Enhanced YOLOR for Accurate and Real-Time Traffic Sign Detection in Autonomous Driving

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: 13 May 2025 | Revised: 10 June 2025 | Accepted: 21 June 2025

DOI: https://doi.org/10.48084/etasr.12102

ABSTRACT

Traffic sign detection systems are vital for improving road safety and supporting autonomous navigation in urban environments. This paper presents a fine-tuned traffic sign detection system based on the You Only Learn One Representation (YOLOR) architecture. The model was trained and evaluated on a dataset comprising 15 traffic sign classes: Green Light, Red Light, Speed Limit 10, Speed Limit 100, Speed Limit 110, Speed Limit 120, Speed Limit 20, Speed Limit 30, Speed Limit 40, Speed Limit 50, Speed Limit 60, Speed Limit 70, Speed Limit 80, Speed Limit 90, and Stop. To enhance detection performance across diverse classes, the model was fine-tuned to accurately detect and classify these elements under varying conditions. The experimental results demonstrate strong detection capabilities, achieving a precision of 87.5%, a recall of 87.8%, a mean Average Precision at IoU 0.5 (mAP@0.5) of 88.5%, and a mAP across IoU thresholds from 0.5 to 0.95 (mAP@0.5:0.95) of 77.7%. These results highlight the effectiveness of the YOLOR-based approach for real-world traffic sign recognition tasks, offering a promising solution for intelligent transportation and autonomous driving applications. Furthermore, the model's competitive performance compared to recent methods reinforces its relevance in current state-of-the-art benchmarks.

Keywords-computer vision; traffic signs detection; fine-tuned YOLOR; autonomous driving

I. INTRODUCTION

The rapid evolution of autonomous systems and intelligent transportation has significantly amplified the demand for accurate and real-time object detection models, particularly in the domain of traffic sign recognition. As traffic signs provide essential information for vehicle navigation and decision-making, their precise detection is critical to ensure safety and compliance in autonomous driving systems. Over recent years, deep learning-based object detectors have demonstrated superior performance in handling complex visual scenes. The YOLO (You Only Look Once) family of models, particularly YOLO-v5, has become widely adopted due to its balance between speed and accuracy [1]. However, while YOLO-v5 remains effective in general-purpose object detection, emerging requirements in safety-critical domains have led to further exploration of more adaptive and unified models.

YOLOR (You Only Learn One Representation) was introduced to unify conventional and implicit knowledge in a single framework, delivering improved generalization across diverse tasks [2]. This approach has gained traction in various domains, such as safety on construction sites [3], aerial imagery analysis [4], bridge defect inspection [5], and drone detection [6]. In particular, the model has also been integrated into active federated learning systems to improve safety in autonomous vehicles [7], showcasing its flexibility and robustness.

In the context of autonomous vehicles, object detection remains a persistent challenge due to dynamic environments, occlusions, and the need for real-time performance. Studies such as [8, 9] have surveyed the challenges and techniques applied in vehicular perception systems. In [10], adaptive detection and tracking frameworks were proposed to accommodate the variability of road scenes, emphasizing the importance of reliable and real-time systems. Several recent studies have explored the application of YOLOR specifically in transportation and safety-related domains. For instance, in [11], YOLOR was applied for real-time ship detection, while in [12], SOAda-YOLOR was proposed as an adaptive variant tailored for the detection of small objects in road scenarios. Likewise, an asynchronous multitask learning framework was introduced in [13], using YOLOR to further optimize detection performance in multi-label environments.

Building on this body of work, this study presents a fine-tuned YOLOR-based model for the detection of traffic signs using a 15-class dataset obtained from Kaggle [14], which consists of annotated images captured from real-world road environments. This dataset features a wide range of traffic sign categories, including: Green Light, Red Light, Speed Limit 10, Speed Limit 100, Speed Limit 110, Speed Limit 120, Speed Limit 20, Speed Limit 30, Speed Limit 40, Speed Limit 50, Speed Limit 60, Speed Limit 70, Speed Limit 80, Speed Limit 90, and Stop. Unlike broader multiclass detection tasks, traffic sign recognition poses unique challenges, such as small object size, similarity among classes, and strict real-time constraints. Utilizing the advanced architecture of YOLOR, this framework achieves robust object detection with minimal computational overhead, positioning it as an ideal solution for edge-based AI applications in intelligent surveillance and real-time traffic monitoring systems. The primary contributions of this study are summarized as follows:

- Develops and deploys an efficient detection model for traffic signs, leveraging the advanced capabilities of the YOLOR architecture.
- Trains and evaluates the model on the Traffic Signs Detection Dataset, showcasing its robustness across different traffic sign classes and environmental conditions.
- Assesses the real-time performance of the model and validates its applicability to ITS and autonomous driving using both quantitative metrics and qualitative visualizations.

II. PROPOSED FINE-TUNED YOLOV11 FOR ROAD SIGN DETECTION

This study presents a comprehensive method for traffic sign detection based on a fine-tuned YOLOR model. The overall workflow, illustrated in Figure 1, begins with the collection of a diverse traffic sign dataset sourced from Kaggle, covering 15 distinct classes, including speed limit indicators, stop signs, and traffic lights. This dataset is employed to fine-tune the YOLOR model, integrating multi-scale feature alignment, attention mechanisms, and prediction refinement strategies to enhance detection accuracy and robustness. The fine-tuned model is subsequently validated in real-world autonomous driving scenarios, demonstrating its ability to detect and classify traffic signs accurately under a wide range of environmental conditions, including varying lighting, occlusions, and partial visibility.

YOLOR is a unified object detection architecture that seamlessly combines implicit and explicit knowledge representations within a single framework. Traditional detectors often separate the processes of feature extraction and task-specific optimization, potentially limiting performance. In contrast, YOLOR introduces a multitask learning paradigm, simultaneously optimizing shared features for primary detection tasks and auxiliary tasks, thus facilitating better generalization. Specifically, YOLOR leverages both conventional convolutional features (explicit representations) and intermediate representations learned through unsupervised tasks (implicit representations). It utilizes implicit knowledge, information that is not directly interpretable but valuable for the task, by embedding intermediate supervision during training. This approach enables the model to capture fine-grained, context-aware features that are critical for tasks such as traffic sign recognition, where subtle visual differences are essential. Moreover, YOLOR enhances feature aggregation through multiscale feature fusion modules and improves bounding-box prediction using refined anchor box mechanisms and adaptive loss functions. Despite its increased complexity, the hybrid design of YOLOR maintains real-time inference speed, achieving a balance between detection accuracy and computational efficiency. These characteristics make YOLOR an ideal choice for embedded systems and autonomous driving applications, where both speed and precision are paramount [2].

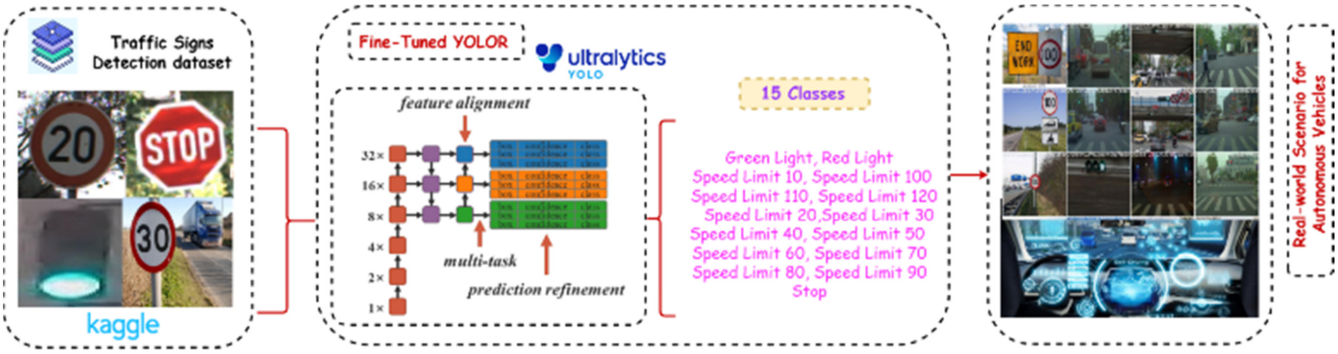

Fig. 1. Overview of the proposed fine-tuned YOLOR-based traffic sign detection pipeline.

III. EXPERIMENTAL RESULTS

A. *Dataset Description*

The Traffic Signs Detection Dataset [14] comprises annotated images containing 15 classes of traffic signs. As shown in Figure 2, the class distribution is notably imbalanced. Classes such as Speed Limit 20, Speed Limit 30, and Stop are more frequently represented, each with close to or above 250 instances, while other classes, such as Speed Limit 10 and Speed Limit 120, are underrepresented. The spatial distribution of the bounding boxes (top-right plot in Figure 2) reveals that most annotated traffic signs tend to appear near the horizontal center of the images, with a slight bias toward the lower half of the frame. The width-height scatter plot (bottom-left) confirms a strong linear relationship, indicating consistency in object proportions and aspect ratios. Most annotations fall within the lower range of width and height values, suggesting that the dataset contains predominantly small-sized traffic signs, a common challenge in real-world detection tasks.

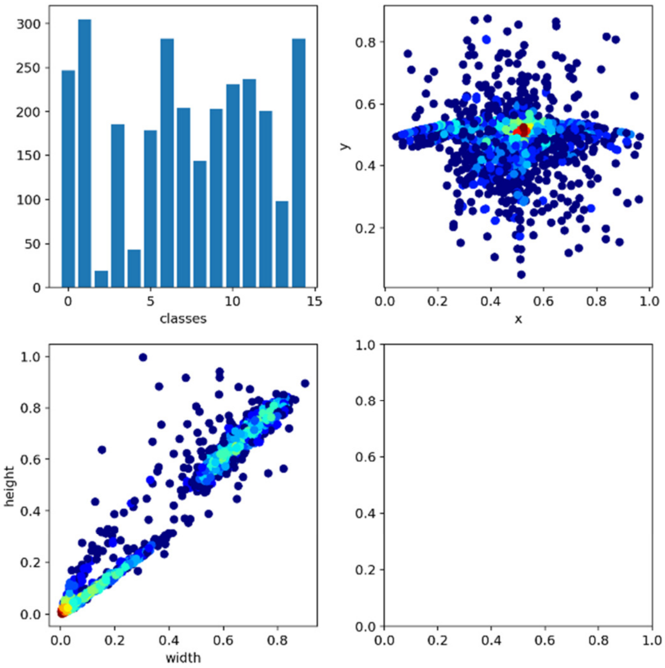

Fig. 2. Dataset labels.

Supporting this, the correlogram in Figure 3 highlights the relationships among bounding box features (x, y, width, height). A strong correlation is observed between width and height, while x and y are less correlated with size attributes, indicating independent positioning from object scale. The distribution plots reveal skewness toward lower values, especially in width and height dimensions.

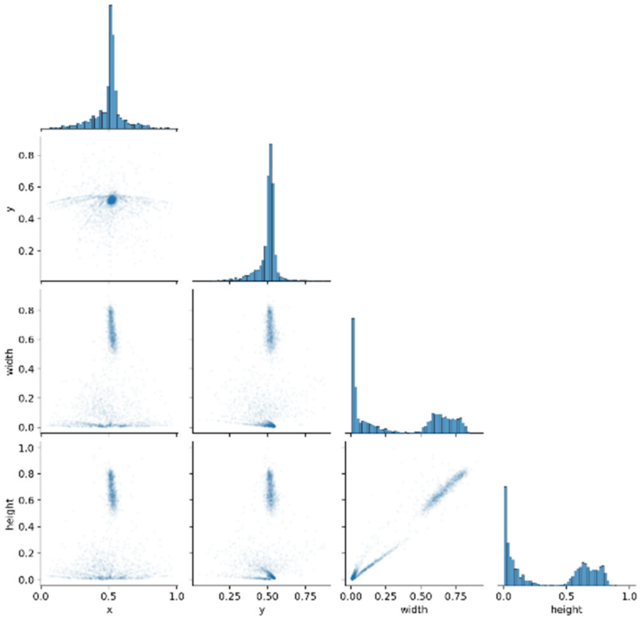

Fig. 3. Dataset correlogram.

This comprehensive dataset analysis emphasizes both the richness and the challenges of the dataset. It is well-suited for training deep object detectors for intelligent transportation systems, especially in scenarios requiring high precision for small object recognition, while also necessitating strategies to mitigate class imbalance and handle scale variability.

B. *Training Platform and Implementation Details*

A PC with Ubuntu 24.04 OS with 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30 GHz CPU, 16 GB of RAM, and an NVIDIA GeForce RTX 3050 GPU was used to train the model. The code implementation of the algorithms was developed in Python, with the Python code for the YOLOR obtained from the official repositories. This ensured the accuracy and practical relevance of the results. The training process was configured with a batch size of 64. Training was carried out over 500 epochs, completed in 7.805 hours.

C. Training Results of YOLOR

Figures 4 and 5 illustrate the training performance of the fine-tuned YOLOR model on the traffic signs detection dataset. As shown in Figure 4, the training and validation losses for bounding box regression (Box), objectness confidence, and classification consistently decreased throughout training epochs, indicating stable convergence. The validation losses remained slightly higher than their training counterparts, suggesting a balanced generalization capability without signs of overfitting. The precision and recall metrics steadily increase, ultimately reaching values above 0.85, which reflects strong detection reliability. Furthermore, mAP@0.5 peaked at approximately 0.885, while the mAP@0.5:0.95 also demonstrated progressive improvement, confirming that the model retains performance even under stricter localization criteria.

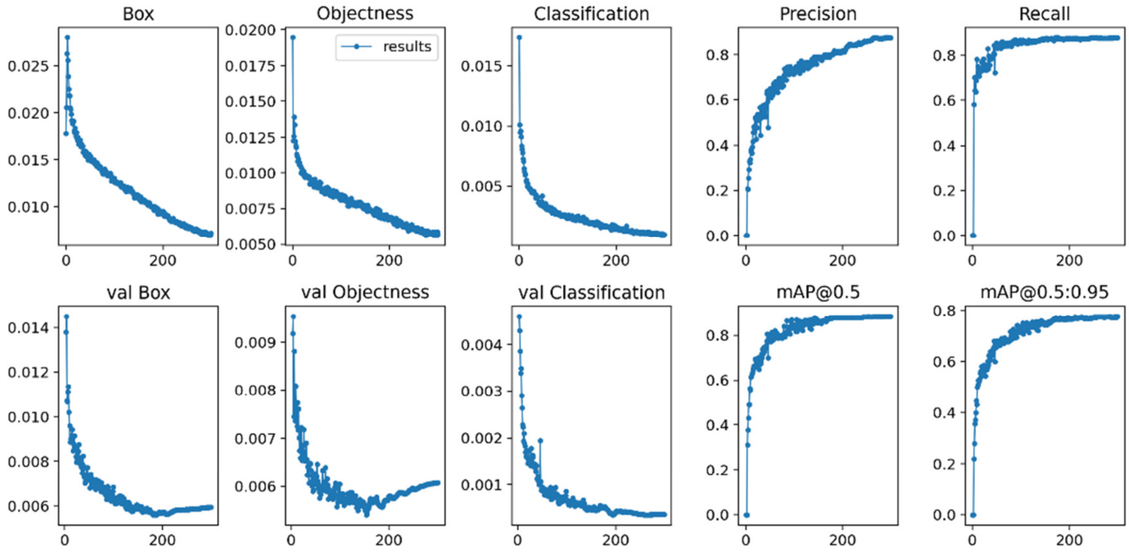

Fig. 4. Training results of YOLOR.

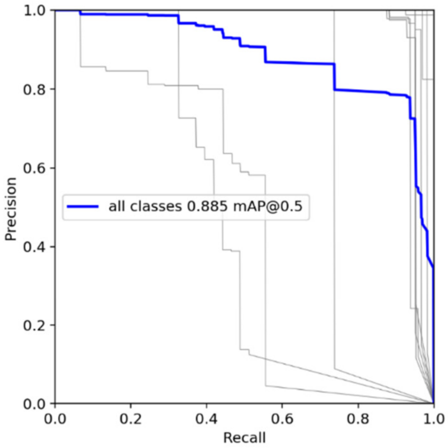

Fig. 5. Precision-Recall (P-R) curve of YOLOR.

Figure 5 presents the Precision-Recall (P-R) curve for the trained model across all 15 classes. The dominant blue curve, representing the aggregated performance across all classes, confirms a high degree of precision even at varying levels of recall, which is characteristic of a well-generalized detector. The mean value of 0.885 mAP@0.5 is consistent with the training metrics observed earlier and indicates a high likelihood of correct predictions in most test instances. The individual class curves, shown in gray, reveal some class-wise variance, implying the need for potential improvements in underrepresented or visually complex categories. Overall, these results highlight YOLOR's strong capacity for accurate and robust multiclass traffic sign detection, even in the presence of small object sizes and class imbalance.

D. Comparative Study

Table I presents the comparative analysis of various object detection models for traffic sign recognition tasks using different datasets. The proposed model achieved strong performance with 87.5% precision, 87.8% recall, 88.5% mAP@0.5, and 77.7% mAP@0.5:0.95, outperforming most of the existing methods. TRD-YOLO [15], evaluated on the TT100K dataset, shows slightly lower results with 86.3% precision, 82.1% recall, 86.5% mAP@0.5, and 65.6% mAP@0.5:0.95. YOLO-CCA [16], using the CCTSDB2021 dataset, has the highest precision (92%) but a lower mAP@0.5:0.95 (59.3%), indicating a steeper drop in performance across IoU thresholds. SegU-Net [17], tested on the GTSDB dataset, yields very high precision (94.6%) but lacks mAP values, limiting full comparison. YOLO-BS [18] achieves the highest mAP@0.5 (90.1%) but has slightly lower recall (80.5%) than the proposed model. Additionally, YOLO-v8s [19] offers good precision (91.7%) on the CCTSDB dataset but lacks data for mAP@0.5:0.95, making it less informative for comprehensive evaluation.

TABLE I. EVALUATION RESULTS OF YOLOR

Study	Model	Dataset	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Our work	YOLOR	Traffic signs detection	87.5	87.8	88.5	77.7
[15]	TRD-YOLO	TT100K	86.3	82.1	86.5	65.6
[16]	YOLO-CCA	CCTSDB2021	92	80.2	86.9	59.3
[17]	SegU-Net	GTSDB	94.6	80.21	-	-
[18]	YOLO-BS	TT100K	87.9	80.5	90.1	70.3
[19]	YOLO-v8s	CCTSDB	91.7	76.2	84.9	-
[20]	YOLO-LLTS	CNTSSS, TT100K-night, CCTSDB2021	88.3	74.9	81.2	60.1

Lastly, YOLO-LLTS [20], designed for low-light conditions and evaluated on the CNTSSS, TT100K-night, CCTSDB2021 datasets, reports a precision of 88.3%, a recall of 74.9%, an mAP@0.5 of 91.2%, and an mAP@0.5:0.95 of 74.6%, showing superior single-threshold accuracy but lower recall and slightly lower consistency across IoU thresholds compared to the proposed model. Overall, the proposed model demonstrates a strong balance across all metrics, particularly excelling in consistent detection performance across IoU thresholds.

IV. CONCLUSION

The results of this study demonstrate that the proposed YOLOR-based model offers a highly effective solution for real-time traffic sign detection. By achieving a strong balance between precision, recall, and mAP scores, particularly a high mAP@0.5:0.95 of 77.7%, the model outperforms or matches several state-of-the-art approaches across key detection metrics. Its performance affirms YOLOR's capability to integrate implicit and explicit knowledge for enhanced object representation, making it well-suited for complex road environments. Compared to existing methods such as TRD-YOLO, YOLO-CCA, and YOLO-BS, the proposed approach shows improved generalization and robustness, especially in maintaining accuracy across varying IoU thresholds. These findings suggest that YOLOR is a reliable and scalable option for deployment in intelligent transportation systems, where both accuracy and computational efficiency are essential. Future work may focus on addressing class imbalance and improving the detection of underrepresented signs to further strengthen the reliability of the system. Additionally, exploring lightweight model variants for edge deployment, incorporating temporal consistency through video-based detection, and extending the model to support multimodal inputs (e.g., LiDAR, GPS) could enhance adaptability and performance in diverse and dynamic traffic scenarios.

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] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "You Only Learn One Representation: Unified Network for Multiple Tasks." arXiv, May 10, 2021, https://doi.org/10.48550/arXiv.2105.04206.

[3] V. T. Tran, T. S. To, T. N. Nguyen, and T. D. Tran, "Safety Helmet Detection at Construction Sites Using YOLOv5 and YOLOR," in *Intelligence of Things: Technologies and Applications*, 2022, pp. 339–347, https://doi.org/10.1007/978-3-031-15063-0_32.

[4] N. Pawar *et al.*, "Miniscule Object Detection in Aerial Images Using YOLOR: A Review," in *Proceedings of International Conference on Communication and Computational Technologies*, 2023, pp. 697–708, https://doi.org/10.1007/978-981-19-3951-8_52.

[5] H. Sun, D. Lu, X. Li, J. Tan, J. Zhao, and D. Hou, "Research on multi-apparent defects detection of concrete bridges based on YOLOR," *Structures*, vol. 65, Jul. 2024, Art. no. 106735, https://doi.org/10.1016/j.istruc.2024.106735.

[6] E. Kizilay and I. Aydin, "A YOLOR Based Visual Detection of Amateur Drones," in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, Chiangrai, Thailand, Mar. 2022, pp. 1446–1449, https://doi.org/10.1109/DASA54658.2022.9765252.

[7] G. Rjoub, J. Bentahar, and Y. A. Joarder, "Active Federated YOLOR Model for Enhancing Autonomous Vehicles Safety," in *Mobile Web and Intelligent Information Systems*, 2022, pp. 49–64, https://doi.org/10.1007/978-3-031-14391-5_4.

[8] A. Balasubramaniam and S. Pasricha, "Object Detection in Autonomous Vehicles: Status and Open Challenges." arXiv, Jan. 19, 2022, https://doi.org/10.48550/arXiv.2201.07706.

[9] M. Masmoudi, H. Ghazzai, M. Frikha, and Y. Massoud, "Object Detection Learning Techniques for Autonomous Vehicle Applications," in *2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, Cairo, Egypt, Sep. 2019, pp. 1–5, https://doi.org/10.1109/ICVES.2019.8906437.

[10] J. E. Hoffmann, H. G. Tosso, M. M. D. Santos, J. F. Justo, A. W. Malik, and A. U. Rahman, "Real-Time Adaptive Object Detection and Tracking for Autonomous Vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 3, pp. 450–459, Sep. 2021, https://doi.org/10.1109/TIV.2020.3037928.

[11] T. Yan, W. Sun, and K. Cui, "Real-time Ship Object Detection with YOLOR," in *Proceedings of the 2022 5th International Conference on Signal Processing and Machine Learning*, Jul. 2022, pp. 203–210, https://doi.org/10.1145/3556384.3556415.

[12] Y. F. Huang, T. J. Liu, C. A. Lin, and K. H. Liu, "SOAda-YOLOR: Small Object Adaptive YOLOR Algorithm for Road Object Detection," in *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Oct. 2023, pp. 1652–1658, https://doi.org/10.1109/APSIPAASC58517.2023.10317144.

[13] C. F. Liou, T. H. Lee, and J. I. Guo, "Asynchronous Multi-Task Learning Based on One Stage YOLOR Algorithm," in *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*, Jun. 2023, pp. 1–5, https://doi.org/10.1109/ISIE51358.2023.10228120.

[14] P. J. Darabi, "Traffic Signs Detection." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/pkdarabi/cardetection.

[15] J. Chu, C. Zhang, M. Yan, H. Zhang, and T. Ge, "TRD-YOLO: A Real-Time, High-Performance Small Traffic Sign Detection Algorithm," *Sensors*, vol. 23, no. 8, Jan. 2023, Art. no. 3871, https://doi.org/10.3390/s23083871.

[16] L. Jiang, P. Zhan, T. Bai, and H. Yu, "YOLO-CCA: A Context-Based Approach for Traffic Sign Detection." arXiv, Dec. 05, 2024, https://doi.org/10.48550/arXiv.2412.04289.

[17] U. Kamal, T. I. Tonmoy, S. Das, and Md. K. Hasan, "Automatic Traffic Sign Detection and Recognition Using SegU-Net and a Modified Tversky Loss Function With L1-Constraint," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1467–1479, Apr. 2020, https://doi.org/10.1109/TITS.2019.2911727.

[18] H. Zhang, M. Liang, and Y. Wang, "YOLO-BS: a traffic sign detection algorithm based on YOLOv8," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, Art. no. 7558, https://doi.org/10.1038/s41598-025-88184-0.

[19] S. Du *et al.*, "TSD-YOLO: Small traffic sign detection based on improved YOLO v8," *IET Image Processing*, vol. 18, no. 11, pp. 2884–2898, 2024, https://doi.org/10.1049/ipr2.13141.

[20] Z. Lin *et al.*, "YOLO-LLTS: Real-Time Low-Light Traffic Sign Detection via Prior-Guided Enhancement and Multi-Branch Feature Interaction." arXiv, Mar. 30, 2025, https://doi.org/10.48550/arXiv.2503.13883.