Revolutionizing Traffic Sign Recognition with YOLOv9 and CNNs

Muteb Alshammari[1†], Aadil Alshammari[2]

[1]Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Arar, Saudi Arabia
[†]Corresponding Author
[2]Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Arar, Saudi Arabia

Abstract
Traffic sign recognition is an essential feature of intelligent transportation systems and Advanced Driver Assistance Systems (ADAS), which are necessary for improving road safety and advancing the development of autonomous cars. This research investigates the incorporation of the YOLOv9 model into traffic sign recognition systems, utilizing its sophisticated functionalities such as Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN) to tackle enduring difficulties in object detection. We employed a publically accessible dataset obtained from Roboflow, which consisted of 3130 images classified into five distinct categories: speed_40, speed_60, stop, green, and red. The dataset was separated into training (68%), validation (21%), and testing (12%) subsets in a methodical manner to ensure a thorough examination. Our comprehensive trials have shown that YOLOv9 obtains a mean Average Precision (mAP@0.5) of 0.959, suggesting exceptional precision and recall for the majority of traffic sign classes. However, there is still potential for improvement specifically in the red traffic sign class. An analysis was conducted on the distribution of instances among different traffic sign categories and the differences in size within the dataset. This analysis aimed to guarantee that the model would perform well in real-world circumstances. The findings validate that YOLOv9 substantially improves the precision and dependability of traffic sign identification, establishing it as a dependable option for implementation in intelligent transportation systems and ADAS. The incorporation of YOLOv9 in real-world traffic sign recognition and classification tasks demonstrates its promise in making roadways safer and more efficient.

Keywords:
Traffic sign recognition, YOLOv9, Image recognition, Convolutional Neural Network, Object detection.

1. Introduction

Traffic sign recognition is an essential element of contemporary intelligent transportation systems, serving a crucial function in guaranteeing road safety and effective traffic control. Precise identification and understanding of traffic signs are crucial for the advancement of self-driving vehicles and advanced driver assistance systems (ADAS) [1]. Over time, researchers have investigated many approaches to improve the precision and effectiveness of traffic sign recognition. Among these approaches, the YOLO (You Only Look Once) family of models has emerged as a notable advancement in the field of object detection.

The development of YOLO models has been a significant advancement in the field of real-time object identification. Beginning with YOLOv1, which offered a consolidated structure for object detection, subsequent versions such as YOLOv5 and YOLOv8 have consistently delivered significant enhancements in terms of speed, accuracy, and computing efficiency. YOLO models have become the favored option for a range of real-time detection applications, including traffic sign recognition, due to these developments [2].

YOLOv9 is a recent development in this series, including novel characteristics that tackle enduring obstacles in object detection. YOLOv9 has advanced features including Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN), resulting in improved performance and efficiency. This makes it an ideal choice for deploying on devices with limited computational resources.

The key technology behind these breakthroughs is Convolutional Neural Networks (CNNs). CNNs have significantly transformed image recognition tasks by offering a resilient framework for extracting features and

Manuscript received August 5, 2024
Manuscript revised August 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.8.2

recognizing patterns. Due to their capacity to acquire hierarchical representations of visual input, they are highly proficient at activities such as traffic sign recognition, which requires the ability to differentiate between different types and circumstances of signs [3].

Precise traffic sign recognition technologies are essential for the secure and effective functioning of autonomous cars and ADAS. Nevertheless, current recognition techniques frequently encounter difficulties such as fluctuating illumination, obstructions, and the existence of diminutive or deteriorated indicators. These issues need the creation of more advanced models that can consistently achieve high accuracy in a wide range of situations.

The use of YOLOv9 with traffic sign recognition systems provides numerous advantages. The model's sophisticated structure improves the precision and speed of detection, while its streamlined design guarantees its suitability for deployment on devices with restricted processing capabilities. This integration enhances both the dependability of traffic sign recognition systems and the overall safety and efficiency of intelligent transportation systems.

This work examines the integration of YOLOv9 into traffic sign recognition systems, emphasizing the advantages of the model and discussing the difficulties encountered by current approaches. Our extensive trials and evaluations confirm that YOLOv9 significantly improves the accuracy and reliability of traffic sign recognition, leading to safer and more efficient roads.

2. Related Work

Yaqin et al. in [4] introduced a lightweight traffic sign detection algorithm based on improved YOLOv5, enhanced with C2f architecture, BiFPN_Concat, SimAm attention mechanism, and Focal EIoU loss function, achieving significant improvements in detection accuracy and speed, with metrics such as an F1 score of 0.8987 and a detection speed of 50 FPS. Implemented on the Android Studio platform, the algorithm demonstrates effective real-time detection under various conditions, validated through comparative and ablation experiments, and successfully deployed on a mobile device.

In [5] Yanjiang et al. proposed EDN-YOLO, an enhanced YOLOv5s-based detection method that uses Efficient-Vit for the backbone network, an efficient decoupled detection head, and an optimized loss function combining CIoU and NWD losses to improve multi-scale traffic sign detection. Experimental results on the GTSDB and TT100K datasets show that EDN-YOLO significantly outperforms the original YOLOv5s, with mAP improvements of 3.1% and 9.6%, respectively, demonstrating its practical significance for real-time traffic sign detection and intelligent transportation systems.

In [6] Xianglong et al. introduced AIF-YOLO, an enhanced traffic sign recognition algorithm based on YOLOv7, incorporating an Adaptive Feature Extraction Module (AFEM) and a Contextual Transformer Efficient Layer Aggregation Network (CoT-ELAN) to improve the recognition of small traffic signs in complex environments. The main findings demonstrate that AIF-YOLO outperforms the baseline YOLOv7 and other popular algorithms in terms of precision, recall, and mean Average Precision (mAP) on the TT100K dataset.

Yaqin et al. in [7] enhanced YOLOv5 for small traffic sign detection by adding a small target detection layer, preserving higher resolution and richer pixel information, and employing an adaptive scaling channel pruning strategy to significantly reduce parameters and computations without compromising accuracy. Experimental results on the TT100k dataset show that the proposed DeployEase-YOLO achieves a 1.3% higher accuracy than YOLOv7 while reducing the model size and computational load, making it suitable for deployment on resource-limited devices.

Baoxiang and Xinwei in [8] presented MSGC-YOLO, an enhanced YOLOv8-based model for traffic sign detection under snowy conditions, incorporating Multi-Scale Group Convolution, an improved small target detection layer, EfficientSlide loss, and Deformable Attention to significantly boost detection accuracy and efficiency. The model achieved a 17.7% and 18.1% increase in mAP@0.5 and mAP@0.5:0.95, respectively, compared to the original YOLOv8, while reducing parameters by 59.6% with minimal accuracy loss, demonstrating robustness across various weather conditions.

Dang et al. in [9] explored traffic sign recognition in Vietnam using YOLOv5 and YOLOv8 models on a dataset comprising images from Can Tho City, Vinh Long province, and the ZaloAI dataset, enhanced through data augmentation. The findings indicate that while YOLOv5 requires more training time and has fewer parameters, it outperforms YOLOv8 in traffic sign recognition tasks.

Selvia et al. in [10] explored the use of YOLOv5 and YOLOv8 deep learning models for detecting and classifying roadway signs under various illumination conditions, demonstrating that both models perform robustly with YOLOv8 achieving slightly higher Mean Average Precision (MAP50) scores ranging from 94.6% to 97.1%. The findings suggest that these models are reliable and adaptable for real-world applications in roadway asset management and Advanced Driving Assistant Systems (ADAS).

A crucial issue for autonomous vehicles, traffic sign detection, is examined by Çınarer in [11] using three YOLOv5 models: YOLOv5s, YOLOv5m, and YOLOv5l. The research assesses the models using three metrics: recall, precision, and mean Average Precision (mAP50). The dataset utilized in the study consists of 741 photos of traffic signs that have been classified into four groups. With a mAP50 score of 98.1%, the YOLOv5m model was the most accurate, while the YOLOv5l model reached a precision of 99.3%. The results show that all three models work great at detecting traffic signs in real-time; their mAP50 values are more than 96%. Research shows that YOLOv5 models can make autonomous vehicles safer and more useful,

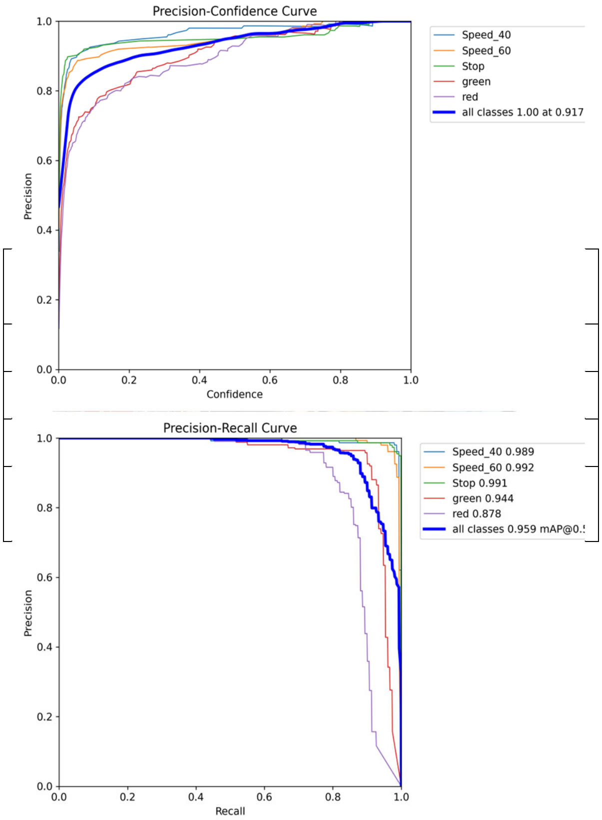


Figure 2: Precision-Confidence, and Precision-Recall Curves

Figure 1: Examples of results

which could lead to more widespread use of the technology and better results in testing.

3. Dataset And Methods

In this section, we outline the dataset utilized for training and evaluating the CNN aimed at detecting traffic signs, along with the methodology used in developing and assessing the model.

A. Dataset

This work uses a publicly available dataset given by Roboflow, an open-source platform for computer vision datasets [12]. Table I shows that the dataset contains 3130 photos divided into 5 categories: speed_40, speed_60, stop, green, and red. The dataset is separated into three subsets: training (68%), validation (21%), and testing (12%). Specifically, 2118 photos are designated for training, 648 for validation, and 364 for testing.

This systematic division ensures a thorough examination of the model's performance at various phases of development. Figure 1 depicts examples of the dataset's classes with various traffic signs, which the model utilizes to learn their appearance.

B. YOLOv9

YOLOv9 is an advanced object detection model that has been successfully used in different fields, demonstrating notable improvements in both performance and efficiency [13], [14]. The system incorporates innovative elements such as a partial attention block to improve the extraction of features and attention mechanisms, resulting in greater accuracy and efficiency in detecting pavement deterioration.

The YOLOv9 model primarily incorporates two groundbreaking concepts: Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN). These improvements seek to tackle major challenges in deep learning, such as data loss across layers and the inefficiency of current designs in preserving complete input information.

PGI is proposed as a remedy for the information bottleneck issue and the constraints of deep supervision in lightweight neural networks. The process involves creating consistent changes in value across additional reversible pathways, guaranteeing that fundamental attributes required for specific objectives are preserved in complex features. This method is designed to be flexible, suitable for different models, and efficient for both small and large models without raising the cost of inference. GELAN is a novel network design that is lightweight and relies on gradient path planning. This method employs traditional convolution operators and provides superior parameter consumption and performance compared to the most advanced depth-wise convolution methods. Furthermore, YOLOv9 has effectively been used in fracture detection tasks, showcasing enhanced performance in comparison to current models by training on extensive datasets and employing data augmentation approaches.

4. RESULTS AND DISCUSSION

The results of our experiments demonstrate the effectiveness of the model in traffic sign recognition tasks. The performance metrics, including precision, recall, and mean Average Precision (mAP), were evaluated to assess the model's accuracy and reliability.

A. Precision-Confidence and Precision-Recall Curves:

Figure 2 illustrates the Precision-Confidence and Precision-Recall curves for the model, which are crucial metrics for assessing the effectiveness of object detection algorithms. The curves illustrate the model's overall accuracy for each class as follows: speed_40 (0.989), speed _60 (0.992), stop (0.991), green (0.944), and red (0.878). The overall mean Average Precision (mAP@0.5) for all classes is 0.959, suggesting an outstanding degree of precision and recall for most of the traffic sign classes, while there is potential for improvement specifically in the red traffic sign class.

The Precision-Recall curve offers valuable insights into the balance between precision and recall. High precision refers to the accuracy of detecting traffic signs, meaning that a majority of the detected signs are correct. On the other hand, high recall shows that the model can identify most of the actual traffic signs. The graphs in Figure 2 demonstrate that YOLOv9 achieves an impressive balance between precision and recall, indicating its ability to identify a significant quantity of traffic signs while minimizing both false positives and false negatives. The balance is of important in applications related to intelligent transportation systems and ADAS, as the occurrence of both missed detections and false alarms can lead to substantial repercussions. The substantial area beneath the Precision-Recall curve provides additional evidence of the model's durability and efficiency in practical traffic sign recognition assignments.

B. Normalized Confusion Matrix:

Figure 3 displays the normalized confusion matrix for the model, offering a complete summary of its classification performance across several traffic sign classes. The matrix consists of cells that indicate the ratio of predictions made for each actual class (rows) compared to the expected class (columns). The diagonal elements correspond to the true positive rates, indicating the instances where the model accurately detects the traffic signs. Conversely, the off-diagonal elements show instances of misclassifications. The elevated values along the diagonal indicate that the model exhibits a robust capability to precisely categorize the majority of traffic signs, with low inaccuracies. Precision is of utmost importance in intelligent transportation systems, as accurate traffic sign recognition is vital for ensuring safety and optimizing efficiency.

The normalized confusion matrix identifies particular regions where the model may require enhancement. For example, when certain off-diagonal elements have comparatively larger values, it suggests that the model sometimes mistakes specific traffic sign classifications. Gaining a comprehension of these misclassifications can offer valuable insights into possible areas for improvement, such as augmenting the training dataset or modifying the model architecture to more effectively differentiate between signs that have a similar appearance. However, the results show that the model was able to identify the correct signs most of the time with very little misclassification. Speed_40 sign was correctly identified 99% of the times, Speed_60 sign was also identified 99%, Stop sign was perfectly identified, and red sign was identified 85%.

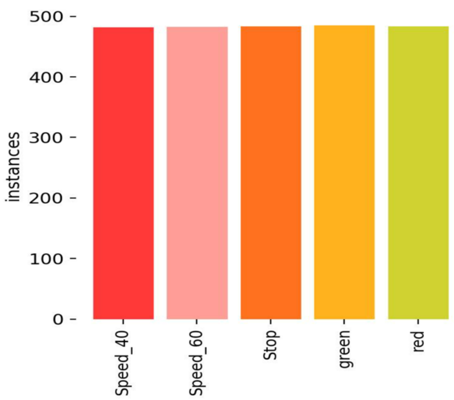
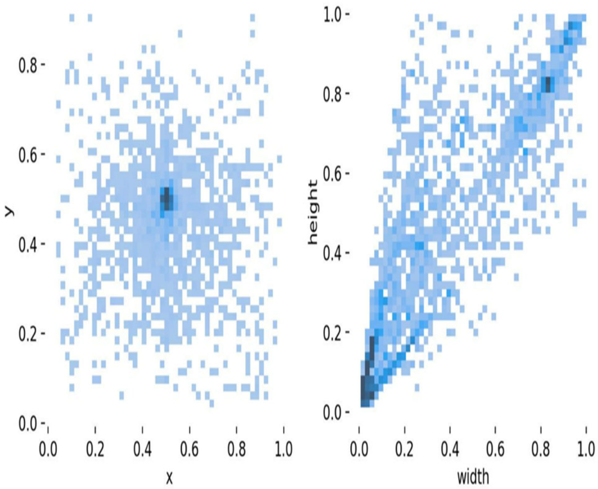


Figure 3: Distribution of Instances Over Categories

C. Distributions of instances:



D. *Figure 4: Confusion Matrix*

Figure 4 depicts the allocation of instances across different types of traffic signs in the dataset employed for training and assessing the model. The significance of this figure lies in its ability to visually depict the distribution of the dataset among the dataset categories. An evenly distributed dataset ensures that the model can effectively train to recognize each category. Striking a balance between accuracy and dependability is crucial in real-world traffic sign identification tasks, as the model needs to

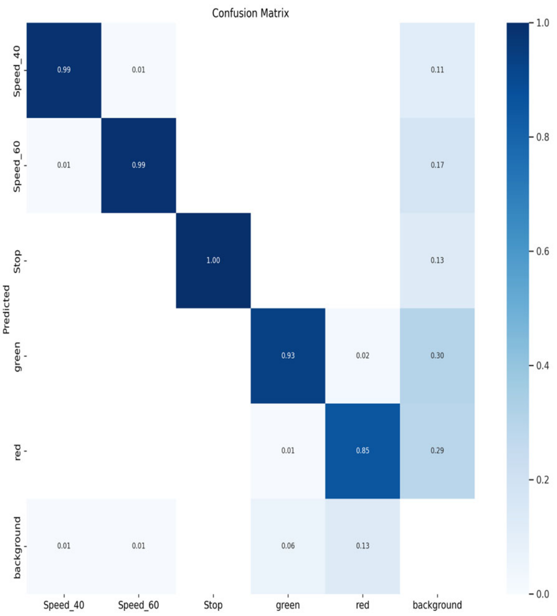


Figure 5: Distribution of width and height of detected traffic signs

accurately detect a diverse range of signs in various settings.

Figure 5 displays the distribution of the width and height of the traffic signs instances in the dataset. The figure is significant as it offers valuable information regarding the range of sizes of traffic signs that the model has identified and categorized. An accurate assessment of the model's capacity to handle signs of varying sizes is facilitated by absorbing the distribution of width and height. This understanding is particularly important in real-world scenarios where traffic signs might exhibit substantial variations in dimensions. The plot on the left shows that higher density of points around the (0.5, 0.5) coordinate which indicates that the traffic signs tend to be detected more frequently around the central area of the image frame. This is true as shown in the examples of the dataset pictures in Figure 1. Whereas the plot on the right reads that most traffic signs are relatively small with respect to the frame, as the highest density of points is concentrated around the bottom left corner. This implies that traffic signs do not usually occupy a large portion of the image frame, which is also true as seen in Figure 1.

5. CONCLUSION

This paper provides a thorough examination of the incorporation of the YOLOv9 model with Convolutional Neural Networks (CNNs) for the purpose of recognizing traffic signs. This is an essential aspect of intelligent transportation systems and Advanced Driver Assistance Systems (ADAS). The study showcases the exceptional performance of YOLOv9 in terms of precision, recall, and mean Average Precision (mAP@0.5) using a dataset from Roboflow. The dataset consists of 3130 photos classified into five different categories of traffic signs. The dataset was systematically partitioned into training, validation, and testing subsets to guarantee a comprehensive assessment of the model.

The inclusion of Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN) in YOLOv9 greatly improves the model's ability to accurately recognize objects and increases its efficiency. The characteristics of YOLOv9 make it well-suited for use on devices that have limited computational capabilities, therefore expanding its usefulness in real-life situations. The model demonstrated a noteworthy mean Average Precision (mAP) of 0.959 at a threshold of 0.5. However, it exhibited room for enhancement in accurately detecting red traffic lights.

When comparing YOLOv9 to other models like YOLOv5 and YOLOv8, it becomes evident that YOLOv9 outperforms them in terms of performance. The study emphasizes the model's capacity to properly manage different sizes and circumstances of traffic signs, making it a dependable option for improving the safety and efficiency of intelligent transportation systems and ADAS.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FFR-2024-1580-07”.

References

[1] M. Bakirci and I. Bayraktar, “Yolov9-enabled vehicle detection for urban security and forensics applications,” in 2024 12th International Symposium on Digital Forensics and Security (ISDFS). IEEE, 2024, pp. 1–6.

[2] S. Liu, “A review of yolo-based traffic sign target detection,” Science and Technology of Engineering, Chemistry and Environmental Protection, vol. 1, no. 7, 2024.

[3] S. Youwai, A. Chaiyaphat, and P. Chaipetch, “Yolo9tr: A lightweight model for pavement damage detection utilizing a generalized efficient layer aggregation network and attention mechanism,” arXiv preprint arXiv:2406.11254, 2024.

[4] Y. Wu, W. Hao, J. Niu, and H. Wang, “Yolo-based lightweight traffic sign detection algorithm and mobile deployment,” 2024.

[5] Y. Han, F. Wang, W. Wang, X. Zhang, and X. Li, “Edn-yolo: Multiscale traffic sign detection method in complex scenes,” Digital Signal Processing, p. 104615, 2024.

[6] X. Luo, L. Chen, and C. Liao, “Aif-yolo: small traffic sign recognition based on yolov7,” 2024.

[7] Y. Li, Z. Zhang, C. Yuan, and J. Hu, “Easily deployable real-time detection method for small traffic signs,” Journal of Intelligent & Fuzzy Systems, no. Preprint, pp. 1–14, 2024.

[8] B. Chen and X. Fan, “Msgc-yolo: An improved lightweight traffic sign detection model under snow conditions,” Mathematics, vol. 12, no. 10, p. 1539, 2024.

[9] P. H. D. Bui, T. T. Nguyen, T. M. Nguyen, and H. T. Nguyen, “An approach for traffic sign recognition with versions of yolo,” in International Conference on Intelligent Systems and Data Science. Springer, 2023, pp. 204–211.

[10] S. Nafaa, K. Ashour, R. Mohamed, H. Essam, D. Emad, M. Elhenawy, H. I. Ashqar, A. A. Hassan, and T. I. Alhadidi, “Advancing roadway sign detection with yolo models and transfer learning,” in 2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI). IEEE, 2024, pp. 1–4.

[11] G. Çınarer, “Deep learning based traffic sign recognition using yolo algorithm,” Duzce¨ Universitesi Bilim ve Teknoloji Dergisi¨ , vol. 12, no. 1, pp. 219–229.

[12] zineb, “zineb robo 2 dataset,” https://universe.roboflow.com/zineblcfgm/zineb robo-2 , jun 2024, visited on 2024-08-03. [Online]. Available: https://universe.roboflow.com/zineb-lcfgm/zineb robo-2

[13] C.-Y. Wang and H.-Y. M. Liao, “YOLOv9: Learning what you want to learn using programmable gradient information,” 2024.

[14] H.-S. Chang, C.-Y. Wang, R. R. Wang, G. Chou, and H.-Y. M. Liao, “YOLOR-based multi-task learning,” arXiv preprint arXiv:2309.16921, 2023.