/10.1109/IVCNZ54163.2021.9653198.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.

[10] Y. Peker, "Traffic Guider Object Detection Dataset by Yasin Peker." Roboflow Universe, 2023, [Online]. Available: https://universe.roboflow.com/yasin-peker-zeih0/traffic-guider.

[11] M. Flores-Calero *et al.*, "Traffic Sign Detection and Recognition Using YOLO Object Detection Algorithm: A Systematic Review," *Mathematics*, vol. 12, no. 2, Jan. 2024, Art. no. 297, https://doi.org/10.3390/math12020297.

[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." arXiv, Apr. 08, 2018, https://doi.org/10.48550/arXiv.1804.02767.

[13] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv, Apr. 23, 2020, https://doi.org/10.48550/arXiv.2004.10934.

[14] M. Bakirci and I. Bayraktar, "The Cutting-Edge YOLO11 for Advanced Aircraft Detection in Synthetic Aperture Radar (SAR) Imagery," in *2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, İstanbul, Turkiye, Dec. 2024, pp. 1–6, https://doi.org/10.1109/ISAS64331.2024.10845222.